© 2017. Published by The Company of Biologists Ltd | Journal of Cell Science (2017) 130, 4180-4192 doi:10.1242/jcs.206854

RESEARCH ARTICLE

Quantitative analysis of multilayer organization of proteins and RNA in nuclear speckles at super resolution

Jingyi Fei[1,2,‡,¶], Mahdieh Jadaliha[3,‡], Tyler S. Harmon[4,§], Isaac T. S. Li[5,§], Boyang Hua[6,*], Qinyu Hao[3], Alex S. Holehouse[4], Matthew Reyer[2], Qinyu Sun[3], Susan M. Freier[7], Rohit V. Pappu[4], Kannanganattu V. Prasanth[3,¶] and Taekjip Ha[6,8,9,10,*]

ABSTRACT

Nuclear speckles are self-assembled organelles composed of RNAs and proteins. They are proposed to act as structural domains that control distinct steps in gene expression, including transcription, splicing and mRNA export. Earlier studies identified differential localization of a few components within the speckles. It was speculated that the spatial organization of speckle components might contribute directly to the order of operations that coordinate distinct processes. Here, by performing multi-color structured illumination microscopy, we characterized the multilayer organization of speckles at a higher resolution. We found that SON and SC35 (also known as SRSF2) localize to the central region of the speckle, whereas *MALAT1* and small nuclear (sn)RNAs are enriched at the speckle periphery. Coarse-grained simulations indicate that the non-random organization arises due to the interplay between favorable sequence-encoded intermolecular interactions of speckle-resident proteins and RNAs. Finally, we observe positive correlation between the total amount of RNA present within a speckle and the speckle size. These results imply that speckle size may be regulated to accommodate RNA accumulation and processing. Accumulation of RNA from various actively transcribed speckle-associated genes could contribute to the observed speckle size variations within a single cell.

KEY WORDS: Long noncoding RNA, Nuclear domain, Splicing factor, Sub-nuclear compartmentalization

INTRODUCTION

Eukaryotic cells contain both membrane-bound and non-membranous organelles. Non-membranous organelles are proposed to form through non-stoichiometric assembly of proteins and RNAs (Banani et al., 2017). These RNA/protein granules are pervasive in both the cytoplasm (e.g. stress granules and P-bodies) (Buchan and Parker, 2009; Decker and Parker, 2012) and the nucleus (e.g. nucleoli, nuclear speckles and paraspeckles) (Dundr, 2012; Mao et al., 2011a; Zhu and Brangwynne, 2015), and are believed to influence the cellular activities. In the mammalian nucleus, nuclear speckles form one of the most conspicuous RNA/protein granules. Fluorescence imaging of nuclear speckle marker proteins (Fu and Maniatis, 1990) typically reveals 20–50 irregular punctate structures per cell, with sizes ranging from 0.5 µm up to a few microns (Spector and Lamond, 2011). In addition to poly(A)-positive RNAs and noncoding RNAs, nuclear speckles contain ∼200 proteins involved in pre-mRNA metabolism, including pre-mRNA splicing, surveillance and RNA export, as well as non-pre-mRNA processing factors such as transcription factors, translation initiation factors and subunits of RNA polymerase II (Hall et al., 2006; Spector and Lamond, 2011).

Many biological processes have been linked to nuclear speckles, although their exact physiological function remains under debate (Spector and Lamond, 2011). Earlier studies suggested that nuclear speckles act as storage and/or assembly sites for pre-mRNA processing factors. These conclusions are supported by the lack of DNA and low level of transcriptional activity inside the speckles (Fakan and van Driel, 2007; Spector and Lamond, 2011). However, other studies have suggested that speckles play an active role in coordinating gene expression (Hall et al., 2006). A subset of genes tend to cluster around nuclear speckles upon transcriptional activation, and mRNAs of several of the speckle-associated genes colocalize with speckles (Brown et al., 2008; Hall et al., 2006; Han et al., 2011; Hu et al., 2010; Khanna et al., 2014; Moen et al., 2004; Shopland et al., 2002; Smith et al., 1999; Yang et al., 2011), implying that nuclear speckles may serve as an active hub, coordinating transcription of these genes with the processing of their pre-mRNAs (Hall et al., 2006). In addition, nuclear speckles may also be involved in post-transcriptional mRNA splicing, mRNA export and quality control for specific splice-defective pre-mRNAs (Ando et al., 2016; Dias et al., 2010; Girard et al., 2012; Hall et al., 2006; Han et al., 2011; Mor et al., 2016).

The internal organization of protein and RNA components within nuclear speckles has been a topic of investigation. Fluorescence and electron microscopy studies have demonstrated the existence of fine structures within speckles. Several constituents appear in the form of sub-speckles (0.2–0.5 µm) within a speckle (Mintz and Spector, 2000) and are composed of clusters of electron-dense granules of 20–25 nm in size, connected by thin fibrils (Thiry, 1995a,b). In addition, a few components have been reported to occupy different locations within nuclear speckles. For example, poly(A)-positive RNA within the nuclear speckles define a broader area compared to the region defined by speckle marker proteins, such as

[1]Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, IL 60637, USA. [2]Institute for Biophysical Dynamics, The University of Chicago, Chicago, IL 60637, USA. [3]Department of Cell and Developmental Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. [4]Department of Biomedical Engineering and Center for Biological Systems Engineering, Washington University in St. Louis, St. Louis, MO 63130, USA. [5]Department of Chemistry, University of British Columbia Okanagan, Kelowna, British Columbia, Canada, V1V 1V7. [6]Center for Biophysics and Quantitative Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. [7]Ionis Pharmaceuticals Inc., Carlsbad, CA 92010, USA. [8]Department of Physics, Center for the Physics of Living Cells, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. [9]Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. [10]Howard Hughes Medical Institute, Johns Hopkins University, Baltimore, MD 21205, USA.
‡,§These authors contributed equally to this work
*Present address: Department of Biophysics and Biophysical Chemistry, Department of Biophysics and Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21205, USA.

¶Author for correspondence (jingyifei@uchicago.edu; kumarp@illinois.edu)

K.V.P., 0000-0003-4587-8362; J.F., 0000-0002-9775-3820

Received 29 May 2017; Accepted 1 November 2017

Journal of Cell Science

Journal of Cell Science (2017) 130, 4180-4192 doi:10.1242/jcs.206854

phosphorylated SR-family splicing regulators [as stained with an anti-SC35 (also known as SRSF2) antibody that preferentially recognizes phosphorylated SR proteins (Fu and Maniatis, 1990)] or splicing co-activator Srm300 (also know as SRRM2) (Carter et al., 1993; Hall et al., 2006). This data suggests that the SC35- or SRm300-stained region within the speckle could form a core domain that is surrounded by a larger shell decorated by poly(A)-positive RNA. In addition, RNA synthesized from several of the speckle-associated genes localized either in speckle periphery or interior (Hall et al., 2006). For example, collagen 1A1-encoding (*COL1A1*) mRNA was shown to localize to the interior core of the speckle, whereas fibronectin (*FN1*) or actin B (*ACTB*) transcripts were predominantly enriched at speckle periphery (Johnson et al., 2000; Shopland et al., 2002). All these observations indicate a non-random organization of at least a subset of the speckle components. Based on these observations, it has been proposed that nuclear speckles are functionally sub-compartmentalized, with transcription and most of the splicing being executed in the speckle periphery, whereas the inner part of the domain plays a vital role in modulating other steps of RNA maturation (Hall et al., 2006). However, a systematic analysis to quantify the relative distribution of various protein and RNA components within the speckle is needed to develop models of the order of operations that might be the hallmarks of speckle functions.

Here, we used multi-color structured illumination microscopy (SIM) to confirm a distinctive multilayer organization of protein and RNA components within nuclear speckles (Gustafsson et al., 2008; Schermelleh et al., 2008). The SIM data reveal a multilayered organization of all labeled components at a higher resolution than in previous studies, with several spliceosome components defining a broader territory compared to splicing regulators and scaffold proteins. A lattice-based computational model is able to reproduce the observed multilayer organization. In this model, spatially organized architectures result from spontaneous self-organization of nuclear speckle components through the interplay of sequence-encoded and solvent-mediated physical interactions between protein and RNA modules. In addition, our data also suggest a role for RNAs in influencing the size of nuclear speckle.

RESULTS

Nuclear speckle components exhibit layered distribution

In order to determine the spatial organization of speckle constituents, we analyzed the distribution of the protein SC35, together with an essential spliceosome component (*U2* snRNA), and a speckle-enriched long non-coding (lnc)RNA (*MALAT1*) (Hutchinson et al., 2007; Ji et al., 2003) in human lung diploid fibroblasts (WI-38 cells). The intra-speckle distribution of these three components were measured by performing 3D SIM using a combination of fluorescence *in situ* hybridization (FISH) (Raj et al., 2008) and immunofluorescence (IF) staining (Fig. 1A; Movie 1). Each of the three components formed small foci within the speckles, potentially corresponding to the sub-speckles. SC35 resided within the core of nuclear speckles, as reported previously (Hall et al., 2006). *U2* and *MALAT1* defined a broader territory (Fig. 1A). In a subset of cells (34±12%, mean±s.d., with variation among biological replicates), this trend was more obvious, and the majority of the speckles in those cells showed a stronger peripheral distribution of *U2* and *MALAT1*, while SC35 occupied the interior region of the speckles (Fig. 1A). We refer to these speckles as the 'demixed' population. In another subset of the cellular population (66±12%) the speckle components did not show a dramatic distinction between the interior and periphery, but *U2* and *MALAT1* still demonstrated a broader radial distribution compared to SC35 (Fig. S1A). We refer to these speckles as the 'mixed' population.

In order to exclude potential artifacts due to the specific fluorophores used to label speckle components, we switched the combination of the fluorophores and components and observed the same phenomena (Fig. S1B). SIM imaging of three additional speckle components [SON protein, *U1* snRNA and U2B″ (also known as SNRPB2) protein] showed that scaffold proteins such as SON localized to the speckle interior when compared to *U1* snRNA and *U2* snRNA-associated U2B″ (Fig. 1B–D).

To obtain a quantitative comparison between different speckle components, we developed an automated approach to analyze the compositional distribution of speckle constituents in thousands of speckles. We first selected individual speckles in 3D by applying an intensity threshold based on the summed intensities from all three channels (Fig. S1C). Since the resolution along the *z*-axis is worse than the resolution in the *xy*-plane in SIM images, we identified the middle *z*-plane of individual speckles based on selected 3D volumes, and only considered the distribution in the associated 2D plane. We determined the geometric center of each speckle and plotted the radial intensity distribution of each component (Fig. S1D). There are two factors that complicated the distribution analysis: (1) the relatively irregular shape of nuclear speckles, and (2) the large variation in the size of speckles compared to other nuclear bodies such as Cajal bodies and paraspeckles, which have more spherical shapes and relatively uniform sizes. We therefore applied two additional thresholds based on 2D area and ellipticity. Thresholding on 2D area excluded oversized speckles that may exist due to the fusion of multiple speckles, while thresholding on ellipticity excluded the speckles that strongly deviated from a spherical shape (Fig. S1E,F).

We superimposed the normalized radial distribution from all selected speckles (demixed and mixed populations) to generate the average radial distribution (Fig. 1E), which revealed a layered distribution of all of the tested speckle components. For comparison, we manually selected cells with obvious demixed and mixed organizations, and plotted the average radial distribution in each category (Fig. S1G). Both the demixed and mixed populations of cells showed layered distribution of speckle components, although the layered distribution was more evident for the demixed case and less evident for the mixed case as compared to the total population-averaged distribution shown in Fig. 1E. The protein SON, a member of the SR-family-like proteins, which has been implicated as a scaffold component in the nuclear speckle (Saitoh et al., 2004; Sharma et al., 2010), exhibited the same radial distribution as SC35. The localization of *U1*, *U2* snRNAs and *MALAT1* were indistinguishable from each other (Fig. 1E). U2B″ (Price et al., 1998; Scherly et al., 1990), a component of *U2* snRNP complex, was mainly present near the peripheral regions of nuclear speckles (Fig. 1E). Considering the radius at which the density of each component accumulated to 50% of the total (Fig. 1F), the outer layer decorated by *U1*, *U2* and *MALAT1*, was on average ∼10–15% larger than the adjacent inner layer occupied by SC35 and SON. Considering the average radius of nuclear speckle is ∼200–500 nm, this 10–15% increase (which corresponds to 20–50 nm) is difficult to resolve using conventional fluorescence microscopy with diffraction-limited resolution. The observed multilayer distribution of speckle components was unlikely to be due to differential accessibility of FISH probes or antibodies because U2B″, labeled by immunostaining, displayed the outermost localization, whereas *COL1A1* mRNA transcripts (described below), labeled by RNA-FISH, displayed interior speckle localization.

Journal of Cell Science (2017) 130, 4180-4192 doi:10.1242/jcs.206854

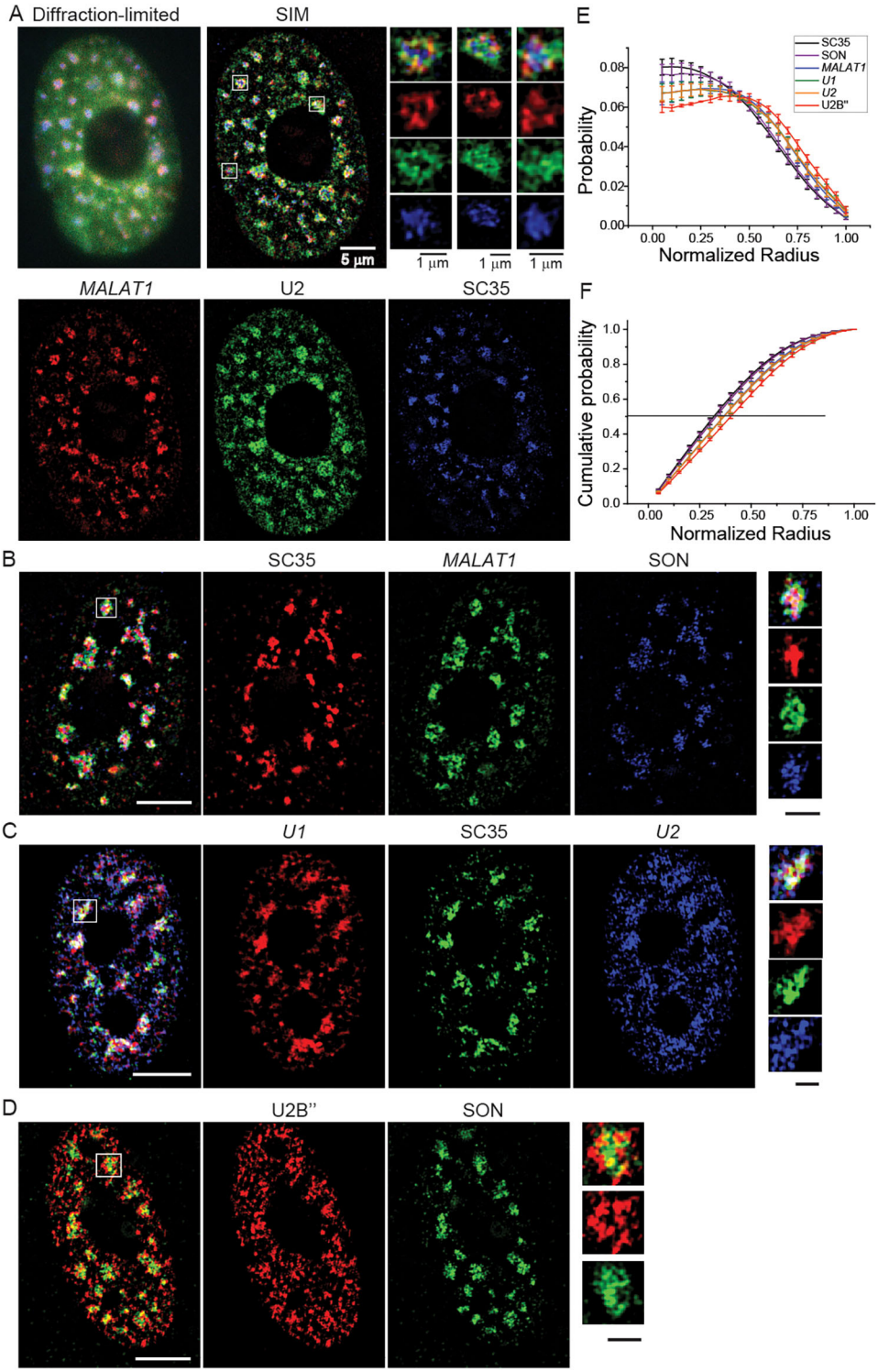


Fig. 1. Nuclear speckle components demonstrate a layered organization. (A) Sample image of *MALAT1* (red), *U2* (green) and SC35 (blue) with diffraction-limited fluorescence microscopy and SIM. Images are rendered in ImageJ for the center *z*-plane of the cell. (B) Combination images of SC35 (red), *MALAT1* (green) and SON (blue). (C) Combination images of *U1* (red), SC35 (green) and *U2* (blue). (D) Combination images of U2B″ (red) and SON (green). (E) Probability density distribution as a function of the radius for each component from the geometric center of the speckle. The radius is normalized to the distance from the center (set to 0) to the boundary of the speckle (set to 1). (F) Cumulative probability distribution as a function of radius for each component from the geometric center of the speckle. Error bars in E and F represent standard deviation from at least three independent measurements. Each measurement contains 150–400 speckles from 15–40 cells on average. Scale bars: 5 µm, cell images; 1 µm, magnified speckle images.

In order to test whether this layered organization shows cell cycle dependence, we performed the same analysis in HeLa cells, which could be synchronized into specific cell cycle stages (Fig. S2A). We imaged cells at the G1/S, S and G2 cell cycle phases (Fig. S2B). We picked G1/S over G1 phase because *MALAT1* is largely dispersed in the nucleoplasm during the G1 phase and is enriched in speckles in the G1/S phase (Tripathi et al., 2013). SC35, *MALAT1* and *U2* displayed similar organizations in speckles in HeLa cells in all tested phases as well as in WI-38 cells (Fig. S2C), suggesting that the layered distribution of speckle components is not limited to a particular cell type or to a specific stage in the cell cycle. In order to determine the stage where proteins such as SON and SC35 define the core of the speckle, we performed co-immunostaining for SON and SC35 in early G1 cells that had just exited mitosis. We found that SON and SC35 had already assembled into nuclear speckles, even in early G1 phase cells (Fig. S2D). The radial distributions of SC35 and SON were found to be the same in all of the cell cycle stages, including early G1, supporting the hypothesis that SC35 and SON define the structural core of nuclear speckles (Fig. S2E).

Journal of Cell Science (2017) 130, 4180-4192 doi:10.1242/jcs.206854

Next, we investigated whether active transcription contributes to the morphology and organization of nuclear speckles. We inhibited RNA polymerase II-mediated transcription by incubating cells with 5,6-dichloro-1-bold β-D-ribofuranosylbenzimidazole (DRB). As reported above (Fig. S2F), *MALAT1* was delocalized from nuclear speckles to the nucleoplasm in transcription-inhibited cells (Bernard et al., 2010). However, the layered organization of other speckle components such as SC35 and *U2* remained approximately the same in control versus transcription-inhibited cells (Fig. S2G), except for a slight decrease in the overall speckle size in transcription-inhibited cells (Fig. 2).

A minimalist computational model reproduces the multilayer organization of speckle components

A recent biophysical study showed that features of the well-established spatial organization of nucleoli can be reproduced *in vitro* using only a small subset of protein and RNA components. Importantly, lattice-based computer simulations also recapitulated this spatial organization (Feric et al., 2016). These simulations used a coarse-grained description of protein and RNA molecules based on the known structural and sequence properties of constituent molecules. Interactions between protein and RNA polymers were modeled by using an effective interaction matrix that captured the balance of polymer–solvent and polymer–polymer interactions. The simulations showed that spatial organization results from the differential affinities of protein and RNA components for one another as compared with their affinity for solvent. We investigated the generality of this approach (Feric et al., 2016) for reproducing the multilayer spatial organization of a five or six component system, in this case a minimalist form of the nuclear speckle. The ability to reproduce the multi-layer architecture of speckles would imply that our experimental observations might be consistent with coexisting demixed phases of SC35, SON, U2B″, *MALAT1* and *U1/U2*.

Following the approach of Feric et al. (2016), protein and RNA molecules were modeled as coarse-grained, lattice-based linear polymers made up of distinct interaction modules, as shown in Fig. 3A. The number and type of modules were determined using knowledge of the sequence architectures and known interaction domains (Fig. S3). The criteria used for mapping the sequences of protein and RNA molecules to lattice interaction modules are summarized in the Materials and Methods section. Each protein module represents a folded domain, an intrinsically disordered region, or a combination of both. The RNA modules represented regions that engage in either module or motif-specific interactions or non-specific interactions with other proteins and RNA molecules (see Fig. 3A for a description of the types of interactions). The effective strengths of pairwise interactions between protein and RNA modules were governed by the parameters of an interaction matrix (see Fig. 3B,D for two examples), which quantified the balance of inter-module and module–solvent interactions. The choice of which elements of the interaction matrix are non-zero was made based on known interaction domains and inferences from the sequence composition of disordered regions. The magnitudes of these energies were chosen based on the expected interaction specificities. Specific binding sites were given the strongest interaction strength, collapsing disordered regions were given the second strongest interactions, and non-specific interactions were given the weakest interactions. As discussed below, given our choice for the structure of the interaction matrix, there are numerous parameters that can reproduce the observed spatial organization of the five- or six-component system. This suggests that the absolute strengths of most interactions require very little fine tuning. Specifically, the pattern of localization for each protein/RNA is mainly determined by the structure of the interaction matrix, while the magnitudes of the interaction terms dictate the extent of segregation.

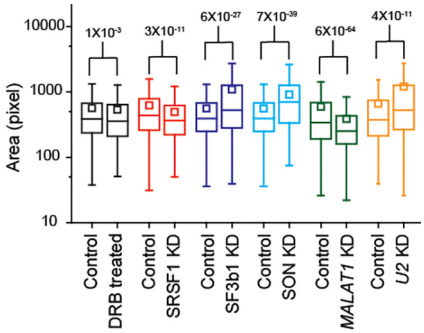


Fig. 2. Comparison of speckle size under various conditions. Size of the speckles as observed by SC35 staining is presented by the area of the speckle in the middle *z* slice, and shown in the box-and-whisker plots. *P*-values (calculated with a Kolmogorov–Smirnov test) are reported above the plots. For all box-and-whisker plots, the bottom and top of the box are the first and third quartiles; the band inside the box reports the median; whisker lines report 1.5× the interquartile range (IQR) and the central square represents the mean value. All speckles from 2–4 independent experiments are combined. Each experiment examined 500–1000 speckles from 30–60 cells (1000–4000 speckles from multiple experiments).

Fig. 3C and E show the radial distributions of molecules within the speckles that were simulated based on the two different interaction matrices shown in Fig. 3B and D, respectively. These two distributions were consistent with the two different types of organization we observed experimentally, as described above (Fig. 1A; Fig. S1A). Fig. 3E represents the mixed case, in which *MALAT1*, *U1* and U2B″ could mix into the core region defined by SON and SC35, whereas Fig. 3C represents the demixed case, in which *MALAT1*, *U1* and U2B″ were excluded from the SON and SC35 core. The change in the architecture of the interaction matrix that was required to discriminate between demixed versus the mixed cases emerged through the promiscuity of the SR modules (Fig. 3B, D, explained below). Considering the relative population of these two distributions, a population average of the simulated distribution (Fig. 3F) correctly reproduced the population average of the distribution observed experimentally (Fig. 1E).

Overall, our computational results are robust to the choices of the actual magnitudes for the effective pairwise affinities in the interaction matrix. The central determinant of the robustness of our results is the structure we prescribe for the interaction matrix. Here, the structure refers to the choice of coarse-grained modules and the choices made for the inter-module interactions, which can be favorable or null. A null interaction implies that the modules prefer to interact with the surrounding solvent and not with one another. In this context, it is worth emphasizing that every module has a finite excluded volume that corresponds to a single lattice site. Specific features of the interaction matrix that are important for producing the experimentally derived radial distributions are: (1) a favorable interaction between SON and SC35 to ensure that these two components make up the core of a multilayered speckle; (2) setting the parameters of the 5′ and 3′ halves of *MALAT1* to have different interaction architectures, which yields robust realization of multilayered speckles in simulations; (3) a favorable heterotypic interaction between the SR modules and *MALAT1*, which is required to observe penetration of *MALAT1* into the core; (4) favorable interactions between *MALAT1* and *U1*/*U2* as well as SR modules,

Journal of Cell Science (2017) 130, 4180-4192 doi:10.1242/jcs.206854

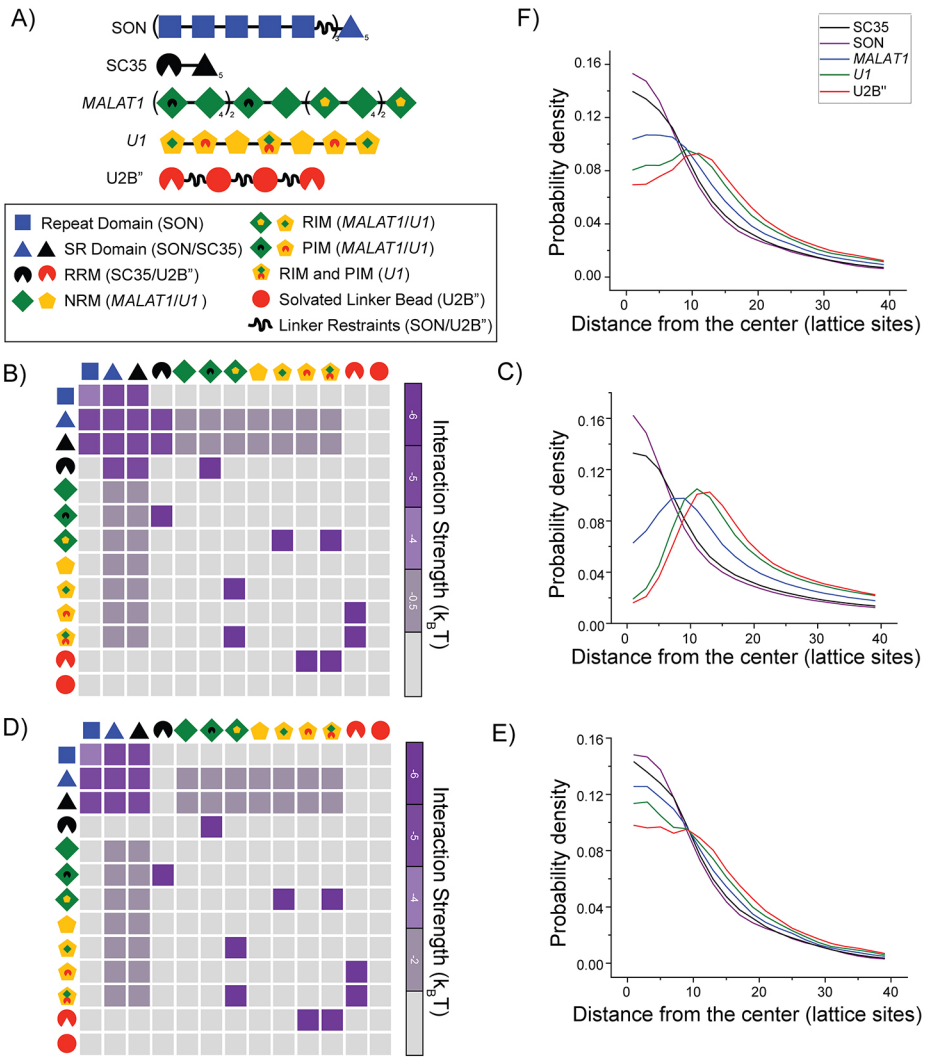


Fig. 3. Multilayer organization is demonstrated in lattice-based computer simulations of a five-component system. (A) Coarse-grained modular architectures of SON, SC35, *MALAT1*, *U1*/*U2* and U2B″ with a key to help identify the different modules within the five molecules. Here, RRM, NRM, RIM, and PIM are acronyms that refer to RNA recognition module, non-specific RNA module, RNA interaction module (on RNA molecules) and protein interaction modules (on RNA molecules). Depending on the macromolecule of interest, connected beads are constrained to either nearest neighbor (straight connectors) or second nearest neighbor (squiggly connectors) lattice sites. (B,D) Interaction matrices are used to model the effective solvent-mediated pairwise interactions between modules. The color bar on the right shows the interaction strength in terms of thermal energy. A gray cell implies that the solvation of the module is preferred over pairwise interaction between the pair of modules. Of special note, the weakest interaction color has different values in these two interaction tables. For the ease of viewing, we have omitted from the energy table and polymer definition for the *MALAT1*–*MALAT1* interaction. Every third *MALAT1* module, starting from the second, can interact with any other *MALAT1* module on such spacing with an interaction strength equal to the *MALAT1*–*U1* interaction. (C,E) Probability density for each of the five components plotted against the distance from the geometric center of the largest cluster. The distance is in units of the number of lattice sites. (F) Combined probability densities from C and E to match the two populations observed experimentally.

which are required for the recruitment of *U1*/*U2* to the shell of the speckle; and finally (5) preferential solvation of the linker modules within U2B″ to target this molecule to the shell of the multilayered speckle. Additionally, the proposed promiscuity of interactions involving SR domains from SC35 and SON offers a mechanism to determine whether mixed or demixed spatial organization is obtained. If the SR domains are less promiscuous and preferentially bind the SC35 RNA recognition motifs (RRMs) they generate a finite interfacial tension between the core and shell of the speckle (Fig. 3B,C), facilitating the demixed architectures. However, if the SR domains are more promiscuous in their interactions, as represented by weaker preferential interactions with SC35 RRMs, but enhanced interactions with RNAs, then the interfacial tension is reduced and the mixed case prevails (Fig. 3D,E).

SR proteins and *MALAT1* influences the localization and organization of proteins and RNAs in nuclear speckles

In order to test the involvement of SR and SR-like proteins in the layered organization of speckle components, we performed SIM analyses in cells that were depleted of SRSF1 or SON proteins (Fig. 4A,B; Fig. S4A). Knockdown of SRSF1 led to the demixed organization between *MALAT1*/*U2* and the SC35 core in almost 100% of the cell population (Fig. 4A and C), with a slight decrease in the speckle size (Fig. 2). In SON-depleted cells, *MALAT1* showed a more homogenous nuclear distribution, as observed previously (Tripathi et al., 2010) (Fig. 4B), and *U2* and SC35 exhibited a strong demixed organization at the periphery (Fig. 4D). Moreover, speckles became spherical in shape with a significant increase in size in SON-depleted cells as compared to control cells (Fig. 2). An earlier study examining SON-depleted cells reported a donut-shaped localization pattern of several speckle components, including the *U1* snRNA-associated protein *U1*-70K (also known as snRNP70), SRSF1, components of exon-junction complex and poly(A)-positive RNA (Sharma et al., 2010). The organization of *U2* observed in our work was consistent with the previous studies of other speckle components, whereas SC35 remained in the speckle core in SON-depleted cells. All of these results strengthen our model that SR domain-mediated interactions could modulate the non-random organization of speckle components.

Several lncRNAs are known to function as scaffold components of RNA–protein assemblies (Mao et al., 2011b; Nakagawa et al., 2012; Souquere et al., 2010). In order to test whether speckle-enriched *MALAT1* influences the organization of other speckle constituents, we imaged SON, SC35 and *U2* in *MALAT1*-depleted cells (Fig. 5A,B; Fig. S4D). Our data, along with that from previous studies (Clemson et al., 2009; Tripathi et al., 2010), showed that *MALAT1* is dispensable for the formation of nuclear speckles even though *MALAT1* can facilitate the recruitment of several of the speckle components to a specific sub-nuclear region (Tripathi et al., 2012). However, *MALAT1*-depleted cells displayed an overall

Journal of Cell Science (2017) 130, 4180-4192 doi:10.1242/jcs.206854

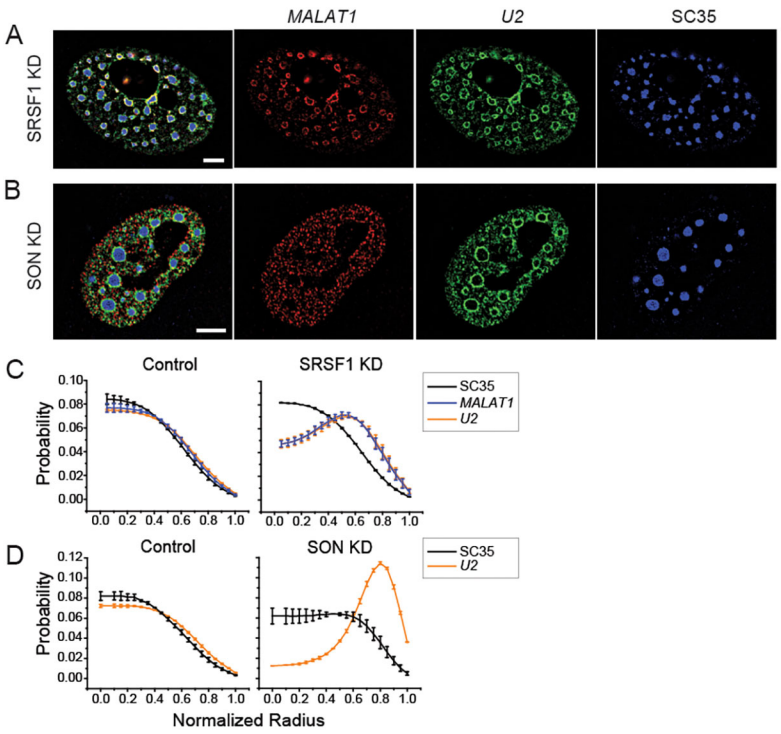


Fig. 4. Effect of SR protein knockdown on speckle organization. Sample image of *MALAT1* (red), *U2* (green) and SC35 (blue) in the WI-38 cells depleted for SRSF1 (A) and SON (B). An example of a corresponding control cell with the same labeling scheme is shown in Fig. S2F. Scale bars: 5 µm. Average radial distribution of *MALAT1*, *U2* and SC35 in WI-38 cells depleted for SRSF1 (C) and SON (B). Error bars represent the standard deviation from two independent measurements. Each measurement contains 500–1000 speckles from 30–60 cells.

decrease in the size of the SC35-stained core area of the speckle (Fig. 2) and a reduced localization of SON to nuclear speckles (Fig. 5A,C). Analysis of SON nuclear localization revealed that the density of SON within the speckle was fivefold higher than in the rest of the nucleus (Fig. 5C). *MALAT1* depletion resulted in an ∼50% reduction in the localization of SON to the speckles.

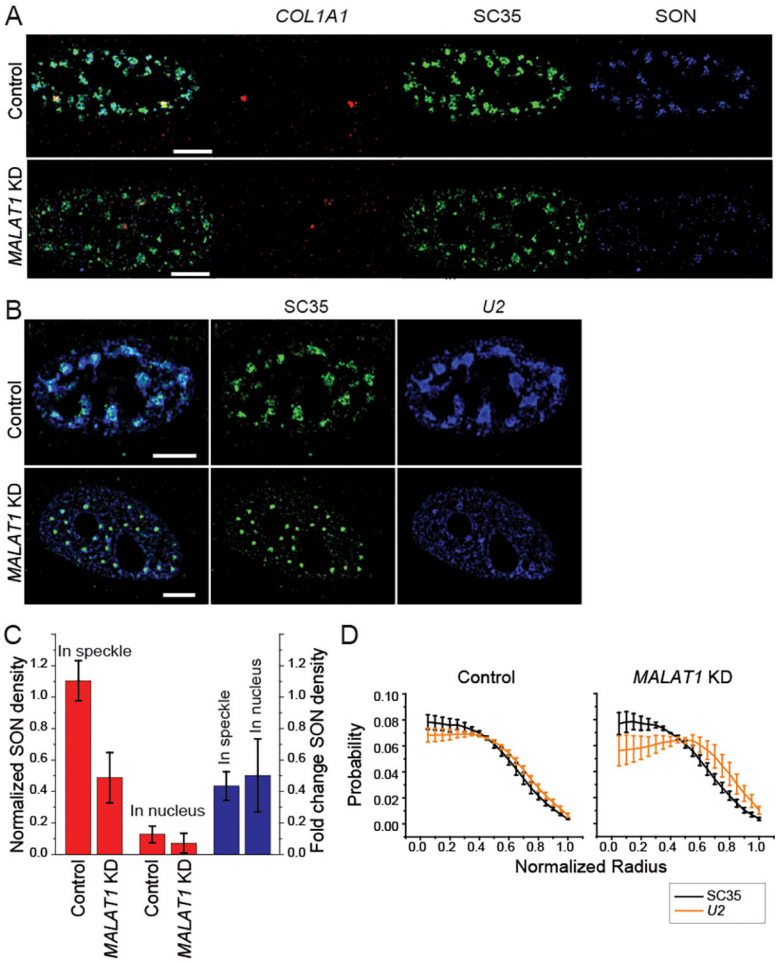


Fig. 5. Effect of *MALAT1* knockdown on speckle. (A) Sample image of *COL1A1* mRNA (red), SC35 (green) and SON (blue) in the control and *MALAT1*-depleted cells. (B) Sample image of SC35 (green) and *U2* (blue) in the control and *MALAT1*-depleted cells. Scale bars: 5 µm. (C) Normalized SON density for the control and *MALAT1*-knockdown (KD) cells. Density is calculated by dividing the total intensity by the volume of either the speckles or the whole nucleus. Error bars report the standard deviation. (D) Average radial distribution of *U2* and SC35 in the control and *MALAT1*-knockdown background. Error bars report the standard deviation. All plots contain data from three or four independent measurements. Each measurement contains 300–1000 speckles from 20–60 cells.

Journal of Cell Science (2017) 130, 4180-4192 doi:10.1242/jcs.206854

However, this reduction in SON density was also observed across the whole nucleus upon *MALAT1* depletion (Fig. 5C), indicating that cellular expression level of SON may be affected by *MALAT1* depletion. As SON is reported to be a nuclear speckle scaffold protein (Sharma et al., 2010), the reduction in the size of nuclear speckles in *MALAT1*-depleted cells may partially be a result of the overall reduction in cellular SON levels.

The layered organization of SC35 and *U2* became slightly more demixed in the absence of *MALAT1* (Fig. 5D). However, we speculate that it is not likely that the more demixed organization of speckle components was due to a direct effect of *MALAT1* depletion because in transcription-inhibited cells, although *MALAT1* was delocalized from the speckles, the relative distribution between SC35 and *U2* remained identical that in to normal cells (Fig. S2G). Given the fact that *MALAT1* knockdown affects the expression and activity of many genes, including that of SR proteins, the minor change in the speckle organization that we observed in *MALAT1*-depleted cells could be due to changes in overall gene expression (Malakar et al., 2017; Tripathi et al., 2010, 2013).

RNA accumulation within the speckles positively regulates the size of nuclear speckles

In order to determine whether speckle-accumulated RNA molecules (i.e. RNA transcripts from speckle-associated genes) preferentially localize to a specific layer within nuclear speckles, we quantified the distribution of RNAs of several speckle-associated genes, including those encoding Collagen, type I, α1 (*COL1A1*), fibronectin (*FN1*), β-actin (*ACTB*) and lamin A/C (*LMNA*). Quantification was achieved by performing RNA-FISH with exon-specific probes along with SC35 staining (Fig. 6A; Fig. S5). These probes hybridize to both spliced and unspliced transcripts. We selected speckles based on the SC35 signal as described above, and calculated the distance (*D*) between the center of each RNA accumulation site (referred to as the RNA site hereafter) to the center of its nearest speckle. If *D* was smaller than the maximum radial distance of that speckle (R_{Max}), we considered the RNA site to be associated with that speckle. Consistent with previous reports (Smith et al., 1999), *COL1A1* and *ACTB* RNAs showed a higher percentage of speckle association than *FN1* and *LMNA* RNAs (Fig. S5D). Among the four genes tested, only the *COL1A1* RNA site was within the interior

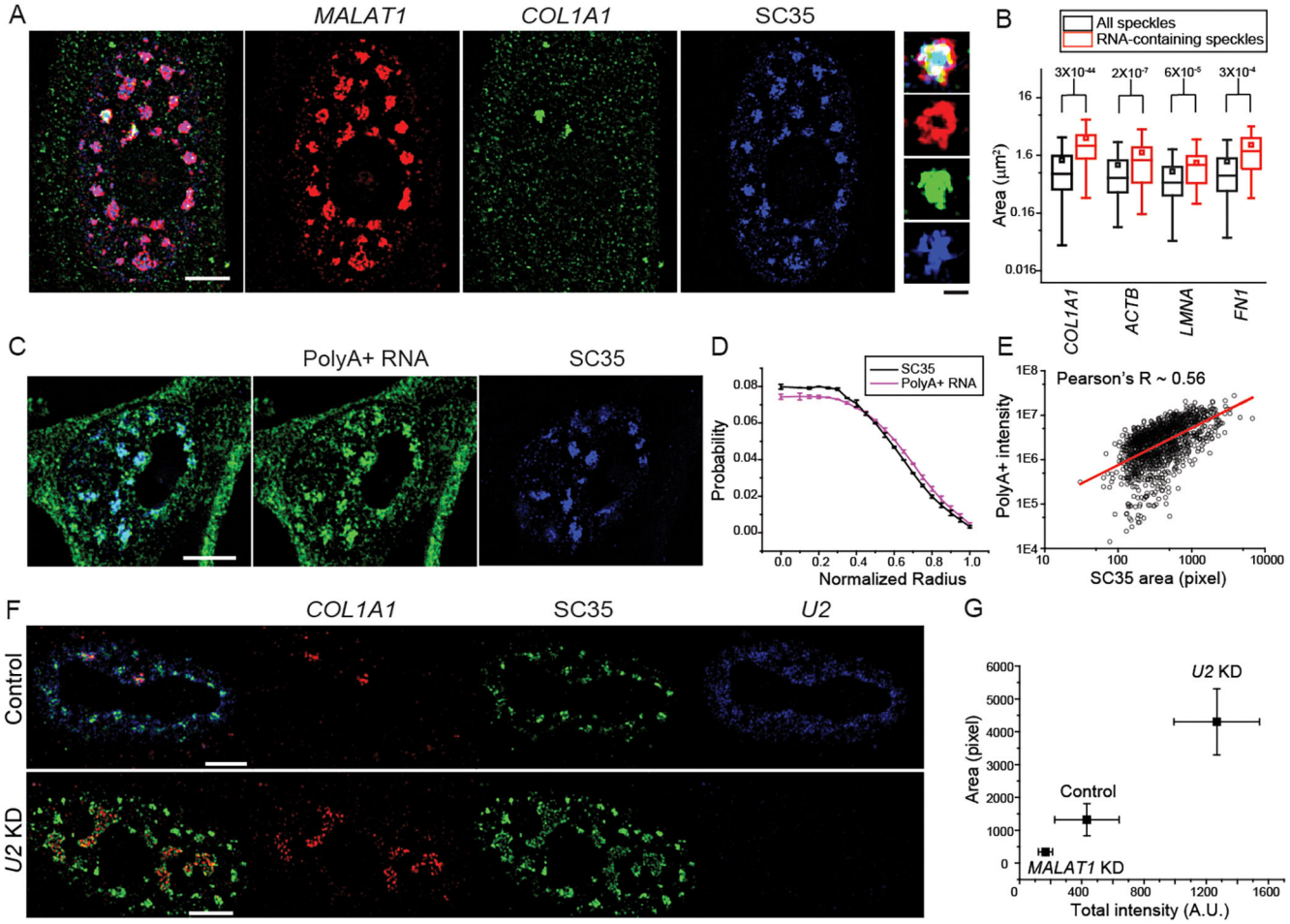


Fig. 6. RNA accumulation contributes to speckle size variation. (A) Sample image of *MALAT1* (red), *COL1A1* RNA (green) and SC35 (blue). (B) Box-and-whisker plot of the areas of RNA accumulation site-associated speckles and all speckles from the same cells. *P*-values (calculated with a Kolmogorov–Smirnov test) are indicated above the plots. All box-and-whisker plots are presented as described in Fig. 2. (C) Sample image of poly(A)-positive (PolyA+) RNA (green) and SC35 (blue). (D) Average radial distribution of poly(A)-positive RNA and SC35. Error bars report standard deviation. (E) Scatter plot of total intensity of poly(A)-positive RNA versus speckle size defined by SC35. The red line designates the linear fitting of the data. (F) Sample image of *COL1A1* mRNA (red), SC35 (green) and *U2* snRNA (blue) in the control and *U2*-knockdown (KD) cells. (G) RNA-associated speckle size versus total RNA intensity in the control, *MALAT1*-knockdown and *U2*-knockdown cells. Error bars report the standard deviation. Scale bars: 5 µm, cell images; 1 µm, magnified speckle images. All plots contain data from 2–4 independent measurements. Each measurement contains data from 300–1000 speckles and 15–70 RNA-containing speckles from 20–60 cells.

Journal of Cell Science (2017) 130, 4180-4192 doi:10.1242/jcs.206854

region of the speckle, with an average relative distance to the center of the associated speckle (defined as $D_{Rel}=D/R_{Max}$) of ∼0.2, whereas the other three RNA sites had an D_{Rel} of ∼0.6, residing within the RNA-rich region (Fig. S5E). As reported previously (Hall et al., 2006), *COL1A1* RNA accumulated within the interior region of the speckle while the *COL1A1* gene remained at the periphery, suggesting that RNA is imported into the speckle interior after being transcribed at the periphery (Fig. S5F). For the four genes tested, the speckles containing RNA sites were larger than the average speckle in the same nucleus (Xing et al., 1995), but to different degrees (Fig. 6B). The increase in speckle size did not correlate with the total cellular levels of a particular RNA determined by using real-time quantitative RT-PCR (RT-qPCR) (Fig. S4B).

In order to test the correlation between the speckle size and the amount of poly(A)-positive RNA that was present within a speckle, we co-stained poly(A)-positive RNA together with SC35 (Fig. 6C). Consistent with the previous results (Carter et al., 1993; Hall et al., 2006), poly(A)-positive RNA defined a broader region than the SC35-decorated inner core of the speckle (Fig. 6D). When plotting the signal from poly(A)-positive RNA as a function of the SC35-defined area, we observed a strong positive correlation between the size of the speckle and the amount of poly(A)-positive RNA present in the speckle (correlation coefficient ∼0.56) (Fig. 6E), again supporting the idea that accumulation of RNA inside the speckle positively correlates with speckle size.

Finally, to determine whether RNA accumulation within a speckle contributes to changes in speckle size, we examined the localization of *COL1A1* RNA in splicing-defective cells and *MALAT1*-depleted cells. Depletion of *U2* snRNA reduced pre-mRNA splicing, as revealed by RT-qPCR analysis performed with intron–exon junction-specific primers against *COL1A1* pre-mRNA (Fig. S4C), and resulted in the accumulation of unspliced transcripts. We labeled *U2* along with *COL1A1* RNA and SC35 in both control and *U2*-depleted cells (Fig. 6F). Because *U2* depletion efficiency showed a large cell-to-cell variation, we selected cells with a knockdown efficiency that was greater than 75% for further analysis (Fig. S6A). Compared to control cells, the speckle size increased in *U2* snRNA-depleted cells (Fig. 2). In addition, *U2*-depleted cells showed an increase in the number of *COL1A1* RNA sites per cell (∼3 RNA sites per cell, compared to ∼1.8 sites per cell in the mock case) along with a concomitant increase in the amount of *COL1A1* RNA accumulated within each speckle (Fig. 6G). Such a dramatic increase in the size of *COL1A1*-containing speckles required a near complete depletion of *U2* snRNA (Fig. 6F) because cells with a partial depletion of *U2* displayed neither an aberrant accumulation of *COL1A1* RNA nor a dramatic change in speckle size (Fig. S6B). In line with our observations under *U2* knockdown, cells depleted for SF3b1 (also known as Sap155; a core component of *U2* snRNP) showed a very similar size expansion for nuclear speckles (Fig. 2; Fig. S6C). Moreover, both in the control and *U2* or SF3b1-depleted cells, EXOSC10 (also known as RRP6 or PM/Scl-100), an essential component of the nuclear exosome, was predominantly localized in the nucleolus and was excluded from nuclear speckles (Fig. S7), suggesting that speckles do not act as sites for the degradation of aberrantly spliced RNAs. *MALAT1*-depleted cells, on the other hand, showed a decrease in (1) the number of *COL1A1* RNA sites per cell (∼1 per cell, compared to ∼1.6 sites per cell in the mock case), (2) the total amount of *COL1A1* RNA at the accumulation site (Fig. 6G), and (3) the cellular levels of un-spliced and spliced *COL1A1* transcripts (Tripathi et al., 2013) (Fig. S4D). All of these results support the hypothesis that the accumulation of RNA contributes to the growth of nuclear speckles.

DISCUSSION

Several non-membranous organelles are known to be internally organized into distinct sub-compartments. For example, nucleoli are organized into at least three distinct subdomains that are responsible for specific stages of ribosome biogenesis and maturation (Boisvert et al., 2007; Pederson, 2011). Recent studies have also revealed non-random organization of components within the paraspeckle sub-nuclear domains (West et al., 2016). The layered structure, characterized by a core-shell architecture, has also been found in cytoplasmic ribonucleoprotein (RNP) bodies (Jain et al., 2016; Wheeler et al., 2016). Our results demonstrate that the layered organization also exists in nuclear speckles, in which SON and SC35 assemble into the core region, *U1*, *U2* and *MALAT* forms a second layer, and U2B″ loosely associates with the peripheral regions of nuclear speckles. Previous studies have revealed that regions of chromatin that contain highly transcribed genes tend to associate with the periphery of nuclear speckles, leading to a model in which pre-mRNA splicing of the speckle-associated genes occurs at the speckle periphery (Brown et al., 2008; Hall et al., 2006; Han et al., 2011; Hu et al., 2010; Khanna et al., 2014; Yang et al., 2011). Our findings that snRNAs and proteins in snRNPs tend to concentrate at the periphery of nuclear speckles further support this model.

Computational simulations suggest that the spatial organization of nuclear speckles observed *in vivo* might emerge from the interplay of favorable sequence-encoded intermolecular interactions and the preferential solvation among the various protein and RNA modules. The interaction matrix used to recapitulate this organization is in accordance with a variety of observations that have been reported in the literature. SC35 and SON are members of SR and SR-like protein families respectively; both proteins possess SR domains that are enriched in arginine and serine repeats and in addition possess an RRM (Sharma et al., 2010; Shepard and Hertel, 2009; Zhong et al., 2009). In addition, SON has distinctive repetitive sequence elements (Fig. S3) that are critical for speckle formation (Sharma et al., 2010). Therefore, the multivalence of SON and SR repeats appear to provide a crucial driving force for the formation of the speckle core region. The multivalence of *MALAT1* enables its interactions with core and shell components through interactions with SR proteins (Änkö et al., 2012; Miyagawa et al., 2012; Sanford et al., 2009; Tripathi et al., 2010) and base pairing with RNAs (Engreitz et al., 2014; Lu et al., 2016; Nguyen et al., 2016), thus promoting the observed multilayer structure. The SR domains are known to undergo extensive and reversible phosphorylation (Lin et al., 2005; Misteli and Spector, 1997; Misteli et al., 1998; Tripathi et al., 2010; Xie et al., 2006), providing a putative mechanism through which the cell could dynamically modulate the 'promiscuity' of the SR domain. Finally, *MALAT1* and snRNAs undergo different levels of modifications (Bringmann and Lührmann, 1987; Liu and Pan, 2016; Liu et al., 2015; Roundtree and He, 2016), which could lead to changes in RNA–protein interactions and therefore contribute to the heterogeneity of the compositional organizations.

Our knockdown experiments have provided hints on the involvement of different speckle constituents in the formation and maintenance of nuclear speckles. SON and SC35 are important for forming the core region of nuclear speckles, which is supported by two lines of evidence: (1) that SON and SC35 assemble into the speckle at early G1 phase, whereas *MALAT1* is dispersed in the nucleoplasm; (2) and that SC35 was localized in the core region of nuclear speckles under all tested perturbation conditions, including upon speckle-resident protein and RNA knockdown, and transcription inhibition. Based on our results, we speculate that SR proteins play an important role in maintaining the ratio of mixed

Journal of Cell Science (2017) 130, 4180-4192 doi:10.1242/jcs.206854

(in which *MALAT*1 and *U2* mix into the core region of the nuclear speckle), and demixed (in which *MALAT*1 and *U2* are excluded from the core region of the nuclear speckles) populations of nuclear speckles. Knockdown of SRSF1 or SON protein changes the populations to being almost 100% demixed. These data support our predictions from computational simulations in that interactions between SR domains, RRMs and RNAs tune the balance between the mixed and demixed modes of organization. SON depletion has an even larger effect on speckle morphology, including inducing an increase in speckle size and a more spherical shape compared to the normally irregular nature of nuclear speckles in control cells. We interpret this observation to indicate that the multivalence of SON may promote changes in the material properties of nuclear speckles, such as the underlying dynamics of the speckles (e.g. liquid-like versus solid-like) (Banani et al., 2016, 2017), an effect which remains to be further investigated. The effect of *MALAT1* depletion on nuclear speckle structure is likely to be an indirect effect, given *MALAT1* is one of the RNA molecules that is most frequently observed to globally interact with nascent transcripts (Engreitz et al., 2014; Lu et al., 2016; Nguyen et al., 2016) and chromosomes (Li et al., 2017). Indeed, we observe that depletion of *MALAT1* affects the cellular level of SON (at the protein level), and *COL1A1* (at the RNA level). We therefore speculate that the changes in the overall structure of nuclear speckles caused by *MALAT1* depletion, including the overall size reduction and the shift towards the more demixed state, are more likely due to the changes in the abundance of nuclear speckle components rather than being a direct effect from *MALAT1*.

Previous studies have demonstrated the localization of the *Hsp70* (*HSPA1A*) gene to the proximity of nuclear speckle upon its activation, as well as transcription-dependent size expansion of *HSPA1A*-associated speckle size upon heat shock (Hu et al., 2010; Khanna et al., 2014). Consistent with these results, our measurements on multiple genes reveal that the transcription of speckle-associated genes and the accumulation of RNA in speckles contribute to the size heterogeneity of nuclear speckles. We find that, in general, speckles size strongly correlate with the total poly (A)-positive RNA signal within the speckles at individual speckle level. As specific examples, speckles containing accumulated transcripts from highly transcribed genes, including *COL1A1*, *ACTB*, *FN1* and *LMNA* are larger than the average speckles. Previous studies have documented that most nuclear speckles are associated with genes and contain RNA transcripts from multiple genes (Hall et al., 2006; Shopland et al., 2002). Our observations here are in agreement with these previous findings, and further reveal a weak positive correlation between the size of speckles containing transcripts from these specific genes and the total signals from the specific transcripts at single speckle level (correlation coefficient ~0.27) (Fig. 7A). We reason that because speckles contain transcripts from multiple genes, the correlation between the speckle size and the amount of RNA accumulated from a specific set of genes is weakened compared to the strong correlation between the speckle size and the amount of accumulated total poly(A)-positive RNA. Nevertheless, the observation that speckles containing transcripts from any one of the four genes show an increase in average size (albeit to different extents) is likely because they are among the most active genes in fibroblast cells. We sorted the RPKM values from RNA-sequencing for all of the annotated genes in fibroblast cells (Table S1) (Marthandan et al., 2016), and found that *COL1A1* is among the top 0.2% abundant, *ACTB* among the top 0.05%, *FN1* among the top 0.2% and *LMNA* among the top 0.8%. However, it should still be noted that the propensity of speckle accumulation is not strictly correlated with the cellular RNA abundance, as shown in the case of *COL1A1* and *ACTB* from imaging and RT-qPCR results. What drives the RNA transcripts to accumulate in nuclear speckles needs to be further investigated. Post-transcriptional processing events, such as pre-mRNA splicing and RNA export, could play a role in mediating the association of RNA with speckles.

Interestingly, when considering all four tested genes at the single speckle level, we also observe a weak negative correlation (correlation coefficient~–0.34) between the speckle size and D_{Rel} (location of the RNA site within the speckle as defined in the Results) (Fig. 7B) and a weak negative correlation (correlation coefficient~–0.32) between the amount of accumulated RNAs and D_{Rel} (Fig. 7C). Including the positive correlation between the speckle size and the amount of accumulated RNAs described above, these correlations suggest that more accumulated RNAs tend to localize toward the center of the associated speckle, resulting in an increase in the speckle size, leading to our proposal of a dynamic interaction between the RNA transcript and the nuclear speckles (Fig. 7D). Once a transcriptionally active gene is recruited to the nuclear speckle periphery, nascent pre-mRNA transcribed from this gene accumulates near speckles. The co-residence of nascent pre-mRNA with snRNPs at the periphery of the nuclear speckle could enhance the effective processing of pre-mRNAs. When a significant amount of pre-mRNA or partially processed pre-mRNA accumulates near a particular speckle, it facilitates the recruitment of more speckle components to that speckle, either from the nucleoplasm or from the neighboring nuclear speckles, leading to the further growth of the speckle (Eils et al., 2000; Hu et al., 2009; Misteli et al., 1997). This dynamic reallocation of nuclear resources is reminiscent of how transcription can modulate nucleoli assembly

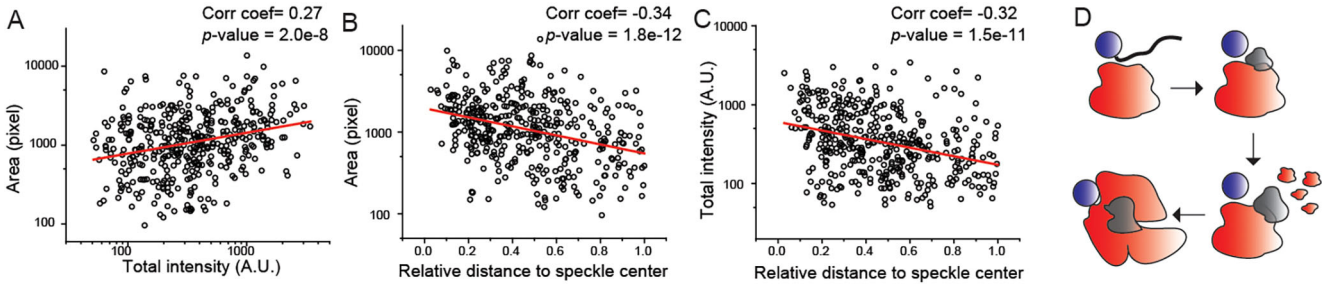


Fig. 7. Correlation between speckle size and total RNA accumulation at single speckle level. (A) Scatter plot of RNA-associated speckle size versus total RNA intensity at the accumulation site for individual speckles. (B) Scatter plot of RNA-associated speckle size versus D_{Rel}. (C) Scatter plot of total RNA intensity versus D_{Rel}. The red lines show the linear fitting of the data. All plots contain data from three or four independent measurements. Each measurement contains 15–70 RNA-containing speckles from 20–50 cells. (D) Model for describing the effect of mRNA accumulation on the speckle size increase: gene (purple); RNA transcript (gray) and speckle (red).

Journal of Cell Science (2017) 130, 4180-4192 doi:10.1242/jcs.206854

(Berry et al., 2015). Since recruitment of new components to a speckle can occur from all directions, the RNA accumulation site shows a relative migration towards the center of the speckle. This is exemplified in the case of *U2* snRNA-depleted cells, where a large amount of *COL1A1* transcripts are enriched at the interior region of the associated speckles (Fig. 6F).

In conclusion, by using super-resolution imaging of multiple nuclear speckles, we characterized in detail a multilayered distribution of several of the nuclear speckle components. This multilayered spatial organization of nuclear speckles might emerge from the interplay of favorable sequence-encoded intermolecular interactions between various speckle components. The morphology and organization of speckles can be directly or indirectly tuned by each of the components. Finally, RNA accumulation in the speckles positively regulates the size of the speckle at the single speckle level.

MATERIALS AND METHODS

Cell culture

WI-38 cells were cultured in minimal essential medium (MEM) supplemented with 1 mM sodium pyruvate, 100 U/ml penicillin, 100 µg/ml streptomycin, 10% fetal bovine serum (FBS) and 1% non-essential amino acid (NEA). HeLa cells were cultured in Dulbecco's modified Eagle's medium (DMEM) containing high glucose, supplemented with 1 mM sodium pyruvate, 100 U/ml penicillin, 100 µg/ml streptomycin and 10% FBS. For imaging, cells were immobilized on poly-L-lysine-treated borosilicate chambered coverglass (Thermo Scientific Nunc Lab-Tek). These cells were obtained from the ATCC and maintained as per their recommendations.

Antisense DNA oligonucleotide- and shRNA- and siRNA-mediated knockdown

MALAT1 and *U2* snRNA were depleted by use of modified DNA (antisense DNA oligonucleotides; ASOs) (Ionis Pharmaceuticals, USA). The ASO is a phosphorothioate modified 5-10-5 MOE gapmer with five 2′-O-methoxy-ethyl nucleotides on either end and ten deoxy-nucleotides in the gap. Oligonucleotides were transfected into cells in two rounds with a gap of 24 h at a final concentration of 100–200 nM. Lipofectamine RNAiMax reagent was used for transfection (Invitrogen, USA). SRSF1 knockdown was performed by stably transducing WI-38 cells with lentiviral-based shRNAs. shRNA was cloned into plko.1 vector using protocols from Addgene (plasmid #10878). A control shRNA clone was purchased from Addgene (plasmid #10879).

SRSF1 shRNA sequences were: SRSF1-sh-F, 5′-CCGGAAGGATTG-TGGAGCACATTTCCTCGAGGAAATGTGCTCCACAATCCTTTTTT-TG-3′ and SRSF1-sh-R, 5′-AATTCAAAAAAAGGATTGTGGAGCA-CATTTCCTCGAGGAAATGTGCTCCACAATCCTT-3′.

Depletion of SON and SF3B1 were performed by using double-stranded siRNAs against SON (Sharma et al., 2010) and SF3B1 (hs.Ri.SF3B1.13.1, IDT, USA). siRNAs were transfected into cells two times with a 24 h interval, at a final concentration of 100 nM for SON and 20 nM for SF3B1 using Lipofectamine RNAi MAX (Invitrogen). Cells were collected 48 h after the second transfection for further analyses.

Cell cycle synchronization

HeLa cell synchronization was carried out by double thymidine block and release. Briefly, cells were cultured in 2 mM thymidine for 24 h as the first block. Then, cells were washed and released for 12 h in fresh medium. Next, the second block was performed by adding 2 mM thymidine for another 24 h to synchronize cells at G1/S. Then, fresh medium was replaced to permit release into the cell cycle; cells were collected at 0 h (G1/S), 4 h (S phase) and 8 h (G2 phase) after release.

Quantitative real-time PCR

Total RNA was extracted using Trizol reagent (Invitrogen) and cDNA was synthesized using a reverse transcriptase kit (Applied Biosystem). Quantitative RT-qPCR was performed with the StepOne plus system to quantify transcript level. Specific primer sets were designed for each gene as follows: MALAT1-F, 5′-GACGGAGGTTGAGATGAAGC-3′ and MALAT1-R, 5′-ATTCGGGGCTCTGTAGTCCT-3′; COL1A1-F, 5′-GTG-CGATGACGTGATCTGTGA-3′ and COL1A1-R, 5′-CGGTGGTTTCTT-GGTCGGT-3′; COL1A1-Intron24-Exon25-F, 5′-TGACCGGCGGGACCC-TAA-3′ and COL1A1-Intron24-Exon25-R, 5′-CATTGTCCTGTCTGCCT-CCCTG-3′.

Western blotting

Immunoblotting was performed by following the protocol described previously (Tripathi et al., 2010). Antibodies used for western blotting (wB) include antibodies against SRSF1 (mIgG, mAb96, WB, 1:1000) (Hanamura et al., 1998), SF3B1/SAP155 (Santa Cruz Biotechnology, sc-514655, 1:200) and α-tubulin (1:10000, Sigma-Aldrich, T5168).

Labeling of FISH probes and secondary antibodies

Chemically synthesized single-molecule (sm)FISH probes were designed by using the Stellaris Probe Designer and ordered from Biosearch Technologies (http://www.biosearchtech.com). Probes were 20 nucleotides in length with GC content of ∼45% and 3′ amine modification. Each probe was synthesized and dissolved individually in water to a final concentration of 100 µM in a 96-well plate. An equal volume of each probe were combined for labeling in 0.1 M sodium bicarbonate (pH 8.5). Alexa Fluor-conjugated succinimidyl ester (Life Technologies) was first dissolved in 1–5 µl DMSO and then mixed with the probe solution. The molar ratio of dye to probe is ∼25:1. The labeling reaction was incubated in the dark at 37°C overnight. Reactions were quenched by adding a 1/9th reaction volume of 3 M sodium acetate (pH 5). Labeled probes were purified from unconjugated free dye, first by ethanol precipitation and then further purified with a P-6 Micro Bio-Spin Column (Bio-Rad). All probes were labeled with more than 80% labeling efficiency. The number of probes used for each RNA is as follows: *MALAT1*, 32 probes; *U2*, 8 probes; *U1*, 6 probes; *COL1A1*, 40 probes; *ACTB*, 36 probes; *LMNA*, 31 probes; *FN1*, 39 probes.

Secondary antibodies against mouse and rabbit IgG (Jackson ImmunoResearch) were labeled with Alexa Fluor succinimidyl ester for immunofluorescence staining; 24 µl of antibody (at 1 mg/ml concentration) was mixed with 3 µl of 10× PBS and 3 µl of 1 M sodium bicarbonate (pH 8.5), and added with 0.001–0.005 mg of Alexa dye. The reaction was incubated at room temperature for 30 min, and labeled antibody was purified on a P-6 Micro Bio-Spin column pre-equilibrated with 1× PBS. The antibody to dye ratio was between 0.5 and 2.5 from batch to batch.

FISH and immunostaining

FISH and immunofluorescence staining were performed as previously described (Raj et al., 2008). Briefly, cells were fixed with 4% paraformaldehyde (Electron Microscopy Sciences) in 1× PBS for 10 min at room temperature and permeabilized with solution containing 0.5% Triton X-100 and 2 mM vanadyl ribonucleoside complexes (VRC New England Labs) in 1× PBS for 10 min on ice. Cells were then temporarily stored in 70% ethanol until hybridization with FISH probes. Cells were washed once with FISH wash solution [10% formamide (Ambion) in 2× saline-sodium citrate (SSC)]. Labeled probes were mixed with 50 µl hybridization buffer [10% dextran sulfate (Sigma-Aldrich), 1 mg/ml yeast tRNA (Sigma-Aldrich), 0.2 mg/ml BSA (Ambion), 2 mM VRC, 10% formamide in 2× SSC]. Labeled probes for all transcripts and *MALAT1* were 4 nM each and for *U1* and *U2* were 10 nM each in the final reaction. Probe solution was then added to the chamber. An 18×18 mm coverslip was used to cover the bottom of the chamber such that the probe solution was evenly distributed to the entire chamber. Hybridization reactions were incubated in the dark at 37°C overnight. On the second day, the cells were washed twice with FISH wash solution.

FISH-labeled cells were then subjected to fixation again at room temperature for 5 min to prevent probe dissociation during immunostaining. Before immunostaining, cells were first incubated with blocking solution [1% BSA (Gibco) in 1× PBS] at room temperature three times with a 10-min incubation each time. Primary antibody was mixed with blocking solution using the following dilution factors: mouse antibody against SC35 (Fu and Maniatis, 1990), 1:500, rabbit antibody against SON (Sharma et al., 2010), 1:500; mouse antibody against U2B″ (clone: 4G3), 1:25 (Habets et al., 1987,

Journal of Cell Science (2017) 130, 4180-4192 doi:10.1242/jcs.206854

1989; Mattaj et al., 1986), mouse anti-EXOSC10 antibody (sc-374595, Santa Cruz Biotechnology), 1:50. 50 µl antibody solution was used for each chamber, and covered with an 18×18 mm glass coverslip during the 1-h incubation at room temperature. Cells were washed with blocking solution three times with a 10-min incubation each time. Labeled secondary antibody was diluted with the blocking solution by 100-fold. Incubation with secondary antibody is performed in the same manner as for primary antibody for ∼40 min. After staining, the cells were washed three times with 1×PBS with a 10-min incubation each time. Cells were stored in 4× SSC at 4°C until imaging.

SIM and imaging reconstruction

Image acquisition was performed on a commercial instrument (SR-SIM Elyra system with Axioobserver Z1 microscope from Zeiss); 488 nm, 561 nm and 632 nm lasers were used for three-color imaging. The exposure time was 100 ms for all three channels. The excitation power was adjusted, but was no larger than 30% of the maximum power, so that signals can be maximized without saturating the camera. For knockdown experiments, the exact same laser power was used for knockdown samples as for corresponding negative controls. SIM images were reconstructed by using commercial software (ZEN 2011 from Zeiss). Fluorescent nanodiamonds (140 nm diameter, Acedemia Sinica, Taipei, Taiwan) were nonspecifically attached to the surface of the chamber and were used for mapping of the different channels.

Data analysis

Custom MATLAB code was developed to automate nuclear speckle selection, processing and analysis (available from the corresponding author upon request). First, grayscale images were constructed from the mean intensity of selected channels. Specifically, for distribution analysis for all components, all three channels were used, and for RNA accumulation site analysis, channels corresponding to RNA and to SC35 were used separately to identify each RNA accumulation site and speckle. An overall fluorescence intensity histogram was produced to assist the determination of a threshold for speckle identification. Nuclear speckles in 2D were identified by a single intensity threshold for each slice in the *z*-stack. The 2D binary images after thresholding were further processed by filling and opening binary operations to remove internal voids and shot noise. After combining all binary images in the *z*-stacks into one 3D binary matrix, individual 3D nuclear speckles were identified and indexed by connected component analysis. Each identified object serves as the 3D mask for subsequent quantitative structural analysis of each nuclear speckle. Each speckle was individually saved with all three fluorescence channels. The center of each speckle was determined from the geometric centroid of its mask, this removes the ambiguity of the interpretation of grayscale intensity and potential biasing effect. The normalized radial distribution function of the intensity of each fluorescent channel is calculated based on this center. The normalized radial distribution carries the most meaningful result when the speckle is shaped closest to a sphere.

Lattice-based computer simulations

The design and implementation of the simulations were identical to that of Feric et al. (Feric et al., 2016). The architectures of coarse-grained protein and RNA molecules were designed to mimic the modular nature of these molecules. SC35 is a 221-residue protein. It has an SR-rich domain (92 residues, 80% S/R) and an RNA-binding domain. The RNA domain has specificity for sites on *MALAT1*.

We modeled SC35 as a linear polymer of seven beads, six of which spanned the SR region and one of which was an RNA-recognition module (RRM) that interacts specifically with sites on *MALAT1*. SON is a 2426-residue protein. Its sequence encompasses charge tracts, nearly perfect repeat regions and an SR-rich region. The mapping of the sequence features to a 20-bead linear polymer was derived from analysis of the sequence features of different regions as described in Fig. S3. This collection of large repetitive blocks of sequences are observed to collapse in simulations, which motivated the architecture of three blocks of five beads that are attractive toward similar beads. The final five beads model the SR-rich domain.

U2B″ is a 225-residue protein comprising two RRMs and a linker enriched in charged residues. Accordingly, this molecule was modeled by using four beads, two corresponding to the RRMs interspersed by two solvated beads corresponding to the linker residues. *U1* and *U2* are 100-nucleotide RNA molecules. These molecules are indistinguishable in the coarse-grained simulations. They were modeled as linear polymers, with seven beads that were either RNA interaction modules (RIMs) or protein interaction modules (PIMs). Finally, *MALAT1*, which is a 6000–7000-nucleotide RNA is known to interact with the SR family of proteins, including SRSF1. Most of the SRSF1 interactions sites are located on the 5′ region of *MALAT1*. Additionally, sites on *MALAT1* also form base pairs with U1/U2, which are located on the 3′ end. Accordingly, *MALAT1* was modeled as a linear di-block copolymer comprising 25 beads and interactions of these beads with other beads that corresponded to protein and RNA interaction modules were as described in Fig. 2. *MALAT1* also encompasses beads that are non-specific RNA modules (NRMs) that interact weakly with SR-rich modules on SON and SC35.

There were 1600 linear polymers in each of the lattice-based Monte Carlo simulations. Of these, 200 corresponded to SON, 400 to SC35, 200 to *MALAT1*, 400 to *U1*/*U2*, and 400 to U2B″. The simulations were initiated with all the molecules dispersed uniformly across the lattice. The system was allowed to evolve via a collection of Monte Carlo moves that were accepted or rejected based on the Metropolis criterion. A bond can form between pairs of modules that occupy adjacent lattice sites if the interaction energy for the pair of modules is negative. Each module can make only one bond at a time.

The probability densities were calculated using an approach that mimics the calculation of densities from experimental data. We accomplished this by identifying the geometric center of the largest cluster in the simulation and analyzing the modules within eight lattice units along the *z*-dimension. We calculated the radial density of modules from the geometric center of the speckle in the central 2D plane through the center of the speckle with a finite thickness. As all three dimensions were equivalent in simulation, the results were an average of repeating this over each of the three dimensions.

Acknowledgements

We thank the core facilities and service at Carl R. Woese Institute for Genomic Biology at University of Illinois Urbana-Champaign for SIM microscopy and imaging reconstruction software, Drs P. Bubulya (Wright State University, USA), Fu X-D (University of San Diego, USA), W. van Venrooij (Radboud University Nijmegen, Netherlands) for the gift of reagents, and Dr. V. Tripathi for technical assistance.

Competing interests

S.M.F. is an employee of Ionis Pharmaceuticals and receives salary from the company.

Author contributions

Conceptualization: J.F., Q.H., R.V.P., K.V.P., T.S.H.; Methodology: J.F., M.J., T.S.H., I.T.S.L., B.H., A.S.H., M.R., Q.S.; Software: J.F., T.S.H., I.T.S.L., B.H., A.S.H., M.R., R.V.P.; Validation: J.F., M.J., T.S.H., I.T.S.L., B.H., A.S.H., M.R., Q.S.; Formal analysis: J.F., M.J., A.S.H., R.V.P., K.V.P.; Investigation: J.F., M.J., I.T.S.L., Q.H., R.V.P., K.V.P., T.H.; Resources: J.F., I.T.S.L., S.M.F., R.V.P.; Data curation: J.F., M.J., I.T.S.L., B.H., Q.H., R.V.P.; Writing - original draft: J.F., S.M.F., R.V.P., K.V.P.; Writing - review & editing: M.J., T.S.H., Q.H., A.S.H., M.R.; Visualization: J.F., M.J., B.H., Q.H., Q.S.; Supervision: J.F., R.V.P., K.V.P., T.H.; Project administration: J.F., K.V.P.; Funding acquisition: K.V.P., T.H.

Funding

J.F. acknowledges support from the National Science Foundation (NSF) (PHY-1430124) and the Searle Scholars Program. I.T.S.L. acknowledges supports from the National Science Foundation (PHY-1430124) and Natural Sciences and Engineering Research Council of Canada (NERSC) (RGPIN201704407). T.H. is a Howard Hughes Medical Institute investigator. K.V.P. acknowledges support from the National Institutes of Health (1RO1GM088252), American Cancer Society and a National Science Foundation EAGER grant. R.V.P. acknowledges support from the National Science Foundation (MCB1614766) and National Institutes of Health (R01NS056114). T.S.H. was a graduate student scholar in the Center for Biological Systems Engineering at the University of Washington and A.S.H. was supported in part by the Bonnie and Kent Lattig scholarship. M.J. is supported by the Cancer Scholars for Translational and Applied Research (C*STAR) graduate educational program from the University of Illinois and Carle Foundation Hospital. Deposited in PMC for release after 12 months.

Supplementary information

Supplementary information available online at
http://jcs.biologists.org/lookup/doi/10.1242/jcs.206854.supplemental

References

Ando, T., Yamayoshi, S., Tomita, Y., Watanabe, S., Watanabe, T. and Kawaoka, Y. (2016). The host protein CLUH participates in the subnuclear transport of influenza virus ribonucleoprotein complexes. *Nature Microbiology* **1**, 16062.

Änkö, M.-L., Müller-McNicoll, M., Brandl, H., Curk, T., Gorup, C., Henry, I., Ule, J. and Neugebauer, K. M. (2012). The RNA-binding landscapes of two SR proteins reveal unique functions and binding to diverse RNA classes. *Genome Biol.* **13**, R17.

Banani, S. F., Rice, A. M., Peeples, W. B., Lin, Y., Jain, S., Parker, R. and Rosen, M. K. (2016). Compositional control of phase-separated cellular bodies. *Cell* **166**, 651-663.

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.

Bernard, D., Prasanth, K.V., Tripathi, V., Colasse, S., Nakamura, T., Xuan, Z., Zhang, M.Q., Sedel, F., Jourdren, L., Coulpier, F., et al. (2010). A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. *EMBO J* **29**, 3082-3093.

Berry, J., Weber, S. C., Vaidya, N., Haataja, M. and Brangwynne, C. P. (2015). RNA transcription modulates phase transition-driven nuclear body assembly. *Proc. Natl. Acad. Sci. USA* **112**, E5237-E5245.

Boisvert, F.-M., van Koningsbruggen, S., Navascués, J. and Lamond, A. I. (2007). The multifunctional nucleolus. *Nat. Rev. Mol. Cell Biol.* **8**, 574-585.

Bringmann, P. and Lührmann, R. (1987). Antibodies specific for N6-methyladenosine react with intact snRNPs U2 and U4/U6. *FEBS Lett.* **213**, 309-315.

Brown, J. M., Green, J., das Neves, R. P., Wallace, H. A. C., Smith, A. J. H., Hughes, J., Gray, N., Taylor, S., Wood, W. G., Higgs, D. R. et al. (2008). Association between active genes occurs at nuclear speckles and is modulated by chromatin environment. *J. Cell Biol.* **182**, 1083-1097.

Buchan, J. R. and Parker, R. (2009). Eukaryotic stress granules: the ins and outs of translation. *Mol. Cell* **36**, 932-941.

Carter, K. C., Bowman, D., Carrington, W., Fogarty, K., McNeil, J. A., Fay, F. S. and Lawrence, J. B. (1993). A three-dimensional view of precursor messenger RNA metabolism within the mammalian nucleus. *Science* **259**, 1330-1335.

Clemson, C. M., Hutchinson, J. N., Sara, S. A., Ensminger, A. W., Fox, A. H., Chess, A. and Lawrence, J. B. (2009). An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Mol. Cell* **33**, 717-726.

Decker, C. J. and Parker, R. (2012). P-bodies and stress granules: possible roles in the control of translation and mRNA degradation. *Cold Spring Harb. Perspect. Biol.* **4**, a012286.

Dias, A. P., Dufu, K., Lei, H. and Reed, R. (2010). A role for TREX components in the release of spliced mRNA from nuclear speckle domains. *Nat. Commun.* **1**, 97.

Dundr, M. (2012). Nuclear bodies: multifunctional companions of the genome. *Curr. Opin. Cell Biol.* **24**, 415-422.

Eils, R., Gerlich, D., Tvaruskó, W., Spector, D. L. and Misteli, T. (2000). Quantitative imaging of pre-mRNA splicing factors in living cells. *Mol. Biol. Cell* **11**, 413-418.

Engreitz, J. M., Sirokman, K., McDonel, P., Shishkin, A. A., Surka, C., Russell, P., Grossman, S. R., Chow, A. Y., Guttman, M. and Lander, E. S. (2014). RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent Pre-mRNAs and chromatin sites. *Cell* **159**, 188-199.

Fakan, S. and van Driel, R. (2007). The perichromatin region: a functional compartment in the nucleus that determines large-scale chromatin folding. *Semin. Cell Dev. Biol.* **18**, 676-681.

Feric, M., Vaidya, N., Harmon, T. S., Mitrea, D. M., Zhu, L., Richardson, T. M., Kriwacki, R. W., Pappu, R. V. and Brangwynne, C. P. (2016). Coexisting liquid phases underlie nucleolar subcompartments. *Cell* **165**, 1686-1697.

Fu, X.-D. and Maniatis, T. (1990). Factor required for mammalian spliceosome assembly is localized to discrete regions in the nucleus. *Nature* **343**, 437-441.

Girard, C., Will, C. L., Peng, J., Makarov, E. M., Kastner, B., Lemm, I., Urlaub, H., Hartmuth, K. and Lührmann, R. (2012). Post-transcriptional spliceosomes are retained in nuclear speckles until splicing completion. *Nat. Commun.* **3**, 994.

Gustafsson, M. G. L., Shao, L., Carlton, P. M., Wang, C. J. R., Golubovskaya, I. N., Cande, W. Z., Agard, D. A. and Sedat, J. W. (2008). Three-dimensional resolution doubling in wide-field fluorescence microscopy by structured illumination. *Biophys. J.* **94**, 4957-4970.

Habets, W. J., Hoet, M. H., De Jong, B. A., Van der Kemp, A. and Van Venrooij, W. J. (1989). Mapping of B cell epitopes on small nuclear ribonucleoproteins that react with human autoantibodies as well as with experimentally-induced mouse monoclonal antibodies. *J. Immunol.* **143**, 2560–6.

Habets, W. J., Sillekens, P. T., Hoet, M. H., Schalken, J. A., Roebroek, A. J., Leunissen, J. A., van de Ven, W. J. and van Venrooij, W. J. (1987). Analysis of a cDNA clone expressing a human autoimmune antigen: full-length sequence of the U2 small nuclear RNA-associated B" antigen. *Proc Natl Acad Sci USA* **84**, 2421–5.

Hall, L. L., Smith, K. P., Byron, M. and Lawrence, J. B. (2006). Molecular anatomy of a speckle. *Anat. Rec. A Discov. Mol. Cell Evol. Biol.* **288A**, 664-675.

Han, J., Xiong, J., Wang, D. and Fu, X.-D. (2011). Pre-mRNA splicing: where and when in the nucleus. *Trends Cell Biol.* **21**, 336-343.

Hanamura, A., Cáceres, J. F., Mayeda, A., Franza, B. R. and Krainer, A. R. (1998). Regulated tissue-specific expression of antagonistic pre-mRNA splicing factors. *RNA* **4**, 430-444.

Hu, Y., Kireev, I., Plutz, M., Ashourian, N. and Belmont, A. S. (2009). Large-scale chromatin structure of inducible genes: transcription on a condensed, linear template. *J. Cell Biol.* **185**, 87-100.

Hu, Y., Plutz, M. and Belmont, A. S. (2010). Hsp70 gene association with nuclear speckles is Hsp70 promoter specific. *J. Cell Biol.* **191**, 711-719.

Hutchinson, J. N., Ensminger, A. W., Clemson, C. M., Lynch, C. R., Lawrence, J. B. and Chess, A. (2007). A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics* **8**, 39.

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.

Ji, P., Diederichs, S., Wang, W., Böing, S., Metzger, R., Schneider, P. M., Tidow, N., Brandt, B., Buerger, H., Bulk, E. et al. (2003). MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene* **22**, 8031-8041.

Johnson, C., Primorac, D., McKinstry, M., McNeil, J., Rowe, D. and Lawrence, J. B. (2000). Tracking COL1A1 RNA in osteogenesis imperfecta. splice-defective transcripts initiate transport from the gene but are retained within the SC35 domain. *J. Cell Biol.* **150**, 417-432.

Khanna, N., Hu, Y. and Belmont, A. S. (2014). HSP70 transgene directed motion to nuclear speckles facilitates heat shock activation. *Curr. Biol.* **24**, 1138-1144.

Li, X., Zhou, B., Chen, L., Gou, L.-T., Li, H. and Fu, X.-D. (2017). GRID-seq reveals the global RNA-chromatin interactome. *Nat. Biotechnol.* **35**, 940-950.

Lin, S., Xiao, R., Sun, P., Xu, X. and Fu, X.-D. (2005). Dephosphorylation-dependent sorting of SR splicing factors during mRNP maturation. *Mol. Cell* **20**, 413-425.

Liu, N. and Pan, T. (2016). N6-methyladenosine–encoded epitranscriptomics. *Nat. Struct. Mol. Biol.* **23**, 98-102.

Liu, N., Dai, Q., Zheng, G., He, C., Parisien, M. and Pan, T. (2015). N(6)-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature* **518**, 560-564.

Lu, Z., Zhang, Q. C., Lee, B. K. J., Flynn, R. A., Smith, M. A., Robinson, J. T., Davidovich, C., Gooding, A. R., Goodrich, K. J., Mattick, J. S. et al. (2016). RNA duplex map in living cells reveals higher-order transcriptome structure. *Cell* **165**, 1267-1279.

Malakar, P., Shilo, A., Mogilevsky, A., Stein, I., Pikarsky, E., Nevo, Y., Benyamini, H., Elgavish, S., Zong, X., Prasanth, K. V. et al. (2017). Long noncoding RNA MALAT1 promotes hepatocellular carcinoma development by SRSF1 upregulation and mTOR activation. *Cancer Res.* **77**, 1155-1167.

Mao, Y. S., Zhang, B. and Spector, D. L. (2011a). Biogenesis and function of nuclear bodies. *Trends Genet.* **27**, 295-306.

Mao, Y. S., Sunwoo, H., Zhang, B. and Spector, D. L. (2011b). Direct visualization of the co-transcriptional assembly of a nuclear body by noncoding RNAs. *Nat. Cell Biol.* **13**, 95-101.

Marthandan, S., Baumgart, M., Priebe, S., Groth, M., Schaer, J., Kaether, C., Guthke, R., Cellerino, A., Platzer, M., Diekmann, S. et al. (2016). Conserved senescence associated genes and pathways in primary human fibroblasts detected by RNA-Seq. *PLoS ONE* **11**, e0154531.

Mattaj, I. W., Habets, W. J. and van Venrooij, W. J. (1986). Monospecific antibodies reveal details of U2 snRNP structure and interaction between U1 and U2 snRNPs. *EMBO J.* **5**, 997-1002.

Mintz, P. J. and Spector, D. L. (2000). Compartmentalization of RNA processing factors within nuclear speckles. *J. Struct. Biol.* **129**, 241-251.

Misteli, T. and Spector, D. L. (1997). Protein phosphorylation and the nuclear organization of pre-mRNA splicing. *Trends Cell Biol.* **7**, 135-138.

Misteli, T., Cáceres, J. F. and Spector, D. L. (1997). The dynamics of a pre-mRNA splicing factor in living cells. *Nature* **387**, 523-527.

Misteli, T., Cáceres, J. F., Clement, J. Q., Krainer, A. R., Wilkinson, M. F. and Spector, D. L. (1998). Serine phosphorylation of SR proteins is required for their recruitment to sites of transcription in vivo. *J. Cell Biol.* **143**, 297-307.

Miyagawa, R., Tano, K., Mizuno, R., Nakamura, Y., Ijiri, K., Rakwal, R., Shibato, J., Masuo, Y., Mayeda, A., Hirose, T. et al. (2012). Identification of cis- and trans-acting factors involved in the localization of MALAT-1 noncoding RNA to nuclear speckles. *RNA* **18**, 738-751.

Moen, P. T., Johnson, C. V., Byron, M., Shopland, L. S., de la Serna, I. L., Imbalzano, A. N. and Lawrence, J. B. (2004). Repositioning of muscle-specific genes relative to the periphery of SC-35 domains during skeletal myogenesis. *Mol. Biol. Cell* **15**, 197-206.

Mor, A., White, A., Zhang, K., Thompson, M., Esparza, M., Muñoz-Moreno, R., Koide, K., Lynch, K. W., García-Sastre, A. and Fontoura, B. M. A. (2016).

Influenza virus mRNA trafficking through host nuclear speckles. *Nat. Microbiol.* **1**, 16069.

Nakagawa, S., Ip, J. Y., Shioi, G., Tripathi, V., Zong, X., Hirose, T. and Prasanth, K. V. (2012). Malat1 is not an essential component of nuclear speckles in mice. *RNA* **18**, 1487-1499.

Nguyen, T. C., Cao, X., Yu, P., Xiao, S., Lu, J., Biase, F. H., Sridhar, B., Huang, N., Zhang, K. and Zhong, S. (2016). Mapping RNA-RNA interactome and RNA structure in vivo by MARIO. *Nat. Commun.* **7**, 12023.

Pederson, T. (2011). The nucleolus. *Cold Spring Harb. Perspect. Biol.* **3**, a000638.

Price, S. R., Evans, P. R. and Nagai, K. (1998). Crystal structure of the spliceosomal U2B″-U2A′ protein complex bound to a fragment of U2 small nuclear RNA. *Nature* **394**, 645-650.

Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. and Tyagi, S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5**, 877-879.

Roundtree, I. A. and He, C. (2016). RNA epigenetics–chemical messages for posttranscriptional gene regulation. *Curr. Opin. Chem. Biol.* **30**, 46-51.

Saitoh, N., Spahr, C. S., Patterson, S. D., Bubulya, P., Neuwald, A. F. and Spector, D. L. (2004). Proteomic analysis of interchromatin granule clusters. *Mol. Biol. Cell* **15**, 3876-3890.

Sanford, J. R., Wang, X., Mort, M., Vanduyn, N., Cooper, D. N., Mooney, S. D., Edenberg, H. J. and Liu, Y. (2009). Splicing factor SFRS1 recognizes a functionally diverse landscape of RNA transcripts. *Genome Res.* **19**, 381-394.

Scherly, D., Dathan, N. A., Boelens, W., van Venrooij, W. J. and Mattaj, I. W. (1990). The U2B' RNP motif as a site of protein-protein interaction. *EMBO J.* **9**, 3675-3681.

Schermelleh, L., Carlton, P. M., Haase, S., Shao, L., Winoto, L., Kner, P., Burke, B., Cardoso, M. C., Agard, D. A., Gustafsson, M. G. L. et al. (2008). Subdiffraction multicolor imaging of the nuclear periphery with 3D structured illumination microscopy. *Science* **320**, 1332-1336.

Sharma, A., Takata, H., Shibahara, K., Bubulya, A. and Bubulya, P. A. (2010). Son is essential for nuclear speckle organization and cell cycle progression. *Mol. Biol. Cell* **21**, 650-663.

Shepard, P. J. and Hertel, K. J. (2009). The SR protein family. *Genome Biol.* **10**, 242.

Shopland, L. S., Johnson, C. V. and Lawrence, J. B. (2002). Evidence that all SC-35 domains contain mRNAs and that transcripts can be structurally constrained within these domains. *J. Struct. Biol.* **140**, 131-139.

Smith, K. P., Moen, P. T., Wydner, K. L., Coleman, J. R. and Lawrence, J. B. (1999). Processing of endogenous pre-mRNAs in association with SC-35 domains is gene specific. *J. Cell Biol.* **144**, 617-629.

Souquere, S., Beauclair, G., Harper, F., Fox, A. and Pierron, G. (2010). Highly ordered spatial organization of the structural long noncoding NEAT1 RNAs within paraspeckle nuclear bodies. *Mol. Biol. Cell* **21**, 4020-4027.

Spector, D. L. and Lamond, A. I. (2011). Nuclear speckles. *Cold Spring Harb. Perspect. Biol.* **3**, a000646.

Thiry, M. (1995a). The interchromatin granules. *Histol. Histopathol.* **10**, 1035-1045.

Thiry, M. (1995b). Behavior of interchromatin granules during the cell cycle. *Eur. J. Cell Biol.* **68**, 14-24.

Tripathi, V., Ellis, J. D., Shen, Z., Song, D. Y., Pan, Q., Watt, A. T., Freier, S. M., Bennett, C. F., Sharma, A., Bubulya, P. A. et al. (2010). The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol. Cell* **39**, 925-938.

Tripathi, V., Song, D. Y., Zong, X., Shevtsov, S. P., Hearn, S., Fu, X.-D., Dundr, M. and Prasanth, K. V. (2012). SRSF1 regulates the assembly of pre-mRNA processing factors in nuclear speckles. *Mol. Biol. Cell* **23**, 3694-3706.

Tripathi, V., Shen, Z., Chakraborty, A., Giri, S., Freier, S. M., Wu, X., Zhang, Y., Gorospe, M., Prasanth, S. G., Lal, A. et al. (2013). Long noncoding RNA MALAT1 controls cell cycle progression by regulating the expression of oncogenic transcription factor B-MYB. *PLoS Genet.* **9**, e1003368.

West, J. A., Mito, M., Kurosaka, S., Takumi, T., Tanegashima, C., Chujo, T., Yanaka, K., Kingston, R. E., Hirose, T., Bond, C. et al. (2016). Structural, super-resolution microscopy analysis of paraspeckle nuclear body organization. *J. Cell Biol.* **214**, 817-830.

Wheeler, J. R., Matheny, T., Jain, S., Abrisch, R. and Parker, R. (2016). Distinct stages in stress granule assembly and disassembly. *Elife* **5**, e18413. .

Xie, S. Q., Martin, S., Guillot, P. V., Bentley, D. L. and Pombo, A. (2006). Splicing speckles are not reservoirs of RNA polymerase II, but contain an inactive form, phosphorylated on serine2 residues of the C-terminal domain. *Mol. Biol. Cell* **17**, 1723-1733.

Xing, Y., Johnson, C. V., Moen, P. T., McNeil, J. A. and Lawrence, J. (1995). Nonrandom gene organization: structural arrangements of specific pre-mRNA transcription and splicing with SC-35 domains. *J. Cell Biol.* **131**, 1635-1647.

Yang, L., Lin, C., Liu, W., Zhang, J., Ohgi, K. A., Grinstein, J. D., Dorrestein, P. C. and Rosenfeld, M. G. (2011). ncRNA- and Pc2 methylation-dependent gene relocation between nuclear structures mediates gene activation programs. *Cell* **147**, 773-788.

Zhong, X.-Y., Wang, P., Han, J., Rosenfeld, M. G. and Fu, X.-D. (2009). SR proteins in vertical integration of gene expression from transcription to RNA processing to translation. *Mol. Cell* **35**, 1-10.

Zhu, L. and Brangwynne, C. P. (2015). Nuclear bodies: the emerging biophysics of nucleoplasmic phases. *Curr. Opin. Cell Biol.* **34**, 23-30.

Supplementary Figures

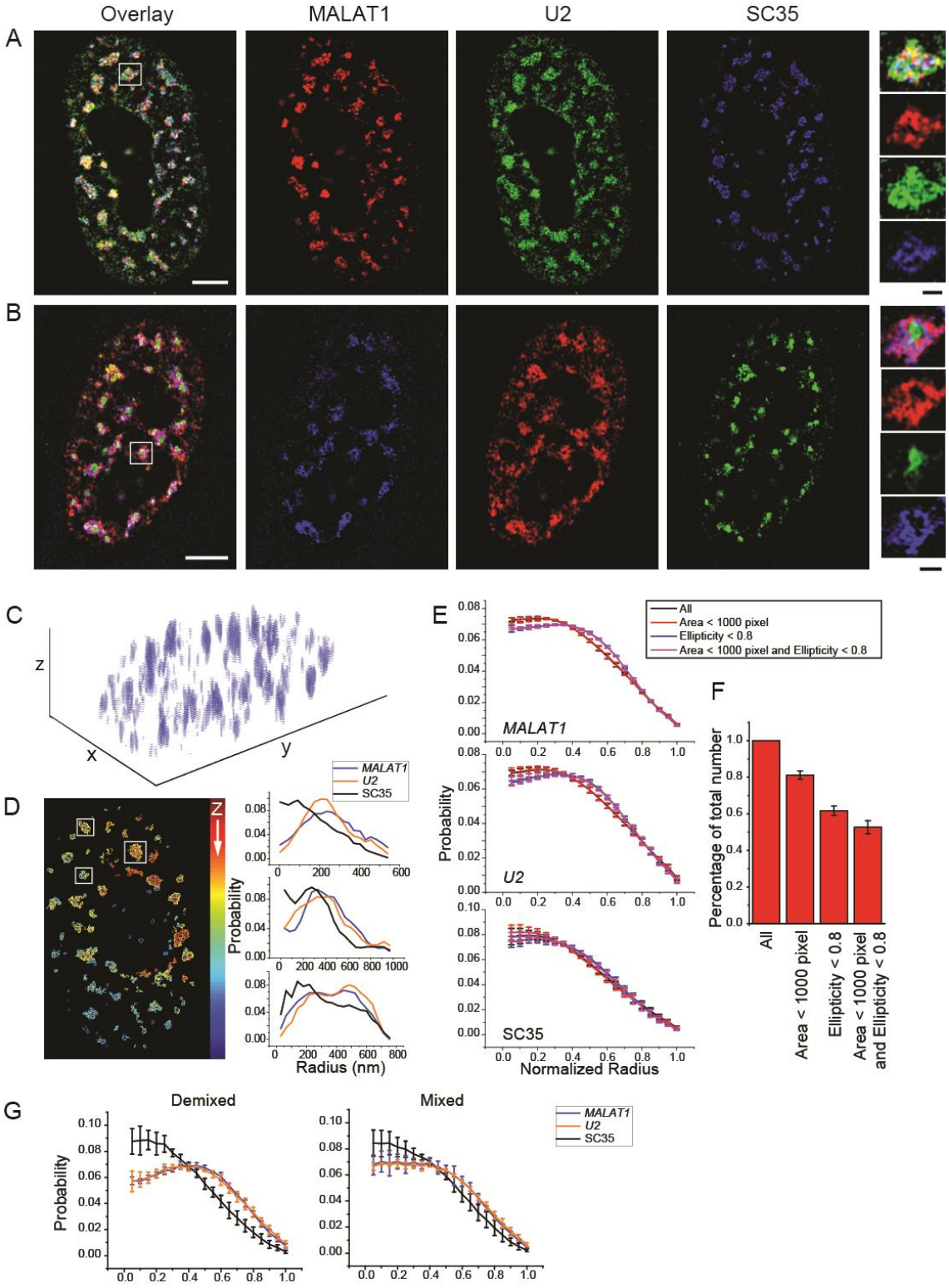


Figure S1. Additional sample images of *MALAT1*, *U2* and SC35 and analysis of radial distribution. (A) Example in which SC35 and *MALAT1/U2* are more mixed compared to the

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).

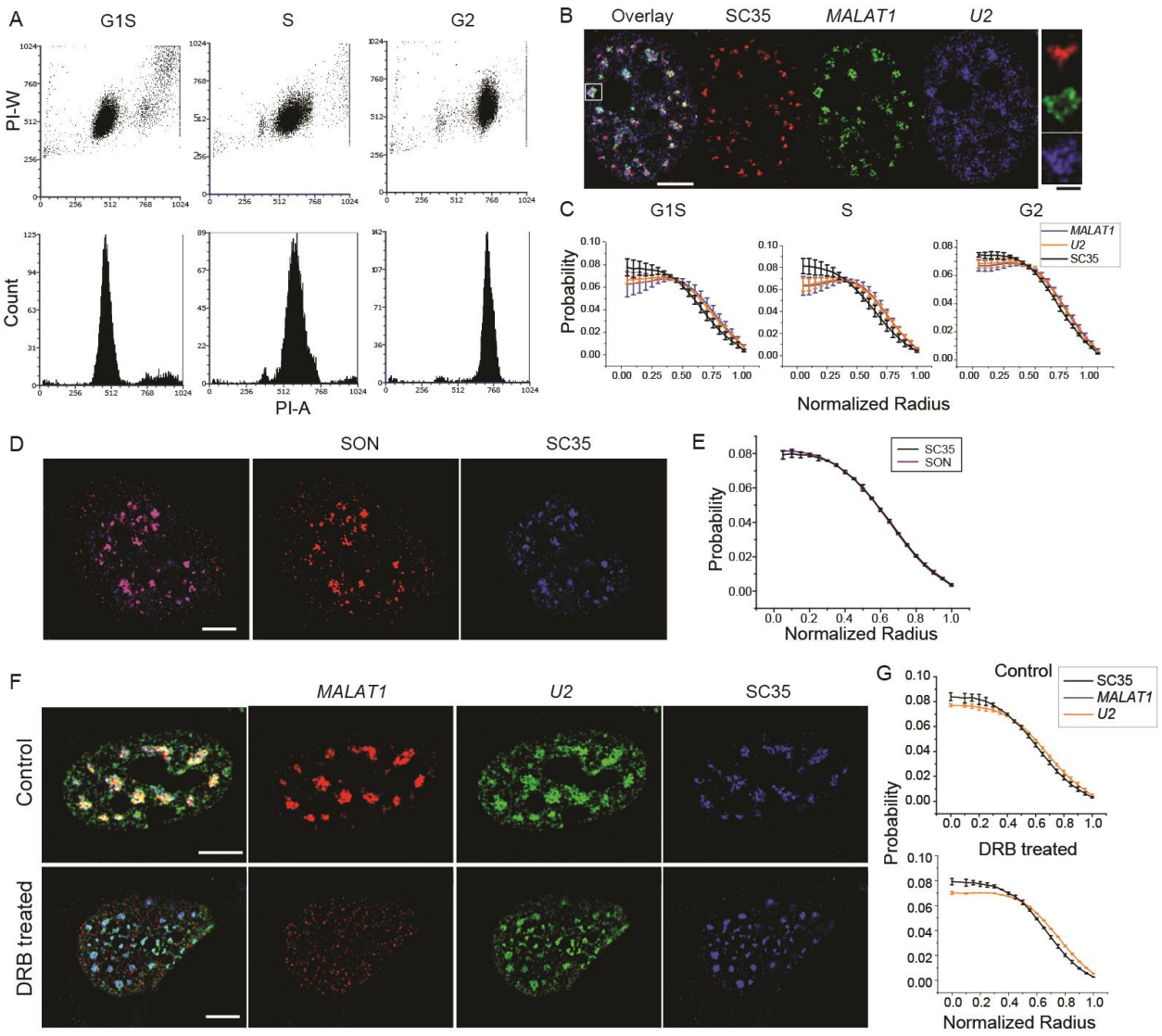


Figure S2. Non-random compositional organization of nuclear speckle during various cell cycle stages and under transcription inhibition. (A) Flow cytometry analyses to demonstrate cell cycle synchronization in HeLa cells (B) Sample image of *MALAT1* (green), *U2* (blue) and SC35 (red) by SIM in S phase 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.

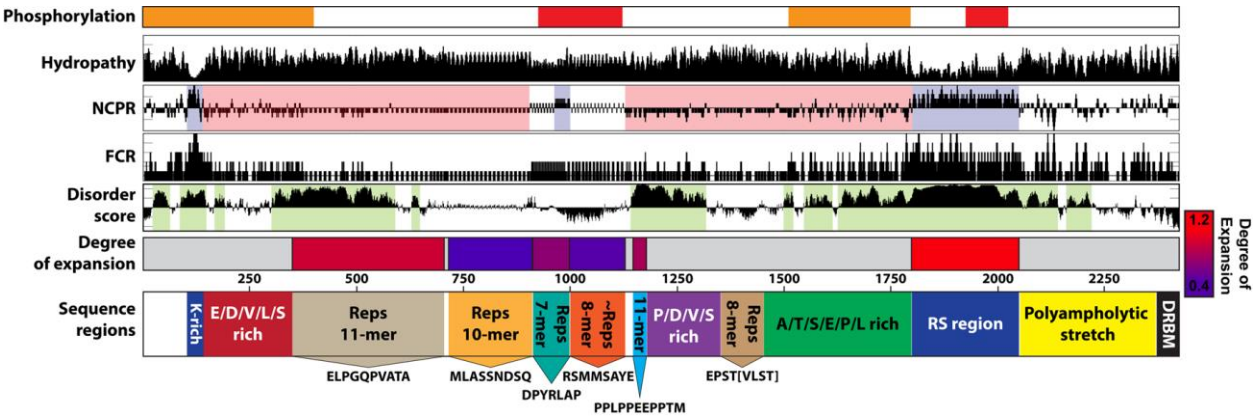


Figure S3. Summary of the sequence attributes of SON that are used to develop the architecture for coarse-graining. The linear sequence properties of SON are quantified using a series of distinct metrics that are organized into sequence-aligned "tracks". The phosphorylation track qualitatively defines 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).

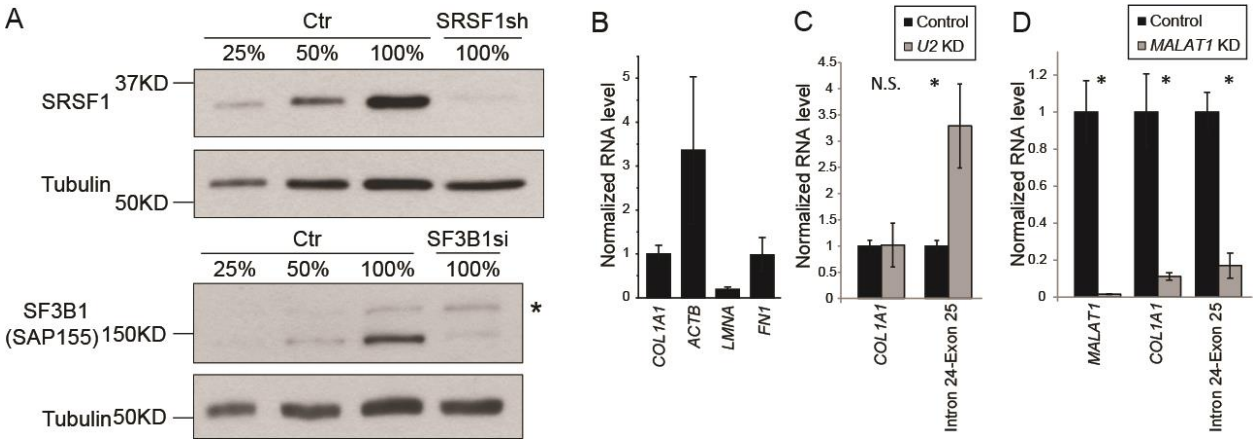


Figure S4. Immunoblot and qPCR analysis for knock-down verification. (A) Immunoblot analysis to depict the efficiency of knock down of SRSF1 and SF3B1 in WI-38 cells treated with SRSF1 shRNA or SF3b1 siRNA. Tubulin is used as loading control. * represents the cross-reacting band. (B) Total cellular mRNA 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.

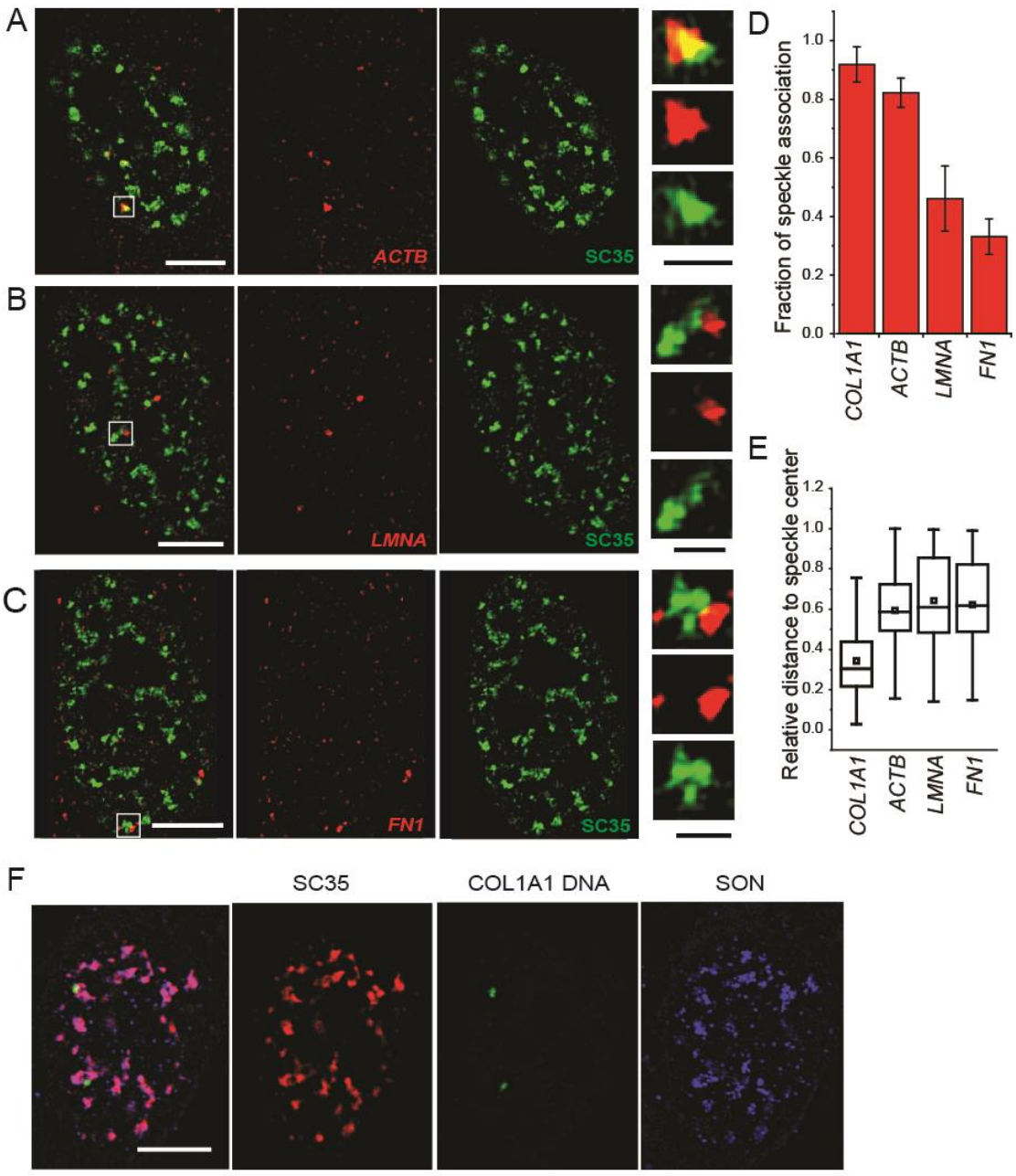


Figure S5. Association of RNA transcripts with nuclear speckles. (A) – (C) Sample images of *ACTB*, *LMNA* and *FN1* RNAs (red) with SC35 (green) respectively. Scale bars represent 5 µm in all cell images and 1 µm all zoomed-in speckle images. (D) Percentage of speckle-associated RNA accumulation sites. 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.

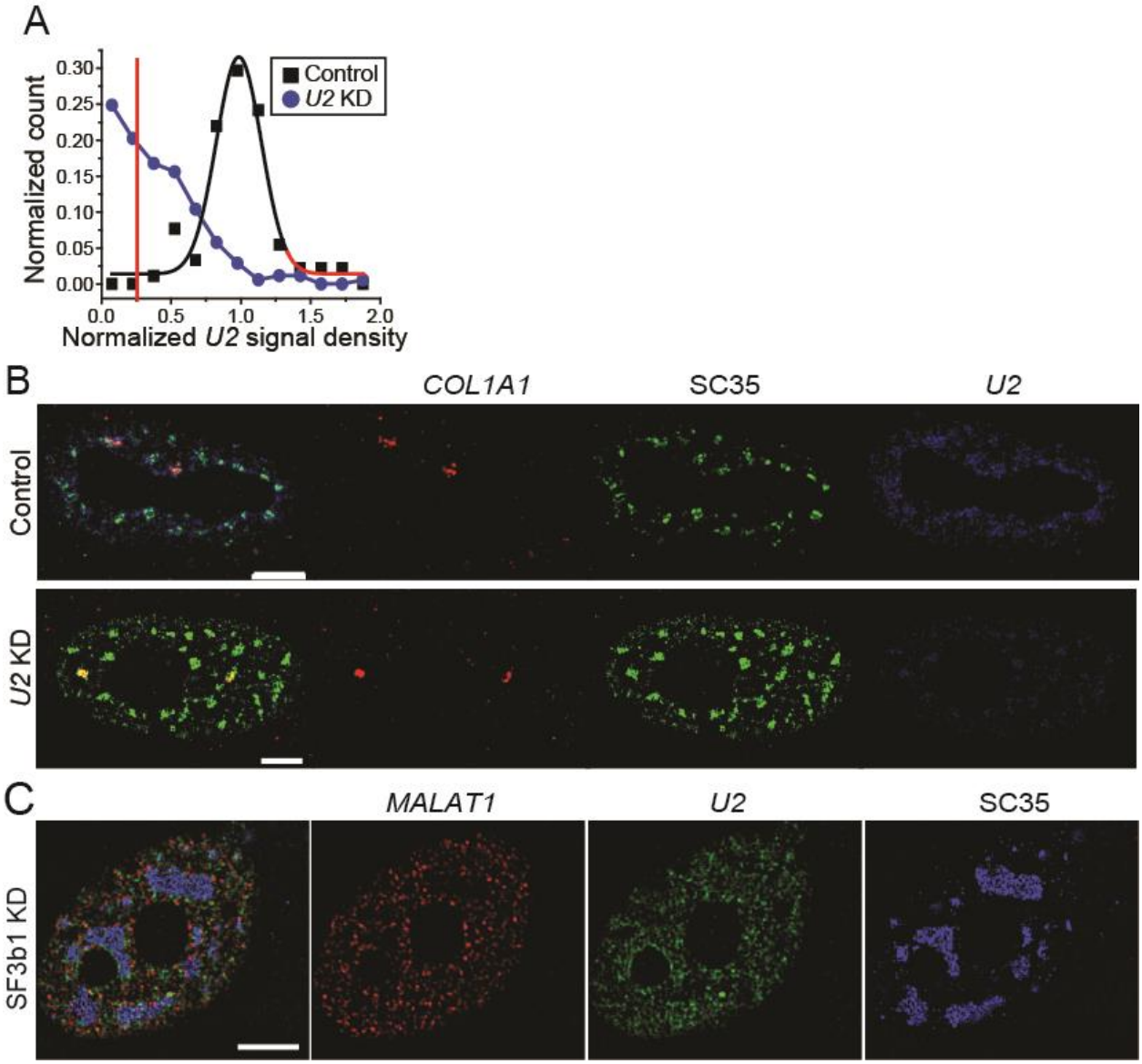


Figure S6. Effect of *U2* and SF3b1 knockdown on nuclear speckles. (A) Normalized *U2* intensity histograms for the control and *U2* knockdown cells. Cells with average *U2* intensity less than 25% of that in the negative control are selected for analysis. (B) Sample image of *COL1A1* RNA (red), SC35 (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.

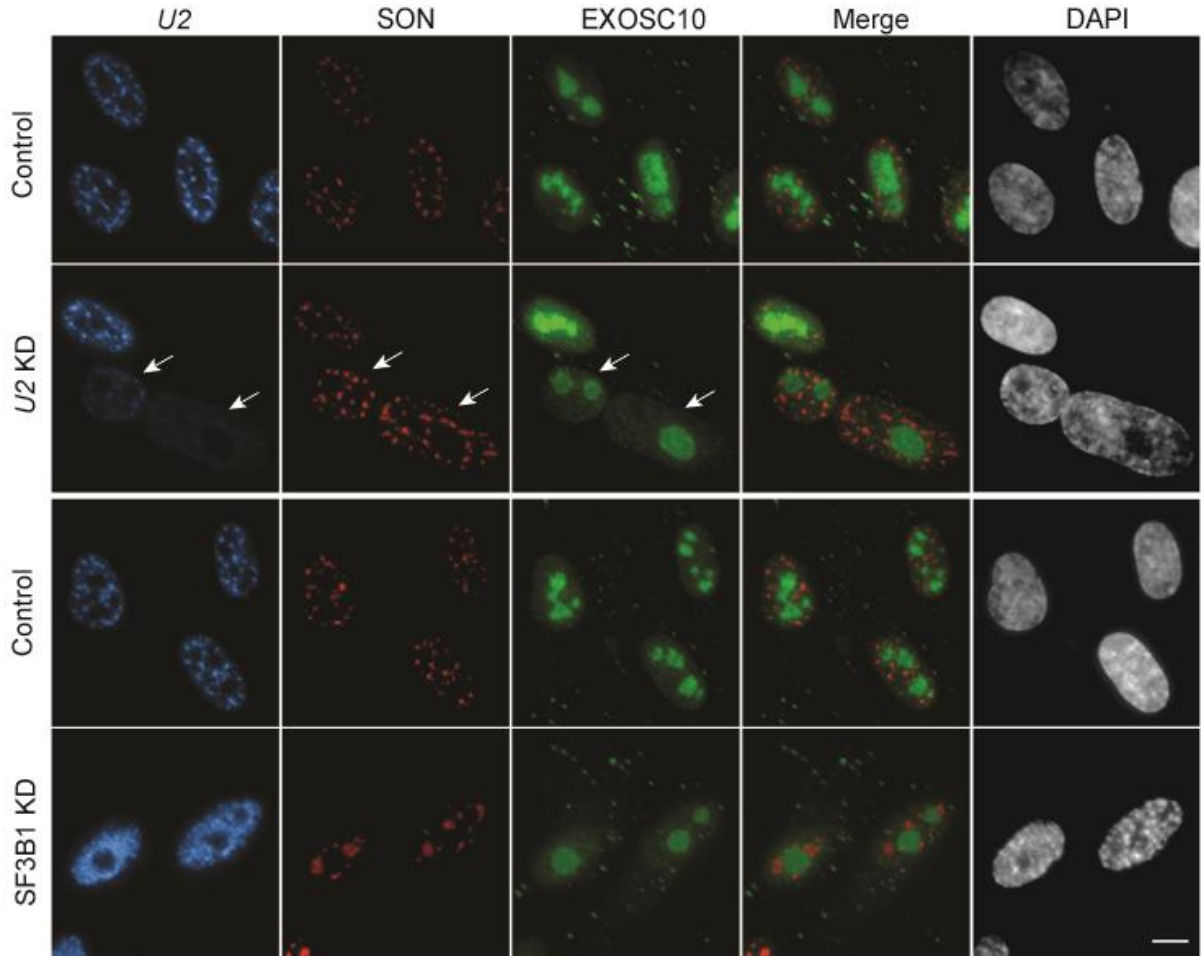


Figure S7. Nuclear speckles do not represent major sites of RNA degradation in the nucleus. Co-FISH and immunostaining to detect *U2* snRNA (blue), SON (red) and EXOSC10 (green) in control, *U2*-depleted and SF3B1-depleted WI-38 cells. DNA is stained with DAPI. Scale bar represents 5 µm. Please note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

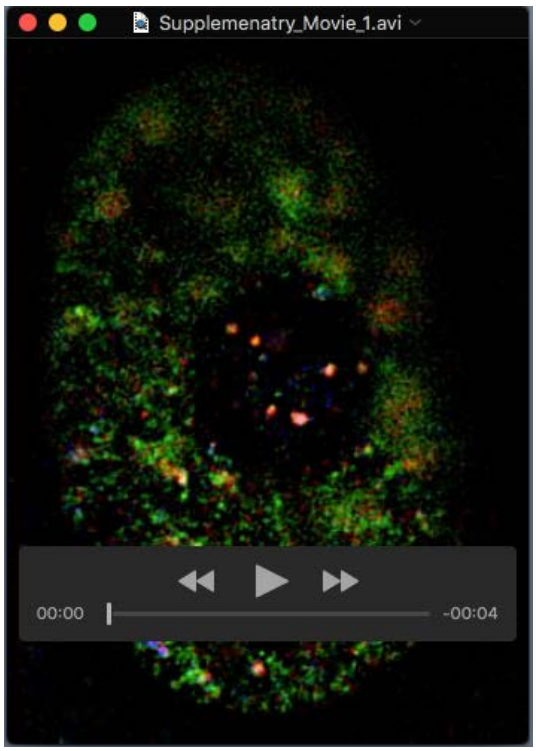
Matlock, M.K., Holehouse, A.S., and Naegle, K.M. (2015). ProteomeScout: a repository and analysis resource for post-translational modifications and proteins. Nucleic Acids Res *43*, D521–30.

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.