ALS-associated FUS mutation reshapes the RNA and protein composition and dynamic of Stress Granules.

Davide Mariani[1,†], Adriano Setti[2,†], Francesco Castagnetti[1], Erika Vitiello[1], Lorenzo Stufera Mecarelli[1,2], Gaia di Timoteo[2], Andrea Giuliani[2], Eleonora Perego[1], Sabrina Zappone[1], Giuseppe Vicidomini[1], Nara Liessi[4], Andrea Armirotti[4] and Irene Bozzoni[1,2,3,*].

1) Center for Human Technologies, Istituto Italiano di Tecnologia, Genoa, Italy;
2) Department of Biology and Biotechnologies "C. Darwin", Sapienza University of Rome, Rome, Italy;
3) Center for Life Nano-& Neuro-Science, Fondazione Istituto Italiano di Tecnologia, Rome, Italy;
4) Analytical Chemistry Lab, Istituto Italiano di Tecnologia, Genoa, Italy.
† These authors contributed equally to this work
* To whom correspondence should be addressed. Tel: +39 0649912202; Email: irene.bozzoni@uniroma1.it

Abstract

Stress Granules (SG) formation is a cellular protection mechanism in harmful conditions, constituting a storage for untranslated mRNAs and RNA-binding proteins (RBPs); however, these condensates can turn into pathological aggregates, that in neurons are related to the onset of neurodegenerative diseases, like Amyotrophic Lateral Sclerosis (ALS). Mutations in the RBP FUS leading to its cytoplasmic mis-localization are causatively linked to ALS, since they mediate its accumulation in SGs, triggering their transition towards cytotoxic inclusions. Here, we describe the SG transcriptome in a neural context and compare with datasets from other systems, identifying both common rules and diversifying characteristics for SG recruitment of transcripts. We demonstrate that SG dynamics and RNA content are strongly modified by the incorporation of aberrantly localized mutant FUS, switching to a more unstructured, AU-rich SG transcriptome and losing a wide population of GC-rich, structured RNAs. We show that these alterations mainly depend on mutant FUS which favors the SG incorporation of its protein interactors and in turn of their target RNAs. Our data give a comprehensive view of the molecular differences between physiological and pathological SG in ALS conditions, showing how a single mutation is sufficient to impact the RNA and protein population of these condensates.

Introduction

A well-defined intracellular compartmentalization is vital for eukaryotic cells, with membrane-bound organelles physically separating specialized molecular processes from the rest of the cellular environment. Alongside these structures, a large number of membrane-less organelles (MLOs) has been identified in recent years as crucial in the regulation of gene expression and metabolic activity of cells (1–3). Lacking a lipid bilayer, these assemblies are separated from the neighboring cytoplasm or nucleoplasm by liquid-liquid demixing, conferring them stability, but also dynamic properties of exchange with the molecular surroundings (4–6). The liquid-like behavior of MLOs, mainly constituted by diverse RNAs and RNA binding proteins (RBPs), results from the summation of several weak interactions involving its constituents (7).

RNA is a pivotal component of such condensates, since the intrinsic characteristics of individual transcripts, such as length, structure and nucleotide composition, are sufficient to influence the number of RNA-RNA and RNA-protein interactions within a granule, thereby affecting its physical properties (8–11). Moreover, it has been observed that the presence of Intrinsically Disordered Regions (IDRs) in RBPs is a molecular determinant of their tendency to phase-separate, both *in vitro* and in living cells (12); IDRs are often coupled to canonical RNA binding domains within the same protein, providing the foundation for establishing multivalent interactions with RNAs and other proteins, thus leading to self-assembling, separated granules (13, 14).

The formation of cytoplasmic Stress Granules (SG) in response to variations in cell homeostasis stimuli is a primordial protection mechanism (15). SG result from the condensation of untranslated, stalled mRNAs and several IDR-containing RBPs, functioning as a temporary reservoir and being readily disassembled as protein translation is reactivated after stress (16). SG assembly is a multi-step event, where initial phase separation of individual ribonucleoparticles leads to the coalescence of stable *cores*, surrounded by a liquid, quickly remodeled *shell* (17).

Post-mitotic cells like neurons are highly susceptible to dysregulation of SG assembly and disassembly (18–20). Mutations in numerous protein components of these condensates are causative of the formation of permanent, insoluble protein-RNA assemblies; these aggregates have been causatively linked to the pathogenesis of neurodegenerative diseases and constitute a hallmark of several motor-neuron diseases, such as Amyotrophic Lateral Sclerosis (ALS) (21, 22).

A paradigmatic example is represented by FUS/TLS, whose mutations account for 4% of familial ALS cases (23). The protein embeds a IDR which is responsible for self-assembly and spontaneous aggregation (24–26); ALS-linked mutations are clustered in the Nuclear Localization Sequence (NLS) reducing its nuclear reimport and resulting in an aberrant cytoplasmic accumulation and subsequent recruitment in SG (27, 28). As mutant FUS accumulates in the cytoplasm of motor

neurons, it triggers the activation of major stress pathways and impacts SG number and dynamics, finally leading to an increased vulnerability of these cells to harmful stimuli (29).

In this work, we focus on the RNA content of SG in a neuronal-like system, demonstrating that neuronal RNAs are specifically enriched in these condensates in response to oxidative stress. We show that the expression of the mis-localized FUS^{P525L} mutant is sufficient to impact the dynamics of SG and to strongly modify their transcriptome by enriching AU-rich RNAs, at the expenses of GC-rich transcripts. The modification of the nucleotide composition caused by the presence of mutant FUS reduces the global structuration propensity of SG RNAs, by recruiting poorly structured transcripts and excluding the more structured ones. Furthermore, we show that the presence of FUS^{P525L} also impacts SG proteome properties; thus, while the direct binding of mutant FUS to RNA is not sufficient to elucidate the altered engagement of transcripts into SG, the differential recruitment of FUS-interacting RBPs contributes the aberrant composition of pathological granules.

RESULTS

Characterization of the RNA composition of neural-like SG

To investigate the RNA composition of neuronal Stress Granules (SG), we adapted the purification protocol described by Khong et al. (30) **(Supplementary Fig. 1A)** on a SK-N-BE neuroblastoma cell line expressing the SG core protein G3BP1 fused to a GFP tag (SK-GFP-G3BP1), Fluorescence microscopy showed that upon oxidative stress these cells display GFP-G3BP1 cytoplasmic foci, confirming that SG were correctly assembled **(Supplementary Fig. 1B)**. Western Blot analysis confirmed the specific enrichment of the endogenous SG marker TIAR in the IP fraction of SK-GFP-G3BP1 only in stress conditions compared to unstressed control, suggesting that the SG are correctly purified (**Supplementary Fig. 1C**). Moreover, after RNA extraction, qRT-PCR was performed on targets previously identified as enriched or depleted in granules of U2OS cells (31). AHNAK, CENPF, HUWE1 and TRIO were selected as positive controls and displayed a strong enrichment in the IP fraction under stress induction, while the negative controls ATP5O, COX8A and LGALS1 were not precipitated **(Supplementary Fig. 1D).**

After these controls**,** we generated four replicates of purified SG and corresponding total RNA inputs and performed total RNA sequencing. Multidimensional scaling (MDS) plot of leading fold change **(Supplementary Fig. 1E)** and Pearson's correlation matrix **(Supplementary Fig. 1F)** confirmed the expected clusterization of the samples. We proceeded to RNA-Sequencing analysis

comparing SG-enriched (SG-enr) versus Input samples (INP). Among the 12318 RNAs detected in the system (>1 FPKM in INP or SG-enr) we identified 1816 enriched ($log_2FC > 1$ and FDR < 0.05) and 2939 depleted ($log_2FC < -1$ and FDR < 0.05) transcripts in the SG core (**Fig. 1A**). As expected, most of the RNAs that we identified as enriched in SG were mRNAs (85% of the enriched fraction) and among the non-coding fraction (15%) the most represented category was long non-coding RNAs (9%) (**Supplementary Fig. 1G).**

In order to identify the discriminating features of the obtained neuroblastoma cells SG transcriptome, we compared it to the ones of osteoblastoma (U2OS) (31) and endocervical adenocarcinoma (SMMC-7721 cell line) (32). At first, considering RNAs detected in all the three systems, we observed a strong correlation of the enrichment scores (log_2FC) among all the three datasets (**Supplementary Fig. 1H**). This evidence suggests not only that SG enrichment scores of shared RNAs are similar between datasets, but also that the methodology is highly reproducible even when applied in different systems. Moreover, using Venn diagrams, we compared the SG transcriptomes and we observed an overlap among the enriched RNAs (**Fig. 1B**). Indeed, the identified common *core* of SG consists of 502 RNAs which represent 27,6%, 26,8%, and 22,9% of SK-N-BE, U2OS and SMMC-7721 datasets, respectively. These transcripts, that characterize the ever-present core of SG in the three systems, are mostly mRNAs (**Fig. 1C**) and are enriched in Gene Ontology (GO) categories related to cytoskeletal and structural components, even neuronal-specific (**Fig. 1D, Supplementary Fig. 1I**).
Noteworthy, although neuronal transcripts are globally less represented in the non-neuronal transcriptomes, (**Supplementary Fig. 1J)**, they generally show a significant enrichment in the SG common *core* (Figure 1D). This could be possibly due to the fact that they tend to be longer than other mRNA categories (e.g. metabolic process, mitochondrion, endoplasmic reticulum and translation, **Fig. 1G**). Indeed, length was confirmed as a relevant feature for SG engagement in any cell type analyzed (**Fig. 1E).**

However, the overlap of the common *core* transcripts is paralleled with strong divergences in the SG RNA content in the different systems, with 37,8%, 40,4% and 52,7% of the transcriptomes being specifically enriched in SK-N-BE, U2OS and SMMC-7721 granules, respectively. Interestingly, the specific part of the SG transcriptome in neuroblastoma cells is enriched for neuronal categories such as “neuron to neuron synapse”, “neuronal cell body” and “axon part”, while those categories are not represented in the specific part of the other systems (**Fig. 1F, Supplementary Fig. 1K, L and M**). This evidence indicates that, in our neural system, transcripts

with neuronal functions have a higher propensity to be included in SG, and this behavior correlates with their length compared to transcripts belonging to other categories (**Fig. 1G).**

Summarizing the features of enriched transcripts, we observed that most of them are mRNAs (**Fig. 1C**), and that the length of the whole transcripts (**Fig. 1E**), or of single mRNA regions like CDS and 3'UTR (**Supplementary Fig. 1N**), is a crucial feature for SG engagement, with a less relevant contribution from 5'UTR length.

However, length and RNA biotype are not the only discriminating features for recruitment into granules. In fact, when we analyzed the nucleotide composition, we observed that the SK-N-BE and SMMC-7721 datasets have a higher GC content than the U2OS set (**Fig. 1E**), suggesting that the inclusion of specific transcripts leads to different overall nucleotide composition and thus could be associated with different physical properties of SG. Noteworthy, these evidences are not dependent on the thresholds selected to define enriched RNAs. Indeed, we observed the same results by stratifying the dataset according to the enrichment score ($\log_2$FC) in SG and comparing groups of equal numerosity (**Supplementary Fig. 1O**).

ALS-related FUS mutation P525L affects SG dynamics and transcriptome composition.

Mis-localization of mutant FUS in the cytoplasm and its accumulation in inclusions is known to be a central hallmark of ALS/FTD-associated mutations. Moreover, it is known that the mis-localized FUS co-localizes with other RBPs in SG as a consequence of oxidative stress, and that this process is possibly associated to aggregation (33). Despite this evidence, few efforts have been made to characterize the implications of mutant FUS localization on the physic-chemical features of SG and their RNA composition. To elucidate this aspect, we generated two stable SK-N-BE cell lines overexpressing either the wild-type FUS protein (FUS^{WT}) or a version of FUS carrying the P525L mutation on the NLS, which is associated to severe and juvenile ALS (FUS^{P525L}) (34), under doxycycline induction. We stimulated FUS overexpression with 50 ng/mL doxycycline obtaining expression levels of ectopic FUS comparable to the endogenous protein (**Supplementary Fig. 2A**). We showed that the overexpressed FUS^{P525L} is physically recruited into SG, as demonstrated by biochemical purification (**Supplementary Fig. 2B).** We also verified the correct FUS localization and observed that, while overexpressed FUS^{WT} is localized completely in the nucleus, FUS^{P525L} is cytoplasmic and co-localizes with the SG core protein G3BP1 in specific foci (**Supplementary Fig. 2C and 2D**). Moreover, we compared the number of SG in both conditions. As reported in other systems (35), we found an increased number of stress granules per cell in presence of FUS^{P525L} (**Supplementary Fig. 2E and 2F**).

We then analyzed GFP-G3BP1 diffusion coefficients in the two conditions, using SPAD-Fluorescence Fluctuation Spectroscopy (SPAD-FFS), to evaluate changes in the dynamics of condensation due to FUS mutation (36). Interestingly, according to the observed condensation process, we found that in both conditions G3BP1 movement velocity coefficient (D_{app}, measures the speed of movement of individual molecules) decreases upon stress induction (**Fig. 2A**). On the other hand, we found a higher confinement coefficient (S_{conf}, a measurement of the fluidity of individual molecules dynamic) in FUS^{P525L} condition compared to WT, suggesting a more confined and less dynamic conformation for SG when the mutant FUS is included (**Fig. 2A)**. Altogether, these data indicate that SG physiology is strongly affected by mutant FUS; in particular, the presence of FUS^{P525L} in SG induces an increased number of granules that are globally less dynamic.

Since we observed variations in SG number and dynamics, we decided to analyze whether the RNA content of these granules is altered in presence of FUS^{P525L}. Therefore, we generated duplicated samples of purified SG and performed total RNA sequencing of matched total RNA inputs in condition of doxycycline induction FUS^{WT} and FUS^{P525L}. Leading Fold change analysis (**Supplementary Fig. 2G)** and dendrogram of relative correlation distances (**Supplementary Fig. 2H)** confirmed that samples show the expected clusterization in INP and SG-enr conditions and by FUS^{WT} or FUS^{P525L} backgrounds. Therefore, for each condition we compared the SG-enriched samples with matched input: in FUS^{WT} condition we identified 2178 enriched ($log_2FC > 1$ and FDR < 0.05) and 2990 depleted RNAs ($log_2FC < -1$ and FDR < 0.05) (**Fig. 2B, left panel**); using the same thresholds, we observed 1510 SG enriched and 2752 depleted RNAs in FUS^{P525L} (**Fig. 2B, right panel**). Besides, we compared the SG transcriptome in all the three datasets we generated: FUS no DOXY (without FUS induction, described above), FUS^{WT} and FUS^{P525L}. First, we observed that 87,5% of no DOXY SG RNAs are enriched also in FUS^{WT} condition, as expected from the mainly nuclear localization of the overexpressed WT protein. Conversely, investigating the impact of mutant FUS, we observed that 822 RNAs enriched in SG in both no DOXY and FUS^{WT} conditions are excluded from SG in presence of FUS^{P525L} (thus, we renamed this group of transcripts "FUS^{P525L} LOSS"). Moreover, we also identified a consistent group of 556 RNAs that are specifically recruited in SG in presence of the mutation, but not in the other two conditions (group renamed as "FUS^{P525L} GAIN"). Despite the strong differences observed, a set of 768 transcripts is maintained in all the three conditions and always engaged in SG ("COMMON" group) (**Fig. 2C**).

Moreover, using "Resampling analysis" (see Material and Methods), we confirmed that the observed alterations of SG transcriptome in presence of FUS^{P525L} is not dependent on the technical

variability of RNA-Seq library sizes (**Supplementary Fig. 2I**), while "Enrichment convergence analysis" (see Material and Methods) confirmed that it is does not depend on the enrichment threshold (log_2FC SG-enr vs INP) used to define the SG transcriptomes (**Supplementary Fig. 2J**). To biochemically validate the quantitatively differential recruitment of RNAs into SG in the different analyzed conditions, we selected some transcripts from the top-enriched in the LOSS and GAIN groups, which we found to be related to neurodegenerative phenotypes and ALS in literature. Through qRT-PCR validation on independent SG purification samples, we confirmed the strong difference in the ratio of recruitment of the selected RNAs due to the expression of mutant FUS (**Supplementary Fig. 2K** and **2L)**.

In order to define the phenotypic impact of the observed reorganization of SG transcriptome, we performed Gene Ontology terms enrichment analysis. Among the categories enriched in the LOSS group, we found RNAs related to cell signaling pathways, and cytoskeletal and neuronal components (**Fig. 2D, top panel, Supplementary Fig. M)**, while we found RNAs related to transcriptional regulation in the GAIN group (**Fig. 2D, bottom panel, Supplementary Fig. N).** Finally, we focused our analysis also on the RNAs that are enriched among all the conditions (COMMON); noteworthy, these RNAs are still engaged in SG in presence of FUS^{P525L}, thus any effect caused by the alteration of SG physiology (different mobility and condensation state) could have an impact on the homeostasis of the COMMON set of RNAs. Interestingly, in this group we found genes related to neural development, cell-cell adhesion and transcription regulators (**Fig. 2D, middle panel, Supplementary Fig. O**).

All these evidences suggest that the ALS-associated FUS^{P525L} mutation has a strong effect on SG dynamics and induces a dramatic reorganization of the SG transcriptome, especially for mRNAs that encode for neuronal factors.

FUS^{P525L} affects SG transcriptome nucleotide composition and structuration propensity.

We then decided to investigate whether FUS mutation also affects the characterizing features of SG enriched RNAs that we previously identified. Thus, we studied how SG enrichment (log_2FC) relates with RNA length, mRNAs UTRs and CDS size and GC content.

As we observed for the no DOXY condition, length is a key feature to determine SG entry for both FUS^{WT} and FUS^{P525L} samples, with longer RNAs being more efficiently recruited (**Fig. 3A**). Moreover, we noticed that the length of CDS and 3'UTR of mRNAs is relevant for SG engagement in all the three conditions, while the contribution of the 5'UTR in this enrichment is less obvious.

However, despite the length of RNAs and mRNAs regions being a common criterion for SG engagement, we could detect some relevant differences: RNAs in the FUS^{P525L} SG are globally longer (**Supplementary Fig. 3A**), and this increased size is mainly due to the length of the 3'UTRs. Indeed, conversely to what is observed for the 3'UTR, both 5'UTR and CDS are slightly shorter in the FUS^{P525L} SG transcriptome (**Supplementary Fig. 3B**).

Furthermore, we also analyzed the fraction of protein coding transcripts and we observed that most of the RNAs identified as enriched were mRNAs, with comparable percentages in all the three groups (80% in FUS^{WT} to 85% FUS^{P525L} and 85% no DOXY) (**Fig. 3A).**

However, we observed the strongest difference in the GC content of the compared transcriptomes, with a consistent reduction of the GC content in the FUS^{P525L} condition compared to no DOXY and FUS^{WT} conditions (**Fig 3A** and **Supplementary Fig. 3C**). Indeed, analyzing the GC content in no DOXY and FUS^{WT} conditions, we observed a low correlation degree with SG enrichment (**Rho=0.065** no DOXY and **Rho=0.04** FUS^{WT} in **Supplementary Fig. 3D**), but when we focused on the FUS^{P525L} condition we strikingly observed a strong anti-correlation (**Rho=-0.3**, **Supplementary Fig. 3D**). This evidence not only highlights that SG enriched RNAs in FUS^{P525L} condition have a greater AU content (**Fig 3A**) compared to the others, but also that this enrichment is a feature associated with RNA entry in SG only in the presence of mutant FUS.

Noteworthy, this evidence is not dependent on the thresholds selected to define enriched RNAs. Indeed, we got the same results by stratifying the dataset according to the enrichments score in SG (**Supplementary Fig. 3E**).

To further investigate the relationship between sequence composition and SG recruitment, we asked whether specific subsequences are particularly represented in SG enriched transcripts. To highlight this aspect, we performed a k-mer (k = 7) enrichment analysis. Analyzing the top 50 k-mers over-represented in SG RNAs, we found that transcripts enriched in both FUS^{WT} and no DOXY conditions are characterized by a higher content of GC-rich k-mers, while the mutant FUS condition preferentially recruits RNAs containing AU-rich k-mers (**Fig. 3B).** Indeed, K-means clustering analysis of subsequence enrichment in SG identified two defined clusters of k-mers (cluster1 and cluster 5), that have divergent patterns of enrichment: the cluster 1, specifically enriched in FUS^{P525L}, displayed AU-rich sequences, while the cluster 5 that is specifically enriched in no DOXY and FUS^{WT} conditions displayed GC-rich ones (**Fig. 3C**).

To better understand the relationship between different RNA composition and the phenotype of reduced SG dynamics observed in FUS^{P525L}, we decided to deeper investigate the properties of the LOSS and GAIN transcripts, which represent the strong changes in transcriptome architecture. To do that, we compared the RNA secondary structures propensity of mRNAs from the two groups. In order to have a more reliable comparison, we sampled an equal number of RNAs from the two groups; moreover, since mRNA functional regions (5'UTR, CDS and 3'UTR) have different structural and interaction properties, we compared mRNAs with similar 5'UTR, CDS and 3'UTR length (Mann Whitney U- Test, p-value > 0.05, **Supplementary Fig 3F, G and H**). We then applied the RNAfold algorithm (37) to predict the minimum free energy (MFE) RNA secondary structure of each transcript, calculating the structuration score as the number of paired nucleotides over the length of the full transcripts or regions of them. Interestingly, we found that RNAs in the GAIN group were globally less structured than the LOSS group (**Fig. 3D and Supplementary Fig. 3I**). Moreover, comparing accessibility (defined as the fraction of un-paired nucleotides over the length of the full transcripts or regions of them) in each mRNA region, we found that CDS displays the stronger contribution to this feature, with 9% increased accessibility in FUS^{P525L} GAIN compared to the FUS^{P525L} LOSS group (**Fig. 3E**).

Moreover, to corroborate this evidence with other approaches, we used CROSS algorithm (38) that is an RNA secondary structure predictor based on a machine learning approach, with parameters trained on RNA structures derived by PARS datasets. We observed a higher prediction score for LOSS RNAs compared to the GAIN group (**Fig. 3F and Supplementary Fig. 3J**), highlighting that the RNAs that lose their enrichment in SG in presence of mutant FUS are more structured. Finally, in accordance with the predictions, also RNA accessibility data derived from DMS-Map-Seq methodology (39) confirmed less single stranded regions in the RNAs of LOSS group compared to the GAIN ones (**Fig. 3G and Supplementary Fig. 3K**). Altogether, these data suggest that FUS^{P525L}-mediated alteration of SG has a strong impact on the global GC content of the granules, and that the RNAs specifically lost in this condition are more structured, while the recruited ones are less prone to generate secondary structures.

FUS^{P525L} mutation alters the binding properties of the SG proteome.

In order to investigate the causal links between the presence of mutant FUS and the observed alterations in SG transcriptome, we focused on its RNA interactome. Indeed, since mutant FUS is almost completely localized in SG and directly binds RNAs, we expected an enrichment of FUS interacting transcripts in FUS^{P525L} SG compared to the WT condition.

We investigated this hypothesis by taking advantage of the endogenous mutant FUS PAR-CLIP experiment performed in iPSC derived motor-neurons (40). Remarkably, even though FUS establishes interactions with a wide fraction of the transcriptome, we found a significant, yet not striking difference in the enrichment of FUS^{P525L} interactions among the LOSS, COMMON and GAIN group, suggesting that FUS direct binding is not sufficient for the recruitment of the detected SG RNAs (**Fig. 4A**). Moreover, to support this idea, we stratified the transcriptome by SG engagement (log_2FC) and we found a similar recruitment of mutant FUS RNA interactors in the FUS^{P525L} condition, compared to the no DOXY and FUS^{WT} ones (**Supplementary Fig. 4A**).

Therefore, we looked for other possible contributions to this remodeling of SG-recruited RNAs, hypothesizing that the presence of mutant FUS could alterate the proteome, thus indirectly impacting on the SG transcriptome. For this reason, aiming to characterize the protein content of granules in FUS^{WT} and FUS^{P525L} condition, we performed SG purification followed by mass spectrometry, detecting 876 and 899 proteins in the two conditions, respectively. In order to assess the proper purification and detection of the SG proteome, we performed a GO enrichment analysis on the proteins identified in FUS^{WT} condition; these samples could be used as a reference control, since the overexpressed FUS^{WT} is almost exclusively nuclear and is more comparable to a basal situation. As expected from already described SG proteomes (17, 41), we found among the most enriched categories those related to translation, RNA processing and RNA binding (**Supplementary Fig. 4B**).

Moreover, we compared the SG proteome in FUS^{WT} condition with other published SG proteomes datasets from HEK293T cells (42) and U2OS cells (17), observing that a consistent portion of the SG localized proteins (77 common *core* proteins) are shared by all the three systems (**Supplementary Fig. 4C)**. Among the common *core* of proteins, we found known SG components such as G3BP2 and CAPRIN1, helicases such as DHX36, RNA processing factors like XRN1 and CNOT1, and also RNA binding proteins such as STAU2, PUM1 and EWSR1 (**Supplementary Fig. 4D**). Despite the common *core*, we also found proteins identified only in our dataset and not in the others that we compared (**Supplementary Fig. 4C**), supporting the idea that the common SG *core* can be accompanied by other proteins that characterize SG proteomes in different systems and conditions (43).

Therefore, we studied the impact of FUS^{P525L} on the SG proteome by comparing the identified proteins from mass spectrometry data in condition of FUS^{WT} or FUS^{P525L} induction. Conversely to what we observed for SG transcriptome, we found strong similarities between the proteome in both

conditions. Indeed, among the 876 proteins detected in FUS^{WT} SG and the 899 proteins in FUS^{P525L} SG, 796 are shared by the two conditions (90,8% of FUS^{WT} and 89% of FUS^{P525L} respectively, **Fig. 4B**), suggesting that the presence of mutant FUS does not cause a dramatic inclusion of new protein species in the SG. However, despite the qualitative similarity in SG protein contents, it is not excluded that the presence of the mutant FUS itself could have an impact on SG proteome organization. Thus, in order to better understand the relationship between FUS protein interactors and the SG proteome, we took advantage of 5 published datasets of FUS protein interactors related to its soluble (44–46), condensed (47) or aggregated form (42). The latter dataset refers to a mis-localized and aggregated form of FUS carrying the R522G mutation, and displayed the highest amount of FUS protein interactors (488 proteins). Interestingly, for every compared dataset we found that a great fraction of the FUS interactors is present in SG in both WT and mutant conditions, ranging from 42% to 86% of each analyzed interactome (**Fig 4B, Supplementary Fig. 4E**), meaning that a high portion of FUS protein interactors are already SG constituents. This evidence supports the hypothesis that mutant FUS is able to interact with a consistent set of proteins engaged in the SG.

Noteworthy, we also investigated the relationship between the inclusion of these proteins and the altered dynamic observed in FUS^{P525L}-embedding SG. We analyzed the protein-protein physical interactions (PPI, STRING database) that are expected to occur in the group of SG FUS interactors (named “SG proteins AND FUS interactors”) and in the remaining fraction of SG proteins (named “SG proteins NOT FUS interactors”). Since these two groups have different amounts of proteins (237 and 559, respectively, **Fig. 4B**), we performed three down-sampling of “SG proteins NOT FUS interactors” group and compared the PPI of “SG proteins AND FUS interactors” with those of the down-sampled groups to obtain a more reliable comparison (**Supplementary Fig. 4F and 4G**). Interestingly, the SG proteins that don’t interact with FUS show around 40% less PPIs than FUS interactors (**Fig. 4C and 4D**). This evidence indicates that mutant FUS localized in SG could affect the hub of protein-protein interactions of a huge set of SG components.

Since many SG protein components and FUS binding interactors are RBPs, we asked whether they could participate and explain the observed strong remodeling in the SG transcriptome upon FUS mutation. We took advantage of RNAact database of protein-RNA interaction predictions based on catRAPID algorithm (48). In order to identify those proteins that are able to preferentially bind the RNAs recruited in the altered SG, we ranked the SG proteome for the preferential binding with the GAIN RNA group compared to the LOSS one. Interestingly, among the SG proteins that are more prone to interact with the GAIN RNAs we found an enrichment of FUS interactors (**Fig. 4E,**

Supplementary Fig. 4H). We noticed a similar trend of enrichment of FUS interactors as preferential binders of the GAIN RNAs over the LOSS ones even when we performed the same analysis using experimental data of RBP-RNA interactions from POSTAR3 database (49) (**Supplementary Fig. 4I**). This evidence suggests that not only FUS itself, but also its protein interactors could be indirectly implicated in the observed SG transcriptome alteration. This hypothesis is also supported by the fact that at least two known FUS interactors, DHX36 (also known as RHAU) and ELAVL1 (HuR), widely studied for their ability to bind AU-rich sequences (50, 51), are actually more recruited in SG in presence of FUS^{P525L}, as demonstrated by Western Blot in comparison with other unperturbed SG markers (**Supplementary Fig. 4J**). Taken together, these data show that the GAIN RNA population is preferentially bound by FUS protein interactors, and that those proteins could be responsible for their recruitment in SG upon FUS mutation.

Finally, to have a more complete overview of the protein interactions implicated in the SG aberration, we focused on the ability of the three identified RNA groups (GAIN, COMMON and LOSS) to bind the whole set of SG proteins. Interestingly, we found that both GAIN and COMMON groups have less propensity to interact with SG proteins compared to the LOSS group (**Fig. 4F**). This result highlights that the RNAs recruited in SG in presence of FUS^{P525L} are globally less prone to establish interactions with the SG proteins, despite their affinity with mutant FUS interactors.

Taken together, our data suggest that the SG transcriptome alteration can be explained by a combination of the direct recruitment of mutant FUS, that leads to SG engagement of its direct RNA interactors, and of the ability of the proteins that are both SG component and FUS interactors to bind AU-rich transcripts, altering their localization. Finally, we showed that the remodeled SG transcriptome has a lower architectural effect on the SG proteome.

DISCUSSION

One of the common hallmarks of many neuropathies is the formation of inclusion bodies in neuronal cells. This aggregation process that is associated with neurodegeneration is often guided by RNA binding proteins (RBPs) which display an altered localization from the nucleus to the cytoplasm. An important paradigm of this pathological mis-localization is the RBP FUS; indeed, many of the known ALS-associated FUS mutations preferentially target its nuclear localization signal (52). Furthermore, under conditions of oxidative stress, the mis-localization of RBPs including FUS often causes their recruitment into Stress Granules (SG). Even though SG recruitment of mis-localized RBP was proposed as a protective buffering mechanism against protein

aggregation (53), however the engagement of these proteins in SG often alters their properties and is thought to be a crucial first step for the disease-related aggregation process (33). As an example, it was reported that TDP-43 inclusion in SG is a crucial step in the aggregation process of the RBP (54). Even though studying the interplay between RBPs and SG is pivotal, especially in neuropathies, little is known about the neuronal composition of SG and how RBPs can affect their properties. To fill this gap, combining granule purification with high throughput RNA sequencing, we characterized the composition of SG RNAs in a human neural system (SK-N-BE, neuroblastoma) and analyzed the impact of the ALS-associated FUS mutation P525L on SG properties.

In agreement with the description of SG RNA composition in U2OS cells (31) and in SMMC-7721 (32), we found that length is the main feature of the SG-enriched RNAs. Since longer RNA molecules usually harbor an increased number of possible interaction sites with both other RNAs and proteins, this feature emphasizes the role of multiple non-specific RNA interactions in the phase separation process (8).

Indeed, it was suggested that liquid-liquid phase separation (LLPS) could be driven by RNA itself: the release from ribosomes during stress event could engage mRNAs in a more complex interaction network, as occurs for proteins misfolding that leads to an increased availability of sites for not-specific protein-protein interactions. This process involves either the exposure of CDSs, allowing the mRNAs to gain non-specific interactions, or the decrease of the continuous processing by the helicase eiF4A, which contributes to resolving transient RNA-RNA interactions (55). The importance of these new interactions involving unmasked mRNAs regions is supported by the observation that both the CDS and the 3'UTR are longer in SG-enriched transcripts. This data regarding overall length and region-specific sizes of mRNAs is further supported across various systems where the SG purification methodology has been employed (31, 32).

Indeed, the methodology based on SG biochemical isolation and immunoprecipitation displayed high reproducibility in different systems (U2OS and SMMC-7721), and it led us to identify a common *core* of diverse SG transcriptomes. However, significant differences in SG composition between SK-N-BE, U2OS, and SMMC-7721 SG transcripts composition were observed. Our evidence on RNA composition is in line with previous observations about stress-, disease- and system-specific variations in the SG protein distribution (43). Moreover, we described GC content as a diversifying feature of SG RNA in different cell types. This GC divergence is important to take into consideration in pathological contexts: since the RNA nucleotide composition is a crucial

property for molecular interactions (56), an altered GC content of SG-recruited RNAs could be at the base of a variable susceptibility to the irreversible aggregation process of these granules in different cell types. Furthermore, we found that SK-N-BE SG-enriched RNAs are related to important neuronal functions and components such as neural development, cytoskeleton, and synapses. In neuronal contexts, mRNAs encoding neuronal factors tend to be longer, having longer 3'UTRs (57, 58); so, the enrichment of neuronal categories in SG might be a fortuitous outcome of their length.

Little is still known about why aggregation phenomena might be linked to neurodegeneration (59). The involvement of neuronal mRNAs that encode proteins with synaptic or axonal functions in neural SG suggests that any disruption in granule physiology affecting mRNAs, such as their retention in aberrant granules that are more resistant to dissolution, could result in substantial disturbances in neural functions or structures. Moreover, FUS mutations affecting the NLS result in its cytoplasmic delocalization and its recruitment into SG under stress. Although the degree of its cytoplasmic localization and SG recruitment correlates with the severity of ALS pathology (28), the impact of mutant FUS mis-localization on SG RNA composition and biophysical properties is still under investigation.

Using SK-N-BE cells expressing FUS^{WT} and FUS^{P525L} under doxycycline control, we observed that the recruitment of mutant FUS to condensates increases their number and alters their dynamics, leading to more aggregated neural SG with reduced molecular mobility. Furthermore, the RNA composition of neural SG is altered by mutant FUS, resulting in the lost enrichment of a set of transcripts, the gain of new ones, and the retention of a set of RNAs also present in the WT condition. Interestingly, part of the RNAs localized in mutant FUS-containing granules are enriched in categories with neuronal and synaptic functions**.** This emphasizes the significance of investigating the impact of altered mRNA recruitment into granules on neuronal cell homeostasis, as it could provide one of the missing links in understanding why aggregation or disruption of SG is associated with neurodegeneration.

Additionally, FUS mis-localization strongly alters the GC-content of recruited RNAs, shifting towards AU-rich sequences. To get a better understanding of the connection between sequence composition and altered SG properties, we compared the secondary structure-forming propensity of SG RNAs lost and gained due to the presence of mutant FUS, revealing that the lost RNAs are more likely to form structures compared to the gained ones. This result is particularly interesting since structures in RNA molecules often mediate specific interactions with proteins (60), and a reduced

RNA structuration in granules could give fewer opportunities for well-ordered interactions with granule proteins, or increase the amount of nonspecific interactions (61). Moreover, using *in vivo* DMS-Map-Seq datasets (39) we observed that RNAs in the GAIN group have a higher proportion of accessible nucleotides. Accessibility is related to structure, but encompasses interactions with proteins and other RNAs as well (62); this suggests that these RNAs within granules are less likely to interact both with themselves and with other granule components.

To comprehend the factors underlying this granule alteration, we investigated the RNA interactome of FUS. However, the presence of FUS itself within the granule and its potential interactions with RNA can only partially account for the observed granule alteration. In fact, specific RNA interactors of mutant FUS, identified in a motoneuronal context and under non-stress conditions (40), are enriched in SG engaging mutant FUS. Therefore, the observed enrichment suggests that a portion of transcripts is already associated with FUS^{P525L} in the absence of stress, and the formation of SG causes their engagement in the condensate. However, the extent of this enrichment cannot fully explain the transcriptome alteration observed in the granule. This has raised the possibility that other RBP partners of FUS, indirectly, could also contribute to the observed RNA compositional changes. Remarkably, we did not observe a strong qualitative difference in the proteome components of SG in the presence of the mutated FUS. However, the similar protein population detected does not exclude that FUS^{P525L} could affect SG proteome arrangement. In fact, using published datasets of known FUS protein interactors (42, 44–47), we found a consistent overlap with already identified SG components. Noteworthy, we observed that some of these FUS protein interactors are more engaged into SG, and their RNA binding ability could contribute to the described transcriptome alterations. Furthermore, we also observed that the FUS-interacting protein components of SG exhibit a significantly higher potential to form protein-protein physical interactions; this characteristic has been previously documented for the aggregated state of FUS^{R522G} interactors (42). Moreover, the physical interactions considered for this analysis include those involved in the formation of protein complexes (STRING database) and extend beyond contacts simply involving the low-complexity domains (LCDs) of proteins. This observation further highlights the importance of protein-protein interactions in the formation of more extensively aggregated and less dynamic SG. Moreover, while aberrant RNAs are physically bound to FUS^{P525L} and have more propensity to interact with FUS binding partners, they generally have reduced interaction capacity with the whole SG proteome.

While it has been suggested that RNA itself possesses aggregating potential (55), and undoubtedly contributes to the phenomena of phase separation, recent evidence supports the idea that RNA can

act as a buffer in neurotoxic aggregation processes. This has been observed for TDP-43 and FUS, where the role of the RNA-binding domain (RRM) and its interaction with RNA is vital in constraining LCD-driven aggregation. This phenomenon has also been demonstrated using ectopic sequences with affinity for TDP-43 (63). Furthermore, this finding aligns with a key paradigm of neurotoxic aggregation as a cytoplasmic event: condensation is influenced by the RNA-to-protein ratio, and in the nucleus this is limited by the abundance of newly transcribed RNA (64).

In line with this hypothesis, we observe that the altered dynamics of stress granules and the increased condensate aggregation state are associated with three main features of SG molecular components: 1) Increased protein-protein interactions.

2) Reduced architectural capacity of RNAs, involving a decreased propensity to form structures.

3) Diminished ability of SG RNAs to interact with the SG proteome.

Our results shed new light on the importance of RNA features in the aggregation process involving SG. This not only suggests that specific and distinct RNA characteristics are linked to pathological granules and aggregation processes, but also introduces intriguing possibilities for understanding how these phenomena can strongly impact neural physiology and be causal for neurodegeneration. Furthermore, these findings lay the groundwork for understanding aggregation processes, highlighting the role of RNA structures and interactions. Moreover, these results indirectly suggest that RNA modifications like m^6A, which are involved in determining structure and interaction with proteins (65), could potentially play a role in the pathological process leading to neurodegeneration by influencing RNA aggregation dynamics. This could be an intriguing field to explore, since recent evidence reports that m^6A methylated RNAs are more abundant in spinal cords of patients affected by ALS (66).

MATERIAL AND METHODS

Cloning and SK-N-BE stable cell line generation

The plasmid containing an N-terminal GFP-tagged form of human G3BP1 was a gentle gift from Roy Parker's Lab (University of Colorado Boulder). The GFP-G3BP1 sequence was subcloned in the ePB-BSD-EIF1a vector (67); briefly, GFP- G3BP1 was PCR-amplified with CloneAmp HiFi PCR Premix (TakaraBio, 639298) and cloned in the acceptor plasmid linearized with BglII and NotI restriction enzymes (Thermo Fisher, FD0084 and FD0596) using the InFusion Cloning kit following manufacturer protocol (TakaraBio, 638910).

Stable cell lines were generated taking advantage of the transposon-based PiggyBac technology. A PiggyBac HyPB transposase expression plasmid (68) was co- transfected with the ePB-EIF1a-GFP-G3BP1 vector with a 1:10 ratio; cells were selected in 10 μg/ml of blasticidin S (Thermo Scientific) for 7 days and then tested for consistent expression of the fluorescent protein. SK-N-BE cell lines expressing FUS^{WT} and FUS^{P525L} under a doxycycline-inducible promoter were already present in the lab (69).

RNA extraction and analysis

RNA extraction was performed with the Direct-zol Miniprep RNA Purification Kit (Zymo Research) with on-column DNase treatment, according to the manufacturer's instructions, with elution in 25 μL of RNase-free water (50 μL for total RNA Input samples). Total RNA was retro-transcribed with PrimeScript™ RT Reagent Kit (Takara) according to the manufacturer's instructions. For qPCR analysis, the PowerUp SYBR Green Master Mix (Thermo Fisher Scientific), following the manufacturer's protocol, in 15 μL reactions, on a QuantStudio 5 qPCR system (Thermo Scientific) with standard cycling conditions. For SG purification experiments, data were expressed as % of input: briefly, the Ct of Input samples for each target were adjusted to account for the dilution by subtracting the $\log_2$ of the dilution factor. Then, the % of input was calculated as 2^ (Adjusted input Ct - IP Ct) for each target.

Protein extraction and Western Blot analysis

Protein extracts was obtained using RIPA buffer, supplied with 1× Complete Protease Inhibitor Cocktail (Merck). Protein concentration was assessed using BCA protein assay (Thermo Scientific). Protein electrophoresis was performed using 4–15% Mini-PROTEAN TGX Precast acrylamide gel (Bio-Rad), and proteins were transferred to 0,45 μm nitrocellulose membrane, using the TransBlot Turbo System (Bio-rad), according to the manufacturer's instructions. Membranes were stained with Ponceau, blocked with 5% non-fat dry milk (Difco skim milk) for 1 h and incubated overnight at 4°C with the following primary antibodies: anti-GFP (Thermo Scientific, A-11122, 1:1000), anti-TIAR (BD Transduction, 610352, 1:1000), anti-FUS (SantaCruz Biotech, sc-47711, 1:2000), anti-GAPDH-HRP (SantaCruz Biotech, sc-47724, 1:1000), anti-FLAG (Merck, F3165, 1:1000), anti-ELAVL1 (SantaCruz Biotech, sc-5261, 1:1000), anti-DHX36 (Proteintech, 13159-1-AP, 1:500). After secondary antibody incubation for 1 hour at room temperature, protein detection was carried out with Immobilon Crescendo Western HRP substrate (Merck) using ChemiDoc™ MP System and images were analyzed using Image Lab™ Software (Bio-Rad).

Fluorescent microscopy analysis

Cells were plated on glass coverslips and treated as previously described. Samples were fixed in 4% paraformaldehyde for 10 min at room temperature (RT), followed by 3 washes in complete PBS. Permeabilization and blocking were performed by incubating the coverslips in PBS supplemented with 2% BSA, 0,2% TRITON X- 100 for 30 min at RT. Samples were then incubated with primary antibody diluted in PBS with 1% BSA for 1 hour at RT, followed by 3 washes for 5 min in PBS. Fluorophore-conjugated secondary antibodies (Thermo Fisher) were diluted 1:300 in PBS with 1% goat/donkey serum and incubated for 45 min at RT. After 3 PBS washes, nuclei were stained with 4',6-Diamidino-2-Phenylindole (DAPI, Sigma; diluted 1ug/ml in PBS) for 3 min at RT and coverslips were mounted with Prolong Diamond Mounting Media (Thermo Fisher, P-36961) on microscope slides. Fixed samples were imaged on a Nikon Instrument A1 Confocal Laser Microscope equipped with 60x and 20x objectives. Confocal images were collected with NIS-Elements AR software (Nikon): ND Acquisition module was used for multipoint acquisition of Z-stack images (150-175nm Z-spacing) of 4 um thickness.

Fluorescence Correlation Spectroscopy (FCS) measurements were performed as described in (36, 70). Cells were plated and FUS expression was induced as previously described; before starting the acquisition, culture medium was replaced with FluoroBriteTM DMEM (Thermo Fisher, A1896701) supplemented with 10% FBS, and sodium arsenite was added at a final concentration of 0.5 mM for the “stressed” samples. Acquisition of GFP/RFP fluorescence signals was obtained with a custom-made SPAD-FCS apparatus from the Molecular Microscopy and Spectroscopy division of the Italian Institute of Technology in Genoa.

SG purification

Stress granules purification for the RNA sequencing experiments was performed according to (30), with minor adjustments. After lysis, the protein concentration of samples was measured through BCA assay (Thermo Scientific); 3 mg of lysate were diluted in 1 mL of SG Lysis buffer and used for the purification, saving 5% as Total RNA input (stored in TRIZOL L/S, Thermo Scientific). SG were pelleted as in (30), while the immunoprecipitation step was performed by incubation with antibody-coupled beads. For each sample, 50 µL of DEPC-treated Protein G Dynabeads (Thermo Scientific) were incubated with 12.5 µg of anti-GFP antibody in rotation at 4°C for 1 hour, then washed thrice in 1 mL of SG Lysis Buffer and finally resuspended in 150 µL of SG Lysis Buffer. Prepared beads were added to the precleared SG-enriched fraction and incubated in rotation at 4°C for 2 hours. Washes and beads elution were performed as in (30). To isolate SG for qRT-PCR

analysis, mass spectrometry and Western Blot validation, Dynabeads were replaced with 25 µL/sample of GFP-Trap beads (Chromotech), which are covalently bound to a highly efficient alpaca-raised nanobody against the tag. For qRT-PCR, beads were washed and RNA was eluted as in (30). To isolate proteins for mass spectrometry and WB analysis, after capture the beads were washed twice with 1 mL Wash Buffer (10 mM TRIS-HCl pH 7.5, 150 mM NaCl, 0.05% Igepal, 0.5 mM EDTA) in rotation for 10 minutes at 4°C, then washed twice with 1 mL High Salt Wash Buffer (10 mM TRIS-HCl pH 7.5, 500 mM NaCl, 0.05% Igepal, 0.5 mM EDTA) in rotation for 10 minutes at 4°C. Beads were resuspended in 40 µL of 1x Laemmli Sample Buffer (Biorad) supplemented with 50 mM DTT, and heated to 85 °C for 10 minutes; the supernatant was collected and used for Western Blot analysis, or further processed for mass spectrometry (see the "SG proteome analysis" section).

SG RNA-Seq and bioinformatic analysis

Purified SGcore RNAs and relative inputs were retrieved in absence of doxy-induction (4 replicates); in condition of doxycycline-induced overexpression of FUS^{P525L} (2 replicates) and FUS^{WT} (2 replicates). RNA libraries for all samples were produced using Stranded Total RNA Prep with Ribo-Zero Plus (Illumina). All samples were sequenced on an Illumina Novaseq 6000 Sequencing system. Trimmomatic (71) and Cutadapt (72) were used to remove adapter sequences and poor quality bases; minimum read length after trimming was set to 35. Reads aligning to rRNAs were filtered out; this first alignment was performed using Bowtie2 software (v2.4.2) (https://bowtie-bio.sourceforge.net/bowtie2/index.shtml). STAR software (73) was used to align reads to GRCh38 genome. PCR duplicates were removed from all samples using MarkDuplicates command from Picard suite (https://broadinstitute.github.io/picard/). Uniquely mapping fragments were counted for each annotated gene (ensemble release 99) using htseq-counts software (74). edgeR software (75) was used to compare SG enriched RNAs to relative input samples. RNAs with $\log_2 FC > 1$ and FDR < 0.05 were defined "enriched" in SG core, while those with $\log_2 FC < -1$ and FDR < 0.05 were defined "depleted"; all the others were labelled as "invariant". SG from SMMC-7721 (32) cell line were re-analyzed using the previously described pipeline. For each expressed gene, the longest isoform was selected as reference. Only protein coding isoforms were taken in consideration for protein coding genes.

In the RNA-Seq analysis, the high variability of the library size of IP samples could affect the number of SG enriched RNAs identified. This could introduce technical bias especially when we need to compare the enriched RNAs of Ips performed in different conditions. To overcome this

bias, we performed a "Resampling analysis" (Supplementary Fig. 2I). We decided to use the observed number of fragments in FUS^{P525L} condition experiment and then we randomly sampled fragments from the alignment files of the other conditions in order to mirror the library size composition of the reference condition. In this way, we obtained samples with similar technical variability and with a simulated similar library size. Then we compared the enriched RNAs resulted by this analysis without the bias introduced by the variability of the resulted library size. Random sampling of fragments was performed with Picard suite using FilterSamReads function and using random subsets of fragments ids list as input (**https://broadinstitute.github.io/picard**).

In order to define the SG enriched RNAs, we used the already specified log_2FC threshold for each condition. However, the high variability of the immunoprecipitation-based experiments could result in high variability in the SG enrichment efficiency. Experiments with low enrichment efficiency could be penalized by a fixed threshold compared to others with a high efficiency, affecting the comparison of the defined enriched RNA sets. To overcome this technical limit, we performed an "Enrichment convergence analysis" (Supplementary Figure 2J). For each analyzed condition, we first ranked the expressed RNAs by their SG enrichment (log_2FC) and then we selected a fixed number of enriched RNAs equal for each dataset. We then compared these three equally sized RNA sets using Venn Diagrams. We repeated this analysis selecting different fixed numbers of enriched RNAs, gradually reducing the inclusion of the noise from the background transcriptome (fixed numbers of top enriched RNAs: 5000, 4500, 4000, 3500, 3000, 2500, 2000, 1500, 1000, 500).

GO term enrichment analysis

Gene ontology was performed using WebGestalt R tool (v0.4.4)(76) applying weighted set cover reduction.

K-mers enrichment analysis

For each analyzed condition (FUS^{P525L}, FUS^{WT} and no DOXY) the SG enriched and depleted RNAs were selected. For each of the two groups the frequency of every possible k-mer of k=7 was calculated and used to compute 7-mer Z-scores. Then the difference between 7-mer Z-scores in the SG enriched group and SG depleted group was calculated and the top 50 k-mers from both sides of differences distributions were selected as top enriched or depleted k-mers in SG. For each condition, sequence logo of the top enriched 50 k-mers were computed using Logomaker Python package (77). After that, the average transcript frequency of each k-mer was calculated for SG enriched RNAs and background distribution (invariant group), and their ratio (fold-change, FC) was

computed. Then, the log_2FC of the top enriched and depleted k-mers of all the analyzed conditions were used for Heatmap representation and clustering performed using ComplexHeatmap R package (https://bioconductor.org/packages/release/bioc/html/ComplexHeatmap.html).

RNA structure prediction and analysis

In RNA structure propensity analysis three indipendents random sampling of 200 mRNAs from GAIN and LOSS groups was performed imposing the similarity of 5'UTR, CDS and 3'UTR regions between groups (Mann Whitney-U Test, PValue > 0.05). RNA secondary structures were predicted using RNAfold algorithm (37) using standard parameters and CROSS algorithm (**http://service.tartaglialab.com/update_submission/736048/9393232b64**) using human PARS data settings. For RNAfold prediction, the structuration score was calculated by the number of structured nucleotides over the length of the region/transcript. For RNA accessibility analysis the previously described procedure of sampling was applied to GAIN and LOSS groups, but selecting 50 mRNAs that were also detected in HEK293T cell line in the DMS-Map-Seq experiment (39). DMS-Map-Seq data were retrieved from RASP database (78).

SG Proteome analysis

After immunoprecipitation and washes (see "SG purification" section), beads were washed in 1 mL of Digestion Buffer (50 mM Sodium Bicarbonate) and resuspended in 120 µL of Digestion Buffer. Disulfide bonds were reduced on-beads by adding DTT to a final concentration of 10 mM and incubating at 56°C for 30 minutes. Proteins were alkylated by adding Iodoacetamide to a final concentration of 20 mM and incubating beads at room temperature for 20 minutes. On-beads trypsin digestion was then performed by resuspending the beads in 200 µL of Digestion Buffer supplemented with 2 µL of 0.5 µg/µL Sequencing Grade Trypsin (Merck). Samples were incubated overnight at 37°C with constant agitation at 750 rpm, then each sample (~8 µg of peptides) was dried using a Speed-Vac and resuspended in 16 µL acetonitrile, 3% added with 0.1% formic acid. Tryptic peptides were desalted on a trapping column then loaded on a Aurora C18 column (75 mm x 250 mm, 1.6 µm particle size - Ion Opticks, Fitzroy, Australia) and separated using a Dionex Ultimate 3000 nano-LC system (Thermo Scientific™). Eluents were A: water + 0.1% formic acid and B: water:acetronitrile=20:80 (*v/v*) + 0.1% formic acid. Injection volume was 2 µL, flow was 0.300 µL/min, the column temperature was 40°C, samples were eluted with a gradient program: 0.0 -1.0 min 3 % B; 1.0 – 5.0 min 3 to 5 % B; 5.0 – 35.0 min 5 to 31 % B; 35.0 – 40.0 min 31 to 44% B, 40.0 – 41.0 min 44.0 – 95 % B, 41.0 – 46.0 min 95% B and 46.0 – 47.0 min back to 3 % B. The column was then reconditioned for 13 min. The total run time was 60.0 min. Separated peptides

were analyzed using an Orbitrap Exploris 480 Mass Spectrometer (Thermo Scientific™), in positive ESI mode. The capillary voltage was set to 2.0 kV, the RF lens was set to 40% and the normalized AGC target % was set to 300, maximum injection time 50 ms. The acquisition was performed in Data Dependent mode (DDA) with a scan range set from 350 to 1500 m/z, resolution 120000. The Intensity threshold defined for a data dependent scan was set to 10000, MS/MS spectra were acquired in HCD mode with a collision energy of 30%. The raw data were processed using Thermo Proteome Discoverer software (doi: 10.3390/proteomes9010015), 1% false discovery rate (FDR) threshold and at least two peptides were used for protein identification. Proteins from Mass Spectrometry data were annotated using biomaRt (ensemble 109 release).

Protein-Protein physical Interactions were retrieved from STRING database (v11.5). Protein-RNA interactions predictions performed by catRAPID software were retrieved from RNAct database (48). For each couple of interacting protein and RNA, the GENCODE basic isoform with the highest interaction score was selected as representative. The predicted preferential binding with GAIN over LOSS group was performed calculating for each SG protein the median normalized score with GAIN RNAs and with LOSS RNAs. Then Z-score were calculated from the ratios of median normalized scores (GAIN / LOSS) of the whole set of SG proteins.

Protein-RNA interactions analysis of experimental datasets performed using data retrieved from POSTAR3 database (49). The preferential binding with GAIN over LOSS group was performed calculating for each RBP available in POSTAR3 database the fraction of GAIN RNAs identified as interactors of the RBP over the fraction of LOSS RNAs interactors.

Quantification and statistical analysis

The distribution and deviation of data shown in the figures of this work, the statistical tests used to calculate significant differences, and the exact value of n (e.g., number of biological replicates of the experiments) are denoted in figure legends. In main figure legends “SD” stands for “standard deviation” and “SEM” stands for “standard error mean”. Significance values were depicted in the figures using the following key legend: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

AUTHOR CONTRIBUTION

I.B., D.M. and A.S. conceptualized the project. I.B., G.V and A.A. supervised and coordinated the experiments. D.M. set up the experimental system. D.M., F.C., and L.S.M. performed molecular biology experiments. A.S. designed and carried out all the bioinformatic analyses. E.V. performed

and analyzed IF experiments. G.d.T. and A.G. analyzed results and discussed data. F.C., E.P. and S.Z. performed the FCS experiments. N.L. performed mass spectrometry. I.B., D.M. and A.S. wrote the original manuscript with contribution of all the authors.

ACKNOWLEDGMENTS

We thank Prof. Alberto Diaspro and Dr. Paolo Bianchini (Nanoscopy & NIC@IIT, Istituto Italiano di Tecnologia), Dr. Michele Oneto (Nikon Imaging Center) and Marco Scotto (Molecular Microscopy and Spectroscopy, Istituto Italiano di Tecnologia) for experimental support in confocal and SIM microscopy. We are also grateful to Dr. Diego Vozzi and the Genomic Facility of CHT for support in RNA sequencing experiments. A special thank goes to the technical and administrative staff of IIT, and to all the members of the RNA Initiative@IIT community for thought-provoking discussions. We also wish to thank Angelo d'Angelo for fruitful discussion regarding bioinformatic analyses. The project was partially supported by grants from ERC-2019-SyG 855923-ASTRA, AIRC IG 2019 Id. 23053, PRIN 2017 2017P352Z4 and "National Center for Gene Therapy and Drugbased on RNA Technology" (CN00000041), NextGenerationEU PNRR MUR, together with intramural IIT funding.

CONFLICT OF INTEREST

G.V. has a personal financial interest (co-founder) in Genoa Instruments, Italy.

FIGURE LEGENDS

Figure 1

A) Upper panel. Summary of the amount of RNAs detected in RNAseq of Stress Granule (SGs) purification experiment, and their partition among enriched, invariant and depleted groups.

Lower panel. Volcano plot showing the log2 fold change (SGenr/Input) and the -log10 p-value of SG enrichment for each RNA detected. Significantly enriched and depleted RNAs (FDR < 0.05, |logFC| > 1) or invariant RNAs are indicated by red or gray dots, respectively.

B) Venn diagram depicting the overlap between SG-enriched RNAs in neuroblastoma (SK-N-BE, this study), osteoblastoma (U2OS) and endocervical adenocarcinoma (SMMC-7721) cell lines.

C) Left panel. Pie chart depicting the fraction of protein coding and noncoding RNAs in the SG common *core*.

Right panel. Set of the 15 top enriched mRNAs and all the ncRNAs enriched the SG common *core*.

D) Dotplot depicting GO over-represented categories in the SG common *core*. Only significant categories (FDR < 0.05) of Cellular Component database were depicted. X-axis represents category enrichment score while y-axis reports the GO category description. Dot size represents the amount of RNAs in
the analyzed group that overlap the category, while color code (green) reports the significativity of the enrichment.

E) Heatmap depicting the association of RNA features with SG enrichment in SK-N-BE (upper panel), SMMC-7721 (middle panel) and U2OS (lower panel) cell lines.
For the defined transcripts group («depleted», «invariant» and «enriched») of each cell line the following characteristics were described: median log2FC (heatmap); GC content (density plot); RNA length (nt,boxplot); transcripts biotypes fraction (barplot).

F) Dotplot depicting GO over-represented categories of the SK-N-BE SG specific RNAs. Graph features are the same as panel D.

G) Heatmap (lower panel) depicting the relationship between RNA length and GO categories enriched in "SG common core". The 10 transcripts groups represented were defined by length-based stratification of the transcriptome. The log2 Odds Ratio depicted in the heatmap shows the depletion (blue) or the enrichment(red) of each GO category in the defined group compared to the remaining fraction of the transcriptome. Boxplot (upper panel) represents the distributions of RNAs length for the 10 transcripts groups.

Figure 2

A) Scatter plot showing measurements of GFP-G3BP1 SpotVariation FCS with the SPAD array detector. In y-axis the S_{conf} measurement indicates the strength of confinement of the molecule, while the x-axis depicts the D_{app} that indicates the velocity of the molecule. Blue dots indicate measurments before oxidative stress induction while white dots indicate measurements after stress induction. FUS WT and FUS P525L conditions are indicated in the left and right panel, respectively.

B) Volcano plot showing for each RNA detected in the RNA-seq experiment in FUS WT (upper panel) or FUS P525L condition (lower panel) the log2 fold change (SGenr/Input) and the -log10 p-value of SG enrichment. Significantly enriched or depleted RNAs (FDR < 0.05, |logFC| > 1) or invariant RNAs are indicated by red or gray dots, respectively.

C) Venn diagrams showing the overlap between SG enriched RNAs in no DOXY, FUS WT and FUS P525L conditions. The amount of RNAs resulted enriched in SGs of each condition was indicated. Arrows indicate the fraction of RNAs indicated as FUS P525L GAIN, COMMON and FUS P525L LOSS.

D) Dotplot depicting GO over-represented categories in the FUS P525L LOSS (upper panel), COMMON (middle panel) and FUS P525L GAIN (lower panel) groups. Only significant categories (FDR < 0.05) of Biological Process database were depicted. X-axis represents category enrichment score while y-axis reports the GO category description. Dot size represents the amount of RNAs in the analyzed group that overlap the category while color code (red) reports the significance of the enrichment.

Figure 3

A) Heatmap depicting the association of RNA features with SG enrichment in FUS P525L, FUS WT and no DOXY conditions. For the defined transcripts group («depleted», «invariant» and «enriched”) of each cell line the following characteristics were described: median log2FC (heatmap); GC content (density plot); RNA length (nt, boxplot); 5’ UTR, CDS and 3’UTR length (nt, boxplot); transcripts biotypes fraction (barplot).

B) Sequence logo of the top 50 7-mers over-represented in SG enriched RNAs compared to SG depleted in no DOXY, FUS WT and FUS P525L conditions.

C) Heatmap showing the clusters of 7mers frequency fold between SG enriched over invariant RNAs in no DOXY, FUS WT and FUS P525L conditions. For each 7mer, the log2 fold between average frequencies in groups was depicted. Only the union of the top 50 over-represented and 50 under-represented 7-mers in each condition was depicted. Pie chart represent the percentage of AU or GC nucleotides in the 7-mers enriched (cluster 1) and depleted (cluster 5) in FUS P525L SG.

D) Boxplot depicting the distribution of RNA structuration degree predicted using RNAfold algorithm between FUS P525L GAIN and LOSS groups. For each RNA, the minimum free energy

(MFE) structure was taken in consideration. The differences between groups were calculated using Mann-Whitney U test. The first sampling was reported. Size of sampling = 200 RNAs per group.

E) Representative heatmap showing the fold between average accessibility of FUS P525L GAIN over LOSS groups. The differences between groups were calculated using Mann-Whitney U test. Only significant differences were depicted (P-value < 0.05). The first sampling was reported. Size of sampling = 200 RNAs per group.

F and G) Boxplot depicting the distribution of RNA structuration score predicted by CROSS algorithm (F) and RNA average accessibility calculated using DMS-Map-Seq data (G) between FUS P525L GAIN and LOSS groups. The differences between groups were calculated using Mann-Whitney U test. The first sampling was reported. Size of sampling = 200 RNAs per group for CROSS (F) and 50 RNAs per group for DMS-Map-Seq (G).

Figure 4

A) Barplot depicting the enrichemnt of endogenous FUS P525L PAR-CLIP interactors retrieved from iPSC derived motorneurons (MNs) in GAIN and COMMON compared to LOSS groups, and the remaining fraction of the transcriptome. Only RNAs detected in both systems (SK-N-BE and MNs) were taken in consideration. Black color refers to the fraction of RNAs not bound by FUS, while white color refers to FUS P525L RNA interactors. The significance of enrichment was calculated with Fisher's exact test (two-tails), after multiple-test correction (FDR).

B) Venn diagram depicting the overlap among SG proteomes in FUS WT and FUS P525L conditions and FUS protein interactors. Arrows indicate the fraction of proteins defined as SG proteins AND FUS interactors or SG proteins NOT FUS interactors.

C) Barplot depicting the enrichment of protein-protein physical interactions (PPI) in the SG proteins AND FUS interactors compared to the three samplings performed from the SG protein NOT FUS interactors group.

D) Network of protein-protein physical interactions in SG proteins AND FUS interactors and in one of the three sampling performed from the SG protein NOT FUS interactors group.

E) Heatmap depicting the SG proteome ranked by interaction propensity for GAIN over the LOSS group. Heatmap colors from blue to red report the ratio of SG proteome average interaction propensity (RNAact, catRAPID) of GAIN over LOSS groups, the barplot on the left reports the Z-score of these ratios. The SG proteome was stratified in 5 groups with first group (bottom) that displayed the highest interaction propensity for the GAIN group, while the last group (top) displayed the highest interaction propensity for the LOSS group.

F) Pie charts depicting the fraction of SG proteins AND FUS interactors and SG proteins NOT FUS interactors in the first (bottom) and last group (top). Bar plot (middle) displays the indicated fraction in every stratification group. Black color refers to SG proteins NOT FUS interactors while white color refers to SG proteins AND FUS interactors. For GAIN and LOSS comparison the matched length sampling were used. The first sampling was depicted. Significance of the enrichment of SG proteins AND FUS interactors in the first group compared to the last was assessed with Fisher's exact test. These data refer to the analysis on RNA-protein interaction made with catRAPID algorithm.

G) Boxplot depicting the distribution of protein-RNA interaction scores retrieved from RNAact database (catRAPID prediciton) among SG proteome and GAIN, COMMON and LOSS RNA groups. The significance between groups was assessed using Mann-Whitney U test.

REFERENCES

1. 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.

2. Das,S., Vera,M., Gandin,V., Singer,R.H. and Tutucci,E. (2021) Intracellular mRNA transport and localized translation. *Nat Rev Mol Cell Biol*, **22**, 483–504.

3. Hirose,T., Ninomiya,K., Nakagawa,S. and Yamazaki,T. (2023) A guide to membraneless organelles and their various roles in gene regulation. *Nat Rev Mol Cell Biol*, **24**, 288–304.

4. Brangwynne,C.P., Eckmann,C.R., Courson,D.S., Rybarska,A., Hoege,C., Gharakhani,J., Jülicher,F. and Hyman,A.A. (2009) Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation. *Science*, **324**, 1729–1732.

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

6. Mitrea,D.M. and Kriwacki,R.W. (2016) Phase separation in biology; functional organization of a higher order. *Cell Commun Signal*, **14**, 1.

7. Gomes,E. and Shorter,J. (2019) The molecular language of membraneless organelles. *Journal of Biological Chemistry*, **294**, 7115–7127.

8. Van Treeck,B. and Parker,R. (2018) Emerging Roles for Intermolecular RNA-RNA Interactions in RNP Assemblies. *Cell*, **174**, 791–802.

9. Garcia-Jove Navarro,M., Kashida,S., Chouaib,R., Souquere,S., Pierron,G., Weil,D. and Gueroui,Z. (2019) RNA is a critical element for the sizing and the composition of phase-separated RNA–protein condensates. *Nat Commun*, **10**, 3230.

10. Roden,C. and Gladfelter,A.S. (2021) RNA contributions to the form and function of biomolecular condensates. *Nat Rev Mol Cell Biol*, **22**, 183–195.

11. Bevilacqua,P.C., Williams,A.M., Chou,H.-L. and Assmann,S.M. (2022) RNA multimerization as an organizing force for liquid–liquid phase separation. *RNA*, **28**, 16–26.

12. Boeynaems,S., Alberti,S., Fawzi,N.L., Mittag,T., Polymenidou,M., Rousseau,F., Schymkowitz,J., Shorter,J., Wolozin,B., Van Den Bosch,L., *et al.* (2018) Protein Phase Separation: A New Phase in Cell Biology. *Trends in Cell Biology*, **28**, 420–435.

13. Molliex,A., Temirov,J., Lee,J., Coughlin,M., Kanagaraj,A.P., Kim,H.J., Mittag,T. and Taylor,J.P. (2015) Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization. *Cell*, **163**, 123–133.

14. Wiedner,H.J. and Giudice,J. (2021) It's not just a phase: function and characteristics of RNA-binding proteins in phase separation. *Nat Struct Mol Biol*, **28**, 465–473.

15. Glauninger,H., Wong Hickernell,C.J., Bard,J.A.M. and Drummond,D.A. (2022) Stressful steps: Progress and challenges in understanding stress-induced mRNA condensation and accumulation in stress granules. *Molecular Cell*, **82**, 2544–2556.

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

17. 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.

18. Shelkovnikova,T.A., Dimasi,P., Kukharsky,M.S., An,H., Quintiero,A., Schirmer,C., Buée,L., Galas,M.-C. and Buchman,V.L. (2017) Chronically stressed or stress-preconditioned neurons fail to maintain stress granule assembly. *Cell Death Dis*, **8**, e2788–e2788.

19. Advani,V.M. and Ivanov,P. (2020) Stress granule subtypes: an emerging link to neurodegeneration. *Cell. Mol. Life Sci.*, **77**, 4827–4845.

20. Jeon,P. and Lee,J.A. (2021) Dr. Jekyll and Mr. Hyde? Physiology and Pathology of Neuronal

Stress Granules. *Front. Cell Dev. Biol.*, **9**, 609698.

21. Wolozin,B. and Ivanov,P. (2019) Stress granules and neurodegeneration. *Nat Rev Neurosci*, **20**, 649–666.

22. Nedelsky,N.B. and Taylor,J.P. (2019) Bridging biophysics and neurology: aberrant phase transitions in neurodegenerative disease. *Nat Rev Neurol*, **15**, 272–286.

23. Volk,A.E., Weishaupt,J.H., Andersen,P.M., Ludolph,A.C. and Kubisch,C. (2018) Current knowledge and recent insights into the genetic basis of amyotrophic lateral sclerosis. *Medizinische Genetik*, **30**, 252–258.

24. 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.

25. Murray,D.T., Kato,M., Lin,Y., Thurber,K.R., Hung,I., McKnight,S.L. and Tycko,R. (2017) Structure of FUS Protein Fibrils and Its Relevance to Self-Assembly and Phase Separation of Low-Complexity Domains. *Cell*, **171**, 615-627.e16.

26. Kato,M. and McKnight,S.L. (2021) The low-complexity domain of the FUS RNA binding protein self-assembles via the mutually exclusive use of two distinct cross-β cores. *Proc. Natl. Acad. Sci. U.S.A.*, **118**, e2114412118.

27. Bosco,D.A., Lemay,N., Ko,H.K., Zhou,H., Burke,C., Kwiatkowski,T.J., Sapp,P., McKenna-Yasek,D., Brown,R.H. and Hayward,L.J. (2010) Mutant FUS proteins that cause amyotrophic lateral sclerosis incorporate into stress granules. *Human Molecular Genetics*, **19**, 4160–4175.

28. Lenzi,J., De Santis,R., de Turris,V., Morlando,M., Laneve,P., Calvo,A., Caliendo,V., Chiò,A., Rosa,A. and Bozzoni,I. (2015) ALS mutant FUS proteins are recruited into stress granules in induced Pluripotent Stem Cells (iPSCs) derived motoneurons. *Disease Models & Mechanisms*, 10.1242/dmm.020099.

29. Szewczyk,B., Günther,R., Japtok,J., Frech,M.J., Naumann,M., Lee,H.O. and Hermann,A. (2023) FUS ALS neurons activate major stress pathways and reduce translation as an early protective mechanism against neurodegeneration. *Cell Reports*, **42**, 112025.

30. Khong,A., Jain,S., Matheny,T., Wheeler,J.R. and Parker,R. (2018) Isolation of mammalian stress granule cores for RNA-Seq analysis. *Methods*, **137**, 49–54.

31. 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. *Molecular Cell*, **68**, 808-820.e5.

32. Chen,S., Zhang,J. and Zhao,F. (2022) Screening Linear and Circular RNA Transcripts from

Stress Granules. *Genomics, Proteomics & Bioinformatics*, 10.1016/j.gpb.2022.01.003.

33. Baradaran-Heravi,Y., Van Broeckhoven,C. and van der Zee,J. (2020) Stress granule mediated protein aggregation and underlying gene defects in the FTD-ALS spectrum. *Neurobiology of Disease*, **134**, 104639.

34. Chiò,A., Restagno,G., Brunetti,M., Ossola,I., Calvo,A., Mora,G., Sabatelli,M., Monsurrò,M.R., Battistini,S., Mandrioli,J., *et al.* (2009) Two Italian kindreds with familial amyotrophic lateral sclerosis due to FUS mutation. *Neurobiology of Aging*, **30**, 1272–1275.

35. Lo Bello,M., Di Fini,F., Notaro,A., Spataro,R., Conforti,F.L. and La Bella,V. (2017) ALS-Related Mutant FUS Protein Is Mislocalized to Cytoplasm and Is Recruited into Stress Granules of Fibroblasts from Asymptomatic FUS P525L Mutation Carriers. *Neurodegenerative Diseases*, **17**, 292–303.

36. Slenders,E., Castello,M., Buttafava,M., Villa,F., Tosi,A., Lanzanò,L., Koho,S.V. and Vicidomini,G. (2021) Confocal-based fluorescence fluctuation spectroscopy with a SPAD array detector. *Light Sci Appl*, **10**, 31.

37. Lorenz,R., Bernhart,S.H., Höner zu Siederdissen,C., Tafer,H., Flamm,C., Stadler,P.F. and Hofacker,I.L. (2011) ViennaRNA Package 2.0. *Algorithms for Molecular Biology*, **6**, 26.

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

39. Rouskin,S., Zubradt,M., Washietl,S., Kellis,M. and Weissman,J.S. (2014) Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature*, **505**, 701–705.

40. De Santis,R., Alfano,V., de Turris,V., Colantoni,A., Santini,L., Garone,M.G., Antonacci,G., Peruzzi,G., Sudria-Lopez,E., Wyler,E., *et al.* (2019) Mutant FUS and ELAVL4 (HuD) Aberrant Crosstalk in Amyotrophic Lateral Sclerosis. *Cell Rep*, **27**, 3818-3831.e5.

41. Asadi,M.R., Sadat Moslehian,M., Sabaie,H., Jalaiei,A., Ghafouri-Fard,S., Taheri,M. and Rezazadeh,M. (2021) Stress Granules and Neurodegenerative Disorders: A Scoping Review. *Frontiers in Aging Neuroscience*, **13**.

42. An,H., Litscher,G., Watanabe,N., Wei,W., Hashimoto,T., Iwatsubo,T., Buchman,V.L. and Shelkovnikova,T.A. (2022) ALS-linked cytoplasmic FUS assemblies are compositionally different from physiological stress granules and sequester hnRNPA3, a novel modifier of FUS toxicity. *Neurobiology of Disease*, **162**, 105585.

43. 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.

44. Groen,E.J.N., Fumoto,K., Blokhuis,A.M., Engelen-Lee,J., Zhou,Y., van den Heuvel,D.M.A., Koppers,M., van Diggelen,F., van Heest,J., Demmers,J.A.A., *et al.* (2013) ALS-associated mutations in FUS disrupt the axonal distribution and function of SMN. *Human Molecular Genetics*, **22**, 3690–3704.

45. Blokhuis,A.M., Koppers,M., Groen,E.J.N., van den Heuvel,D.M.A., Dini Modigliani,S., Anink,J.J., Fumoto,K., van Diggelen,F., Snelting,A., Sodaar,P., *et al.* (2016) Comparative interactomics analysis of different ALS-associated proteins identifies converging molecular pathways. *Acta Neuropathol*, **132**, 175–196.

46. Kamelgarn,M., Chen,J., Kuang,L., Arenas,A., Zhai,J., Zhu,H. and Gal,J. (2016) Proteomic analysis of FUS interacting proteins provides insights into FUS function and its role in ALS. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, **1862**, 2004–2014.

47. Reber,S., Jutzi,D., Lindsay,H., Devoy,A., Mechtersheimer,J., Levone,B.R., Domanski,M., Bentmann,E., Dormann,D., Mühlemann,O., *et al.* (2021) The phase separation-dependent FUS interactome reveals nuclear and cytoplasmic function of liquid–liquid phase separation. *Nucleic Acids Research*, **49**, 7713–7731.

48. Lang,B., Armaos,A. and Tartaglia,G.G. (2019) RNAct: Protein–RNA interaction predictions for model organisms with supporting experimental data. *Nucleic Acids Research*, **47**, D601–D606.

49. Zhao,W., Zhang,S., Zhu,Y., Xi,X., Bao,P., Ma,Z., Kapral,T.H., Chen,S., Zagrovic,B., Yang,Y.T., *et al.* (2022) POSTAR3: an updated platform for exploring post-transcriptional regulation coordinated by RNA-binding proteins. *Nucleic Acids Research*, **50**, D287–D294.

50. Tran,H., Schilling,M., Wirbelauer,C., Hess,D. and Nagamine,Y. (2004) Facilitation of mRNA Deadenylation and Decay by the Exosome-Bound, DExH Protein RHAU. *Molecular Cell*, **13**, 101–111.

51. Gallouzi,I.-E., Brennan,C.M., Stenberg,M.G., Swanson,M.S., Eversole,A., Maizels,N. and Steitz,J.A. (2000) HuR binding to cytoplasmic mRNA is perturbed by heat shock. *Proceedings of the National Academy of Sciences*, **97**, 3073–3078.

52. Deng,H., Gao,K. and Jankovic,J. (2014) The role of FUS gene variants in neurodegenerative diseases. *Nat Rev Neurol*, **10**, 337–348.

53. Wolozin,B. (2012) Regulated protein aggregation: stress granules and neurodegeneration. *Molecular Neurodegeneration*, **7**, 56.

54. Gasset-Rosa,F., Lu,S., Yu,H., Chen,C., Melamed,Z., Guo,L., Shorter,J., Cruz,S.D. and Cleveland,D.W. (2019) Cytoplasmic TDP-43 De-mixing Independent of Stress Granules Drives Inhibition of Nuclear Import, Loss of Nuclear TDP-43, and Cell Death. *Neuron*, **102**, 339-357.e7.

55. Ripin,N. and Parker,R. (2022) Are stress granules the RNA analogs of misfolded protein

aggregates? *RNA*, **28**, 67–75.

56. Adinolfi,M., Pietrosanto,M., Parca,L., Ausiello,G., Ferrè,F. and Helmer-Citterich,M. (2019) Discovering sequence and structure landscapes in RNA interaction motifs. *Nucleic Acids Research*, **47**, 4958–4969.

57. An,J.J., Gharami,K., Liao,G.-Y., Woo,N.H., Lau,A.G., Vanevski,F., Torre,E.R., Jones,K.R., Feng,Y., Lu,B., *et al.* (2008) Distinct Role of Long 3′ UTR BDNF mRNA in Spine Morphology and Synaptic Plasticity in Hippocampal Neurons. *Cell*, **134**, 175–187.

58. Tushev,G., Glock,C., Heumüller,M., Biever,A., Jovanovic,M. and Schuman,E.M. (2018) Alternative 3′ UTRs Modify the Localization, Regulatory Potential, Stability, and Plasticity of mRNAs in Neuronal Compartments. *Neuron*, **98**, 495-511.e6.

59. Wilson,D.M., Cookson,M.R., Van Den Bosch,L., Zetterberg,H., Holtzman,D.M. and Dewachter,I. (2023) Hallmarks of neurodegenerative diseases. *Cell*, **186**, 693–714.

60. Sanchez de Groot,N., Armaos,A., Graña-Montes,R., Alriquet,M., Calloni,G., Vabulas,R.M. and Tartaglia,G.G. (2019) RNA structure drives interaction with proteins. *Nat Commun*, **10**, 3246.

61. Vandelli,A., Cid Samper,F., Torrent Burgas,M., Sanchez de Groot,N. and Tartaglia,G.G. (2022) The Interplay Between Disordered Regions in RNAs and Proteins Modulates Interactions Within Stress Granules and Processing Bodies. *Journal of Molecular Biology*, **434**, 167159.

62. Miladi,M., Montaseri,S., Backofen,R. and Raden,M. (2019) Integration of accessibility data from structure probing into RNA–RNA interaction prediction. *Bioinformatics*, **35**, 2862–2864.

63. Mann,J.R., Gleixner,A.M., Mauna,J.C., Gomes,E., DeChellis-Marks,M.R., Needham,P.G., Copley,K.E., Hurtle,B., Portz,B., Pyles,N.J., *et al.* (2019) RNA Binding Antagonizes Neurotoxic Phase Transitions of TDP-43. *Neuron*, **102**, 321-338.e8.

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

65. Liu,N., Zhou,K.I., Parisien,M., Dai,Q., Diatchenko,L. and Pan,T. (2017) N6-methyladenosine alters RNA structure to regulate binding of a low-complexity protein. *Nucleic Acids Research*, **45**, 6051–6063.

66. McMillan,M., Gomez,N., Hsieh,C., Bekier,M., Li,X., Miguez,R., Tank,E.M.H. and Barmada,S.J. (2023) RNA methylation influences TDP43 binding and disease pathogenesis in models of amyotrophic lateral sclerosis and frontotemporal dementia. *Molecular Cell*, **83**, 219-236.e7.

67. Martone,J., Mariani,D., Santini,T., Setti,A., Shamloo,S., Colantoni,A., Capparelli,F., Paiardini,A., Dimartino,D., Morlando,M., *et al.* (2020) SMaRT lncRNA controls translation of a G-

quadruplex-containing mRNA antagonizing the DHX36 helicase. *EMBO reports*, **21**, e49942.

68. Yusa,K., Zhou,L., Li,M.A., Bradley,A. and Craig,N.L. (2011) A hyperactive piggyBac transposase for mammalian applications. *Proceedings of the National Academy of Sciences*, **108**, 1531–1536.

69. Morlando,M., Dini Modigliani,S., Torrelli,G., Rosa,A., Di Carlo,V., Caffarelli,E. and Bozzoni,I. (2012) FUS stimulates microRNA biogenesis by facilitating co-transcriptional Drosha recruitment. *The EMBO Journal*, **31**, 4502–4510.

70. Perego,E., Zappone,S., Castagnetti,F., Mariani,D., Vitiello,E., Rupert,J., Zacco,E., Tartaglia,G.G., Bozzoni,I., Slenders,E., *et al.* (2023) Content-enriched fluorescence lifetime fluctuation spectroscopy to study bio-molecular condensate formation. 10.1101/2023.06.09.544221.

71. Bolger,A.M., Lohse,M. and Usadel,B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

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

73. Dobin,A., Davis,C.A., Schlesinger,F., Drenkow,J., Zaleski,C., Jha,S., Batut,P., Chaisson,M. and Gingeras,T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.

74. Anders,S., Pyl,P.T. and Huber,W. (2015) HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166–169.

75. Robinson,M.D., McCarthy,D.J. and Smyth,G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.

76. Liao,Y., Wang,J., Jaehnig,E.J., Shi,Z. and Zhang,B. (2019) WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Research*, **47**, W199–W205.

77. Tareen,A. and Kinney,J.B. (2020) Logomaker: beautiful sequence logos in Python. *Bioinformatics*, **36**, 2272–2274.

78. Li,P., Zhou,X., Xu,K. and Zhang,Q.C. (2021) RASP: an atlas of transcriptome-wide RNA secondary structure probing data. *Nucleic Acids Research*, **49**, D183–D191.

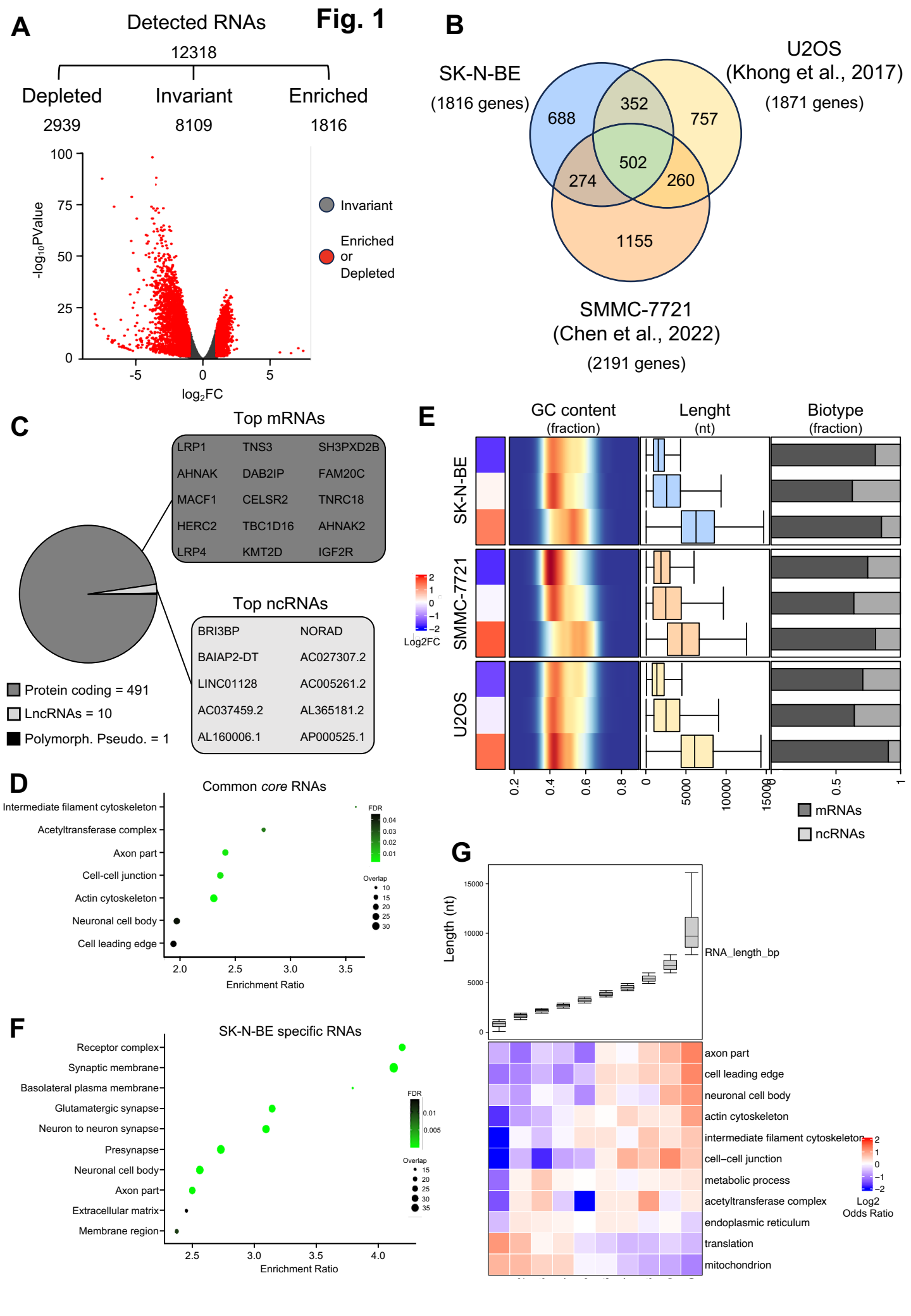
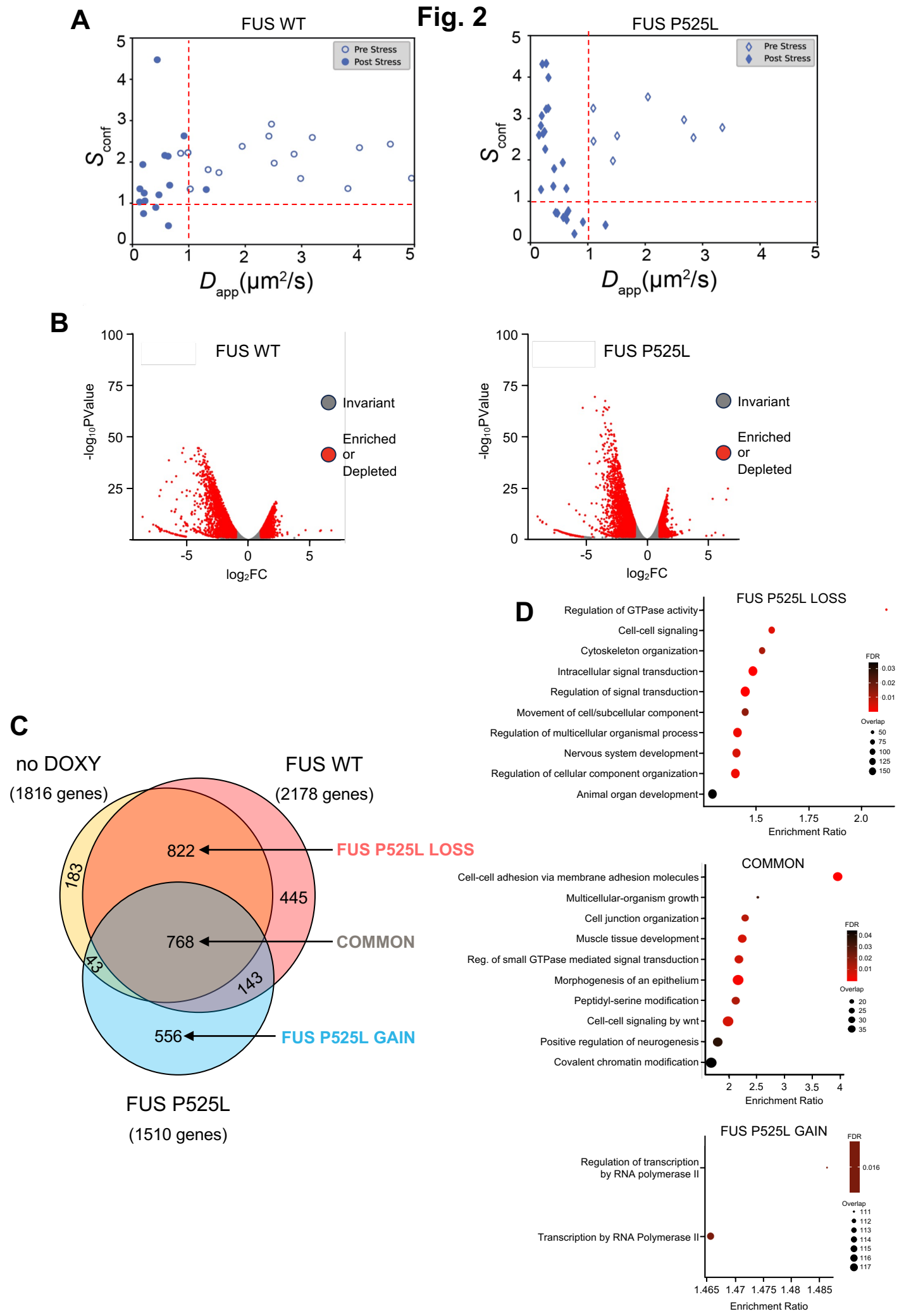


Fig. 1
A
Detected RNAs
12318
Depleted
2939
Invariant
8109
Enriched
1816
-log10PValue
log2FC
Invariant
Enriched or Depleted
B
SK-N-BE
(1816 genes)
U2OS
(Khong et al., 2017)
(1871 genes)
SMMC-7721
(Chen et al., 2022)
(2191 genes)
688
352
757
502
274
260
1155
C
Top mRNAs
LRP1 TNS3 SH3PXD2B
AHNAK DAB2IP FAM20C
MACF1 CELSR2 TNRC18
HERC2 TBC1D16 AHNAK2
LRP4 KMT2D IGF2R
Top ncRNAs
BRI3BP NORAD
BAIAP2-DT AC027307.2
LINC01128 AC005261.2
AC037459.2 AL365181.2
AL160006.1 AP000525.1
Protein coding = 491
LncRNAs = 10
Polymorph. Pseudo. = 1
D
Common core RNAs
Intermediate filament cytoskeleton
Acetyltransferase complex
Axon part
Cell-cell junction
Actin cytoskeleton
Neuronal cell body
Cell leading edge
Enrichment Ratio
FDR
Overlap
E
GC content (fraction)
Lenght (nt)
Biotype (fraction)
SK-N-BE
SMMC-7721
U2OS
Log2FC
mRNAs
ncRNAs
F
SK-N-BE specific RNAs
Receptor complex
Synaptic membrane
Basolateral plasma membrane
Glutamatergic synapse
Neuron to neuron synapse
Presynapse
Neuronal cell body
Axon part
Extracellular matrix
Membrane region
Enrichment Ratio
FDR
Overlap
G
Length (nt)
RNA_length_bp
axon part
cell leading edge
neuronal cell body
actin cytoskeleton
intermediate filament cytoskeleton
cell–cell junction
metabolic process
acetyltransferase complex
endoplasmic reticulum
translation
mitochondrion
Log2 Odds Ratio

Fig. 2


A
FUS WT
Pre Stress
Post Stress
S_{conf}
D_{app}(μm²/s)
FUS P525L
B
FUS WT
Invariant
Enriched or Depleted
-log10PValue
log2FC
FUS P525L
C
no DOXY
(1816 genes)
FUS WT
(2178 genes)
822
FUS P525L LOSS
183
445
768
COMMON
43
143
556
FUS P525L GAIN
FUS P525L
(1510 genes)
D
FUS P525L LOSS
Regulation of GTPase activity
Cell-cell signaling
Cytoskeleton organization
Intracellular signal transduction
Regulation of signal transduction
Movement of cell/subcellular component
Regulation of multicellular organismal process
Nervous system development
Regulation of cellular component organization
Animal organ development
FDR
Overlap
Enrichment Ratio
COMMON
Cell-cell adhesion via membrane adhesion molecules
Multicellular-organism growth
Cell junction organization
Muscle tissue development
Reg. of small GTPase mediated signal transduction
Morphogenesis of an epithelium
Peptidyl-serine modification
Cell-cell signaling by wnt
Positive regulation of neurogenesis
Covalent chromatin modification
FUS P525L GAIN
Regulation of transcription by RNA polymerase II
Transcription by RNA Polymerase II

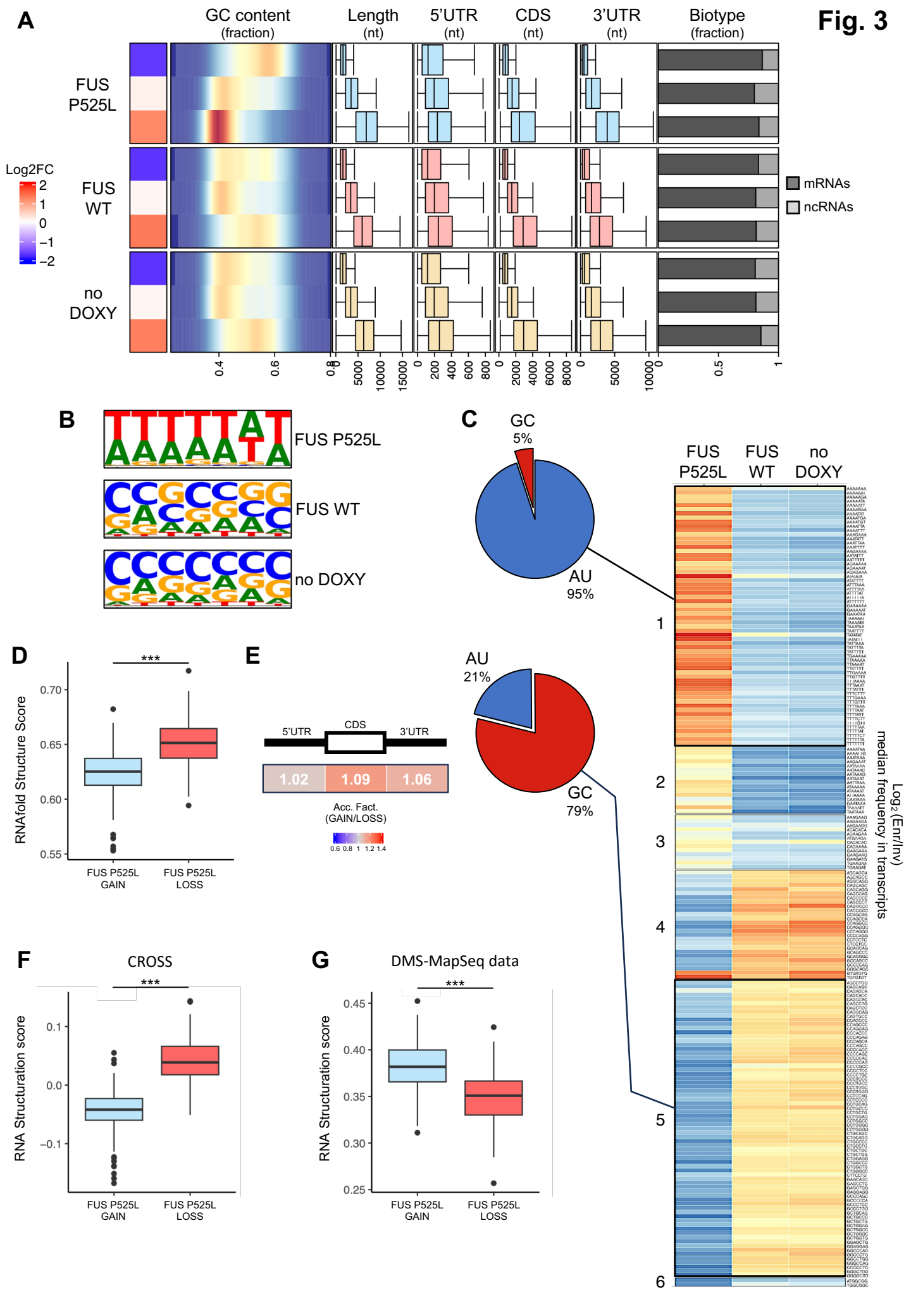
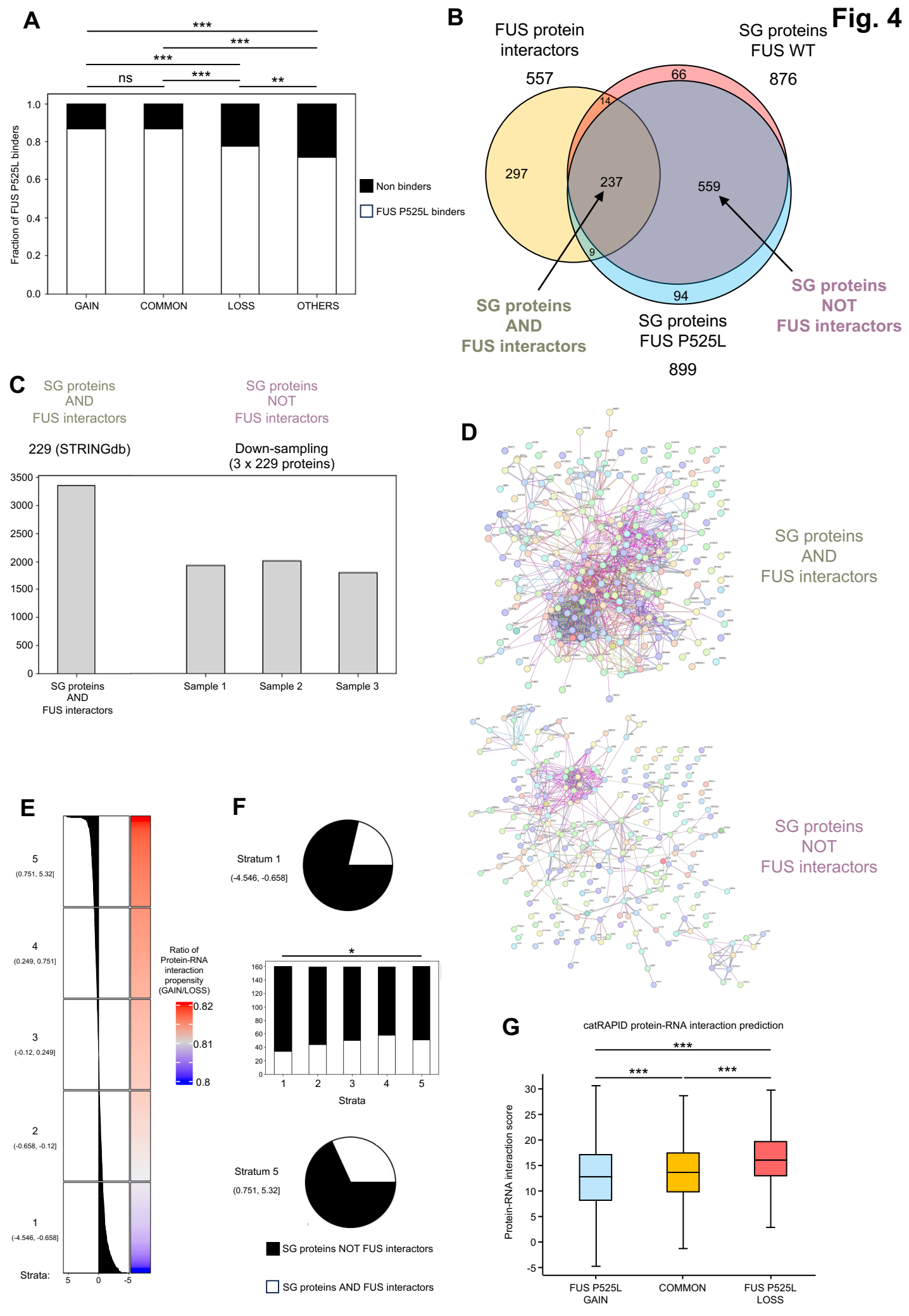


Fig. 3
A
GC content (fraction)
Length (nt)
5'UTR (nt)
CDS (nt)
3'UTR (nt)
Biotype (fraction)
FUS P525L
FUS WT
no DOXY
Log2FC
mRNAs
ncRNAs
B
FUS P525L
FUS WT
no DOXY
C
GC 5%
AU 95%
AU 21%
GC 79%
FUS P525L
FUS WT
no DOXY
median frequency in transcripts
Log2 (Enr/Inv)
D
RNAfold Structure Score
FUS P525L GAIN
FUS P525L LOSS
E
5'UTR
CDS
3'UTR
1.02
1.09
1.06
Acc. Fact. (GAIN/LOSS)
F
CROSS
RNA Structuration score
FUS P525L GAIN
FUS P525L LOSS
G
DMS-MapSeq data
RNA Structuration score
FUS P525L GAIN
FUS P525L LOSS

Fig. 4


A
***
***
***
ns
***
**
Fraction of FUS P525L binders
1.0
0.8
0.6
0.4
0.2
0.0
GAIN
COMMON
LOSS
OTHERS
Non binders
FUS P525L binders
B
FUS protein interactors
557
SG proteins FUS WT
876
66
14
297
237
559
9
94
SG proteins AND FUS interactors
SG proteins FUS P525L
899
SG proteins NOT FUS interactors
C
SG proteins AND FUS interactors
SG proteins NOT FUS interactors
229 (STRINGdb)
Down-sampling (3 x 229 proteins)
3500
3000
2500
2000
1500
1000
500
0
SG proteins AND FUS interactors
Sample 1
Sample 2
Sample 3
D
SG proteins AND FUS interactors
SG proteins NOT FUS interactors
E
5
(0.751, 5.32]
4
(0.249, 0.751]
3
(-0.12, 0.249]
2
(-0.658, -0.12]
1
(-4.546, -0.658]
Strata:
5
0
-5
Ratio of Protein-RNA interaction propensity (GAIN/LOSS)
0.82
0.81
0.8
F
Stratum 1
(-4.546, -0.658]
*
160
140
120
100
80
60
40
20
0
1
2
3
4
5
Strata
Stratum 5
(0.751, 5.32]
SG proteins NOT FUS interactors
SG proteins AND FUS interactors
G
catRAPID protein-RNA interaction prediction
***
***
***
Protein-RNA interaction score
30
25
20
15
10
5
0
-5
FUS P525L GAIN
COMMON
FUS P525L LOSS