PNAS

Supporting Information for

Immiscible proteins compete for RNA binding to order condensate layers.

Wilton T. Snead, Mary K. Skillicorn, Krishna Shrinivas, Amy S. Gladfelter
Corresponding authors: Amy S. Gladfelter, Krishna Shrinivas, Wilton T. Snead
Email: amy.gladfelter@duke.edu, krishna@northwestern.edu, wilton.snead@northwestern.edu

This PDF file includes:

SI Methods

Cluster profiles

Analysis of the data was mainly done as post-simulation calculations except for ones employed in Hoomd-Blue during simulation including logs of potential energy, kinetic energy, and temperature[1]. Post-simulation calculations were performed using a combination of python packages including *numpy*, *freud*, and *matplotlib*. Data was stored in GSD files and read through the gsd module on python. Generally analysis included determining clusters with a mix of protein and RNA together using the freud cluster module under a neighboring threshold of 1.5 unit distance[2]. For each cluster determined, properties including the center of mass, size and radius of gyration (rg) were extracted, and used for the specific analysis explained below. Each cluster property including radial density plots (RDP) and maximum intensity plots (MIP) are averaged over the largest clusters in each trajectory, cumulating from all specified frames, then averaged over multiple runs (n = 10).

Radial density plots

For each frame used for analysis, all specified clusters with protein and RNA are reported with a rg and center of mass. From the center of mass to 2*rg of clusters, all particles' distance from the center for a specific species (5´, middle, 3´, NONO, FUS, and TDP-43) were organized into bins (number of bins=10). These bins were then divided by the representative spherical volume to give a number density. The radial density for each species was plotted over the normalized distance from the center of the cluster. This was visualized using matplotlib in Python.

Core enrichment analysis

For a designated model, the average density for the middle region of NEAT1 was divided by the averaged sum of the 5´ and 3´ end densities. Calculating this parameter for each bin of the radial density profile provided the enrichment for the middle region of NEAT1 along the radial profile of a cluster. An enrichment value greater than one indicates more of the middle region compared to the ends in that region (Fig. S9C). To compare core enrichment of the NEAT1 middle region across different models (TDP-43 mutant models described in the next section), we averaged the enrichment values in the bins to the left of the core/shell boundary (Fig. S9C, D). This boundary was defined for the revised model as the portion of the profile in which the enrichment of the NEAT1 middle region was greater than one, corresponding to the first two bins of ten (i.e. 0.4*rg; Fig. S9C). This core/shell boundary position was held constant during core enrichment analysis across all other models.

TDP-43 mutant models

TDP-43 mutations and loss from the nucleus are frequently observed in the neurodegenerative diseases amyotrophic lateral sclerosis (ALS) and frontotemporal lobar degeneration (FTLD)[3]. Because it is beyond the limits of our coarse-grained model to capture the fine-grained effects of sequence-specific TDP-43 mutations, we adopted a simpler effective approach. Specifically, we applied changes to the revised model (Fig. 5A) that mimicked common phenotypes resulting from mutations in TDP-43. Phenotypes with higher aggregation[4] or a reduced propensity to condense[5] were represented in the model as an increase ($\epsilon_{TDP43-TDP43} = 3.5\,kT$) or decrease ($\epsilon_{TDP43-TDP43} = 1\,kT$) in TDP-43 self interactions, respectively. For mutations related to TDP-43 loss from the nucleus[3,4,6], we performed simulations with TDP-43 copy number reduced to 1000 or set to zero (Table S1). Importantly, previous work revealed that reduction in the TDP-43 nuclear concentration correlates with higher transcription of NEAT1[3]. To examine the impacts of increased NEAT1 expression, we simulated two additional cases in which NEAT1 chains were increased (from 20 to 40) alone or in combination with a reduction in TDP-43 copy number. In all TDP-43 mutant models, trajectories were run under the same simulation parameters described above (Table S1).

Maximum intensity plots

For each frame used for analysis all specified clusters with protein and RNA are reported with its rg and center of mass. From the center to 2*rg of clusters, the plane of either XY, YZ, or ZX is divided into a 30x30 grid. For a plane's orthogonal axis it is divided into three hundred slices. For each slice, the 30x30 grid is filled with a specific species local density that is calculated from freud's local density module (r_max=1 and diameter=1)[2]. Over all slices, each cell in the grid records its largest density. The maximum intensity plots for each species is visualized using matplotlib heat maps and normalized to each species highest recorded density.

Visualization of simulation data

Figures were generated using Visual Molecular Dynamics (VMD)[7]. Images were rendered using Tachyon (Internal) with RNA monomers as AOChalky and proteins as AOShiny into tga files. The python package, *pillow*, was used to convert the tga files into pdfs.

Pair correlation analysis of paraspeckle-enriched RBP motifs

We cross-referenced a list of paraspeckle-enriched proteins (ref[8]) with a compendium of RBP motifs generated from ENCODE eCLIP data (ref[9]) to obtain the paraspeckle protein motifs shown in Fig. S11A. To estimate the sequence

similarity between pairs of motifs, Pearson pair correlations between motif PWMs were calculated. When two motifs were of unequal length, a window of size equal to the length of the shorter motif was used to generate multiple sub-motifs from the longer motif, in sliding increments of one. The highest pair correlation identified for all such sub-motifs was kept for analysis. For clustering of pair correlations of motif sequence, the seaborn clustermap package was used with a calculated euclidean distance and hierarchical clustering of the nearest point (Fig. S11B). Motif distributions along NEAT1 (Fig. S11C) were calculated as described in the previous section (RBP motif analysis of NEAT1). Pearson pair correlations of motif distributions in NEAT1 were computed and plotted, but values were not clustered (Fig. S11D). Rather, proteins were ordered following the same clustered pattern identified in the motif sequence pair correlation plot.

Secondary structure analysis of NEAT1

RNA secondary structure prediction was performed using the ViennaRNA package for Python version 2.7.0 (https://www.tbi.univie.ac.at/RNA/ViennaRNA/doc/html/index.html). 1000 nt fragments of NEAT1, tiled every 100 nt, were folded and the minimum free energy (MFE) of the optimal secondary structure was obtained and plotted. To estimate pair probability of each NEAT1 fragment, matrices of base pair probabilities were obtained and summed along the rows to obtain the total probability of base pairing for each nucleotide in the fragment. The average ± standard error of the mean of the resulting list of pair probabilities was used for plotting.

Analysis of quantitative proteomics data

Previously published quantitative proteomics datasets were obtained for human tissues[10], multiple cultured cell lines[11], and matched orthologs in primary skin fibroblasts from different mammalian species[12]. Briefly, from each of these data, protein counts were averaged across replicates and normalized to count of FUS. These effective expression ratios are reported in Fig. 3C and S5C.

Protein purification and labeling

All proteins were expressed as 6x/7xHis-MBP fusion constructs. After transforming bacterial expression plasmids into BL21 *E. coli* (NEB C2527H), cells were grown at 37 °C until reaching OD600 0.6-0.8 and proteins were induced overnight at 18 °C following addition of IPTG (0.5 mM for FUS, NONO, SFPQ, and TDP-43 and 0.4 mM for TEV). Cells were harvested by centrifugation at 14,000 rcf at 4 °C the following day. Cells were resuspended in lysis buffer (described below for each protein) and lysed using probe sonication (QSonica Q500 with 20 kHz converter and 1/4" microtip). Lysate was clarified by centrifugation at 27,000 rcf for 30 min at 4 °C. Following purification, all proteins were verified for purity and size using SDS-PAGE. The following subsections provide purification details for each protein.

FUS purification

Lysis buffer consisted of 50 mM HEPES pH 7.4, 1.5 M NaCl, 10% glycerol, 20 mM imidazole, 5 mM β-mercaptoethanol (BME), and EDTA-free protease inhibitor cocktail (Pierce A32965). Clarified lysate was mixed with 0.4-0.5 mL of washed, packed HisPur cobalt resin (Thermo Scientific 89965) per 1 L of cells for 30-60 min at 4 °C. After protein binding, resin was transferred into a gravity flow column and washed with approximately 80-100 resin bed volumes of lysis buffer without protease inhibitor cocktail. Protein was eluted in elution buffer consisting of 50 mM HEPES pH 7.4, 150 mM NaCl, 10% glycerol, 200 mM imidazole, and 5 mM BME. Eluted protein was dialyzed overnight using a 3 mL Slide-A-Lyzer 20K MWCO dialysis cassette (Thermo Scientific 66003) with two rounds of 1 L dialysis buffer consisting of 20 mM HEPES pH 7.4, 150 mM NaCl, 5% glycerol, and 5 mM BME.

TDP-43 purification

Lysis buffer consisted of 50 mM HEPES pH 7.4, 1.5 M NaCl, 10% glycerol, 20 mM imidazole, 5 mM BME, and EDTA-free protease inhibitor cocktail. Protein was purified from clarified lysate using an AKTA go fast protein liquid chromatography (FPLC) system (Cytiva). Lysate was loaded onto a 5 mL HisTrap HP column (Cytiva 17524801) equilibrated with lysis buffer, and the column was washed with 10 column volumes of wash buffer consisting of 50 mM HEPES pH 7.4, 1.5 M NaCl, 40 mM imidazole, and 5 mM BME. Protein was eluted with a step to 100% elution buffer consisting of 50 mM HEPES pH 7.4, 1 M NaCl, 500 mM imidazole, and 5 mM BME. Eluted protein was further purified using size exclusion chromatography (SEC) over a Superdex 200 Increase 10/300 GL column (Cytiva 28990944) equilibrated with storage buffer consisting of 20 mM HEPES pH 7.4, 300 mM NaCl, and 5 mM BME.

NONO and SFPQ purification

Lysis buffer consisted of 50 mM HEPES pH 7.4, 1.5 M NaCl, 10% glycerol, 20 mM imidazole, 250 mM L-arginine, 0.2% Triton X-100, 5 mM BME, and EDTA-free protease inhibitor cocktail. Proteins were purified from clarified lysates using FPLC following a similar procedure described for TDP-43. Wash buffer consisted of 50 mM HEPES pH 7.4, 1.5 M NaCl, 40 mM imidazole, 250 mM L-arginine, and 5 mM BME. Elution buffer consisted of 50 mM HEPES pH 7.4, 500 mM NaCl, 500 mM imidazole, 250 mM L-arginine, and 5 mM BME. Eluted proteins were further purified using SEC as described for TDP-43. SEC storage buffer consisted of 20 mM HEPES pH 7.4, 500 mM NaCl, and 5 mM BME.

TEV protease purification

Lysis buffer consisted of 25 mM HEPES pH 7.4, 300 mM KCl, 10% glycerol, and 20 mM imidazole. HisPur cobalt resin binding and gravity column washing was performed following a similar procedure described for FUS. Resin was further washed using approximately 100 resin bed volumes of lysis buffer containing 10 mM ATP-$MgCl_2$, followed by an additional 80-100 resin bed volumes of lysis buffer without ATP-$MgCl_2$. Elution buffer consisted of 25 mM HEPES pH 7.4, 150 mM KCl, 10% glycerol, and 250 mM imidazole. Eluted protein was dialyzed overnight using a 3 mL Slide-A-Lyzer 10K MWCO dialysis cassette (Thermo Scientific 66455) with two rounds of 0.5 L dialysis buffer consisting of 25 mM HEPES pH 7.4, 150 mM KCl, 10% glycerol, and 5 mM BME.

Protein labeling

Approximately 0.3-0.5 mL of purified proteins were fluorescently labeled with amine-reactive, NHS ester-conjugated Atto 488 or Atto 594 dyes (Sigma 41698-1MG-F and 08741-1MG-F). After resuspending the reactive dye in anhydrous DMSO to a stock concentration of 10 mM, dye was added to the protein at a dye:protein molar ratio of 1:1. The dye was allowed to conjugate to the protein for 20 min at 25 °C, and the mixture was transferred to a 0.5 mL Slide-A-Lyzer 20K MWCO dialysis cassette (Thermo Scientific 66005) and dialyzed with two rounds of 0.5 L dialysis buffer at 4 °C to remove unconjugated dye.

Measuring protein and dye concentrations

For FUS, TDP-43, and TEV, protein and conjugated dye concentrations were measured using a Nanodrop spectrophotometer. A260/A280 absorbance ratios were found to be in the range of 0.60-0.65, indicating relatively minor absorbance contributions from residual nucleic acids. For NONO and SFPQ, protein concentrations were measured using the Qubit Protein BR Assay Kit (Invitrogen A50669) according to the manufacturer's instructions. Small aliquots of labeled and unlabeled proteins were snap frozen in liquid nitrogen and stored at -80 °C.

In vitro transcription of NEAT1 RNA fragments

Linear DNA templates for *in vitro* transcription were generated via PCR from respective plasmids using the following forward and reverse primers: 5´-TAATACGACTCACTATAGGGAGTTCTACAGTCCGACGATC-3´ and 5´-GCCTTGGCACCCGAGAATTCCA-3´. *In vitro* transcription was performed using the HiScribe T7 kit (NEB E2040S) following the manufacturer's instructions using approximately 1 µg of DNA template per 20 µL reaction. 20 U of murine RNase inhibitor (NEB M0314L) was also included in each reaction. To generate fluorescently labeled RNAs, 0.1 µL of Cy5-UTP (Cytiva PA55026) or 1 µL of Atto 488-UTP (Jena Bioscience NU-821-488) were included per 20 µL reaction. Following incubation at 37 °C for 2 h, DNA templates were removed by adding 2 µL of RQ1 RNase-free DNase (Promega M6101) per 20 µL reaction and incubating at 37 °C for 20 min. RNA was precipitated by adding LiCl to a final concentration of approximately 830 mM and transferring to -80 °C for at least 15 min. Precipitated RNA was pelleted by centrifugation at 21,300 rcf for 10 min at 4 °C. Supernatant was removed and RNA pellets were washed at least three times with 0.5 mL of RNase-free 80% ethanol. Following resuspension in RNase-free water, RNA concentration was measured using a Nanodrop spectrophotometer and verified for purity and size using a denaturing agarose gel and ssRNA ladder (NEB N0362S). Small aliquots of RNA were stored at -80 °C.

SI Figures

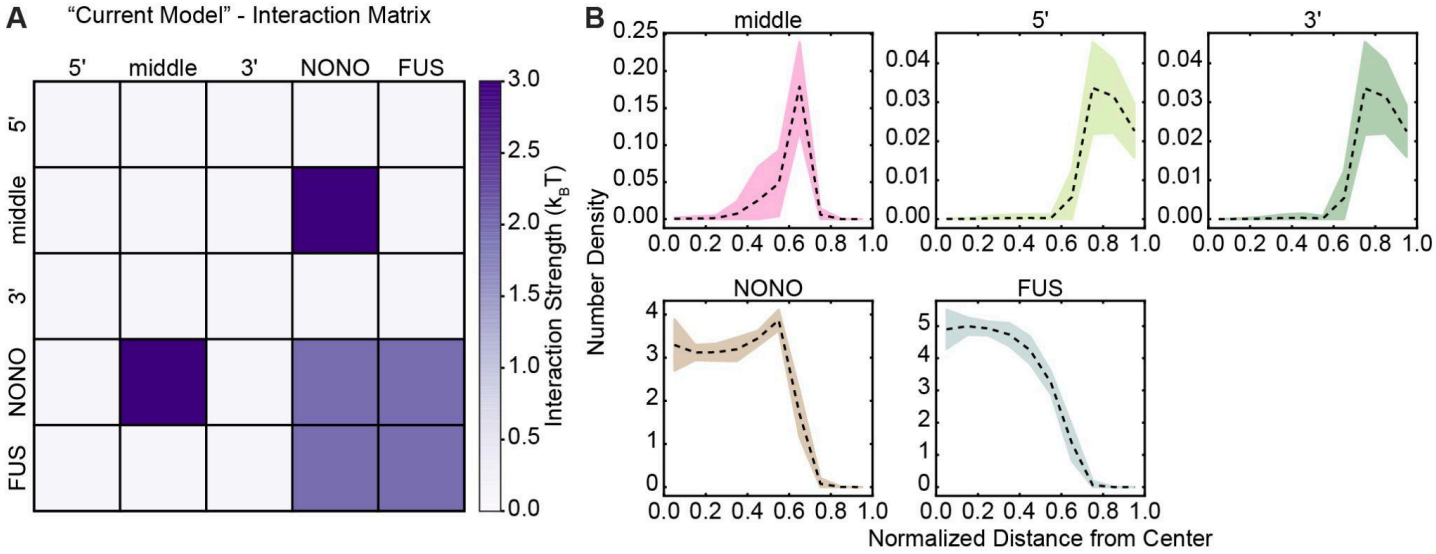

Fig. S1. Simulating the current model of paraspeckle assembly. (**A**) Matrix of interactions used in simulations. (**B**) Radial density profiles of NEAT1 segments (top) and proteins (bottom). Dashed lines and shaded areas represent average and standard deviation of the largest clusters from n=10 simulations.

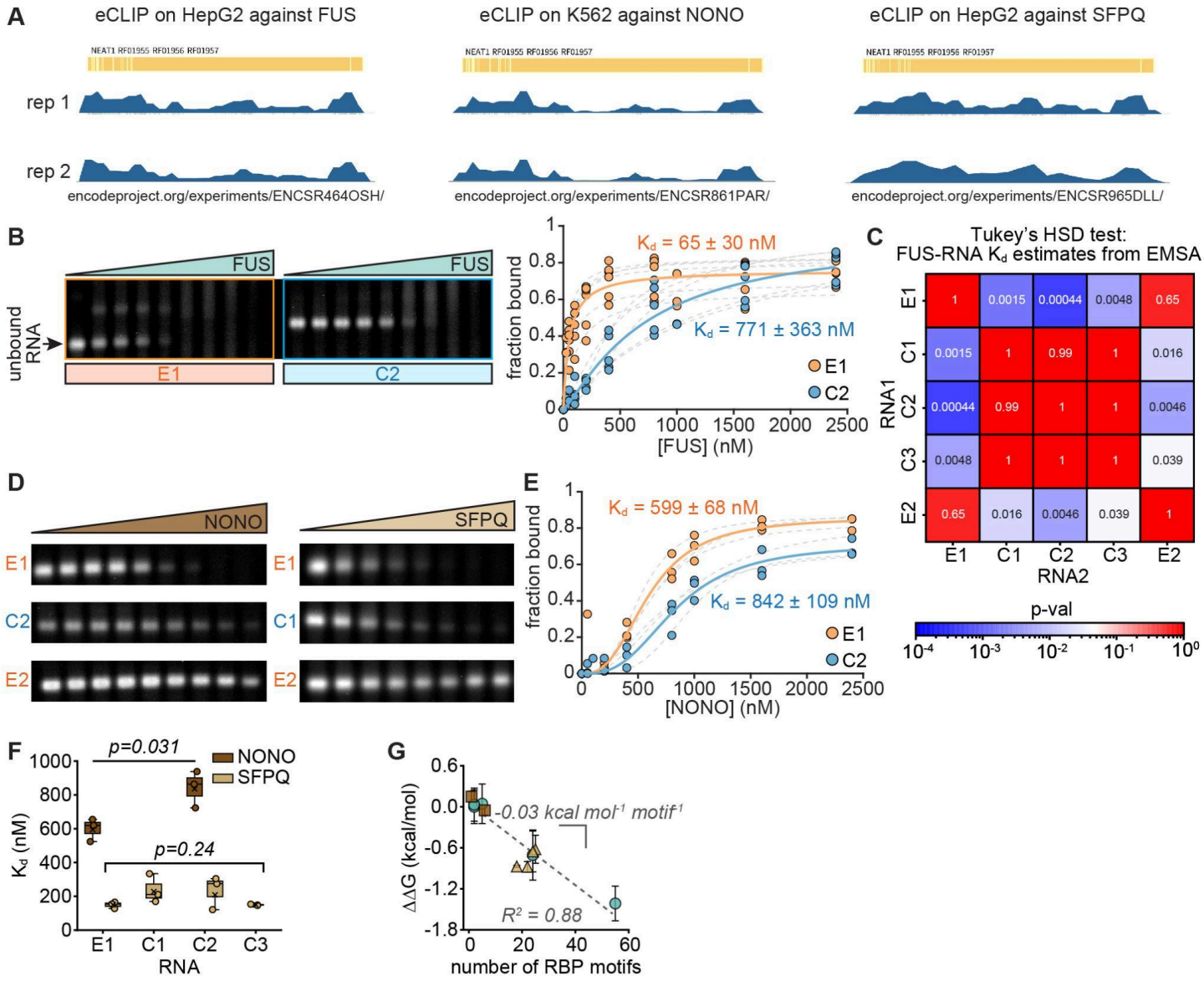

Fig. S2. Paraspeckle core RBPs display unexpected binding to NEAT1. (**A**) NEAT1 eCLIP data from ENCODE for the indicated cell type against the indicated protein. (**B**) Background-subtracted image of unbound RNA in FUS EMSA gel with E1 and C2 fragments (separate replicate from the one displayed in Fig. 2B). Plot: Fraction of bound E1 or C2 RNA as a function of FUS concentration. Gray dashed curves and solid curves in color indicate Langmuir isotherm fits to individual replicates and pooled data from all replicates, respectively. n=6 and 5 replicates for E1 and C2, respectively. (**C**) Heatmap of pairwise p-values (Tukey's HSD test) for the indicated RNAs in FUS EMSA experiments. Significant values ($p<0.05$) are blue, non-significant values are red. (**D**) Background-subtracted images of unbound RNA in EMSA gels with the indicated proteins and RNAs. For each protein, the three sets of lanes were cropped from the same gel, with equal contrast settings applied. Poor NONO and SFPQ binding to E2 hindered K_d estimation. (**E**) Fraction of bound E1 or C2 RNA as a function of NONO concentration, formatted similar to (B). (**F**) NONO-RNA and SFPQ-RNA K_d estimates from EMSAs. Circles and x's indicate K_d estimates from individual replicates and pooled data, respectively. Boxes indicate IQRs with medians as bisecting lines and whiskers as 1.5*IQR. n=3 replicates per RNA in (E, F). p-values in (F) from two-tailed Student's t-test and one-way ANOVA for NONO and SFPQ data, respectively. (**G**) Data from Fig. 2C, expressed as $\Delta\Delta G$, as a function of motif abundance. $\Delta\Delta G$ values were computed relative to the overall median of FUS binding to C1, C2, and C3. Dashed gray line indicates linear regression.

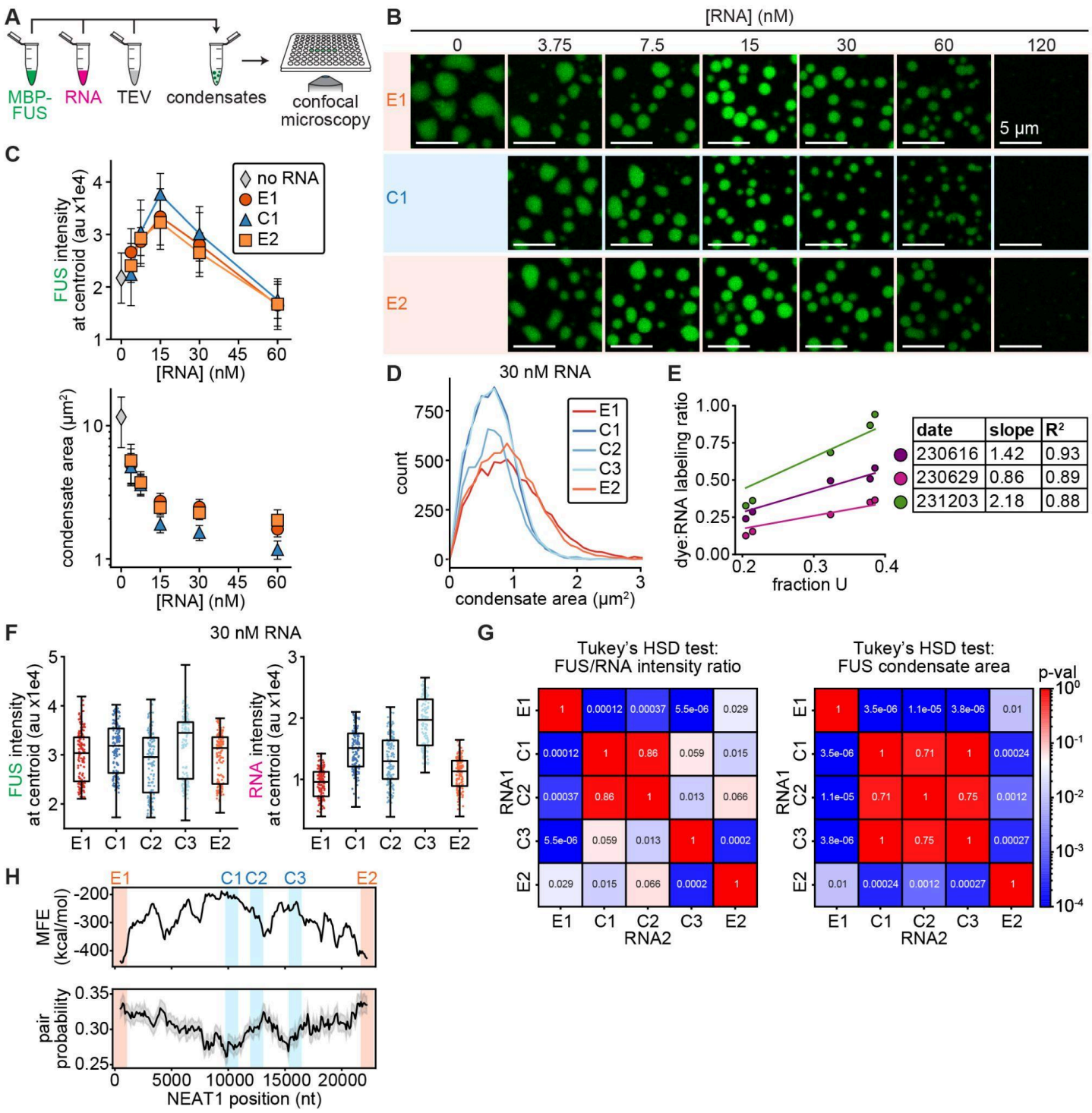

Fig. S3. FUS condensates assembled with middle region RNAs are smaller owing to stronger RNA-RNA interactions in the dense phase. (**A**) Schematic of *in vitro* condensation experiments. (**B**) Confocal slices of condensates assembled with 2 µM FUS (green) and RNA at the indicated concentrations after 3 h at 25 °C. FUS channel is contrasted equally in all images. (**C**) FUS intensity at condensate centroids (upper) and condensate area (lower) as a function of RNA concentration for the three indicated RNAs. Data points and error bars indicate mean ± one standard deviation for the 50 largest condensates in each experimental replicate. n = 3 replicates per RNA (150 condensates total per RNA). Note that at RNA concentrations of 15 nM and higher, condensate area is consistently lower for C1 compared to E1 and E2. (**D**) Histograms of condensate area, assembled with 2 µM FUS and 30 nM RNA. n = 6648, 7669, 5589, 7810, 6937 condensates for E1, C1, C2, C3, E2, each pooled from 3 replicates per RNA. (**E**) Dye:RNA labeling ratio as a function of uridine (U) fraction for the five RNAs E1, C1, C2, C3, and E2. Data represent three RNA preps from different days, each using the same reaction master mix for the five RNAs. Lines indicate linear regression with y-intercept set to 0. Table summarizes fit slopes and R^2 values. (**F**) Boxplots of FUS (left) and RNA (right) intensity at condensate centroids for condensates assembled with 2 µM FUS and 30 nM RNA. Data represent the 50 largest condensates from each

experimental replicate. n = 3 replicates per RNA (150 condensates total per RNA). Circles indicate individual condensates, boxes indicate IQR with medians as bisecting lines and whiskers as 1.5*IQR. (**G**) Heatmaps of pairwise p-values (Tukey's HSD test) for the indicated RNAs in condensate experiments with 2 μM FUS and 30 nM RNA. Left: FUS/RNA intensity ratio. Right: condensate area. Significant values ($p<0.05$) are blue, non-significant values are red. (**H**) Minimum free energy (MFE) of the optimal secondary structure (top) and base pair probability (bottom) within 1000 nt windows of NEAT1, tiled every 100 nt from 5′ to 3′. Bottom plot indicates the average pair probability of all nt's within each window, shaded error bar indicates standard error of the mean. The higher MFE values and lower pair probabilities of middle region RNAs suggest that the middle region is less structured and more single-stranded. RNA structure prediction was performed using the ViennaRNA package for Python (Methods). Orange and blue vertical bars indicate the five NEAT1 fragments used in this study.

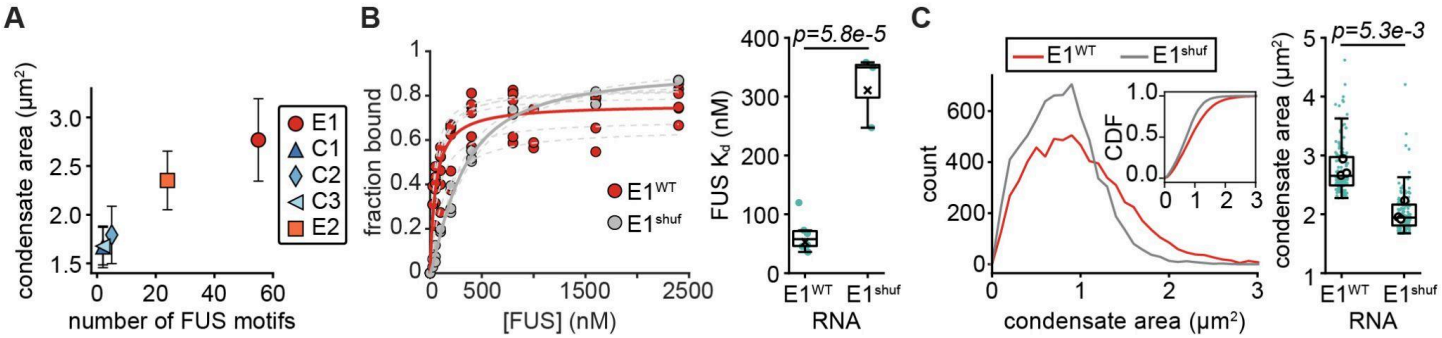

Fig. S4. Condensate area scales with FUS-NEAT1 affinity. (**A**) Condensate area as a function of the number of FUS motifs within each RNA. Data points and error bars indicate mean ± one standard deviation for the 50 largest condensates in each experimental replicate. n = 3 replicates per RNA (150 condensates total per RNA). (**B**) Left: fraction of bound $E1^{WT}$ or $E1^{shuf}$ RNA as a function of FUS concentration from EMSAs. Gray dashed curves indicate Langmuir isotherm fits to individual replicates, solid curves in color indicate fits to pooled data from all replicates. $E1^{WT}$ data repeated from Fig. S2B. Right: boxplots of FUS K_d estimates. Circles and x's indicate K_d estimates from individual replicates and pooled data, respectively. Boxes indicate IQRs with medians as bisecting lines and whiskers as 1.5*IQR. n = 6 and 3 replicates for $E1^{WT}$ and $E1^{shuf}$. (**C**) Left: histograms of condensate area from experiments corresponding to images in Fig. 2H. n = 6648 and 7116 condensates for $E1^{WT}$ and $E1^{shuf}$, each pooled from 3 replicates per RNA. Right: boxplots of FUS condensate area. Points indicate individual droplets, white circles indicate replicate means. Boxes indicate IQRs with medians as bisecting lines and whiskers as 1.5*IQR. p-values in (B, C) from two-tailed Student's t-test; t-test in (C) based on replicate means (white circles).

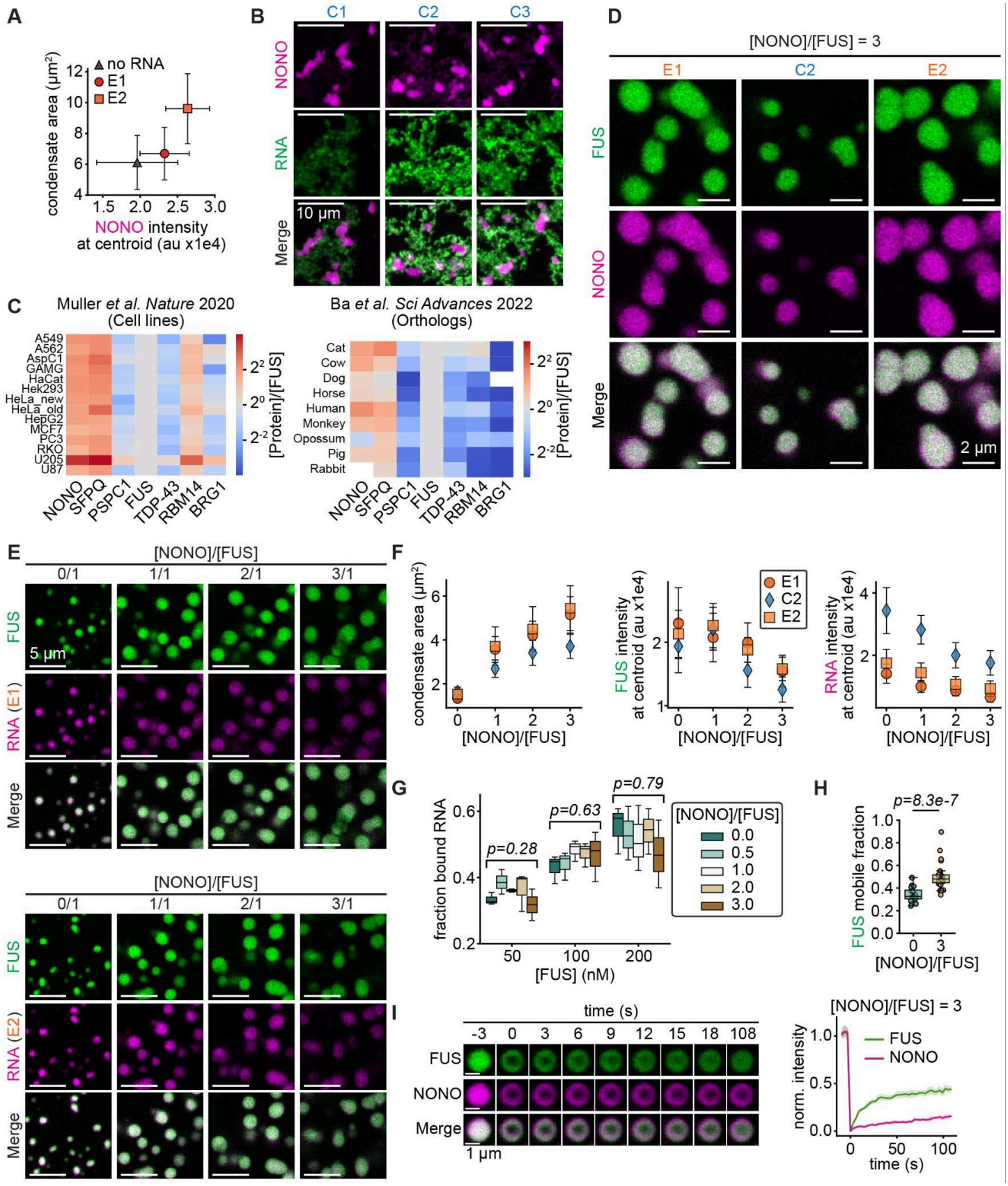

Fig. S5. NONO/FUS stoichiometry rewires the properties of FUS-RNA condensates. (**A**) Condensate area as a function of NONO intensity at condensate centroids for condensates assembled with 2 μM NONO + 15 nM of the indicated RNA. (**B**) Confocal maximum intensity z-projections (all z-stacks contain 60 frames) of condensates and networks assembled with 2 μM NONO (magenta) + 15 nM of the indicated RNA (green) after 3 h at 25 °C. NONO and RNA channels each contrasted equally in all images. (**C**) Heat maps of paraspeckle protein abundance relative to FUS in the indicated cell lines (left, data from ref[11]) and in fibroblasts from the indicated mammalian species (right, data from ref[12]). (**D**) Confocal slices of

condensates assembled with 1 μM FUS (green) + 15 nM of the indicated RNA (unlabeled) + 3 μM NONO (magenta) after 3 h at 25 °C. FUS and NONO channels each contrasted equally in all images. (**E**) Confocal slices of condensates assembled with 1 μM FUS (green) + 15 nM of the indicated RNA (magenta) + NONO (unlabeled) at the indicated concentrations after 3 h at 25 °C. FUS and RNA channels each contrasted equally in all images. (**F**) Condensate area (left) and FUS (middle) and RNA (right) intensity at condensate centroids as a function of NONO/FUS stoichiometry with the indicated RNAs, from experiments corresponding to (E). Data points and error bars in (A, F) indicate mean ± one standard deviation for the 50 largest condensates from each experimental replicate. n = 3 replicates per RNA (150 condensates total per RNA). (**G**) Fraction of bound E1 RNA from EMSAs performed with the indicated FUS concentrations in the presence of the indicated NONO/FUS stoichiometries. n = 3 replicates per condition. p-values from one-way ANOVA performed on the indicated groups. No groups contain pairs that are significantly different from each other. (**H**) FUS mobile fraction estimates from FRAP fits. Circles and x's indicate estimates from fits to individual droplets and average data, respectively. n = 24 and 25 condensates for NONO/FUS stoichiometries of 0 and 3, respectively, each pooled from 3 replicates per condition. p-value from two-tailed Student's t-test. Boxes in (G, H) indicate IQR with medians as bisecting lines and whiskers as 1.5*IQR. (**I**) Confocal slices of FUS (green) and NONO (magenta) photobleaching recovery within droplets assembled with 1 μM FUS + 15 nM E2 RNA + 3 μM NONO. Plot shows normalized FUS and NONO intensity recovery profiles as a function of time for experiments corresponding to the images on the left. FUS profile repeated from Fig. 3F. Curves and shaded error bars indicate average ± 95% CI. n = 25 condensates pooled from 3 replicates.

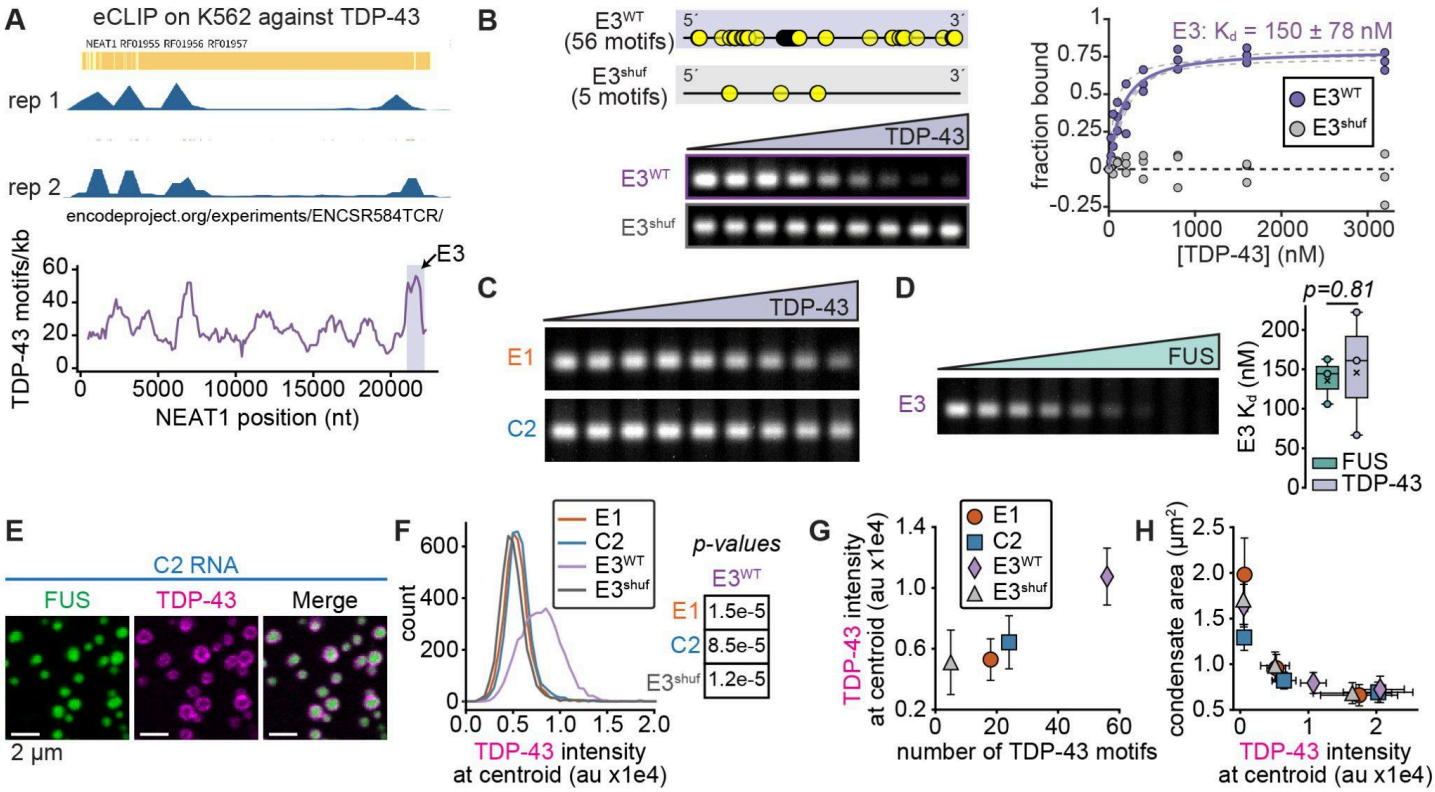

Fig. S6. TDP-43 binds abundant cognate motifs in a 3´-proximal NEAT1 fragment to oppose FUS condensation. (**A**) Top: NEAT1 eCLIP data from ENCODE for K562 cells against TDP-43. Bottom: the number of TDP-43 motifs within 1000 nt windows of NEAT1, tiled every 100 nt from 5´ to 3´. Plot repeated from Fig. 4A. Purple vertical bar indicates the E3 fragment. (**B**) Schematic of TDP-43 motifs in $E3^{WT}$ and a shuffled E3 sequence, $E3^{shuf}$. Below: background-subtracted image of un-bound $E3^{WT}$ and $E3^{shuf}$ RNA in a TDP-43 EMSA gel. Lanes cropped from the same gel, with equal contrast settings applied. Plot: Fraction of bound $E3^{WT}$ or $E3^{shuf}$ RNA as a function of TDP-43 concentration. Gray dashed curves and solid curve in color indicate Langmuir isotherm fits to individual replicates and pooled data from all replicates, respectively. n = 3 replicates for each RNA. Undetectable TDP-43 binding to $E3^{shuf}$ over the tested concentration range hindered Langmuir isotherm fitting. (**C**) Background-subtracted images of unbound RNA in TDP-43 EMSA gels for E1 and C2 RNAs. Each set of lanes was cropped from a different gel, but equal contrast settings were applied to both images. Poor TDP-43 binding to E1 and C2 hindered K_d estimation. (**D**) Left: background-subtracted image of unbound RNA in FUS EMSA gel for $E3^{WT}$. Right: FUS and TDP-43 K_d estimates for $E3^{WT}$. Circles and x's indicate K_d estimates from individual replicates and pooled data, respectively. Boxes indicate IQRs with medians as bisecting lines and whiskers as 1.5*IQR. n = 3 replicates for each protein. p-value from two-tailed Student's t-test. (**E**) Confocal slices of condensates co-assembled with 2 μM FUS (green) + 0.2 μM TDP-43 (magenta) + 15 nM of unlabeled C2 RNA after 3 h at 25 °C. (**F**) Histograms of TDP-43 intensity at the centroids of condensates from experiments corresponding to (E) and Fig. 4B. n = 3796, 3854, 4044, 3753 condensates for E1, C2, $E3^{WT}$, $E3^{shuf}$, each pooled from 3 replicates per RNA. p-values from Tukey's HSD test; comparisons based on replicate means. Non-significant pairwise p-values not shown. (**G**) TDP-43 intensity at condensate centroids, from experiments corresponding to (E) and Fig. 4B, plotted as a function of the number of TDP-43 motifs per RNA. (**H**) Condensate area as a function of TDP-43 intensity at condensate centroids. Data are from experiments in which TDP-43 was added at concentrations of 0, 0.2, and 0.5 μM, with fixed FUS and RNA concentrations indicated in (E). Data points and error bars in (G, H) indicate mean ± one standard deviation for the 50 largest condensates from each experimental replicate. n = 3 replicates per RNA (150 condensates total per RNA).

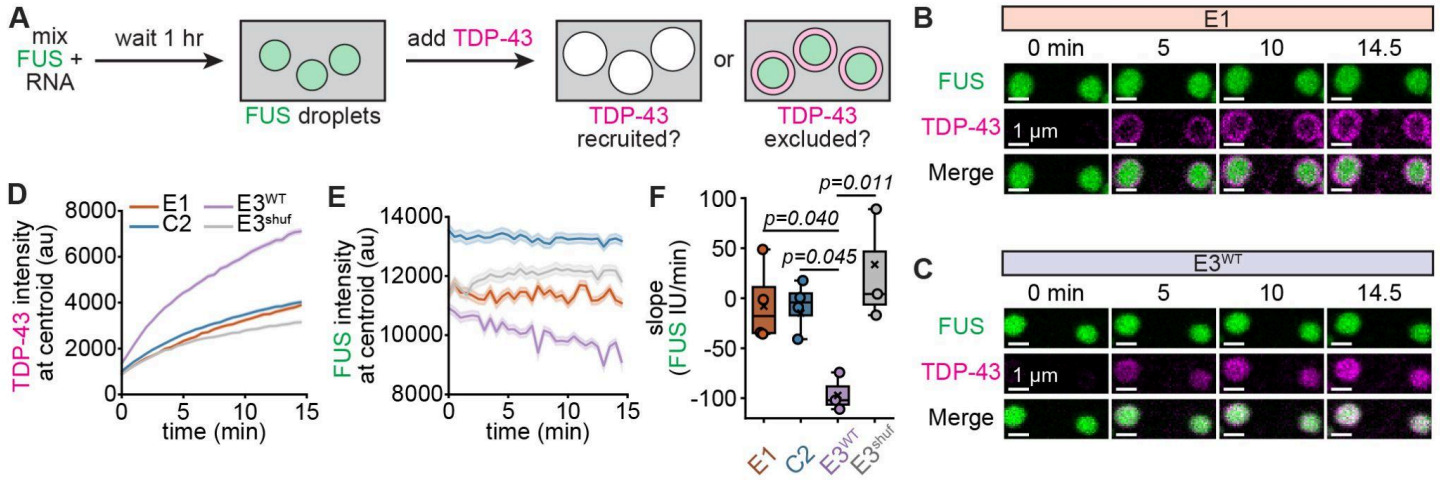

Fig. S7. TDP-43 evicts FUS from pre-assembled condensates with E3^{WT} RNA. (**A**) Workflow for examining TDP-43 recruitment to pre-assembled FUS + NEAT1 condensates. (**B, C**) Confocal time series of condensates pre-assembled with 2 μM FUS (green) + 15 nM of either E1 (B) or E3^{WT} (C) RNAs following addition of 2 μM TDP-43 (magenta) at the indicated time points. (**D, E**) TDP-43 (D) and FUS (E) intensity at the centroids of FUS condensates assembled with the indicated RNAs, plotted as a function of time after TDP-43 addition. Curves and shaded error bars indicate average ± 95% CI. Each time point contains a different number of condensates, owing to gradual sedimentation of new condensates into the imaging plane during the time course. Specifically, n = 1068, 1213, 1056, 887 condensates (first frame) and n = 1416, 1557, 1355, 1082 condensates (last frame) for E1, C2, E3^{WT}, E3^{shuf} (each pooled from 3-4 replicates per RNA). (**F**) Slope of FUS intensity over time following addition of TDP-43 (IU: intensity units), quantified from data in (E). Circles and x's indicate slopes from fits to individual replicates and pooled data, respectively. Boxes indicate IQR with medians as bisecting lines and whiskers as 1.5*IQR. n = 4, 4, 3, 3 replicates for E1, C2, E3^{WT}, E3^{shuf}. p-values from Tukey's HSD test; non-significant pairwise p-values not shown. The more negative slope for E3^{WT} indicates that FUS was evicted from condensates by TDP-43 more strongly compared to other RNAs.

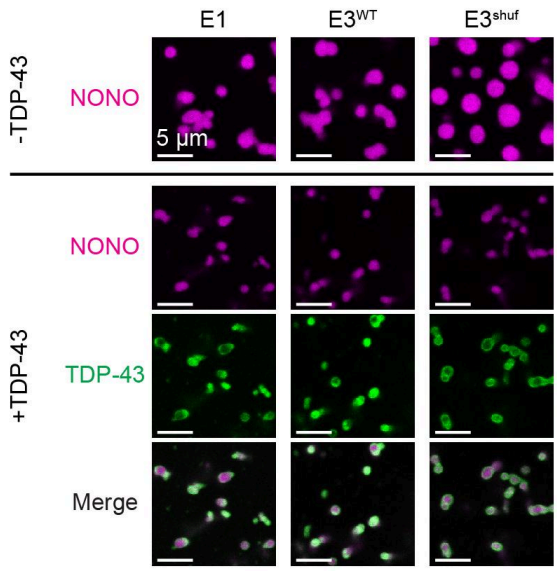

Fig. S8. TDP-43 opposes NONO condensation by forming surfactant-like layers on condensate surfaces. Confocal slices of condensates assembled with 2 µM NONO (magenta) + 15 nM of the indicated RNAs ± 0.2 µM TDP-43 (green) after 3 h at 25 °C. NONO and TDP-43 channels are each contrasted equally across all images.

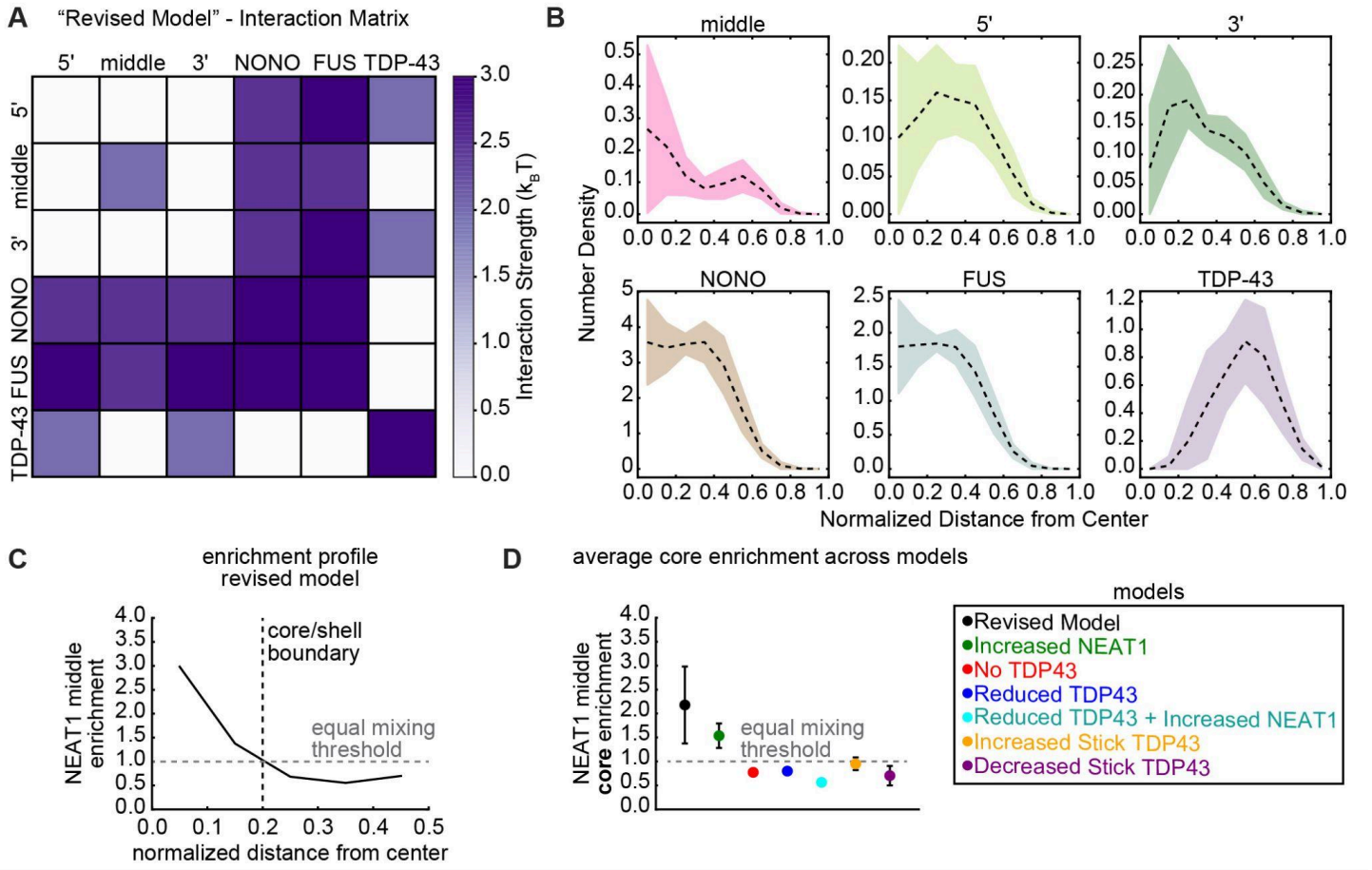

Fig. S9. A revised model relates the observed network of biomolecular interactions to paraspeckle assembly. (**A**) Matrix of interactions used in simulations. (**B**) Radial density profiles of NEAT1 segments (top) and proteins (bottom). Dashed lines and shaded areas represent average and standard deviation of the largest clusters from n = 10 simulations. (**C**) Radial enrichment profile of the NEAT1 middle region, computed as the density of the middle region divided by the average density of the 5′/3′ ends (Methods). Black, vertical dashed line indicates the boundary demarcating core and shell. Gray, horizontal dashed line indicates the threshold at which the NEAT1 middle and end segments show equal mixing. (**D**) Average enrichment of the NEAT1 middle region in the core, defined as the portion of the curve to the left of the vertical dashed line in (C). Plot shows average ± standard deviation of middle region core enrichment for different models that approximately mimic the indicated, disease-associated TDP-43 mutations or aging-related loss of TDP-43 from the nucleus. n = 10 simulations per model. All perturbation models decrease the core enrichment of the NEAT1 middle region, supporting a key role for TDP-43 in establishing the core/shell organization of NEAT1.

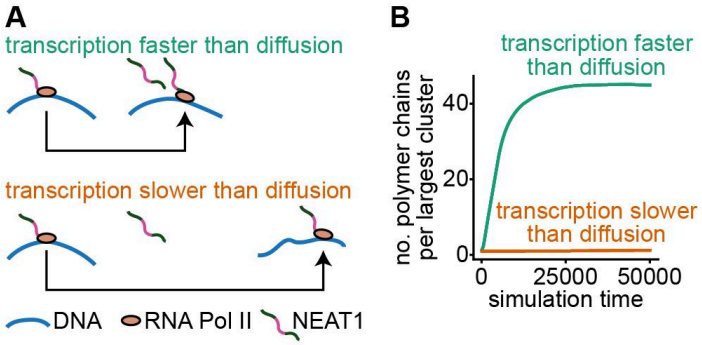

Fig. S10. Transcriptional dynamics determine the number of NEAT1 transcripts incorporated into clusters. (**A**) Schematic of the two model scenarios. Upper: when transcription is faster than chromatin diffusion (i.e. burst transcription), multiple transcripts accumulate within close spatial proximity to each other. Lower: when transcription is slower than chromatin diffusion (i.e. latent or paused transcription), transcripts disperse over a relatively large volume. (**B**) The number of NEAT1 polymer chains within the largest cluster in simulations as a function of time. See Methods for simulation details.

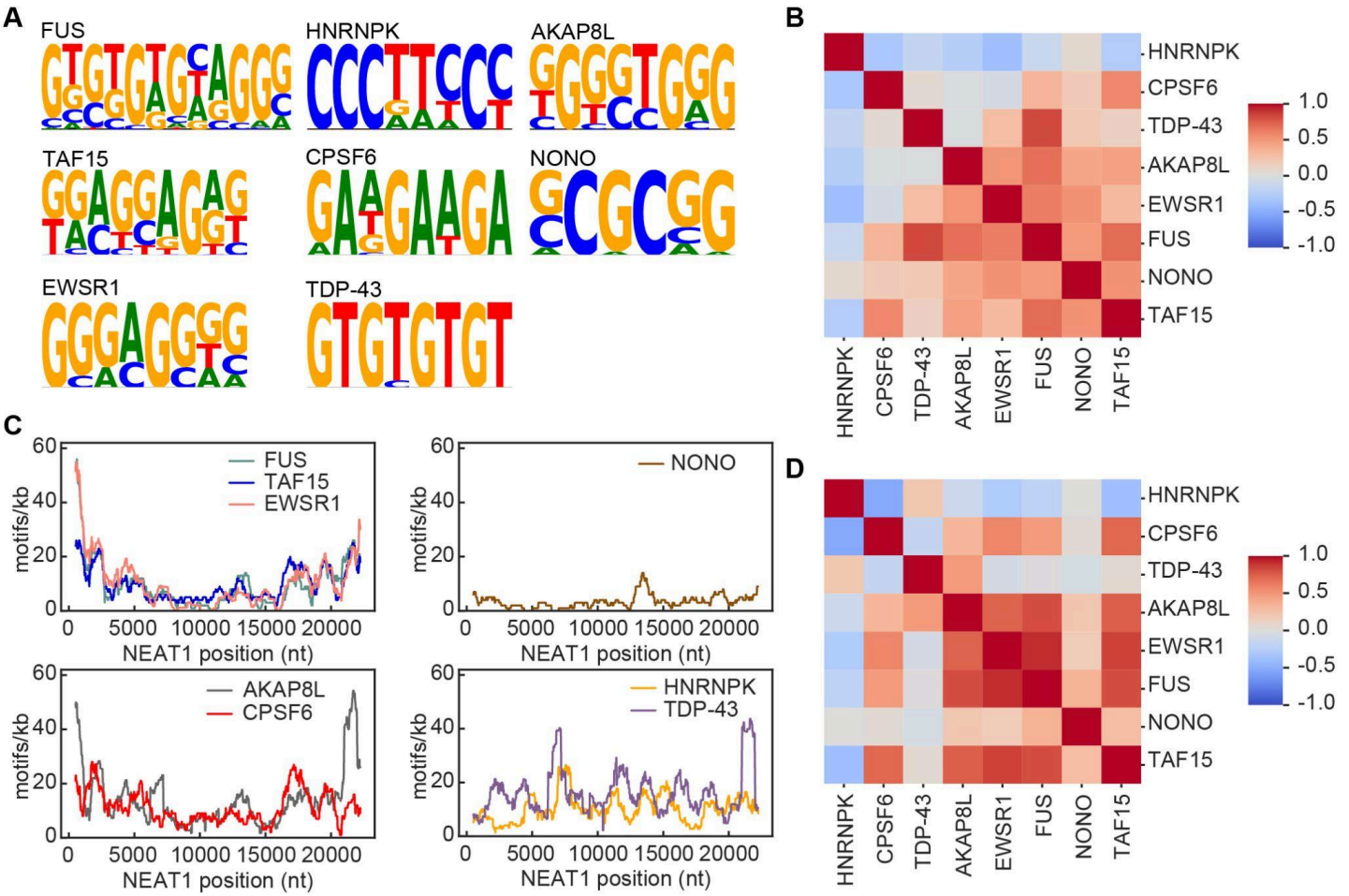

Fig. S11. Some paraspeckle-enriched proteins display similar motif distributions in NEAT1 without necessarily showing similar motif sequence preferences. (**A**) RNA-binding motifs for paraspeckle-enriched proteins[8], obtained from a compendium of motifs generated from ENCODE eCLIP data in ref[9]. FUS, NONO, and TDP-43 motifs used in this paper were obtained from the same compendium. (**B**) Clustered heatmap of correlations between pairs of motif sequences (Methods). (**C**) The number of motifs for each indicated protein within 1000 nt windows of NEAT1, tiled every 100 nt from 5′ to 3′. Similar to Fig. 2A. (**D**) Heatmap of correlations between pairs of motif *distributions* along NEAT1 from (C) (Methods). Proteins follow the same order as (B) to assist comparison of the two datasets.

Table S1. Interaction parameters and key model input values used in coarse-grained simulations of the current (i.e. “old”) model (Fig. 1 and S1A) and revised (i.e. “new”) model (Fig. 5 and S9A) of paraspeckle assembly. Also included are model parameters for simulations mimicking TDP-43 mutations or aging-associated nuclear loss of TDP-43.

Inputs	New Model	Old Model	No TDP-43	Reduced TDP-43	Increase NEAT1	Reduced TDP-43 + Increase NEAT1	Decrease Stick TDP-43	Increase Stick TDP-43
Simulation Set-Ups								
Number of NEAT1 copies	20	20	20	20	40	40	20	20
Number of monomers in NEAT1	45	45	45	45	45	45	45	45
Number of monomers for 5'/3' end	15	15	15	15	15	15	15	15
Number of monomers for middle	15	15	15	15	15	15	15	15
Length of box (x,y,z)	150	150	150	150	150	150	150	150
Number of proteins total	12000	12000	9000	10000	12000	10000	12000	12000
Number of NONO proteins	6000	6000	6000	6000	6000	6000	6000	6000
Number of FUS proteins	3000	6000	3000	3000	3000	3000	3000	3000
Number of TDP43 proteins	3000	0	0	1000	3000	1000	3000	3000
Lennard Jones Well-Depths								
Middle-Middle NEAT1	2.0	0.1	2.0	2.0	2.0	2.0	2.0	2.0
5'-5', 3'-3',5'-3' NEAT1	0.01	0.1	0.01	0.01	0.01	0.01	0.01	0.01
5'-M , 3'-Middle NEAT1	0.01	0.1	0.01	0.01	0.01	0.01	0.01	0.01
NONO-NONO	3.0	2	3.0	3.0	3.0	3.0	3.0	3.0
FUS-FUS	3.0	2	3.0	3.0	3.0	3.0	3.0	3.0
TDP-43-TDP-43	3.0	N/A	3.0	3.0	3.0	3.0	1.0	3.5
NONO-FUS	3.0	2	3.0	3.0	3.0	3.0	3.0	3.0
NONO-TDP-43	0.01	N/A	0.01	0.01	0.01	0.01	0.01	0.01
FUS-TDP-43	0.01	N/A	0.01	0.01	0.01	0.01	0.01	0.01
NONO-Middle NEAT1	2.5	3	2.5	2.5	2.5	2.5	2.5	2.5
FUS-Middle NEAT1	2.5	0.1	2.5	2.5	2.5	2.5	2.5	2.5
TDP-43-Middle NEAT1	0.01	N/A	0.01	0.01	0.01	0.01	0.01	0.01
NONO-5' NEAT1	2.5	0.1	2.5	2.5	2.5	2.5	2.5	2.5

NONO-3' NEAT1	2.5	0.1	2.5	2.5	2.5	2.5	2.5	2.5
FUS-5' NEAT1	3.0	0.1	3.0	3.0	3.0	3.0	3.0	3.0
FUS-3' NEAT1	3.0	0.1	3.0	3.0	3.0	3.0	3.0	3.0
TDP-43-5' NEAT1	2.0	N/A	2.0	2.0	2.0	2.0	2.0	2.0
TDP-43-3' NEAT1	2.0	N/A	2.0	2.0	2.0	2.0	2.0	2.0
Lennard Jones Sigma								
RNA-RNA Sigma	1.0	1	1.0	1.0	1.0	1.0	1.0	1.0
Protein-RNA Sigma	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75
Protein-Protein Sigma	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5

SI References

1. Anderson, J.A., Glaser, J., and Glotzer, S.C. (2020). HOOMD-blue: A Python package for high-performance molecular dynamics and hard particle Monte Carlo simulations. Comput. Mater. Sci. *173*, 109363. https://doi.org/10.1016/j.commatsci.2019.109363.

2. Ramasubramani, V., Dice, B.D., Harper, E.S., Spellings, M.P., Anderson, J.A., and Glotzer, S.C. (2020). freud: A software suite for high throughput analysis of particle simulation data. Comput. Phys. Commun. *254*, 107275. https://doi.org/10.1016/j.cpc.2020.107275.

3. Shelkovnikova, T.A., Kukharsky, M.S., An, H., Dimasi, P., Alexeeva, S., Shabir, O., Heath, P.R., and Buchman, V.L. (2018). Protective paraspeckle hyper-assembly downstream of TDP-43 loss of function in amyotrophic lateral sclerosis. Mol. Neurodegener. *13*, 30. https://doi.org/10.1186/s13024-018-0263-7.

4. Neumann, M., Sampathu, D.M., Kwong, L.K., Truax, A.C., Micsenyi, M.C., Chou, T.T., Bruce, J., Schuck, T., Grossman, M., Clark, C.M., et al. (2006). Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Science *314*, 130–133. https://doi.org/10.1126/science.1134108.

5. Lang, R., Hodgson, R.E., and Shelkovnikova, T.A. (2024). TDP-43 in nuclear condensates: where, how, and why. Biochem. Soc. Trans. *52*, 1809–1825. https://doi.org/10.1042/BST20231447.

6. Versluys, L., Ervilha Pereira, P., Schuermans, N., De Paepe, B., De Bleecker, J.L., Bogaert, E., and Dermaut, B. (2022). Expanding the TDP-43 Proteinopathy Pathway From Neurons to Muscle: Physiological and Pathophysiological Functions. Front. Neurosci. *16*. https://doi.org/10.3389/fnins.2022.815765.

7. Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: Visual molecular dynamics. J. Mol. Graph. *14*, 33–38. https://doi.org/10.1016/0263-7855(96)00018-5.

8. Fox, A.H., Nakagawa, S., Hirose, T., and Bond, C.S. (2018). Paraspeckles: Where Long Noncoding RNA Meets Phase Separation. Trends Biochem Sci *43*, 124–135. https://doi.org/10.1016/j.tibs.2017.12.001.

9. Katsantoni, M., van Nimwegen, E., and Zavolan, M. (2023). Improved analysis of (e)CLIP data with RCRUNCH yields a compendium of RNA-binding protein binding sites and motifs. Genome Biol *24*, 77. https://doi.org/10.1186/s13059-023-02913-0.

10. Jiang, L., Wang, M., Lin, S., Jian, R., Li, X., Chan, J., Dong, G., Fang, H., Robinson, A.E., G. TEx Consortium, et al. (2020). A Quantitative Proteome Map of the Human Body. Cell *183*, 269-283 e19. https://doi.org/10.1016/j.cell.2020.08.036.

11. Müller, J.B., Geyer, P.E., Colaço, A.R., Treit, P.V., Strauss, M.T., Oroshi, M., Doll, S., Virreira Winter, S., Bader, J.M., Köhler, N., et al. (2020). The proteome landscape of the kingdoms of life. Nature *582*, 592–596. https://doi.org/10.1038/s41586-020-2402-x.

12. Ba, Q., Hei, Y., Dighe, A., Li, W., Maziarz, J., Pak, I., Wang, S., Wagner, G.P., and Liu, Y. (2022). Proteotype coevolution and quantitative diversity across 11 mammalian species. Sci. Adv. *8*, eabn0756. https://doi.org/10.1126/sciadv.abn0756.