Supplementary Figures

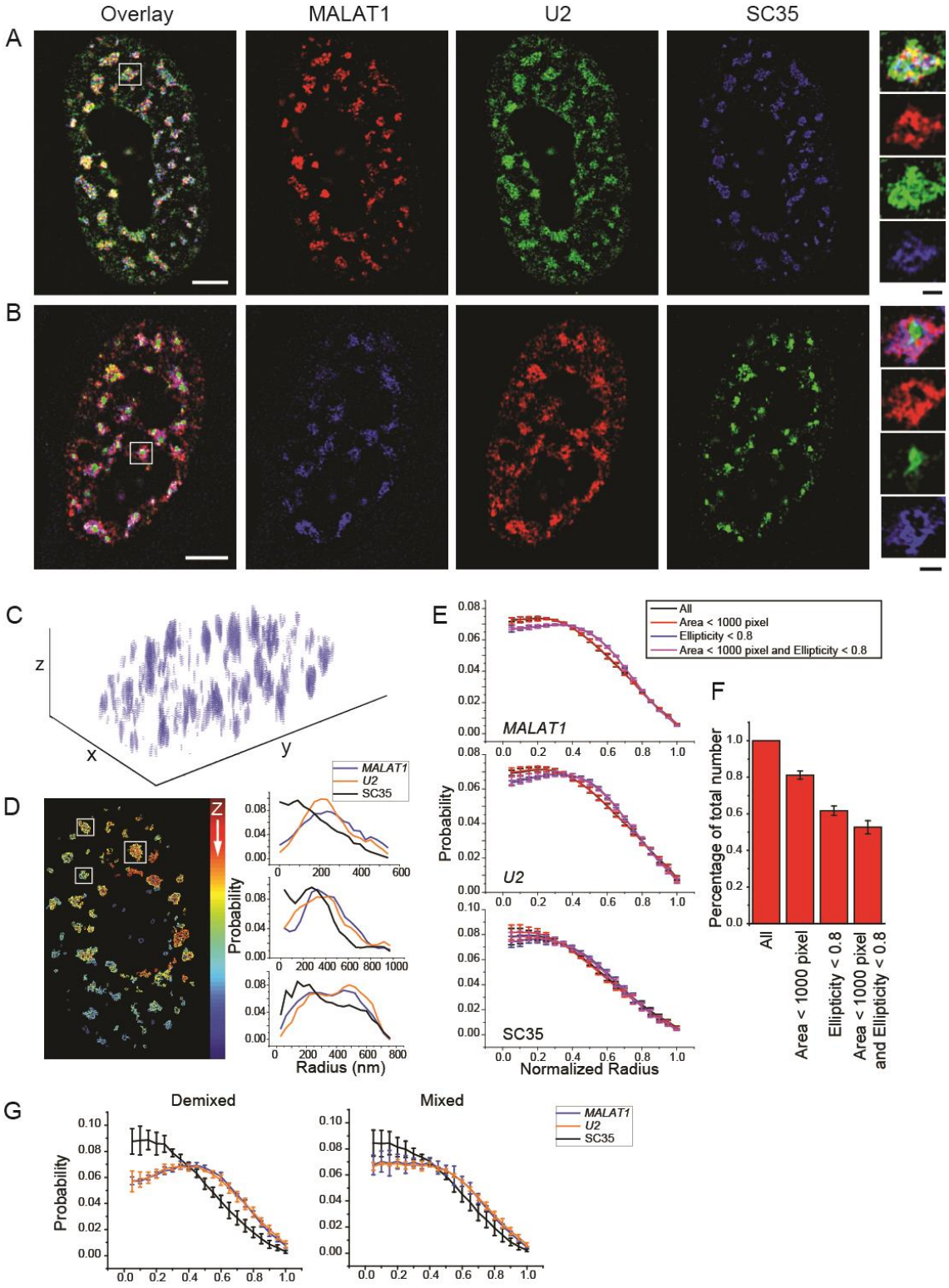


Figure S1. Additional sample images of *MALAT1*, *U2* and SC35 and analysis of radial distribution. (A) Example in which SC35 and *MALAT1/U2* are more mixed compared to the

example for the demixing case shown in Fig. 1A. (B) Labeling scheme of *MALAT1*, *U2* and SC35 is changed compared to the example in Fig. 1A. Scale bars represent 5 µm in cell images and 1 µm zoomed-in speckle images. (C) 3D rendering of individual speckles identified by applying intensity threshold. (D) Contour plot of selected speckle, with color bar representing different z planes. Each speckle is considered first as a 3D sphere. The middle z plane of the 3D sphere is isolated. The intensity density of each components is plotted as a function of radius from the geometric center of each speckle to report the radial distribution of each component. (E) Superposition of the radial distributions from all speckles with different cutoffs of area and/or ellipticity. (F) Percentage of selected speckles after area and/or ellipticity cutoffs. (G) Probability density distribution as a function of radius for *MALAT1*, *U2* and SC35 from the geometric center of the speckle from selected speckles showing a demixing or mixing organization respectively. Radius is normalized to the distance from the center (set to 0) to the boundary of the speckle (set to 1).

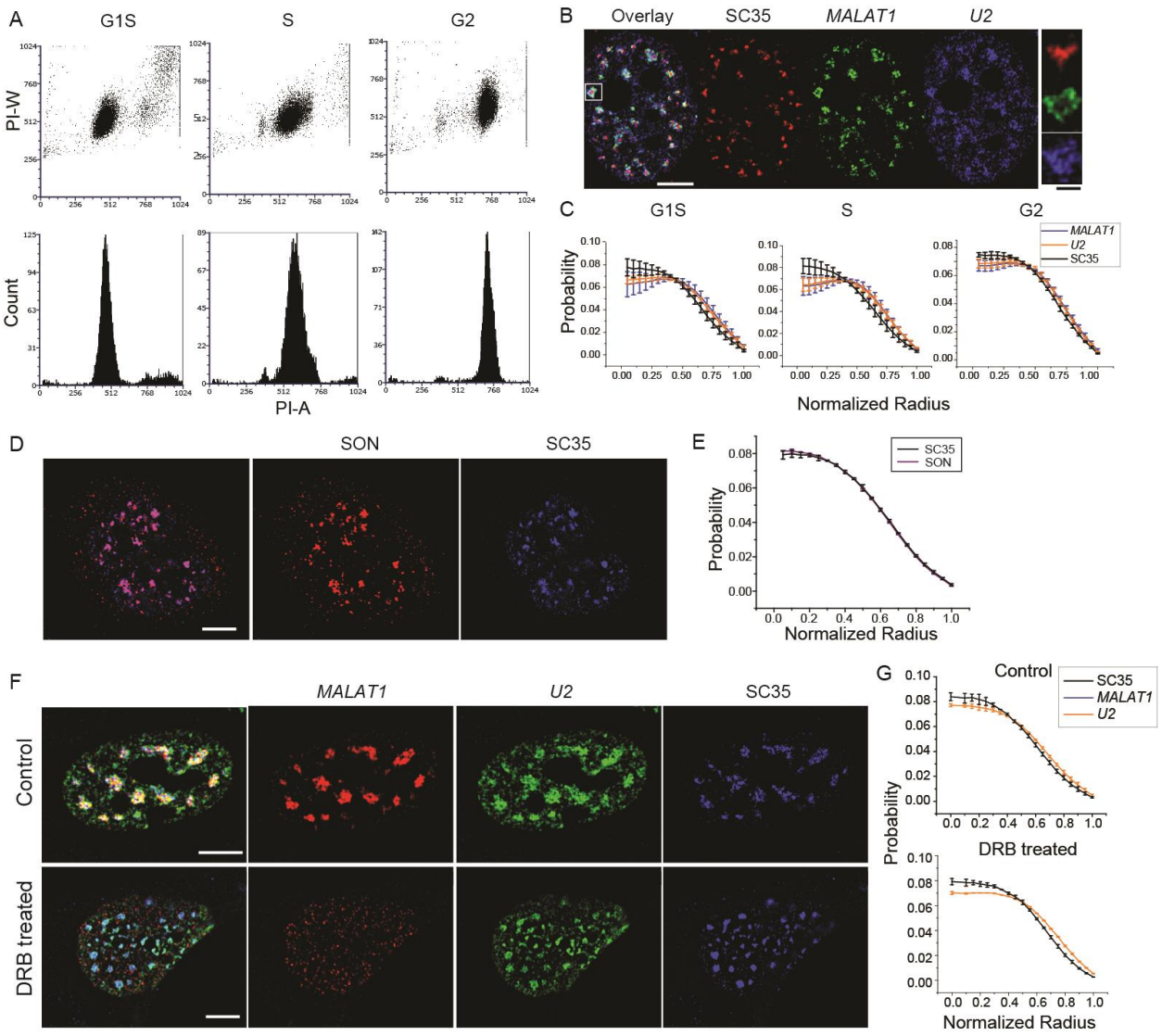


Figure S2. Non-random compositional organization of nuclear speckle during various cell cycle stages and under transcription inhibition. (A) Flow cytometry analyses to demonstrate cell cycle synchronization in HeLa cells (B) Sample image of *MALAT1* (green), *U2* (blue) and SC35 (red) by SIM in S phase HeLa cells (C) Normalized radial distribution of *MALAT1*, *U2* and SC35 in G1/S, S and G2 phases. Histograms are built by combing speckles from four independent measurements. Each measurement contains 200-600 speckles from 20-40 cells. (D) Hela cells are synchronized to G1 phase, and SON (red) and SC35 (blue) are stained with immunofluorescence and imaged by SIM. (E) Average radial distribution of SON and SC35. Error bars represents standard deviation from two independent data measurements. Each measurement contains 500-1000 speckles from 30-60 cells. (F) Sample image of *MALAT1* (red), *U2* (green) and SC35 (blue) by SIM in control and transcription-inhibited (DRB-treated) WI-38 cells. (G) Average radial distribution of *MALAT1*, *U2* and SC35 in control and transcription-inhibited cells. Error bars represents standard deviation from two independent measurements. Each measurement contains 500-1000 speckles from 30-60 cells. Scale bars represent 5 μm in cell images and 1 μm zoomed-in speckle images.

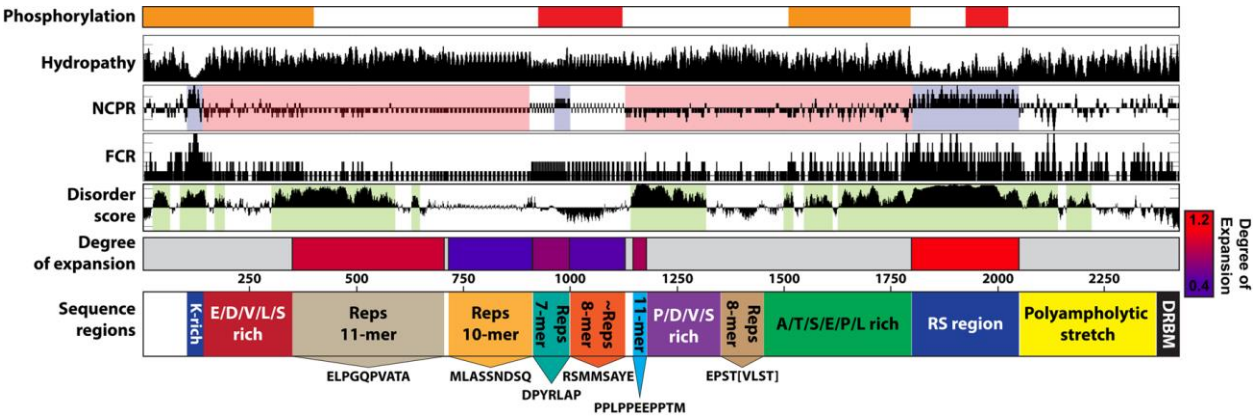


Figure S3. Summary of the sequence attributes of SON that are used to develop the architecture for coarse-graining. The linear sequence properties of SON are quantified using a series of distinct metrics that are organized into sequence-aligned "tracks". The phosphorylation track qualitatively defines the regions that are predicted to undergo multi-site phosphorylation, based on data from the ProteomeScout database (Matlock et al., 2015). The hydropathy track measures the local sequence hydrophobicity as calculated over a five-residue sliding window based on the Kyte-Doolittle hydrophobicity scores (Kyte and Doolittle, 1982). The net charge per residue (NCPR track) measures the local net-charge per residue associated with a five-residue sliding window. Extended stretches with a non-zero net charge are highlighted in blue (positive) or red (negative). The fraction of charged residues (FCR track) uses a five-residue sliding window to measure the local density of charged residues irrespective of the sign of the charge. The disorder score track defines the IUPred-long (Dosztányi et al., 2005) disorder prediction score for the local region. Extended regions with an IUpred score of >0.5 are highlighted in green. The degree of expansion measures the normalized radius of gyration from all-atom simulations of multiple tandem repeats extracted from a defined region. The calculated radii of gyration were normalized by values obtained for atomistic self-avoiding random walks that were obtained for the matching sequence stretch. This score provides a qualitative description of the propensity for this region of SON to be conformationally expanded or collapsed. For context, folded proteins typically have a degree of expansion of 0.3-0.4 (i.e., lower than any of the regions examined here). The RS region towards the C-terminus of SON has a degree of expansion of ~1.2. This means that it is substantially more expanded than a self-avoiding random walk, as would be expected for a strong polyelectrolyte (Mao et al., 2013). The sequence regions track at the bottom characterizes specific local repeats across the protein. Consensus repeats were identified using local sequence alignments. Analyses of NCPR, FCR and hydropathy analysis were performed using localCIDER (Holehouse et al., 2017). All atom simulations were performed using the CAMPARI Monte Carlo modeling suite (http://campari.sourceforge.net) and the ABSINTH implicit forcefield paradigm (Vitalis and Pappu, 2014). Simulations were set up with the approach described in Pak *et al.* (Pak et al., 2016).

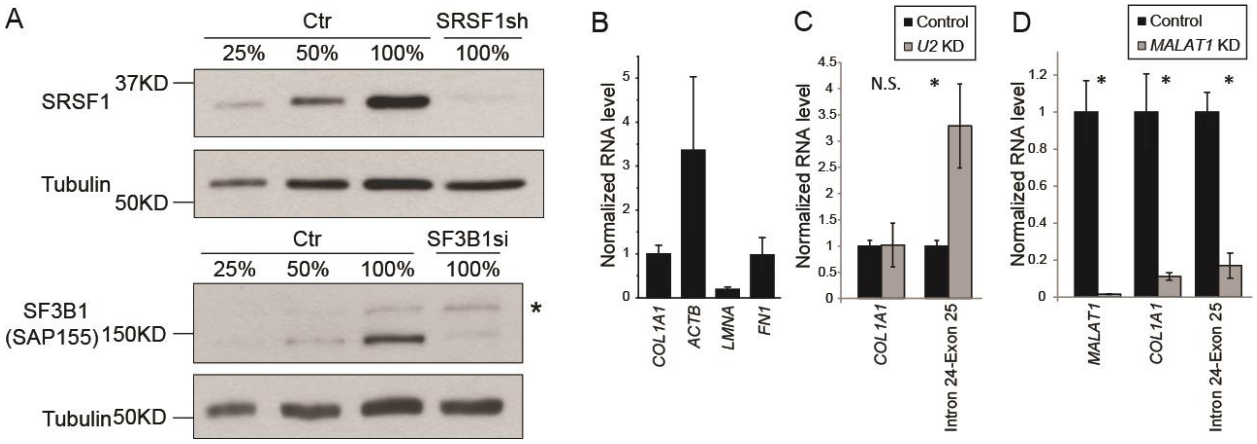


Figure S4. Immunoblot and qPCR analysis for knock-down verification. (A) Immunoblot analysis to depict the efficiency of knock down of SRSF1 and SF3B1 in WI-38 cells treated with SRSF1 shRNA or SF3b1 siRNA. Tubulin is used as loading control. * represents the cross-reacting band. (B) Total cellular mRNA level for four genes. mRNA abundance is normalized to *COL1A1* mRNA level. (C) and (D) RT-qPCR measurements of cellular level of *COL1A1* mRNA and intron 24-exon 25 junction region of *COL1A1* pre-mRNA in *U2* depleted and *MALAT1* depleted cell. *MALAT1* knockdown efficiency is also measured by RT-qPCR in (D). Data are normalized to the negative control cells and error bars report the standard deviations of three replicates.

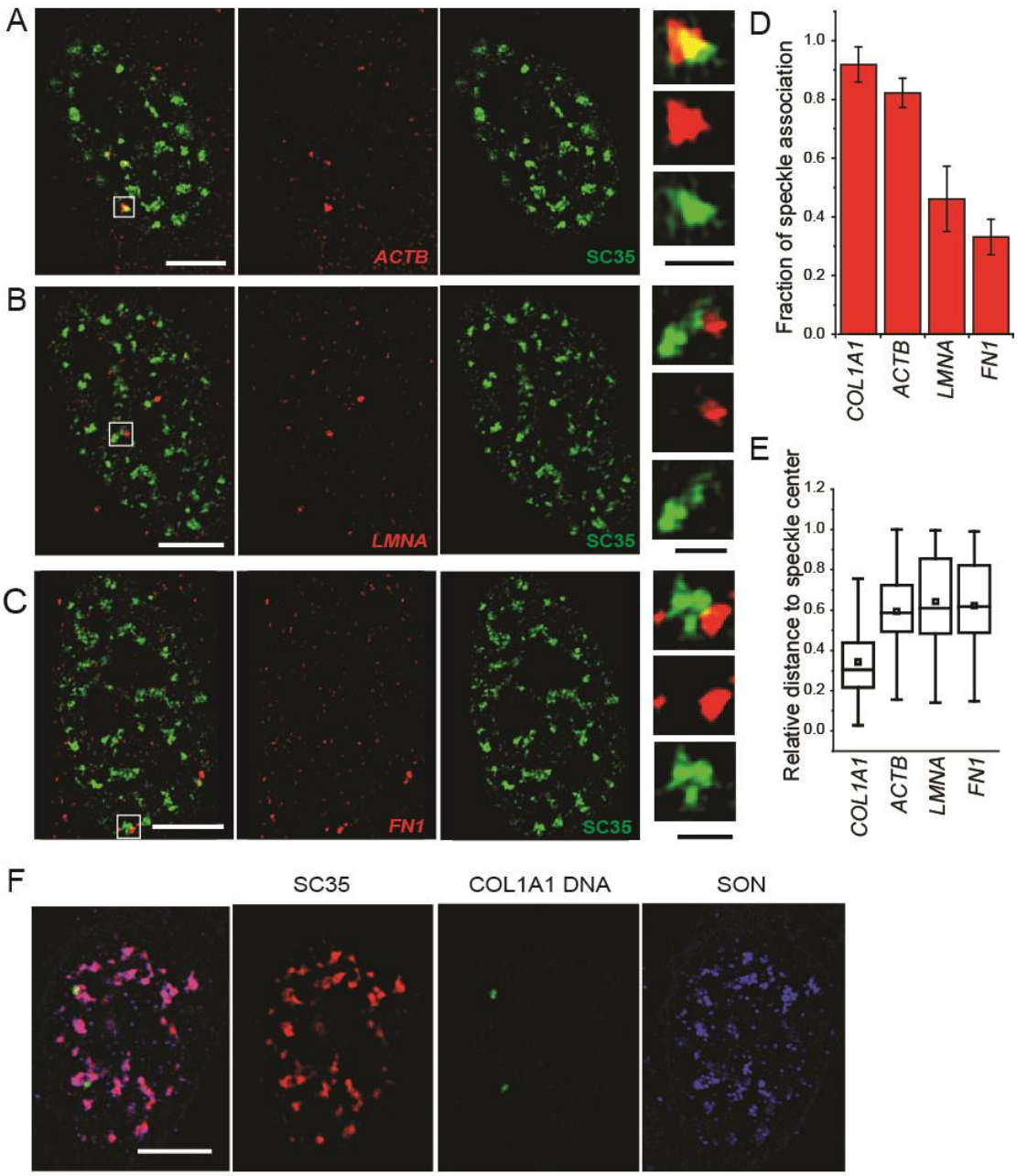


Figure S5. Association of RNA transcripts with nuclear speckles. (A) – (C) Sample images of *ACTB*, *LMNA* and *FN1* RNAs (red) with SC35 (green) respectively. Scale bars represent 5 µm in all cell images and 1 µm all zoomed-in speckle images. (D) Percentage of speckle-associated RNA accumulation sites. Error bars report the standard deviation. (E) Box-and-whisker plot of relative distance between the centers of RNA accumulated sites and the centers of the associated speckles, normalized by the radius of the speckle (D_{Rel} as explained in the text). (F) Association of COL1A1 gene with nuclear speckles. Sample images of COL1A1 gene (green) labeled with Texas Red by DNA FISH with SC35 (red) and SON (blue). Scale bars represent 5 µm. Plots in (D) and (E) are generated from 3-4 independent measurements. Each measurement contains 15-75 RNA-containing speckles from 20-50 cells.

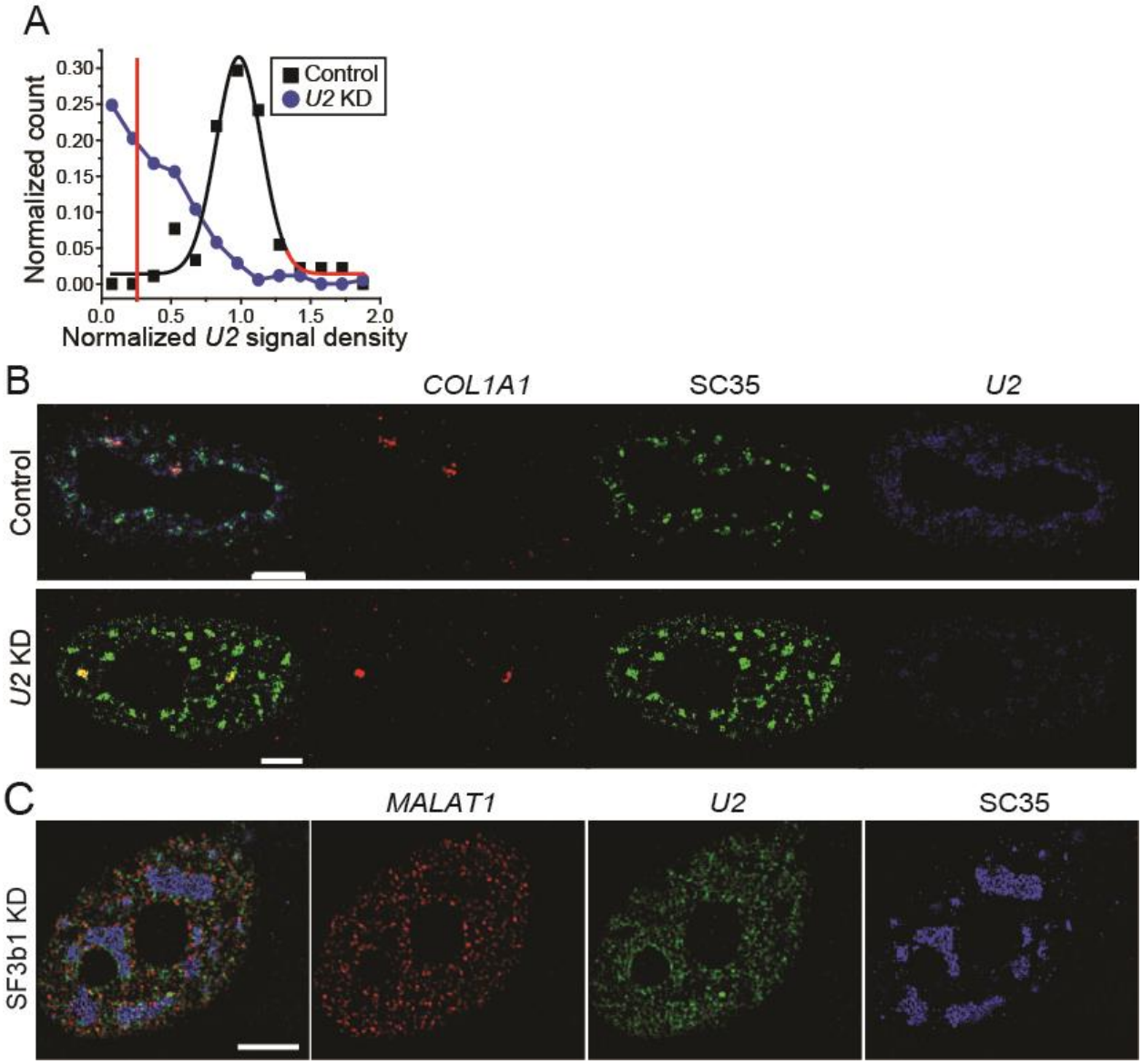


Figure S6. Effect of *U2* and SF3b1 knockdown on nuclear speckles. (A) Normalized *U2* intensity histograms for the control and *U2* knockdown cells. Cells with average *U2* intensity less than 25% of that in the negative control are selected for analysis. (B) Sample image of *COL1A1* RNA (red), SC35 (green) and *U2* snRNA (blue) in the control and *U2* partially knockdown cells. (C) Sample image of *MALAT1* (red), *U2* snRNA (green) and SC35 (blue) in the SF3B1-depleted WI-38 cells. The example of corresponding control cell with the same labeling scheme is shown in Fig. S5A. Scale bars represent 5 µm.

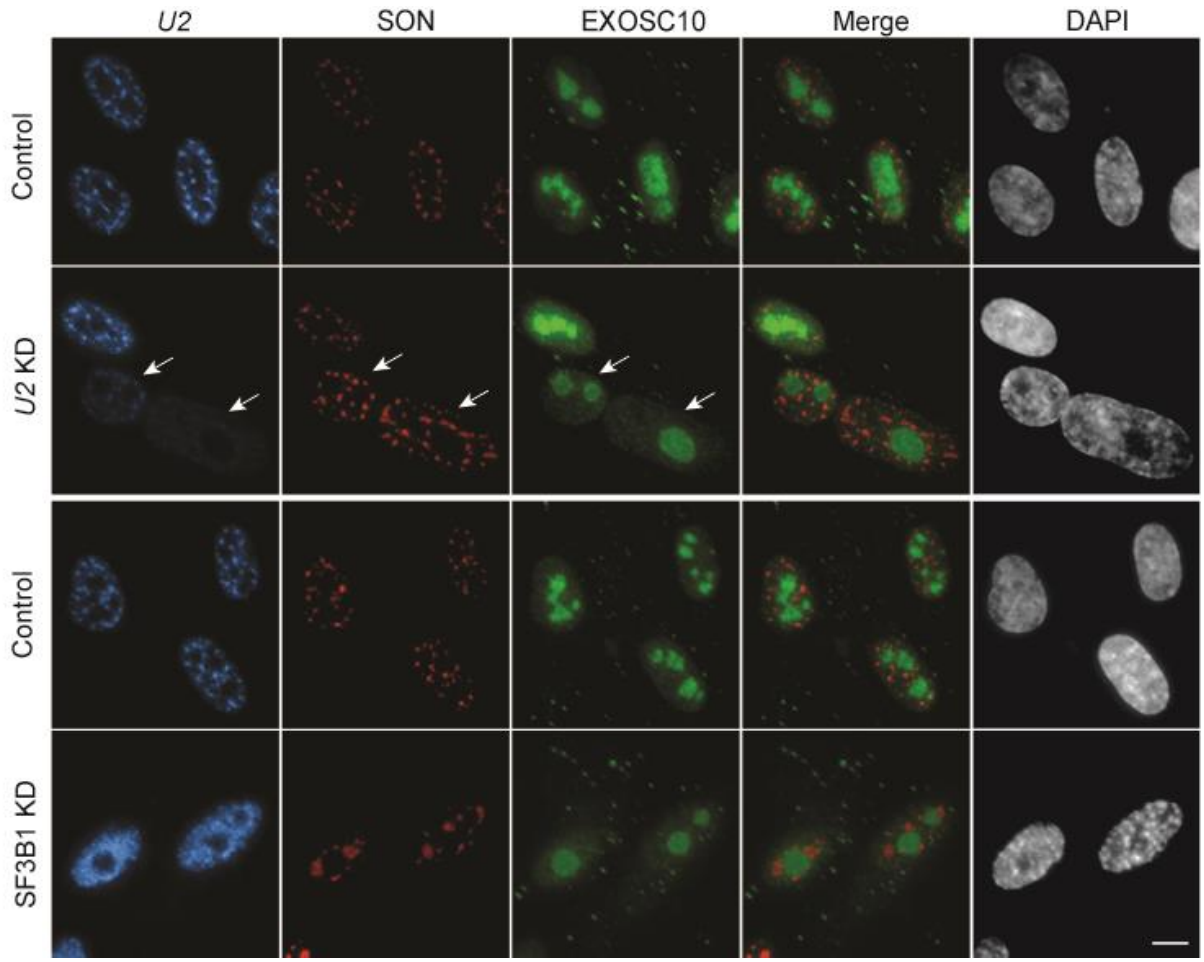


Figure S7. Nuclear speckles do not represent major sites of RNA degradation in the nucleus. Co-FISH and immunostaining to detect *U2* snRNA (blue), SON (red) and EXOSC10 (green) in control, *U2*-depleted and SF3B1-depleted WI-38 cells. DNA is stained with DAPI. Scale bar represents 5 µm. Please note that cells depleted of *U2* snRNA (arrow) show more prominent SON-labeled speckles.

References

Dosztányi, Z., Csizmok, V., Tompa, P., and Simon, I. (2005). IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics *21*, 3433–3434.

Holehouse, A.S., Das, R.K., Ahad, J.N., Richardson, M.O.G., and Pappu, R.V. (2017). CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins. Biophys J *112*, 16–21.

Kyte, J., and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. J Mol Biol *157*, 105–132.

Mao, A.H., Lyle, N., and Pappu, R.V. (2013). Describing sequence-ensemble relationships for intrinsically disordered proteins. Biochem J *449*, 307–318.

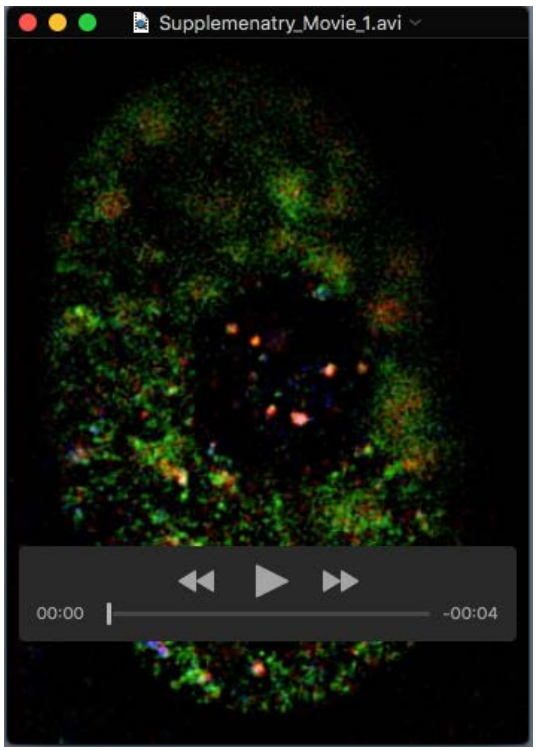
Matlock, M.K., Holehouse, A.S., and Naegle, K.M. (2015). ProteomeScout: a repository and analysis resource for post-translational modifications and proteins. Nucleic Acids Res *43*, D521–30.

Pak, C.W., Kosno, M., Holehouse, A.S., Padrick, S.B., Mittal, A., Ali, R., Yunus, A.A., Liu, D.R., Pappu, R.V., and Rosen, M.K. (2016). Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein. Mol Cell *63*, 72–85.

Vitalis, A., and Pappu, R.V. (2014). A simple molecular mechanics integrator in mixed rigid body and dihedral angle space. J Chem Phys *141*, 034105.

Table S1. RNA-sequencing data for all of the annotated genes in human fibroblasts sorted by RPKM value. Data was collected from a study by Marthandan et al. (Marthandan et al., 2016 Plos One 11(5):e0154531. doi: 10.1371/journal.pone.0154531).

Click here to Download Table S1



Movie S1. Representative movie of nuclear speckle stained with MALAT1 (red), U2 (green) and SC35 (blue). Movie is shown from the top to the bottom of the cell in z direction.