Nuclear speckle-localized RNAs exhibit preferential positioning and orientation

Sneha Paul[1†], Mauricio A. Arias[2†], Li Wen[3‡], Susan E. Liao[2‡], Jiacheng Zhang[4], Xiaoshu Wang[5], Oded Regev[2*], Jingyi Fei[1,6*]

[1] Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, IL, 60636
[2] Courant Institute of Mathematical Sciences, New York University, New York, NY, 10012
[3] Department of Physics, The University of Chicago, Chicago, IL, 60636
[4] Graduate Program in Biophysical Sciences, The University of Chicago, Chicago, IL, 60636
[5] The College, The University of Chicago, Chicago, IL, 60636
[6] Institute for Biophysical Dynamics, The University of Chicago, Chicago, IL, 60636

[†] These authors contributed equally to this work
[‡] These authors contributed equally to this work

*Correspondence: jingyifei@uchicago.edu (JF, Lead contact), regev@cims.nyu.edu (OR)

Abstract (150 words)

Nuclear speckles are a type of membraneless organelle in higher eukaryotic nuclei. Multivalent interactions between residing proteins and RNAs not only drive the formation, but also the layered compositional organization of nuclear speckles. Here, using super-resolution imaging, we demonstrate that nuclear speckle-localized RNA transcripts exhibit preferential intra-speckle positioning and radial orientation. Specifically, we show that transcripts containing a region enriched in SR protein binding motifs and another region enriched in hnRNP binding motifs localize to the outer shell. Moreover, they are oriented so that the SR enriched region is closer to the speckle center relative to the hnRNP enriched region. This differential intra-speckle positioning of RNA transcripts is driven by multivalent interactions between the two motifs and SR/hnRNP proteins respectively, two protein families that exhibit distinct subcellular localization relative to nuclear speckles. Such intra-speckle RNA positioning and orientation might explain the importance of RNAs in maintaining the structural integrity of nuclear speckles, and point at their possible functional roles in RNA splicing.

Main (4500 words)

Eukaryotic cells contain many membraneless organelles with distinct nuclear[1–3] or cytoplasmic[4,5] localizations. These membraneless organelles generally contain RNAs, RNA binding proteins (RBPs) and ribonucleoprotein assemblies[6–8]. Multivalent interactions between RNA and protein components drive the formation of these organelles through liquid-liquid phase separation[9–12]. Constituent molecules are dynamically exchanged between the membraneless organelles and the surrounding media[13–16]. Dynamic localization of RNAs and proteins to these subcellular compartments has a profound impact on gene expression and cellular activities[17–19] and can provide a novel way for stress response and adaptation[18,20,21].

Nuclear speckles are a type of membraneless organelle in higher eukaryotic cells. There are between 20 and 50 of them per cell[22,23] and their size ranges from a few hundred nanometers to a few microns[23]. Nuclear speckles are rich in snRNP components, certain splicing factors, including SR proteins (a family of RBPs named for containing regions with repetitive serine and arginine residues), polyadenylated RNAs, and certain long noncoding RNAs (lncRNA)[24–27]. While initially proposed to serve as storage sites for splicing factors[2,28], recent evidence suggests they actively participate in gene regulation[29,30]. Nuclear speckles are found near actively transcribed genes[24,31–33] suggesting a functional role in enhancing transcription[24,34,35] and facilitating co-transcriptional splicing[30,36]. Furthermore, nuclear speckles play important roles in several RNA processing steps[37], splicing quality control[38,39], and RNA export[40]. Finally, nuclear speckles are also suggested to help maintain adequate intra-chromosomal interactions and 3D-genome organization[41].

The formation of nuclear speckles requires the scaffold protein SON[42,43] and the target of the SC35 mAB antibody, which was previously mistakenly identified as SRSF2, but recently identified as SRRM2[43]. Multivalent interactions between low complexity regions of these scaffold proteins and other nuclear speckle-resident proteins and RNAs play critical roles in phase separating nuclear speckles from the nucleoplasm[7,44,45]. Using super-resolution imaging, we previously demonstrated that the constituent proteins and RNAs show differential intra-speckle spatial organization[45]. Specifically, SON and SRRM2 form the core layer of speckles, while spliceosomal components, including small nuclear RNAs (snRNAs) and spliceosomal proteins, and nuclear speckle localized lncRNA, *MALAT1*, are enriched in the outer layer. Computational simulation suggests that this layered spatial organization is mediated by sequence-specific intermolecular interactions of resident proteins and RNAs in the speckles. A more recent study identified an additional nuclear speckle scaffold protein, REST corepressor 2 (RCOR2), which localizes to the core domain of nuclear speckles[46], and further demonstrated that RNAs either assist or stabilize RCOR2 localization to the speckle core. Collectively, these reports[45,46] highlight that multivalent interactions between scaffold proteins, other nuclear speckle-resident proteins and RNAs play critical roles in not only phase separating nuclear speckles from the nucleoplasm, but also in promoting layered spatial organization of nuclear speckle components.

While certain RBPs (such as SR proteins) are enriched in nuclear speckles[47–50], other RBPs (such as heterogeneous nuclear ribonucleoprotein (hnRNP)) do not exhibit any enrichment or might be slightly depleted[49,51–53]. Such differential localization of the nuclear proteome can create a unique chemical environment at the surface of nuclear speckles, with different subsets and concentrations of RBPs inside and outside speckles. These differences, in turn, might lead to positional biases for different regions of RNA molecules. Specifically, a transcript containing regions with favorable interactions for different groups of RBPs, should align with the positions of the interacting proteins with respect to the nuclear speckle. In other words, regions interacting with proteins outside the nuclear speckle will tend to localize towards the outside, while regions interacting with proteins inside the nuclear speckle will tend to localize towards the inside. In this way, the position of RNA transcripts would be driven to the outer shell or the periphery of the nuclear speckle, and the orientation of the RNA molecule will be constrained (**Fig. 1**).

In this work, we tested this intra-speckle RNA positioning model (**Fig. 1**) using SRSF1 and SRSF7 from the SR protein family and hnRNPA1 from the hnRNP family as examples. Supporting our

hypothesis, we found that SRSF1/7 motif-rich regions in RNA transcripts were positioned closer to the speckle center, whereas hnRNPA1 motif-rich regions were positioned closer to the periphery. Our results provide evidence that multivalent RNA-RBP interactions drive RNAs to adopt preferential positioning and orientation within membraneless organelles, and point at their potential functional roles in RNA splicing.

Results

SRSF1 and hnRNPA1 proteins exhibit distinct localization relative to nuclear speckles

Confirming previous results, we found that SRSF1 protein was consistently enriched in nuclear speckles[47,48,54] (**Fig. S1a**). For hnRNPA1 proteins, imaging showed they were relatively homogenously distributed throughout the nucleus, with slight depletion in some nuclear speckles (**Fig. S1b**). While we cannot conclusively demonstrate depletion of hnRNPA1 proteins in speckles, our results and previous work strongly suggest that they are not enriched[51,53]. In summary, the relative concentrations of SRSF1 and hnRNPA1 inside versus outside of nuclear speckles are distinct, and we chose this SR-hnRNP protein pair for further analysis.

SRSF1 and hnRNPA1 binding motifs on RNA transcripts display differential intra-speckle positioning

We designed a three-exon reporter construct (WT_SRSF1) containing SRSF1 motifs in the middle exon and hnRNPA1 motifs in the second intron (**Fig. 2a**). To enhance the RNA-RBP interaction strength, we introduced multiple SRSF1 and hnRNPA1 binding motifs in those regions. The constructs were transfected into a HeLa cell line with stably expressed Tet-regulated transactivator Tet-On 3G, and RNA expression was induced with doxycycline. Using fluorescence in situ hybridization (FISH), we labeled the SRSF1 and hnRNPA1 motifs on the RNA transcripts with CF568 and Alexa Fluor 647 (AF647), respectively. Nuclear speckles were stained with Alexa Fluor 488 (AF488) labeled antibody against the scaffold protein SRRM2[43]. At 30 min induction, the FISH signals for both SRSF1 and hnRNPA1 motifs from the WT_SRSF1 construct were mostly localized to nuclear speckles (**Fig. S2a**), as observed under diffraction-limited epifluorescence imaging. At 2 h induction, a significant portion of the FISH signal for the SRSF1 motifs was localized to the cytoplasm, corresponding to the spliced and exported mRNAs, whereas the residual nuclear-localized signals from the SRSF1 and hnRNPA1 motifs remained localized to nuclear speckles, corresponding to the pre-mRNA transcripts. Using a reverse transcription PCR (RT-PCR) assay, we verified that both introns of WT_SRSF1 were spliced with the middle exon (containing SR motifs) included (**Fig. S3a**), consistent with the epifluorescence imaging results. Both imaging and RT-PCR data verified that the RNAs from the reporter constructs were processed and exported appropriately.

To ensure that we were imaging RNA transcripts containing both SRSF1 and hnRNPA1 motifs, rather than spliced RNA products, we (1) introduced a mutation at the 3' splice site of the second intron (MUT_SRSF1) (**Fig. 2a**), which causes the splicing reaction to stall[55,56], and, alternatively, (2) used the splicing inhibitor Pladienolide B to arrest splicing of the WT_SRSF1 construct[57]. Using the RT-PCR assay, we verified that the second intron in the MUT_SRSF1 construct was not spliced even 2 h after induction (**Fig. S3a**), leaving the SRSF1 and hnRNPA1 motifs on the same partially spliced RNA. In addition, we verified that splicing of the WT_SRSF1 pre-mRNA was

inhibited in the presence of Pladieonolide B (**Fig. S3a**) at 30 min and 2 h inductions. Epifluorescence imaging also confirmed that the RNA remained localized in nuclear speckles under all these conditions (**Fig. S2**).

Since the SR and hnRNP motifs are adjacent to each other on the RNA transcript, we expect the difference in their intra-speckle positioning (if any) to be very small in terms of absolute distance. We therefore performed super-resolution imaging using single molecule localization microscopy (SMLM). We imaged RNAs from WT_SRSF1 in the presence of Pladieonolide B (**Fig. 2b**), and from MUT_SRSF1 without Pladieonolide B (**Fig. 2c**). An epifluorescence image of SRRM2 was recorded before the SMLM imaging on the same cell. For quantitative comparison of the two RBP binding motifs in the nuclear speckle, we first selected nuclear speckles containing associated RNA signals by using intensity thresholding on the sum of all three channels, namely the two RNA FISH signal channels and the nuclear speckle marker channel. The radial distributions of the signal intensity from the two RNA motifs were then determined as a function of the normalized distance to the geometric center of the speckle (D_{Center}, **Fig. 2d**). Supporting our hypothesis, this analysis demonstrated that the SRSF1 motif-rich region is distributed closer to the interior of the speckle compared to the hnRNPA1 motif-rich region (**Fig. 2e-h**). As an alternative analysis, we calculated the radial intensity distributions of the two RNA signals as a function of the normalized distance to the edge of the speckle (D_{Edge}, **Fig. 2i**). In agreement with our first analysis, this analysis demonstrated that the SRSF1 motif-rich region is distributed further away from the edge of the speckle, *i.e.*, closer to the center of the speckle, compared to the hnRNPA1 motif-rich region (**Fig. 2j-m**).

To rule out imaging artifacts, we reversed the FISH labelling scheme; *i.e.,* SRSF1 motifs were labeled with AF647 and hnRNPA1 motifs were labeled with CF568. We observed the same trend in differential intra-speckle positioning of the two motifs, using both analysis methods **(Fig. S4a-d**).

WT_SRSF1 and MUT_SRSF1 contain repeat sequences of SR and hnRNP binding motifs. To rule out the possibility that this intra-speckle RNA positioning is an artifact due to these repetitions, we designed constructs WT_NR and MUT_NR where we randomly mutated 30% of the nucleotides in the SR and hnRNP motif-containing regions. These constructs are still enriched in SRSF1 motifs in the exon and in hnRNPA1 motifs in the intron, but were devoid of repeats, as measured by BLAST (**Fig. S5a-c**). RNA transcripts from these non-repeat constructs demonstrated the same splicing behavior **(Fig. S3b)**. Importantly, they also demonstrated similar intra-speckle positioning (**Fig. S5d-g**). The relative difference between hnRNPA1 and SRSF1 motifs was lower in MUT_NR as compared to MUT_SRSF1, which can be attributed to the slight decrease in strength of the two motifs in MUT_NR **(Fig. S5c)**. These results confirmed that the intra-speckle positioning of the two motifs is driven by the presence of SRSF1 and hnRNPA1 binding motifs, and not because of the sequence repetition.

Together, these results strongly suggest that the SRSF1 motif-rich and hnRNPA1 motif-rich regions on the same RNA transcripts are not uniformly distributed within nuclear speckles. Instead, they have differential intra-speckle positioning, with SRSF1 motif-rich regions positioned closer to the speckle center compared with hnRNPA1 motif-rich regions.

RNA transcripts with a combination of SRSF7 and hnRNPA1 motifs exhibit similar intra-speckle positioning

To test whether RNA containing other combinations of SR and hnRNP motifs exhibits similar positioning and orientation within nuclear speckles, we replaced the SRSF1 motifs in WT_SRSF1 and MUT_SRSF1 with SRSF7 motifs to generate WT_SRSF7 and MUT_SRSF7 (**Fig. 3a**). The splicing behavior of these constructs was similar to the SRSF1 constructs (**Fig. S3c**). The radial intensity distributions with respect to the center and edge showed that SRSF7 motif-rich region were closer to the center of nuclear speckles than the hnRNPA1 motif-rich region (**Fig. 3b-i**), the same trend as observed for WT_SRSF1 and MUT_SRSF1. These results demonstrate that the intra-speckle positioning of SR and hnRNP motifs is not specific to the SRSF1-hnRNPA1 combination.

RNA-RBP interaction strength determines RNA positioning in nuclear speckle

As we hypothesize that the mechanism underlying the intra-speckle RNA positioning and orientation is the multivalent interactions between RNAs with RBPs residing inside and outside nuclear speckles, we expect that changing the RNA-RBP interaction strengths will modulate RNA intra-speckle positioning. To test this hypothesis, we knocked down specific SR or hnRNP proteins. We expect that weakening RNA-SR protein interaction will lead to the migration of RNAs towards the speckle periphery, whereas weakening RNA-hnRNP protein interaction will lead to the migration of RNAs towards the speckle interior.

Using MUT_SRSF1 as a representative, we measured the intra-speckle positioning of the SRSF1 motif-rich and hnRNPA1 motif-rich regions upon siRNA-mediated knockdown of SRSF1 and hnRNPA1 proteins. Knocking down of these proteins did not cause any significant change in the morphology of nuclear speckles. We achieved 83±5% and 57±8% knockdown efficiency of SRSF1 and hnRNPA1 proteins, respectively (**Fig. S6**). Two-color SMLM imaging was performed under these conditions following the same labelling scheme (CF568 for SR motifs and A647 for hnRNP motifs). Immunofluorescence from an antibody against SON protein was used to visualize nuclear speckles. To choose cells with efficient protein knockdown, hnRNPA1 and SRSF1 proteins were stained with their respective antibodies and imaged with a 750 nm laser. Cells showing significant reduction in immunofluorescence signal compared to cells treated with scramble siRNA were selected.

Confirming our hypothesis, SRSF1 knockdown caused significant migration of RNA transcripts towards the speckle periphery (**Fig. 4**). Conversely, hnRNPA1 knockdown caused migration of the RNA transcripts towards the speckle interior (**Fig. 4**). The trends were reproducible when we reversed the FISH labelling scheme of the RNA transcripts; *i.e.,* SRSF1 motifs were labeled with AF647 and hnRNPA1 motifs were labeled with CF568 **(Fig. S4e-l)**. In addition, knockdown of a particular protein did not exclusively alter the positioning of that specific motif on the RNA; instead, RNA transcripts migrated inwards or outwards as an entity. In other words, the radial orientation of the RNAs remained same under partial knockdown of the proteins.

We extended our siRNA-mediated knockdown experiments to MUT_SRSF7. In contrast to MUT_SRSF1, SRSF1 knockdown only caused minor outward movement of RNA transcripts from MUT_SRSF7 construct. This is consistent with the fact that MUT_SRSF7's middle exon does not

contain any SRSF1 motifs and should thus be less sensitive to SRSF1 protein knockdown. On the other hand, as expected, hnRNPA1 knockdown still led to a similar magnitude of change in positioning of RNAs towards the speckle center in both constructs (**Fig. S7**). The minor outward migration of MUT_SRSF7 RNA transcripts might be explained by a slight downregulation (10±3%) of SRSF7 when knocking down SRSF1 (**Fig. S6**).

These observations collectively suggest that while the strength of RNA-RBP interactions drive intra-speckle RNA positioning, the RNA orientation remains robust under partial RBP knockdown.

Multivalent interactions with SRSF1 and hnRNPA1 proteins explain the localization of *MALAT1* to the outer shell of nuclear speckles

Using super-resolution imaging, we previously showed that *MALAT1* lncRNA localizes to the outer shell of nuclear speckles[45]. We therefore tested whether interactions between *MALAT1* and SRSF1/hnRNPA1 proteins can explain its intra-speckle localization. We computationally predicted a long SR motif-rich region in *MALAT1*, but failed to identify an hnRNP motif-rich region that is sufficiently long and free of splice site (**Fig. S8**). We therefore performed single-color SMLM imaging on *MALAT1* labeled with AF647-conjugated FISH probes spanning the entire transcript. We used cells treated with siRNAs against SRSF1, hnRNPA1, or scramble siRNA. Speckles were stained with antibody against SON protein. Cells with efficient protein knockdown were chosen as described above.

Consistent with the previous study, *MALAT1* in cells treated with scramble siRNA localized to the outer shell of nuclear speckles **(Fig. 5)**. Upon SRSF1 knockdown, *MALAT1* migrated to the periphery of the nuclear speckle, and formed distinct 'ring-like' structures around the speckles[45] **(Fig. 5)**. Conversely, when hnRNPA1 was downregulated, *MALAT1* migrated towards the center of the speckle **(Fig. 5)**. These observations strengthen our hypothesis that RNA-RBP interactions drive differential intra-speckle localization of RNA.

A toy model reproduces the intra-speckle RNA positioning and orientation

Finally, we computationally tested whether RNA-RBP interactions, or binding energies inside versus outside nuclear speckles are sufficient for explaining the observed intra-speckle RNA positioning and orientation. In this toy model, we created four lattice sites for the 2-block RNA molecule (corresponding to the SR and hnRNP motif-rich regions) to occupy (**Fig. 6a**), with two sites inside the speckle and two sites outside. We then considered all six possible RNA configurations. Specifically, the positions of the SRSF1 and hnRNPA1 motif-rich regions can be both inside, both outside, or straddle the speckle interface (one inside and one outside), each with two orientations (SRSF1 motif-rich region facing inside or hnRNPA1 motif-rich region facing inside). In each configuration, the RNA motifs can be bound or unbound by the corresponding protein, defined as four "binding states". The relative population of the binding states can be estimated by the dissociation constants (K_d) of each RNA-RBP pair, and the concentration of these RBPs in each location. The combination of the configurations and binding states generated 24 energy states in total (**Table S2**). The intra-speckle positioning of the two regions on the RNA was then estimated using a Boltzmann distribution.

Our imaging results demonstrated that hnRNPA1 proteins were relatively homogenously distributed throughout the nucleus, with slight depletion in some nuclear speckles (**Fig. S1b**). Therefore, we started with a model in which SRSF1 proteins were enriched in nuclear speckle, whereas hnRNAP1 proteins were uniformly distributed in the nucleoplasm. We estimated the nuclear speckle enrichment of SRSF1 using reported concentration[58,59] and our imaging data (**Supporting Text**). We approximated binding energies using *in vitro*-measured K_d (~100 nM)[60,61]. Our simulation recapitulated the differential intra-speckle positioning of the SRSF1 and hnRNPA1 motif-rich regions, even if hnRNPA1 was uniformly distributed (**Fig. S9a-c**). However, while this model recapitulated the effect of SRSF1 knockdown on the RNA positional change, it did not recapitulate the effect of hnRNPA1 knockdown. We then considered that the binding of these two proteins might be competitive, as illustrated in previous studies[62–64], and added an energy penalty term if both proteins bind at the same time (**Supporting Text**). This toy model including competition between the two proteins fully recapitulated the experimental observations (**Fig. 6b-d**).

We also considered the scenario in which SRSF1 proteins were still enriched in the speckles, whereas hnRNPA1 proteins were depleted in nuclear speckles. The model also recapitulated our experimental observations even if the binding of these two proteins were independent (no competitive binding) (**Fig. S9d-f**).

Finally, we also considered that the *in vivo* K_d of these proteins might be larger than the *in vitro* measured numbers due to competition from other cellular proteins. Increasing K_d by both 10- and 100-fold led to similar trend in intra-speckle RNA positioning and orientation (**Fig. S10**), suggesting that this RNA-RBP interaction-driven intra-speckle positioning might apply to a broad range of binding affinities.

Discussion

Nuclear speckles are prominent phase-separated membraneless organelles in eukaryotic nuclei. The formation and structural integrity of nuclear speckles relies on low complexity, intrinsically disordered regions of the scaffold proteins SON and SRRM2[42,43,45,65]. Additionally, in nuclear speckles, multivalent interactions between resident RNAs and proteins promote a layered organization of these components[45,46]. In this study, we further proposed an intra-speckle RNA positioning model (**Fig. 1**). We demonstrated that, due to multivalent interactions between RNA motifs and RBPs, RNA transcripts can exhibit preferential positioning and radial orientation inside speckles.

By performing SMLM on reporter constructs, we showed that RNA motifs on the same RNA transcript can have differential positioning, with an SR motif-rich region closer to the speckle center than an hnRNP motif-rich region. This differential intra-speckle RNA positioning is governed by interactions between RNA motifs with SR and hnRNP proteins that have different nuclear localizations relative to nuclear speckles (**Fig. 1**). Moreover, tuning the RNA-RBP interaction strengths changes relative RNA positioning: SRSF1 knockdown causes a more outward localization, whereas hnRNPA1 knockdown leads to inward migration of RNA transcripts, further supporting the intra-speckle RNA positioning model (**Fig. 1**). Finally, the model can also explain intra-speckle positioning of an endogenous nuclear retained lncRNA, *MALAT1*[45].

Using a toy model, we demonstrated that the binding energies for RNA-RBP interactions inside versus outside nuclear speckles were sufficient for explaining the observed intra-speckle RNA positioning and orientation, when there is a difference in concentration of the RBPs inside and outside the speckle, or a difference in the effective concentration due to competitive binding between the two RBPs. Interestingly, when using *in vitro* measured K_d of the SRSF1 and hnRNPA1 proteins (~100 nM), the changes in the intra-speckle RNA positioning is predicted to be minor, or much smaller than the experimentally measured difference, upon knocking down of these proteins (**Fig. 6b-d**). The minor response is as expected given the *in vivo* concentrations of these RBPs (on the order of tens of μM) and *in vitro* measured K_d, binding sites will still be mostly bound by the RBPs even with >90% knockdown efficiency. Indeed, the intra-speckle positioning of the RNA motifs became more sensitive when increasing the K_d (**Fig. S10**), more closely mimicking the experimentally measured values. These simulations indicate that RNA-RBP interactions might be considerably weaker compared to the *in vitro* measured values due to competition from other cellular proteins. While we only tested RNAs containing specific SR and hnRNP motifs, these simulations suggest that RNA containing regions enriched in binding motifs of other RBPs could exhibit similar intra-speckle positioning and orientation, when the pair of the RBPs demonstrate differential localization with respect to nuclear speckles.

Our findings suggest that RNA transcripts with multivalent interactions with RBPs inside and outside nuclear speckles may help maintain the structural integrity of nuclear speckles. The lncRNA *NEAT1* is known to act as a scaffold in the assembly of paraspeckles through appropriate RNA-RBP interactions[66–69]. Specifically, *NEAT1* forms a characteristic core-shell spheroidal structure with the 5'- and 3'- ends at the shell and the internal region at the core[70]. In contrast, the formation of nuclear speckles does not rely on any specific RNA species, such as *MALAT1*, but is instead initiated by assembly of RBPs like SRRM2 and SON[42,43,71]. However, RNAs can still act as a driving force for this assembly[29,72,73]. A recent study showed that depletion of nuclear RNAs leads to the loss of nuclear speckles and causes SON and SRRM2 to reorganize into a few large protein aggregates[74]. A separate report demonstrates that the lysine/serine/arginine rich (KSR) region of SRRM2, which drives speckle formation *via* phase separation, can form condensates *in vitro* in the presence of total RNA alone, without addition of molecular crowders[71]. These studies highlight that while any specific RNA species may be dispensable for nuclear speckle formation, RNA molecules are in general needed for the structural integrity of nuclear speckles. The requirement of RNA molecules is consistent with the observation that nuclear speckles tend to be associated with actively transcribed genes[31–33,75–80]. Interestingly, a recent work demonstrated that the MEG-3 protein serves as a Pickering agent to maintain an appropriate size distribution of P granules in *C. elegans*, by localizing to their surface and reducing surface tension[81]. Considering the implications from these studies, our observation that RNAs exhibit preferential intra-speckle positioning and orientation can possibly explain the importance of RNA molecules in maintaining structural integrity of nuclear speckles. Specifically, by favorable interactions between different RNA motifs and proteins in and out of nuclear speckles, the RNA molecules might be oriented similarly to amphiphilic polymers at the water-oil interface. Such positioning and orientation of speckle-residing RNAs might explain the roles of RNAs in preventing formation of protein-based aggregates by SON and SRRM2 and maintaining proper speckle morphology[74].

The intra-speckle positioning and orientation of the pre-mRNA might also help spatially coordinate or facilitate splicing reaction. SR and hnRNP proteins are important splicing regulators that help determine splicing outcomes[82–87]. Along with the difference in their localization in the nucleus, SR proteins and hnRNPs show antagonistic effects on splicing[88–90]. SR motifs tend to enhance splicing if placed in exons and repress splicing if placed in introns[91,92]. In contrast, hnRNP motifs tend to repress splicing in exons and enhance splicing in introns[92,93]. These splicing factors are known to affect splicing outcomes by recognizing splicing regulatory elements[94], but the mechanisms underlying their context-dependent actions are not well understood. It was previously hypothesized that the spatial localization of RNAs with respect to nuclear speckles might explain the context-dependent effect of SR and hnRNP proteins[95]. Specifically, the interactions of exonic SR motif-rich sequences with SR proteins concentrated in speckles and of intronic hnRNP motif-rich sequences with hnRNPs distributed in the surrounding nucleoplasm, position RNA transcripts at the outer shell of nuclear speckles with the exon towards inside and intron towards outside. This intra-speckle RNA positioning and orientation may provide better overlap with the spliceosomal components, which are also localized to the outer shell of the nuclear speckles[45], thereby favoring the splicing reaction. Our observation of intra-speckle RNA positioning and orientation supports this hypothesis. However, future experiments are needed to further demonstrate the correlation between intra-speckle RNA positioning with splicing outcomes.

Methods

Plasmid design and construction

Plasmid design was based on our earlier work[96,97]. Briefly, an SR motif-rich region consists of 15 repeats of an 8nt SR-binding sequence (either SRSF1 or SRSF7) separated by an 8nt reference sequence. The SR motif-rich region is flanked by strong 3' and 5' splice sites. Downstream of the 5' splice site, an hnRNP motif-rich region with 24 repeats was designed similarly, with the 8nt hnRNP A1 binding sequence chosen for affinity and specificity based on RNAcompete data[98].

Constructs WT_SRSF1, MUT_SRSF1, WT_SRSF7 and MUT_SRSF7 were generated in three steps. In the first step, the SR- and hnRNP motif-rich regions were generated separately producing two intermediate plasmids[97], by following a previously published PCR-free cloning approach[99]. Briefly, type IIS enzymes (BsaI, New England Biolabs #R3733, and BsmBI, New England Biolabs #R0739) were used to iteratively concatenate sequence modules. In the second step, the two intermediate plasmids were combined with a plasmid containing the 5' splice site using the same stepwise approach. Hence, an intermediate plasmid was obtained containing the SR motif-rich region followed by a 5' splice site and then the hnRNP motif-rich region. In the third step, the assembled sequences were transferred to the target plasmids (WT, MUT) by using a different set of type IIS enzymes (BbvI, New England Biolabs #R0173, and BfuAI, New England Biolabs #R0701). Importantly, these plasmids contain a tetracycline responsive promoter[97,100].

The non-repeat constructs WT_NR and MUT_NR were obtained from WT_SRSF1 and MUT_SRSF1 by mutating each nucleotide in the SR and hnRNP motif-rich regions with a probability of 37.5%. Doing so is likely to introduce unwanted splice site sequences, and to abolish too many of the RBP binding motifs. We therefore ranked *in silico* 10,000 candidate sequences

for each region (SR rich and hnRNP rich) and picked sequences that (1) contain no predicted splice site sequence[101,102], and (2) keep a similar enrichment of SR or hnRNP motifs as scored using our previous machine learning model (**Fig. S5b-c**)[103]. The final sequences contained over 30% mutations compared to the original ones. Using BLAST of the sequence against itself, we verified the absence of any residual repeats. Gene fragments were then synthesized using gBlocks™ (IDT, USA) and cloned into the same target plasmids.

Cell culture, transfection and drug treatments

HeLa Tet-On cells (TaKaRa) were cultured in high glucose (4.5 g/L) containing Dulbecco's Modified Eagle Medium (DMEM, Gibco) supplemented with 10% fetal bovine serum (FBS, Gibco), 1 mM sodium pyruvate (Gibco), 50 U/mL penicillin-streptomycin (Gibco). Cells were grown at 37 °C in a humidified environment containing 5% CO_2. For imaging, cells were seeded in an eight-well imaging chamber (#1.5 cover glass, Cellvis) and grown overnight to 70-80% confluency before transfection.

For transfecting each well, 0.6 µL of Lipofectamine 3000 reagent (Invitrogen) was diluted in pre-warmed 15 µL reduced serum minimum essential medium (opti-MEM, Gibco) and vortexed briefly. In another tube, 200 ng of plasmid DNA and 0.4 µL P3000 reagent (Invitrogen) were diluted in 15 µL pre warmed opti-MEM and vortexed briefly. The two solutions were mixed, vortexed briefly and incubated for 15 min at room temperature. The cell culture medium was replaced with pre-warmed DMEM containing 10% Tet system approved FBS (Tet-free medium, TaKaRa) and 30 µL of DNA-lipid complex was added to each well. The medium was replaced with fresh Tet-free medium 6-8 h after transfection and incubated overnight.

Transcription induction of the transfected construct was done 24 h after transfection using 2 µg/mL doxycycline (Santa Cruz # sc-204734B) in Tet-free medium. For samples where splicing inhibition was required, cells were treated with 100 nM Pladienolide B (Plad B, Cayman) in Tet-free medium for 4 h.

RT-PCR for biochemical assays

RNA was extracted 20 h after transfection using QIAgen RNeasy minikits (#74104) in a QIAcube following the manufacturer's protocol. DNA was removed using TURBO DNase (Thermofisher #AM2238) in a 30 uL reaction. RNA was then quantified using a Nano Drop One (Thermofisher #ND-ONE-W) and the concentration adjusted to 60 ng/uL. For reverse transcription, 200 ng of RNA were used for a 10 uL reaction using SuperScript IV (Thermofisher #18090010) according to the manufacturer's protocol. Appropriate RT primers (**Table S1**) were added at a final concentration of 100 nM. PCR reactions were carried out in a Veriti 96-well thermocycler (Applied Biosystems #4375305) using a Phusion High Fidelity kit (New England Biolabs # E0553L). The reverse transcription product was diluted 5-fold with water and 2 uL were used in a 25 uL PCR reaction according to the manufacturer's instructions. PCR was run for 21 or 22 cycles allowing 2 min for extension. The PCR product was run in 1.5 % agarose gels and quantified in a BioRad gel documentation system after post-staining with Ethidium Bromide and destaining.

siRNA-mediated knockdown

SRSF1 and hnRNPA1 knockdown was performed using double-stranded siRNAs against SRSF1 (hs.Ri.SRSF1.13.2, IDT, USA) and hnRNPA1 (hs.Ri.HNRNPA1.13.2, IDT, USA). A scrambled double-stranded siRNA (DsiRNA, IDT, USA) was used as negative control. Cells were seeded in an eight-well imaging chamber and grown to 60-70% confluency. For each well, 1.5 µL Lipofectamine RNAiMax reagent (Invitrogen) was diluted in 25 µL pre-warmed opti-MEM and vortexed briefly. In a separate tube, 0.5 µL siRNA (10 µM) was diluted in 25 µL pre-warmed opti-MEM and vortexed. The two solutions were then mixed, vortexed and incubated at room temperature for 5 min. 25 uL was then added to each well after replacing the cell culture medium with Tet-free medium. 6-8 h after transfection, fresh Tet-free medium was added. For the MALAT1 experiments, two rounds of siRNA mediated hnRNPA1 knockdown were done with a 24 h interval. For the experiments with constructs, single siRNA knockdown was done using Lipofectamine RNAiMax reagent followed by plasmid transfection using Lipofectamine 3000 with an interval of 24 h.

qPCR quantification for knockdown efficiency

HeLa Tet-On cells were grown in a 12-well plate and siRNA mediated knockdown was performed. Cells were collected 48 h after knockdown first knockdown. RNA extraction was done using the Qiagen RNeasy kit (Qiagen, #75144) following the provided protocol. cDNA was synthesized using an iScript cDNA synthesis kit (BioRad). 0.5-1 µg RNA template was used and the reaction was performed in a thermal cycler (Applied Biosystems) as follows: priming for 5 min at 25 °C, reverse transcription (RT) for 20 min at 46 °C, RT inactivation for 1 min at 95 °C and then held at 4 °C. For qPCR, 2 µL of cDNA was mixed with 2 µL of forward and reverse primers (2.5 µM each, **Table S1**,) and 1x SYBR Green Supermix (BioRad) for a final reaction volume of 20 µL. The qPCR reactions were performed using the CFX real-time PCR system (Bio-Rad) as follows: pre-incubation of 95 ºC for 30 s, followed by 40 cycles consisting of 95 ºC for 10 s and 60 ºC for 30 s. The reactions were then subjected to melting curve analysis: 95 ºC for 10 s, 65 ºC for 5 s followed by 0.5 ºC increments to 95 ºC for 5 s. The data was analysed with the BioRad CFX Maestro software.

Labelling of FISH probes and secondary antibodies

FISH probes were designed using the Stellaris Probe Designer and purchased from IDT, USA. Probes were 18-20 nucleotides long with a GC content between 45%-55%. The probes targeting motifs with SRSF1, SRSF7 and hnRNPA1 in WT_SRSF1, MUT_SRSF1, WT_SRSF7 and MUT_SRSF7 were purchased with 3' amine modification. For the other probes, amine modification was added using terminal transferase (TdT) enzymatic reaction[104]. For a 60 µL reaction volume, 40 µL of pooled oligonucleotides (100 µM) were mixed with 12 µL ddUTPs (1 mM, New England Biolabs), 2.4 µL TdT (20000 U/mL, New England Biolabs, #M0315L) in 1x TdT buffer (New England Biolabs) and incubated overnight at 37 °C in a PCR thermocycler (Applied Biosystems). The modified probes were purified using a P-6 Micro Bio-Spin Column (Bio-Rad).

For fluorophore conjugation, amine modified probes were dissolved in 0.1 M sodium bicarbonate (pH 8.5). Alexa Fluor (Invitrogen) and CF568 (Sigma Aldrich)-conjugated succinimidyl ester was dissolved in 0.5-4 µL DMSO and mixed with the probe solution. The dye: probe molar ratio was

25:1 approximately[45]. The labelling reaction was incubated overnight in dark at 37 °C. To quench the reaction, 1/9th reaction volume of 3M sodium acetate (pH 5) was added. Labelled probes were precipitated overnight with ethanol (~ 2.5 times the reaction volume) and then passed through a P-6 Micro Bio-Spin column to remove unconjugated free dye. The labelling efficiency of all probes were above 75%. The exact sequences of the probes are provided in **Table S1**.

Secondary antibodies against mouse (Jackson ImmunoResearch, #715-005-150) or rabbit (Jackson ImmunoResearch, #711-005-152) were labelled with Alexa Fluor succinimidyl ester. 24 µL antibody (1 mg/mL) was mixed with 3 µL 10x PBS and 3 µL sodium bicarbonate (1 M, pH 8.5). 0.001-0.003 mg of Alexa dye was added to the above solution and the reaction was incubated for 1 h at room temperature. Labelled antibody was purified using a P-6 Micro Bio-Spin column equilibrated with 1x PBS. 0.8 to 2.2 dye per antibody was typically achieved.

RNA FISH and immunostaining

RNA FISH and immunostaining were performed according to previously published protocol[105]. Cells were fixed with 4% paraformaldehyde (PFA, Electron Microscopy Sciences) in 1x PBS for 10 min at room temperature. Permeabilization was done with a solution containing 0.5% Triton X-100 (Thermo Scientific) and 2 mM vanadyl ribonucleoside complexes (Sigma-Aldrich, #R3380) in 1x PBS for 10 min on ice. Cells were washed 3 times with 1x PBS for at least 5 min after fixation and permeabilization. Cells were stored in 70% ethanol at 4 °C until hybridization with FISH probes. Cells were washed with 2x saline-sodium citrate (SSC) two times followed by a final wash with FISH wash solution (10% formamide (Ambion, #AM9342) in 2x SSC). 125 µL of hybridization buffer (FISH wash solution and 10% dextran sulphate (Sigma-Aldrich) containing 5 nM of each labelled probe and 10 mM dithiothreitol (Sigma-Aldrich) was added to each well of the imaging chamber. The hybridization reaction was incubated overnight at 37 °C in the dark. The following day, cells were washed with FISH was solution for 30 min at 37 °C.

To prevent dissociation of probes during immunostaining, cells were again fixed with 4% PFA for 10 min at room temperature. After washing with 1x PBS, cells were treated with blocking solution (0.1% ultrapure BSA (Invitrogen, #AM2618) in 1x PBS) for 30 min at room temperature. Solutions of primary antibodies were prepared in blocking solution using the following dilutions: mouse antibody against SRRM2 (1:2000, Sigma Aldrich, #S4045), mouse antibody against SRSF1 (1:250, Invitrogen, #32-4600), mouse antibody against hnRNPA1 (1:400, Invitrogen, #MA1-26736), rabbit antibody against SON (1:200, Invitrogen, #PA5-54814). 125 µL of the solution was added to each well and incubated at room temperature for 1 h. Cells were washed with 1x PBS three times with 5 min incubation each time. Labelled secondary antibodies were diluted 200-fold in blocking solution, 125 µL added to each well and incubated for 1 h at room temperature. Cells were washed with 1x PBS for 3 times with at least 5 min incubation time and stored in 4x SSC at 4 °C until imaging.

Imaging and image reconstruction

Diffraction limited epi imaging was performed using a Nikon TiE microscope with a CFI HP TIRF objective (100X, NA 1.49, Nikon), and an EMCCD (Andor, iXon Ultra 888). Imaging was performed using an imaging buffer containing Tris-HCl (50 mM, pH 8), 10% glucose, 2x SSC, glucose oxidase (0.5 mg/mL, Sigma-Aldrich) and catalase (67 µg/mL, Sigma-Aldrich). FISH

signals on the RNAs were imaged using the 647 nm laser (Cobolt MLD) and 561 nm laser (Coherent Obis). The immunofluorescence signal on nuclear speckle marker proteins was imaged using a 488 nm laser (Cobolt MLD). For the knockdown experiments, a 750 nm laser (Shanghai Dream Lasers Technology) was used to look at SRSF1 or hnRNPA1 protein levels stained with Alexa Fluor 750. Images were then processed in Fiji (ImageJ)[106] for further analysis.

2D-SMLM was performed on the same microscope, objective and EMCCD. Fluorescent TetraSpeck beads (0.1 µm, Invitrogen) were diluted 500-fold in 1x PBS, added to each well and incubated for 15 min at room temperature. After washing with 1x PBS to remove unattached beads, the same imaging buffer (as above) with additional 100 mM β-mercaptoethanol (BME, 14.3 M, Sigma-Aldrich) was added to the imaging chamber. For two-color STORM, movies were collected for the Alexa Fluor 647 and CF-568 channels sequentially using JOBS module in the NIS software. Briefly, the 647 nm (~ 40 mW) and 561 nm laser (~ 85 mW) were used to excite Alexa Fluor 647 and CF-568 fluorophore, respectively. A 405 nm laser (CL2000, Crystal Laser) was used for activation of fluorophores from 'off' to 'on' state. The acquisition was performed with 3 frames of 647 or 561 nm laser excitation followed by 1 frame of 405 nm laser excitation, using an exposure time of 42 ms. The laser power of 405 nm laser was adjusted during the acquisition to maintain a reasonable density of fluorophores in the 'on-state'. The maximum 405 nm laser power used with 647 and 561 lasers was ~2.2 mW and ~4 mW, respectively. A total of 15000 frames were recorded for each 647 and 561 channels. Before performing SMLM imaging on a selected cell, an epi image of the same cell with at least one bead present in the region of interest was taken for channel alignment.

SMLM image reconstruction was performed using the Thunderstorm[107] ImageJ plugin. For approximate localization of molecules, 'local maximum' method was used with the peak intensity threshold 2 times the standard deviation of the residual background. To determine sub-pixel localization of molecules, the Point Spread Function (Integrated Gaussian) method was using with fitting radius of 3 pixels (pixel size = 130 nm) and initial sigma as 1.6 pixels. The 'connectivity' was set to '8-neighborhood'. The images were then corrected for translational drift using the cross-correlation method and a bin size of 20-25. Finally, spots with xy-uncertainty more than 45 nm were filtered out. Images were then rendered with 5x magnification and lateral shifts 5.

Data analysis

A custom MATLAB code that we previously developed[45] was modified for radial distribution analysis on reconstructed SMLM images. Briefly, grayscale images were created from the mean intensity of all three fluorescence channels. Nuclear Speckles were identified by applying an appropriate intensity threshold on the grayscale image. Inappropriately fragmented nuclear speckles were removed from the final analysis by applying a size cutoff. Further processing was done on the 2D binary images by filling and opening binary operations to remove internal voids and shot noise. Each identified nuclear speckle was indexed in region of interest (ROI), and the geometric centroid of the mask served as the centre of each speckle. Additional thresholds on 2D-area and ellipticity were applied to discard abnormally large (fused) nuclear speckles and speckles that largely deviated from a spherical shape, respectively. An area cut-off of 5000 pixels (at the pixel size of 26 nm) and ellipticity cut-off of 0.8 worked best for our analysis. For calculating the distance of the RNA motifs to the center of the nuclear speckle, the normalized radial

distribution of intensity of each channel was calculated from the defined centre of the speckle. The mean distance of the 647 nm and 561 nm channels (reporting the RNA signals) was calculated for each nuclear speckle, normalized by the size of the speckle (intensity weighted average radius of SRRMS/SON signal), and represented as box plots. For calculating the distance of the RNA motifs to the edge of the nuclear speckle, MATLAB built-in function *bwboundaries* was used to trace the exterior boundaries of nuclear speckles. For each pixel, distance to the edge is defined as the distance between that pixel and the nearest pixel on the boundary. The same procedures as described above were performed to obtain the normalized radial distribution functions and box plots with respect to the edge of speckles.

Acknowledgements

This project was supported by the NIH Director's New Innovator Award (1DP2GM128185-01) to JF, a Simons Investigator Award and NSF MCB-2226731 to OR, and a Life Sciences Research Foundation Fellowship from Additional Ventures to SEL.

Author contributions

JF, OR, MAA and SEL conceived the project. SP, MAA, SEL and JZ performed the experiments. SP, LW, JZ and XW did the data analysis. LW and JF developed the toy model. SP, MAA, LW, OR and JF wrote the manuscript.

References

1. Dundr, M. Nuclear bodies: multifunctional companions of the genome. *Curr. Opin. Cell Biol.* **24**, 415–422 (2012).
2. Mao, Y. S., Zhang, B. & Spector, D. L. Biogenesis and function of nuclear bodies. *Trends Genet. TIG* **27**, 295–306 (2011).
3. Zhu, L. & Brangwynne, C. P. Nuclear bodies: the emerging biophysics of nucleoplasmic phases. *Curr. Opin. Cell Biol.* **34**, 23–30 (2015).
4. Buchan, J. R. & Parker, R. Eukaryotic stress granules: the ins and outs of translation. *Mol. Cell* **36**, 932–941 (2009).
5. Decker, C. J. & Parker, R. P-bodies and stress granules: possible roles in the control of translation and mRNA degradation. *Cold Spring Harb. Perspect. Biol.* **4**, a012286 (2012).
6. Orti, F., Navarro, A. M., Rabinovich, A., Wodak, S. J. & Marino-Buslje, C. Insight into membraneless organelles and their associated proteins: Drivers, Clients and Regulators. *Comput. Struct. Biotechnol. J.* **19**, 3964–3977 (2021).
7. Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).
8. Banani, S. F. *et al.* Compositional Control of Phase-Separated Cellular Bodies. *Cell* **166**, 651–663 (2016).
9. Hyman, A. A., Weber, C. A. & Jülicher, F. Liquid-liquid phase separation in biology. *Annu. Rev. Cell Dev. Biol.* **30**, 39–58 (2014).
10. Guo, Q., Shi, X. & Wang, X. RNA and liquid-liquid phase separation. *Non-Coding RNA Res.* **6**, 92–99 (2021).
11. Molliex, A. *et al.* Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* **163**, 123–133 (2015).
12. Lin, Y., Protter, D. S. W., Rosen, M. K. & Parker, R. Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. *Mol. Cell* **60**, 208–219 (2015).
13. Politz, J. C. R. *et al.* Rapid, diffusional shuttling of poly(A) RNA between nuclear speckles and the nucleoplasm. *Mol. Biol. Cell* **17**, 1239–1249 (2006).
14. Misteli, T., Cáceres, J. F. & Spector, D. L. The dynamics of a pre-mRNA splicing factor in living cells. *Nature* **387**, 523–527 (1997).
15. Ishihama, Y., Tadakuma, H., Tani, T. & Funatsu, T. The dynamics of pre-mRNAs and poly(A)+ RNA at speckles in living cells revealed by iFRAP studies. *Exp. Cell Res.* **314**, 748–762 (2008).
16. Kim, J., Han, K. Y., Khanna, N., Ha, T. & Belmont, A. S. Nuclear speckle fusion via long-range directional motion regulates speckle morphology after transcriptional inhibition. *J. Cell Sci.* **132**, (2019).
17. Sharp, P. A., Chakraborty, A. K., Henninger, J. E. & Young, R. A. RNA in formation and regulation of transcriptional condensates. *RNA N. Y. N* **28**, 52–57 (2022).

18. Drino, A. & Schaefer, M. R. RNAs, Phase Separation, and Membrane-Less Organelles: Are Post-Transcriptional Modifications Modulating Organelle Dynamics? *BioEssays News Rev. Mol. Cell. Dev. Biol.* **40**, e1800085 (2018).
19. Statello, L., Guo, C.-J., Chen, L.-L. & Huarte, M. Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.* **22**, 96–118 (2021).
20. Campos-Melo, D., Hawley, Z. C. E., Droppelmann, C. A. & Strong, M. J. The Integral Role of RNA in Stress Granule Formation and Function. *Front. Cell Dev. Biol.* **9**, 621779 (2021).
21. Rabouille, C. & Alberti, S. Cell adaptation upon stress: the emerging role of membrane-less compartments. *Curr. Opin. Cell Biol.* **47**, 34–42 (2017).
22. Fu, X. D. & Maniatis, T. Factor required for mammalian spliceosome assembly is localized to discrete regions in the nucleus. *Nature* **343**, 437–441 (1990).
23. Spector, D. L. & Lamond, A. I. Nuclear Speckles. *Cold Spring Harb. Perspect. Biol.* **3**, a000646 (2011).
24. Chen, Y. & Belmont, A. S. Genome organization around nuclear speckles. *Curr. Opin. Genet. Dev.* **55**, 91–99 (2019).
25. Hutchinson, J. N. *et al.* A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics* **8**, 39 (2007).
26. Shepard, P. J. & Hertel, K. J. The SR protein family. *Genome Biol.* **10**, 242 (2009).
27. Ilık, İ. A. & Aktaş, T. Nuclear speckles: dynamic hubs of gene expression regulation. *FEBS J.* (2021) doi:10.1111/febs.16117.
28. Fakan, S. & van Driel, R. The perichromatin region: a functional compartment in the nucleus that determines large-scale chromatin folding. *Semin. Cell Dev. Biol.* **18**, 676–681 (2007).
29. Faber, G. P., Nadav-Eliyahu, S. & Shav-Tal, Y. Nuclear speckles - a driving force in gene expression. *J. Cell Sci.* **135**, jcs259594 (2022).
30. Hall, L. L., Smith, K. P., Byron, M. & Lawrence, J. B. The Molecular Anatomy of a Speckle. *Anat. Rec. A. Discov. Mol. Cell. Evol. Biol.* **288**, 664–675 (2006).
31. Brown, J. M. *et al.* Association between active genes occurs at nuclear speckles and is modulated by chromatin environment. *J. Cell Biol.* **182**, 1083–1097 (2008).
32. Quinodoz, S. A. *et al.* RNA promotes the formation of spatial compartments in the nucleus. *Cell* S0092-8674(21)01230–7 (2021) doi:10.1016/j.cell.2021.10.014.
33. Chen, Y. *et al.* Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. *J. Cell Biol.* **217**, 4025–4048 (2018).
34. Ha, M. Transcription boosting by nuclear speckles. *Nat. Rev. Mol. Cell Biol.* **21**, 64–65 (2020).
35. Kim, J., Venkata, N. C., Hernandez Gonzalez, G. A., Khanna, N. & Belmont, A. S. Gene expression amplification by nuclear speckle association. *J. Cell Biol.* **219**, (2020).
36. Brody, Y. *et al.* The in vivo kinetics of RNA polymerase II elongation during co-transcriptional splicing. *PLoS Biol.* **9**, e1000573 (2011).

37. Galganski, L., Urbanek, M. O. & Krzyzosiak, W. J. Nuclear speckles: molecular organization, biological function and role in disease. *Nucleic Acids Res.* **45**, 10350–10368 (2017).
38. Girard, C. *et al.* Post-transcriptional spliceosomes are retained in nuclear speckles until splicing completion. *Nat. Commun.* **3**, 994 (2012).
39. Johnson, C. *et al.* 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 (2000).
40. Dias, A. P., Dufu, K., Lei, H. & Reed, R. A role for TREX components in the release of spliced mRNA from nuclear speckle domains. *Nat. Commun.* **1**, 97 (2010).
41. Hu, S., Lv, P., Yan, Z. & Wen, B. Disruption of nuclear speckles reduces chromatin interactions in active compartments. *Epigenetics Chromatin* **12**, 43 (2019).
42. Sharma, A., Takata, H., Shibahara, K., Bubulya, A. & Bubulya, P. A. Son is essential for nuclear speckle organization and cell cycle progression. *Mol. Biol. Cell* **21**, 650–663 (2010).
43. Ilik, İ. A. *et al.* SON and SRRM2 are essential for nuclear speckle formation. *eLife* **9**, e60579 (2020).
44. Kato, M. *et al.* Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753–767 (2012).
45. Fei, J. *et al.* Quantitative analysis of multilayer organization of proteins and RNA in nuclear speckles at super resolution. *J. Cell Sci.* **130**, 4180–4192 (2017).
46. Rivera, C. *et al.* Revealing RCOR2 as a regulatory component of nuclear speckles. *Epigenetics Chromatin* **14**, 51 (2021).
47. Neugebauer, K. M. & Roth, M. B. Distribution of pre-mRNA splicing factors at sites of RNA polymerase II transcription. *Genes Dev.* **11**, 1148–1159 (1997).
48. Tripathi, V. *et al.* SRSF1 regulates the assembly of pre-mRNA processing factors in nuclear speckles. *Mol. Biol. Cell* **23**, 3694–3706 (2012).
49. Saitoh, N. *et al.* Proteomic analysis of interchromatin granule clusters. *Mol. Biol. Cell* **15**, 3876–3890 (2004).
50. Mintz, P. J., Patterson, S. D., Neuwald, A. F., Spahr, C. S. & Spector, D. L. Purification and biochemical characterization of interchromatin granule clusters. *EMBO J.* **18**, 4308–4320 (1999).
51. Dopie, J., Sweredoski, M. J., Moradian, A. & Belmont, A. S. Tyramide signal amplification mass spectrometry (TSA-MS) ratio identifies nuclear speckle proteins. *J. Cell Biol.* **219**, (2020).
52. Thompson, M. G. *et al.* Viral-induced alternative splicing of host genes promotes influenza replication. *eLife* **9**, e55500 (2020).
53. Mattern, K. A., van der Kraan, I., Schul, W., de Jong, L. & van Driel, R. Spatial organization of four hnRNP proteins in relation to sites of transcription, to nuclear speckles, and to each

other in interphase nuclei and nuclear matrices of HeLa cells. *Exp. Cell Res.* **246**, 461–470 (1999).

54. Barutcu, A. R. *et al.* Systematic mapping of nuclear domain-associated transcripts reveals speckles and lamina as hubs of functionally distinct retained introns. *Mol. Cell* **82**, 1035-1052.e9 (2022).
55. Das, R. *et al.* Functional coupling of RNAP II transcription to spliceosome assembly. *Genes Dev.* **20**, 1100–1109 (2006).
56. Gozani, O., Patton, J. G. & Reed, R. A novel set of spliceosome-associated proteins and the essential splicing factor PSF bind stably to pre-mRNA prior to catalytic step II of the splicing reaction. *EMBO J.* **13**, 3356–3367 (1994).
57. Martínez-Montiel, N., Rosas-Murrieta, N. H., Martínez-Montiel, M., Gaspariano-Cholula, M. P. & Martínez-Contreras, R. D. Microbial and Natural Metabolites That Inhibit Splicing: A Powerful Alternative for Cancer Treatment. *BioMed Res. Int.* **2016**, 3681094 (2016).
58. Maharana, S. *et al.* RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science* **360**, 918–921 (2018).
59. Hein, M. Y. *et al.* A human interactome in three quantitative dimensions organized by stoichiometries and abundances. *Cell* **163**, 712–723 (2015).
60. Cho, S. *et al.* Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 8233–8238 (2011).
61. Burd, C. G. & Dreyfuss, G. RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. *EMBO J.* **13**, 1197–1204 (1994).
62. Cléry, A. *et al.* Isolated pseudo–RNA-recognition motifs of SR proteins can regulate splicing using a noncanonical mode of RNA recognition. *Proc. Natl. Acad. Sci.* **110**, E2802–E2811 (2013).
63. Zahler, A. M., Damgaard, C. K., Kjems, J. & Caputi, M. SC35 and heterogeneous nuclear ribonucleoprotein A/B proteins bind to a juxtaposed exonic splicing enhancer/exonic splicing silencer element to regulate HIV-1 tat exon 2 splicing. *J. Biol. Chem.* **279**, 10077–10084 (2004).
64. Eperon, I. C. *et al.* Selection of alternative 5’ splice sites: role of U1 snRNP and models for the antagonistic effects of SF2/ASF and hnRNP A1. *Mol. Cell. Biol.* **20**, 8303–8318 (2000).
65. Greig, J. A. *et al.* Arginine-Enriched Mixed-Charge Domains Provide Cohesion for Nuclear Speckle Condensation. *Mol. Cell* **77**, 1237-1250.e4 (2020).
66. Fox, A. H., Nakagawa, S., Hirose, T. & Bond, C. S. Paraspeckles: Where Long Noncoding RNA Meets Phase Separation. *Trends Biochem. Sci.* **43**, 124–135 (2018).
67. Sasaki, Y. T. F., Ideue, T., Sano, M., Mituyama, T. & Hirose, T. MENepsilon/beta noncoding RNAs are essential for structural integrity of nuclear paraspeckles. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 2525–2530 (2009).

68. Clemson, C. M. *et al.* An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Mol. Cell* **33**, 717–726 (2009).
69. Sunwoo, H. *et al.* MEN epsilon/beta nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. *Genome Res.* **19**, 347–359 (2009).
70. West, J. A. *et al.* Structural, super-resolution microscopy analysis of paraspeckle nuclear body organization. *J. Cell Biol.* **214**, 817–830 (2016).
71. Xu, S. *et al.* SRRM2 organizes splicing condensates to regulate alternative splicing. *Nucleic Acids Res.* **50**, 8599–8614 (2022).
72. Enukashvily, N. I., Dobrynin, M. A. & Chubar, A. V. RNA-seeded membraneless bodies: Role of tandemly repeated RNA. *Adv. Protein Chem. Struct. Biol.* **126**, 151–193 (2021).
73. Bhat, P., Honson, D. & Guttman, M. Nuclear compartmentalization as a mechanism of quantitative control of gene expression. *Nat. Rev. Mol. Cell Biol.* **22**, 653–670 (2021).
74. Decker, C. J., Burke, J. M., Mulvaney, P. K. & Parker, R. RNA is required for the integrity of multiple nuclear and cytoplasmic membrane-less RNP granules. *EMBO J.* **41**, e110137 (2022).
75. Khanna, N., Hu, Y. & Belmont, A. S. HSP70 transgene directed motion to nuclear speckles facilitates heat shock activation. *Curr. Biol. CB* **24**, 1138–1144 (2014).
76. Hu, Y., Plutz, M. & Belmont, A. S. Hsp70 gene association with nuclear speckles is Hsp70 promoter specific. *J. Cell Biol.* **191**, 711–719 (2010).
77. Bahar Halpern, K. *et al.* Nuclear Retention of mRNA in Mammalian Tissues. *Cell Rep.* **13**, 2653–2662 (2015).
78. Han, J., Xiong, J., Wang, D. & Fu, X.-D. Pre-mRNA splicing: where and when in the nucleus. *Trends Cell Biol.* **21**, 336–343 (2011).
79. Moen, P. T. *et al.* Repositioning of Muscle-specific Genes Relative to the Periphery of SC-35 Domains during Skeletal Myogenesis. *Mol. Biol. Cell* **15**, 197–206 (2004).
80. Yang, L. *et al.* ncRNA- and Pc2 methylation-dependent gene relocation between nuclear structures mediates gene activation programs. *Cell* **147**, 773–788 (2011).
81. Folkmann, A. W., Putnam, A., Lee, C. F. & Seydoux, G. Regulation of biomolecular condensates by interfacial protein clusters. *Science* **373**, 1218–1224 (2021).
82. Zhu, J., Mayeda, A. & Krainer, A. R. Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. *Mol. Cell* **8**, 1351–1361 (2001).
83. Zahler, A. M., Lane, W. S., Stolk, J. A. & Roth, M. B. SR proteins: a conserved family of pre-mRNA splicing factors. *Genes Dev.* **6**, 837–847 (1992).
84. Geuens, T., Bouhy, D. & Timmerman, V. The hnRNP family: insights into their role in health and disease. *Hum. Genet.* **135**, 851–867 (2016).
85. Matera, A. G. & Wang, Z. A day in the life of the spliceosome. *Nat. Rev. Mol. Cell Biol.* **15**, 108–121 (2014).

86. Hertel, K. J. Combinatorial control of exon recognition. *J. Biol. Chem.* **283**, 1211–1215 (2008).
87. Wahl, M. C., Will, C. L. & Lührmann, R. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**, 701–718 (2009).
88. Erkelenz, S. *et al.* Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms. *RNA N. Y. N* **19**, 96–102 (2013).
89. Jobbins, A. M. *et al.* The mechanisms of a mammalian splicing enhancer. *Nucleic Acids Res.* **46**, 2145–2158 (2018).
90. Busch, A. & Hertel, K. J. Evolution of SR protein and hnRNP splicing regulatory factors. *Wiley Interdiscip. Rev. RNA* **3**, 1–12 (2012).
91. Zhou, Z. & Fu, X.-D. Regulation of splicing by SR proteins and SR protein-specific kinases. *Chromosoma* **122**, 191–207 (2013).
92. Fu, X.-D. & Ares, M. Context-dependent control of alternative splicing by RNA-binding proteins. *Nat. Rev. Genet.* **15**, 689–701 (2014).
93. Smith, C. W. & Valcárcel, J. Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem. Sci.* **25**, 381–388 (2000).
94. Fairbrother, W. G., Yeh, R.-F., Sharp, P. A. & Burge, C. B. Predictive identification of exonic splicing enhancers in human genes. *Science* **297**, 1007–1013 (2002).
95. Liao, S. E. & Regev, O. Splicing at the phase-separated nuclear speckle interface: a model. *Nucleic Acids Res.* **49**, 636–645 (2021).
96. Zhang, X. H.-F., Arias, M. A., Ke, S. & Chasin, L. A. Splicing of designer exons reveals unexpected complexity in pre-mRNA splicing. *RNA N. Y. N* **15**, 367–376 (2009).
97. Arias, M. A., Lubkin, A. & Chasin, L. A. Splicing of designer exons informs a biophysical model for exon definition. *RNA* **21**, 213–229 (2015).
98. Ray, D. *et al.* Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. *Nat. Biotechnol.* **27**, 667–670 (2009).
99. Scior, A., Preissler, S., Koch, M. & Deuerling, E. Directed PCR-free engineering of highly repetitive DNA sequences. *BMC Biotechnol.* **11**, 87 (2011).
100. Gossen, M. & Bujard, H. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 5547–5551 (1992).
101. Yeo, G. & Burge, C. B. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol. J. Comput. Mol. Cell Biol.* **11**, 377–394 (2004).
102. Wong, M. S., Kinney, J. B. & Krainer, A. R. Quantitative Activity Profile and Context Dependence of All Human 5' Splice Sites. *Mol. Cell* **71**, 1012-1026.e3 (2018).
103. Liao, S. E., Sudarshan, M. & Regev, O. Machine learning for discovery: deciphering RNA splicing logic. *bioRxiv* 2022.10.01.510472 (2022) doi:10.1101/2022.10.01.510472.
104. Gaspar, I., Wippich, F. & Ephrussi, A. Enzymatic production of single-molecule FISH and RNA capture probes. *RNA N. Y. N* **23**, 1582–1591 (2017).

105. Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5**, 877–879 (2008).
106. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
107. Ovesný, M., Křížek, P., Borkovec, J., Svindrych, Z. & Hagen, G. M. ThunderSTORM: a comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging. *Bioinforma. Oxf. Engl.* **30**, 2389–2390 (2014).

Figures.

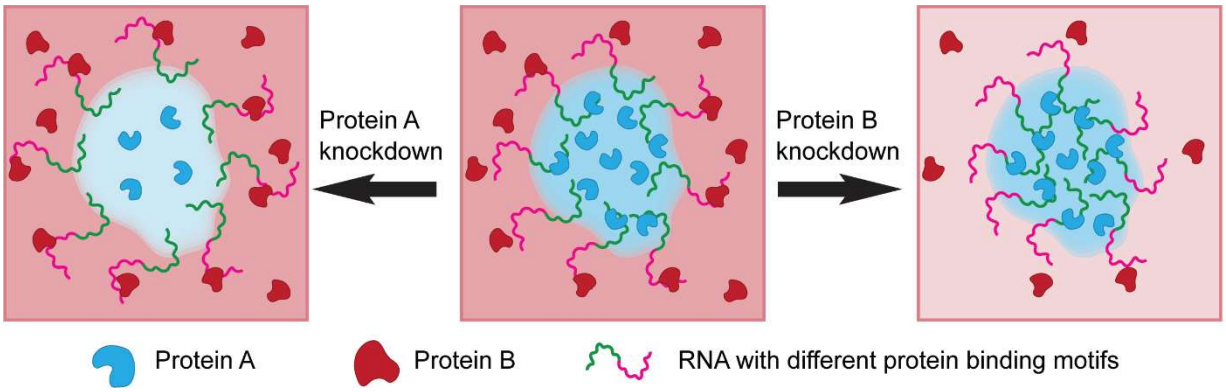

Figure 1. Intra-speckle RNA positioning model. Nuclear proteins show differential localization with respect to nuclear speckles, such that protein A is enriched in speckles and protein B is distributed in the nucleoplasm. As a result, the RNAs containing RBP binding motifs for both proteins A and B will be driven to the outer shell of the nuclear speckles. In addition, the region of RNA interacting with protein A will be positioned relatively closer to the center of nuclear speckles and the region interacting with protein B will be positioned relatively towards the periphery of speckles. Knocking down protein A will lead to outward migration of RNA transcripts (left) and knocking down protein B will lead to inward migration of the RNA (right) due to change in the strength of RNA-RBP interactions. In the present study, protein A is SRSF1, and protein B is hnRNPA1.

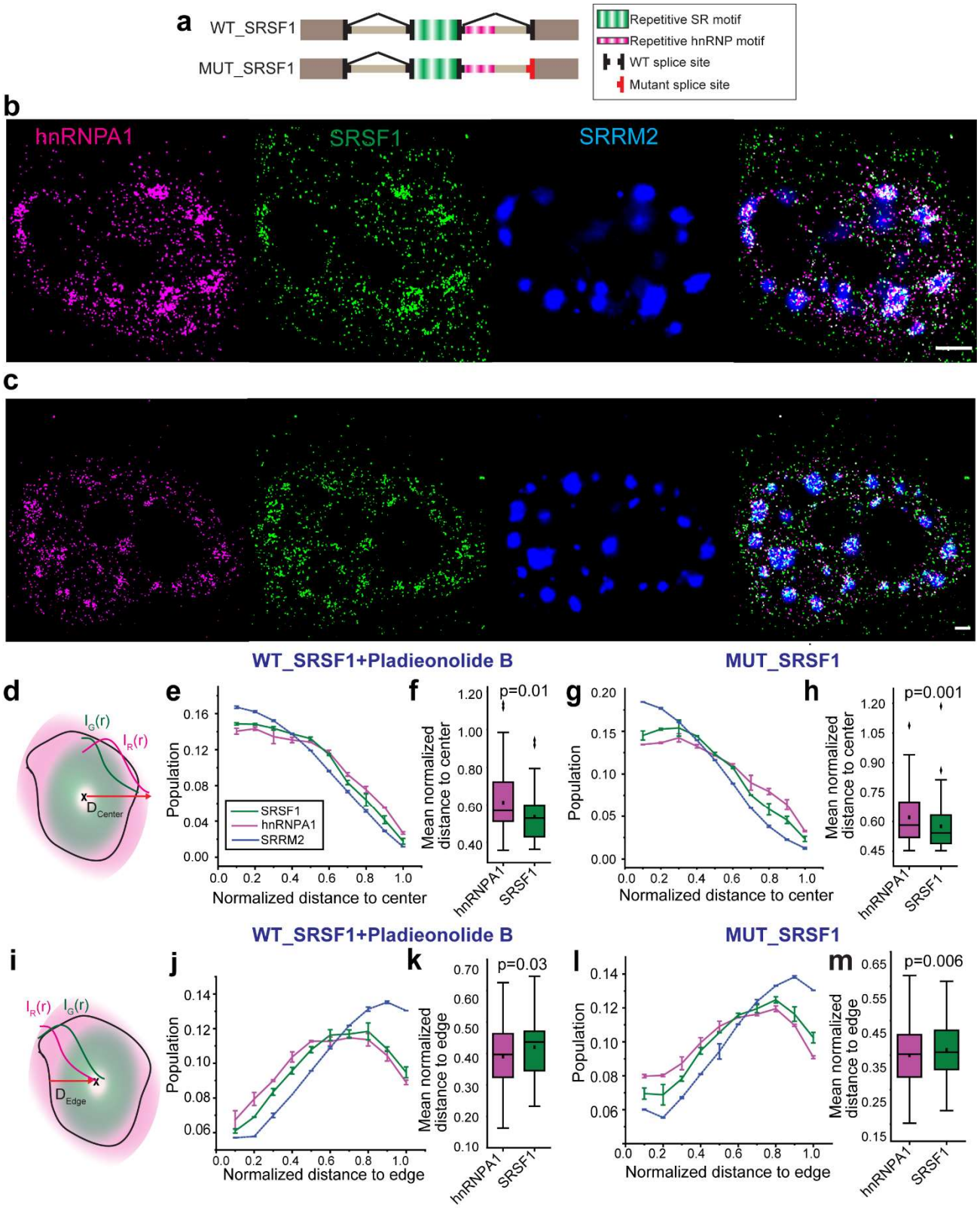

Figure 2. SMLM imaging and analysis of intra-speckle positioning of pre-mRNA containing SRSF1 motifs in exon and hnRNPA1 motifs in intron. (a) Schematic illustration of WT_SRSF1

and MUT_SRSF1 constructs. (b) Representative image of WT_SRSF1 in the presence of Pladienolide B. (c) Representative image of MUT_SRSF1 in the absence of Pladienolide B. FISH signals corresponding to hnRNPA1 (labeled with AF647) and SRSF1 (labeled with CF568) motifs in the RNAs are shown in magenta and green respectively. Immunostaining of SRRM2 is shown in blue. Scale bars represent 5 μm. (d) Calculation of the distribution of FISH signal as a function of the distance to the center of the nuclear speckle. Due to size differences among nuclear speckles, distances are all normalized from 0 (center) to 1 (edge) to build the overlaid distribution. Population distribution of SRSF1 and hnRNPA1 motif signals as a function of the normalized distance to the center of the speckle for WT_SRSF1 in the presence of Pladienolide B (e), and MUT_SRSF1 in the absence of Pladienolide B (g). Box plot of the population-weighted mean normalized distance of SRSF1 and hnRNPA1 signal to the center of speckle for each speckle for WT_SRSF1 in the presence of Pladienolide B (f), and MUT_SRSF1 in the absence of Pladienolide B (h). (i) Calculation of the distribution of FISH signal as a function of the distance to the edge of the nuclear speckle. The distances are all normalized from 0 (edge) to 1 (center) to build the overlaid distribution. Population distribution of SRSF1 and hnRNPA1 motif signals as a function of the normalized distance to the edge of the speckle for WT_SRSF1 in the presence of Pladienolide B (j), and MUT_SRSF1 in the absence of Pladienolide B (l). Box plot of the population-weighted mean normalized distance of SRSF1 and hnRNPA1 signal to the edge of speckle for each speckle for WT_SRSF1 in the presence of Pladienolide B (k), and MUT_SRSF1 in the absence of Pladienolide B (m). Error bars in the population vs. distance plots report the standard deviation from three replicates, each containing at least 60-90 nuclear speckles from 4-6 cells. p-values in the box plots are calculated with paired sample Wilcoxon signed rank test. Description of box plots: center line reports the median; dot inside box reports the mean; box limits are upper and lower quartiles; whiskers are 1.5x interquartile range; and points outside box are outliers.

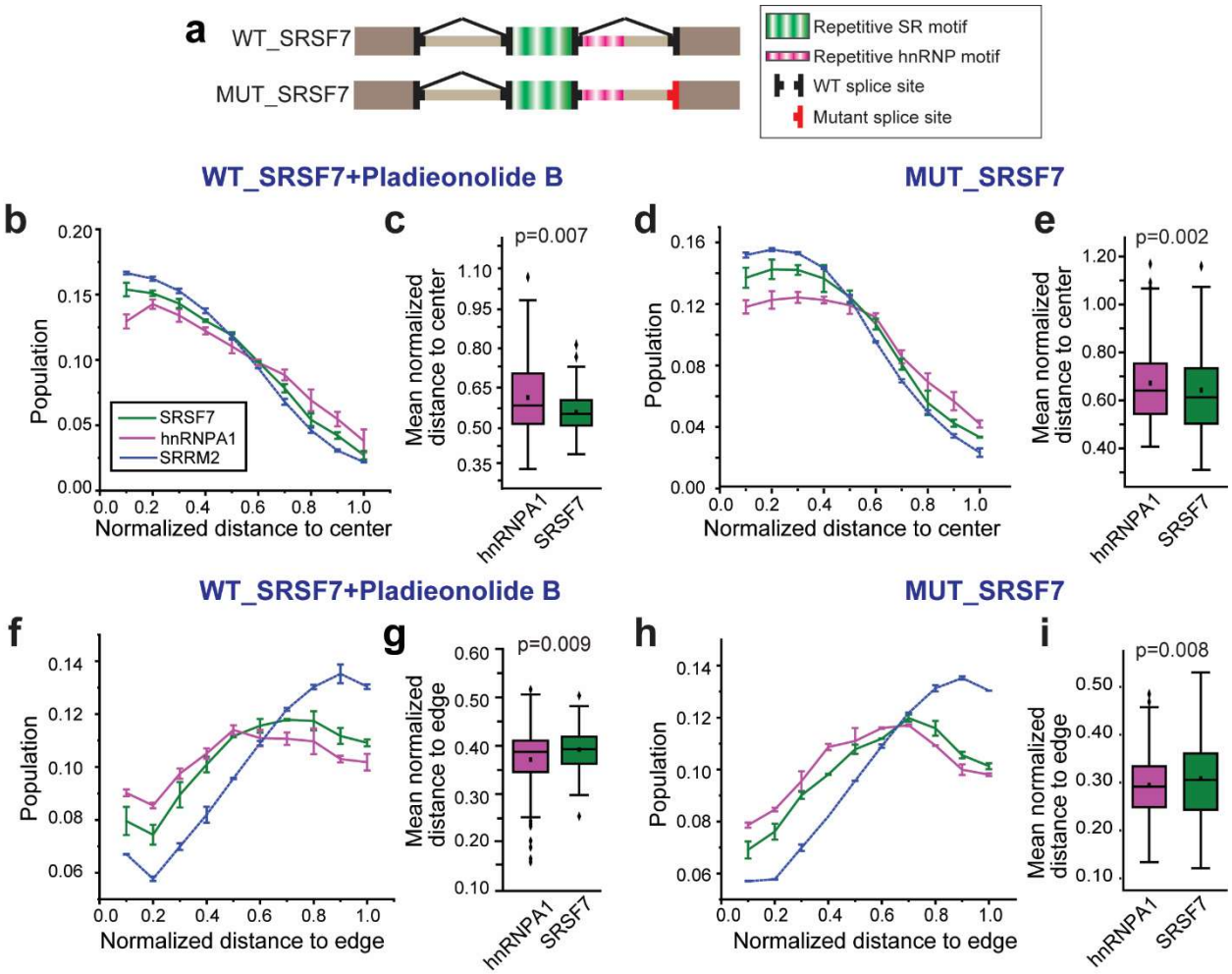

Figure 3. Intra-speckle positioning of pre-mRNA containing SRSF7 motifs in exon and hnRNPA1 motifs in intron. (a) Schematic illustration of WT_SRSF7 and MUT_SRSF7 constructs. Population distribution of SRSF7 and hnRNPA1 motif signals as a function of the normalized distance to the center of the speckle for WT_SRSF7 in the presence of Pladienolide B (b), and MUT_SRSF7 in the absence of Pladienolide B (d). Box plot of the population-weighted mean normalized distance of SRSF7 and hnRNPA1 signal to the center of speckle for each speckle for WT_SRSF7 in the presence of Pladienolide B (c), and MUT_SRSF7 in the absence of Pladienolide B (e). Population distribution of SRSF7 and hnRNPA1 motif signals as a function of the normalized distance to the edge of the speckle for WT_SRSF7 in the presence of Pladienolide B (f), and MUT_SRSF7 in the absence of Pladienolide B (h). Box plot of the population-weighted mean normalized distance of SRSF7 and hnRNPA1 signal to the edge of speckle for each speckle for WT_SRSF7 in the presence of Pladienolide B (g), and MUT_SRSF7 in the absence of Pladienolide B (i). Error bars in the population vs. distance plots report the standard deviation from three replicates, each containing at least 48-72 nuclear speckles from 4-6 cells. p-values in the box plots are calculated with paired sample Wilcoxon signed rank test. Description of box plots is the same as in Fig. 2.

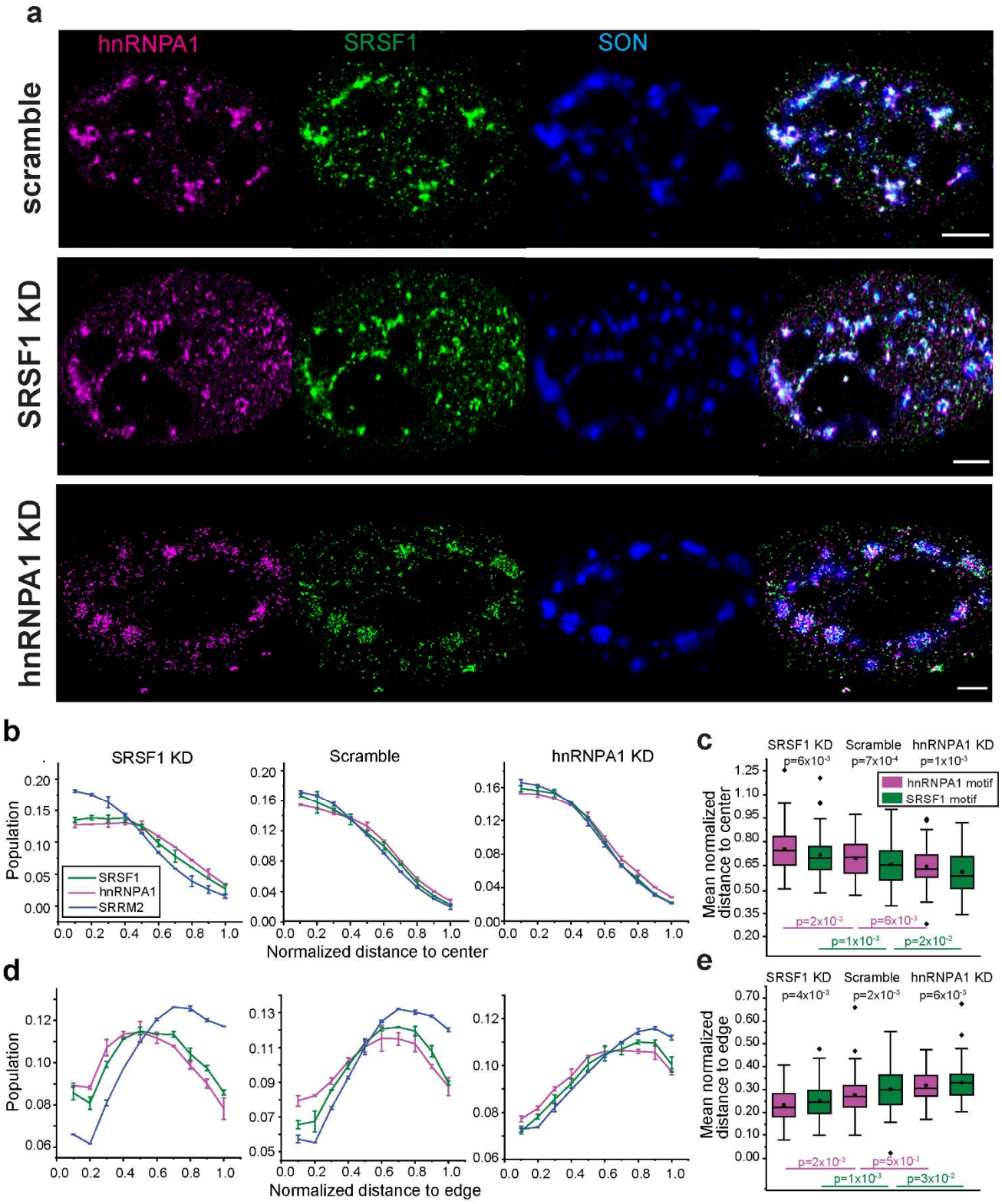

Figure 4. Effect of SRSF1 and hnRNPA1 knockdown on the intra-speckle positioning of pre-mRNA containing SRSF1 motifs in exon and hnRNPA1 motifs in intron. (a) Representative image of MUT_SRSF1 treated with scramble siRNA, siRNA against SRSF1 and siRNA against

hnRNPA1. FISH signals corresponding to hnRNPA1 and SRSF1 motifs in the RNAs are shown in magenta and green respectively. Immunostaining of SON is shown in blue. Scale bars represent 5 μm. Population distribution of SRSF7 and hnRNPA1 motif signals as a function of the normalized distance to the center (b), and to the edge (d) of the speckle. Box plot of the population-weighted mean normalized distance of SRSF7 and hnRNPA1 signal to the center (c), and to the edge (e) for each speckle for MUT_SRSF1. Error bars in the population vs. distance plots report the standard deviation from three replicates, each containing at least 75-105 nuclear speckles from 5-7 cells. p-values in the box plots are calculated with paired sample Wilcoxon signed rank test (black) and two sample t-test (magenta and green). Description of box plots is the same as in Fig. 2.

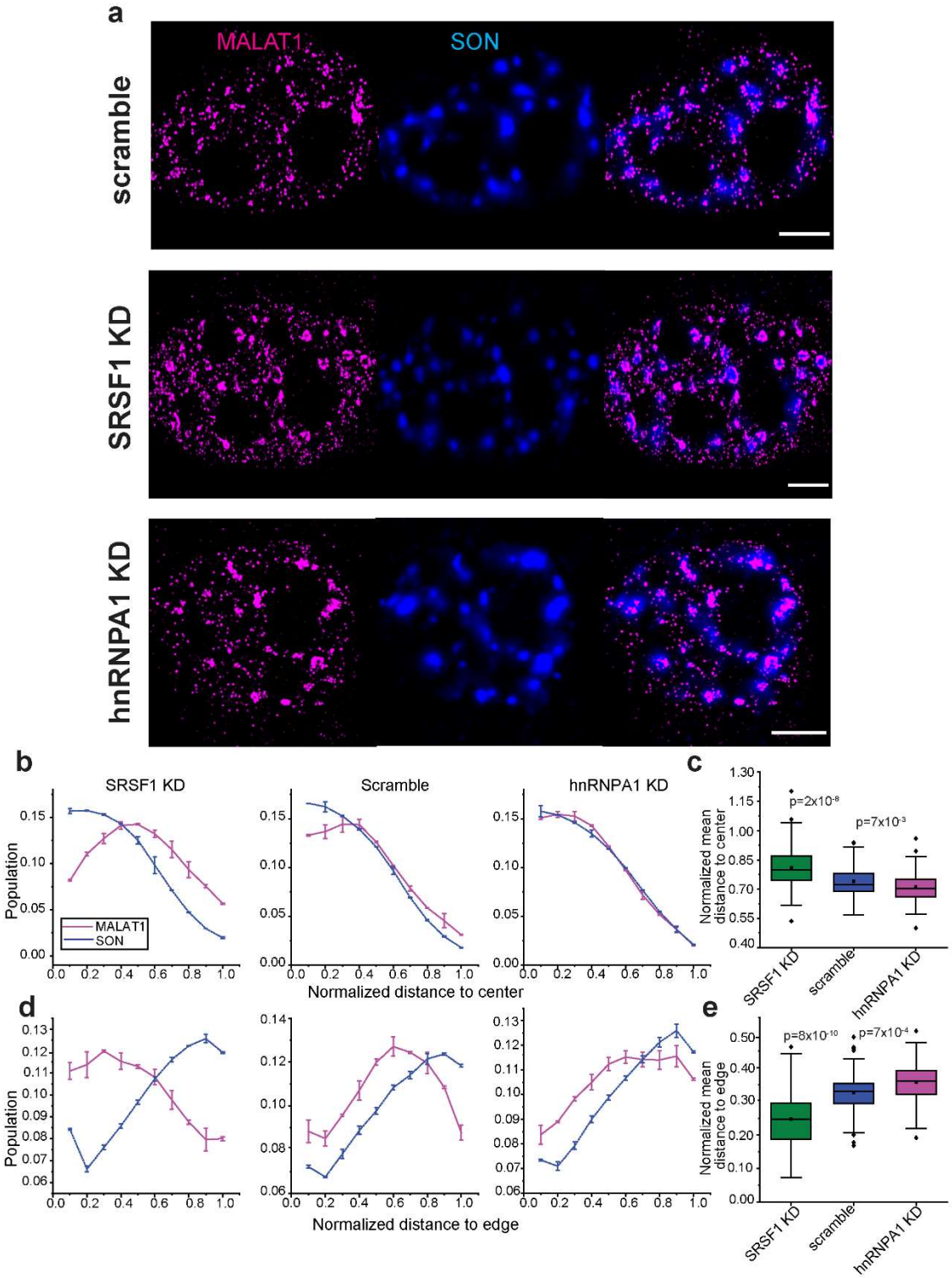

Figure 5. Effect of SRSF1 and hnRNPA1 knockdown on the intra-speckle positioning of *MALAT1*. (a) Representative image of *MALAT1* treated with scramble siRNA, siRNA against SRSF1, and siRNA against hnRNPA1. FISH signal from *MALAT1* is shown in magenta. Immunostaining of SON is shown in blue. Scale bars represent 5 μm. Population distribution of *MALAT1* signal as a

function of the normalized distance to the center (b), and to the edge (d) of the speckle. Box plot of the population-weighted mean normalized distance of *MALAT1* signal to the center (c), and to the edge (e) of the speckle for each speckle. Error bars in the population vs. distance plots report the standard deviation from two replicates, each containing at least 100-140 nuclear speckles from 5-7 cells. p-values in the box plots are calculated with two sample t-test. Description of box plots is the same quantities as in Fig. 2.

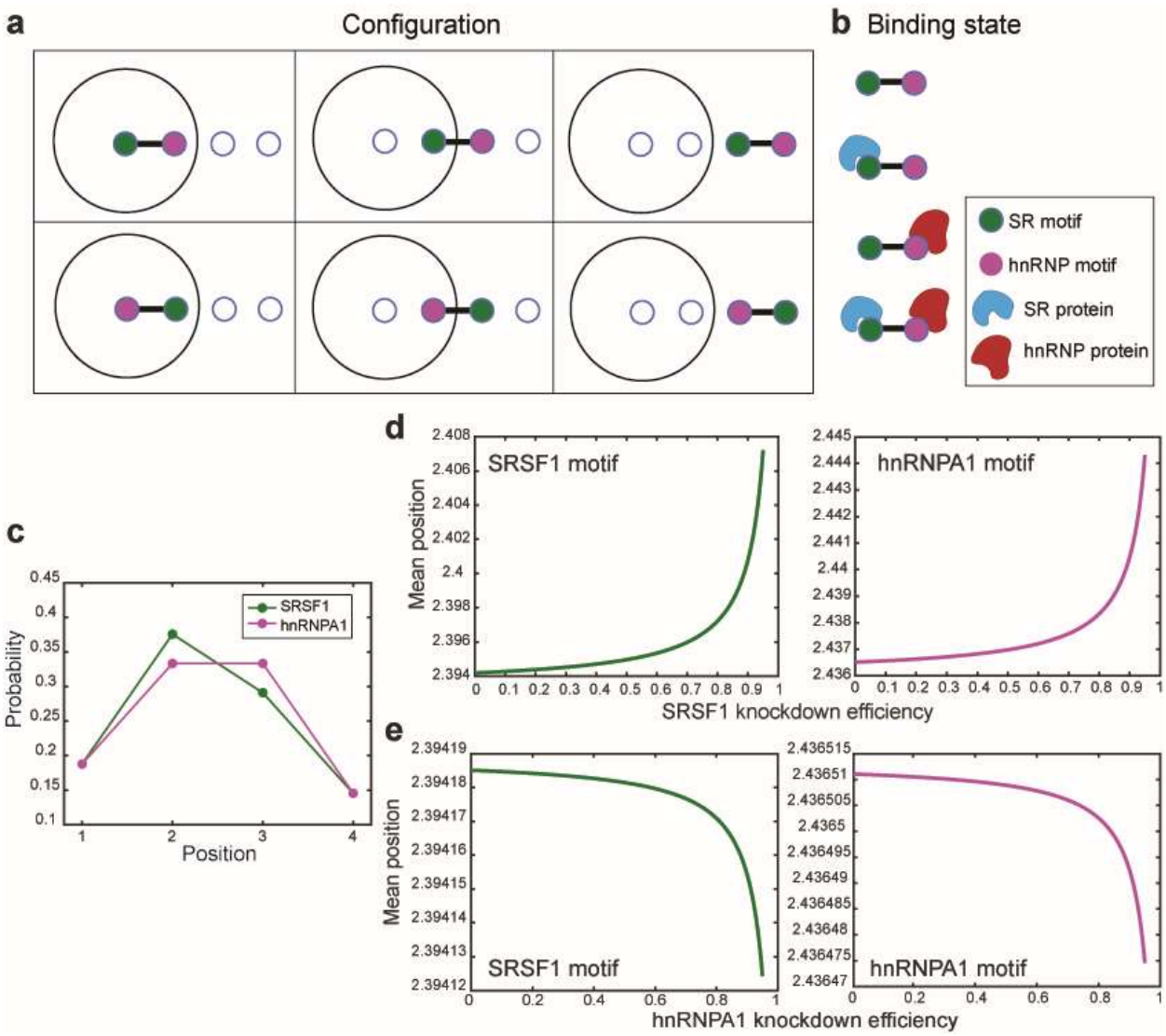

Figure 6. A toy model taking into account of competitive binding between SR and hnRNP proteins recapitulated the intra-speckle positioning of SRSF1 and hnRNPA1 motif-rich region of the RNA. (a) Graphical representation of the 6 configurations (3 positions and 2 orientations) of the RNA molecule. (b) Graphical representation of 4 binding states corresponding to SRSF1 and hnRNPA1 motif being either bound or unbound by the corresponding RBPs. (c) Probability distribution of position of the SRSF1 and hnRNPA1 motif-rich regions, as predicted by a toy model taking into account competition between SRSF1 and hnRNPA1 proteins as described in Supporting Text. Mean positions of SRSF1 and hnRNPA1 motif determined by this model are plotted as a function of both SRSF1 and hnRNPA1 knockdown efficiency in (d) and (e).