

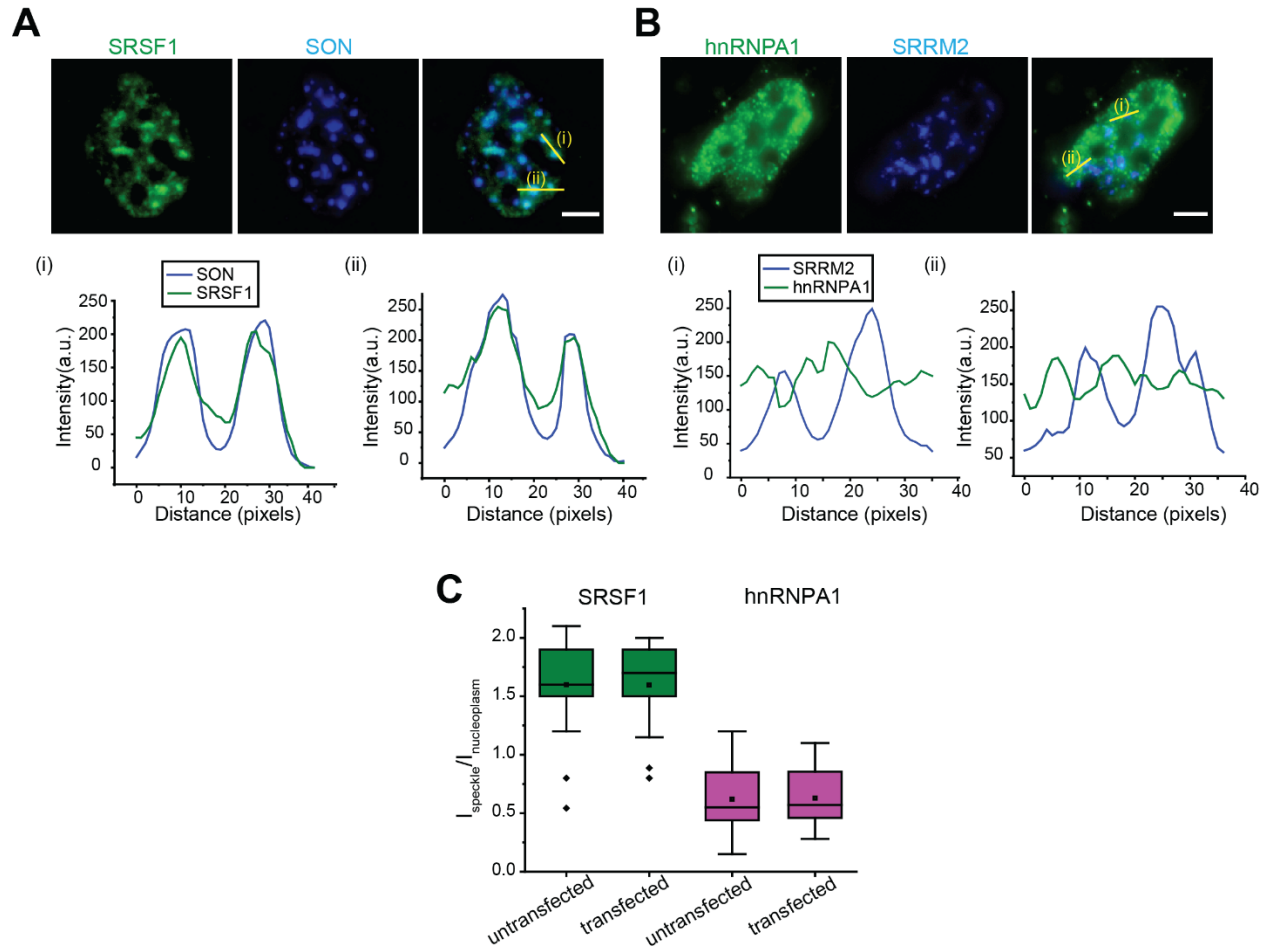
**iScience, Volume 27**

## **Supplemental information**

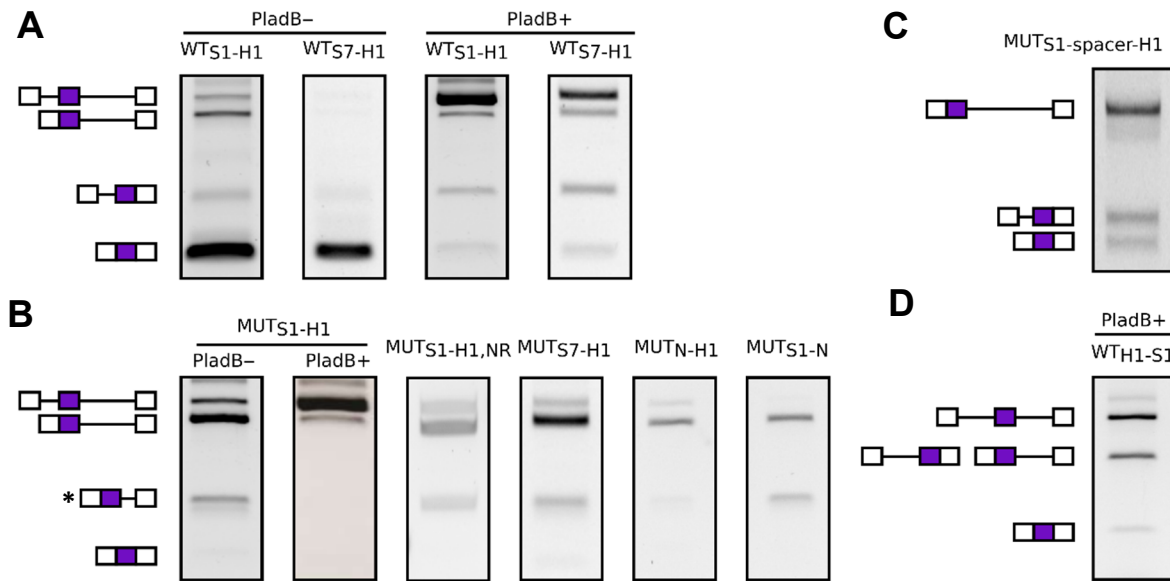
### **RNA molecules display distinctive organization at nuclear speckles**

**Sneha Paul, Mauricio A. Arias, Li Wen, Susan E. Liao, Jiacheng Zhang, Xiaoshu Wang, Oded Regev, and Jingyi Fei**

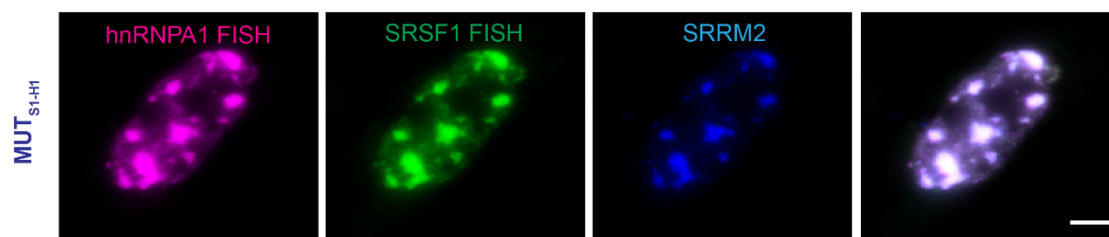
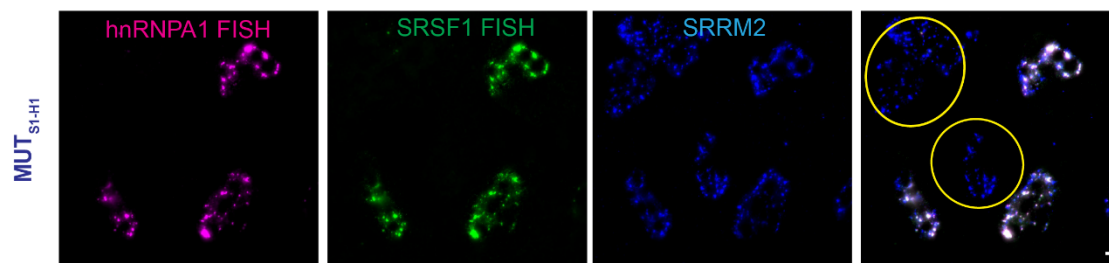
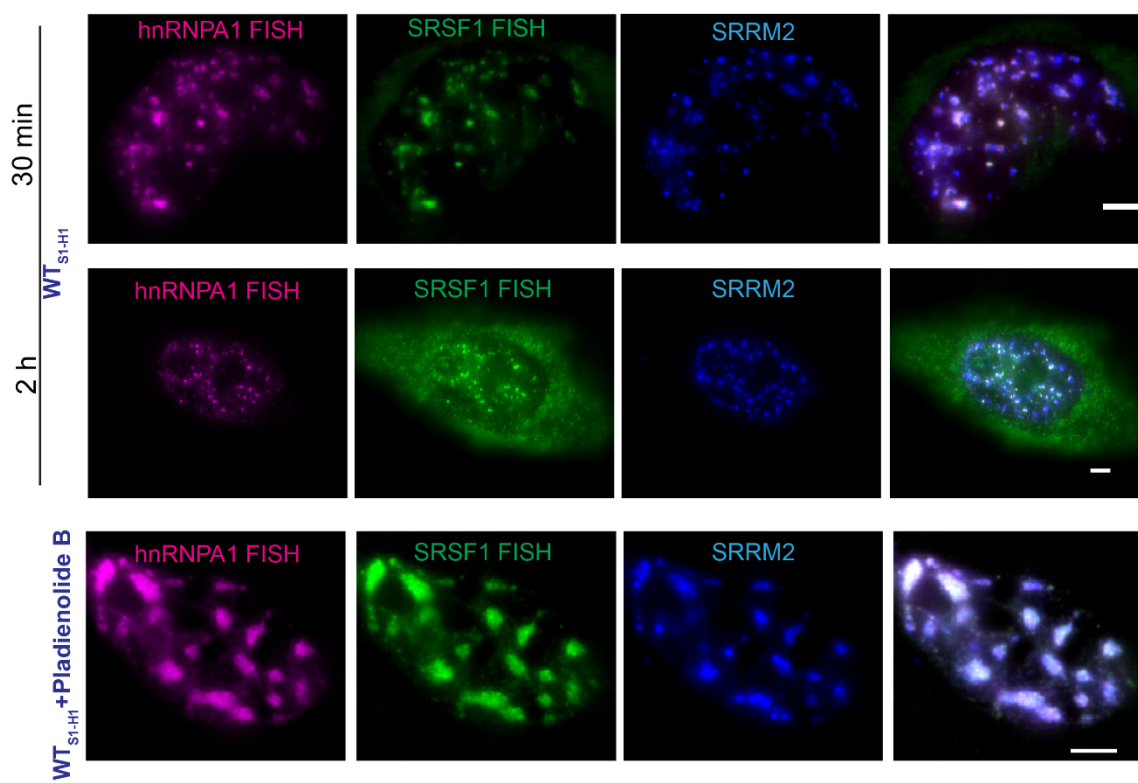
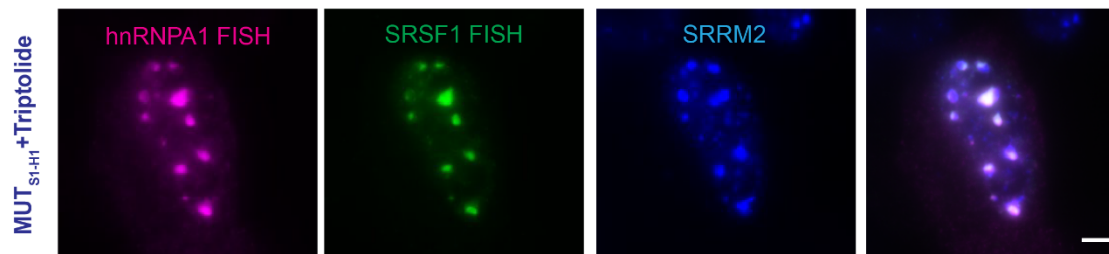
## Supplemental figures



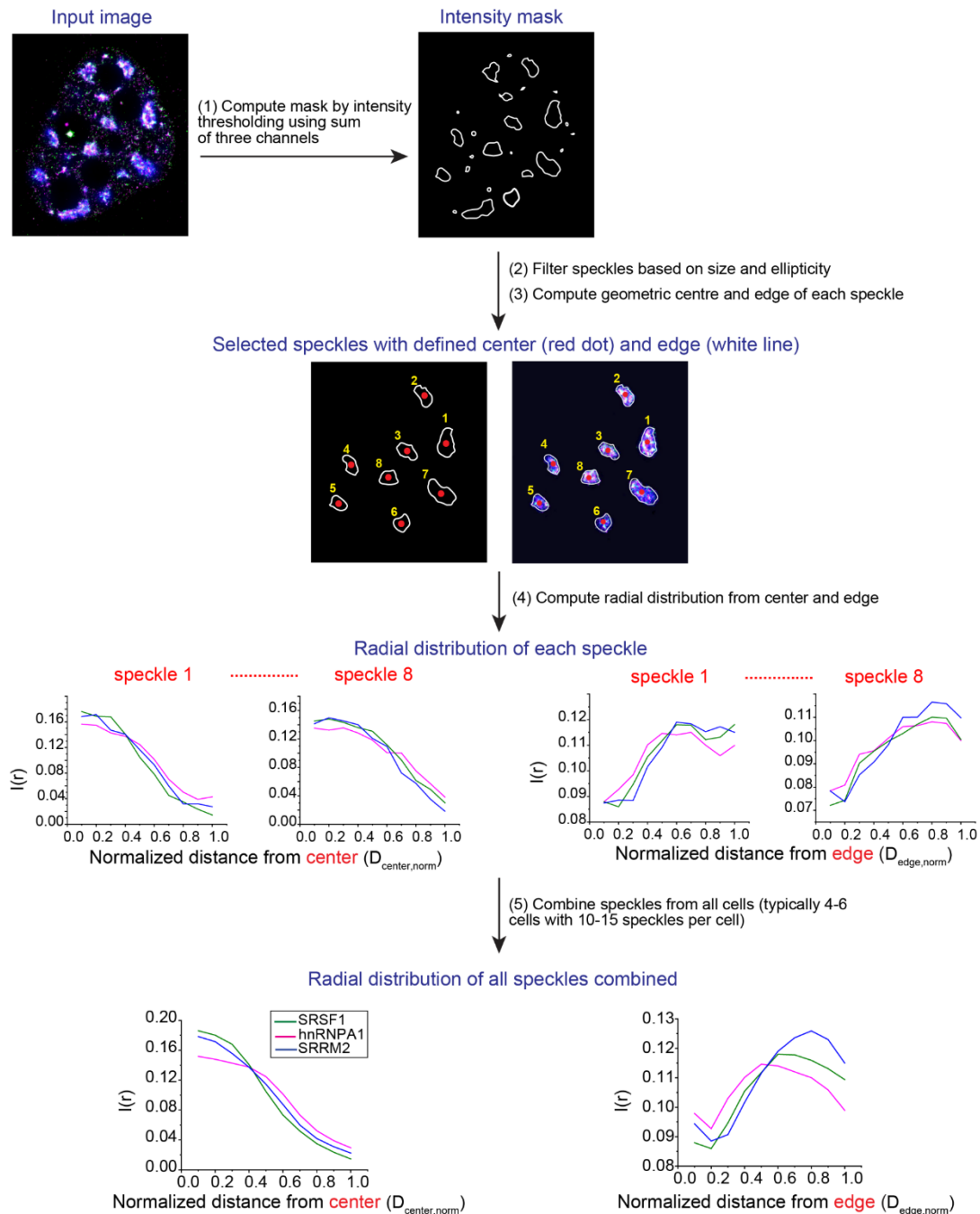
**Figure S1. Diffraction-limited epifluorescence images of SRSF1 and hnRNP A1 proteins, related to Figure 1.** SRSF1 (A) and hnRNP A1 (B) proteins are labeled by immunofluorescence staining using CF568, together with SON or SRRM2 protein stained with AF488. Representative speckles are highlighted to demonstrate distributions of SRSF1 and hnRNP A1 relative to nuclear speckles. (C) Comparison of SRSF1 and hnRNP A1 enrichment in nuclear speckles from ~100 speckles from 10 transfected cells and 10 untransfected cells from the same experimental batch.  $I_{\text{speckle}}$  and  $I_{\text{nucleoplasm}}$  represent the mean intensities of proteins in the nuclear speckles and nucleoplasm, respectively. SRSF1 proteins are enriched in the speckles. hnRNP A1 proteins showed a lower abundance in most nuclear speckles than the surrounding nucleoplasm. Scale bars represent 5  $\mu\text{m}$ . Description of box plot: 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.



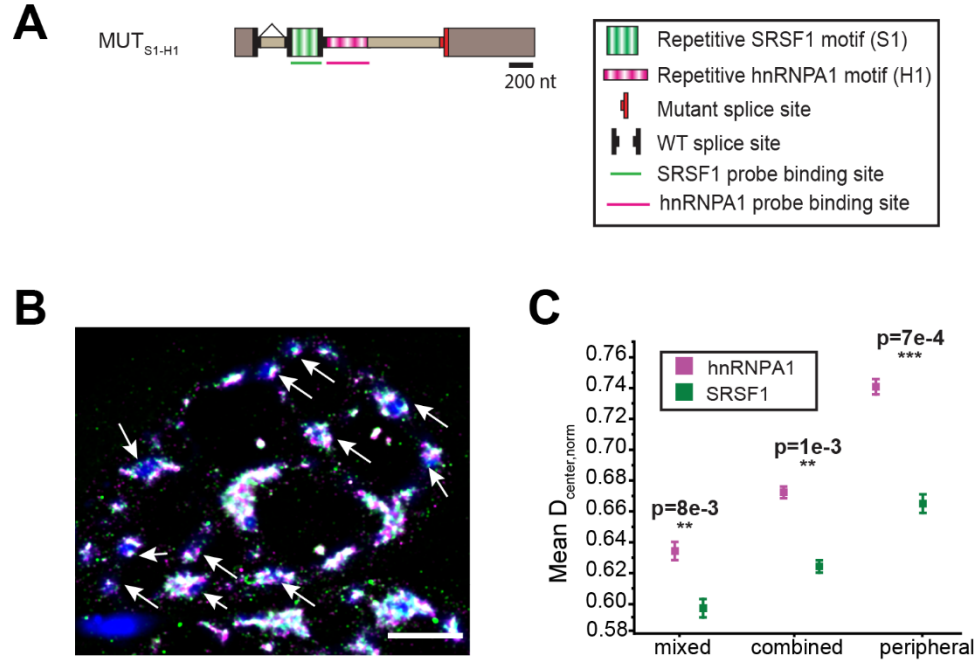
**Figure S2. Reverse-transcription PCR (RT-PCR) assay to test splicing outcomes, related to Figures 2-4.** (A) WT<sub>S1-H1</sub> and WT<sub>S7-H1</sub> in the absence and presence of Pladienolide B. These constructs are spliced with the middle exon included in the absence of Pladienolide B, while in the presence of Pladienolide B, splicing is inhibited. (B) MUT<sub>S1-H1</sub>, MUT<sub>S1-H1,NR</sub>, MUT<sub>S7-H1</sub>, MUT<sub>N-H1</sub> and MUT<sub>S1-N</sub> constructs with a 3' splice site mutation. In these constructs, the middle exon and its downstream intron are mostly not spliced. A minor band (\*) corresponding to the activation of a cryptic 3' splice site in the second intron is visible. Splicing is completely abolished when treating MUT<sub>S1-H1</sub> with Pladienolide B. (C) MUT<sub>S1-spacer-H1</sub> construct with a 3' splice site mutation and a 720-nt long spacer between the SRSF1 and hnRNPA1 motif-rich regions. The middle exon and its downstream intron are mostly not spliced. (D) WT<sub>H1-S1</sub> in the presence of Pladienolide B. Splicing is mostly inhibited in the presence of Pladienolide B. The middle band is a mixture of two products of nearly identical sizes (intron 1 retained and intron 2 retained). Induction time for all constructs is 2 h.

**A****B****C****D**

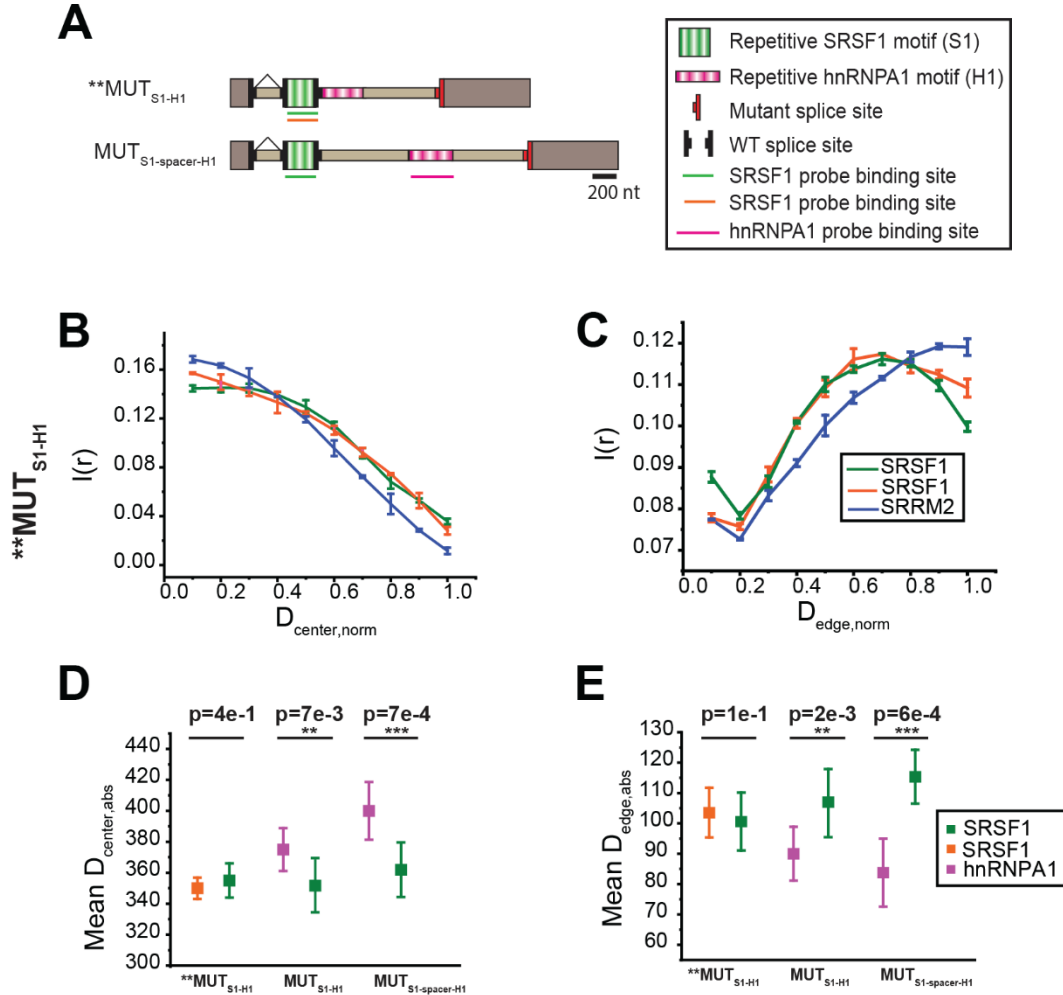
**Figure S3. Representative epifluorescence images of RNAs expressed from the reporter constructs, related to Figures 2-3.** FISH signals corresponding to hnRNPA1 and SRSF1 protein binding motifs in the RNAs are shown in magenta and green respectively. Immunostaining of SRRM2 is shown in blue. (A) 2 h Induction of MUT<sub>S1-H1</sub> in the absence of Pladienolide B. (B) Specificity of RNA FISH probes tested with transfected cells and cells with unsuccessful transfection as control (highlighted in yellow). The cells with unsuccessful transfection have negligible FISH signal as compared to the transfected cells. (C) 30 min and 2 h Induction of WT<sub>S1-H1</sub> in the absence and presence of Pladienolide B. (D) 2 h Induction of MUT<sub>S1-H1</sub> in the presence of Triptolide. Scale bars represent 5  $\mu$ m.



**Figure S4. Data analysis pipeline, related to Figures 2-6.** Input images are in the form of red (R), green (G) and blue (B) where R and G channels represent RNA FISH signals and B corresponds to SRRM2 or SON immunostaining signal. A composite RGB image is generated and then converted to a grayscale image. Using the sum of intensity from RGB channels, a mask is generated through intensity thresholding. Following this a size filter (lower limit 300 pixels, upper limit 5000 pixels) and an ellipticity filter (cutoff 0.8) are applied. After defining the geometric center and edge of each selected speckle, the radial distribution is calculated for each speckle. Finally, the radial distribution of each selected speckle is averaged to generate the final plots.

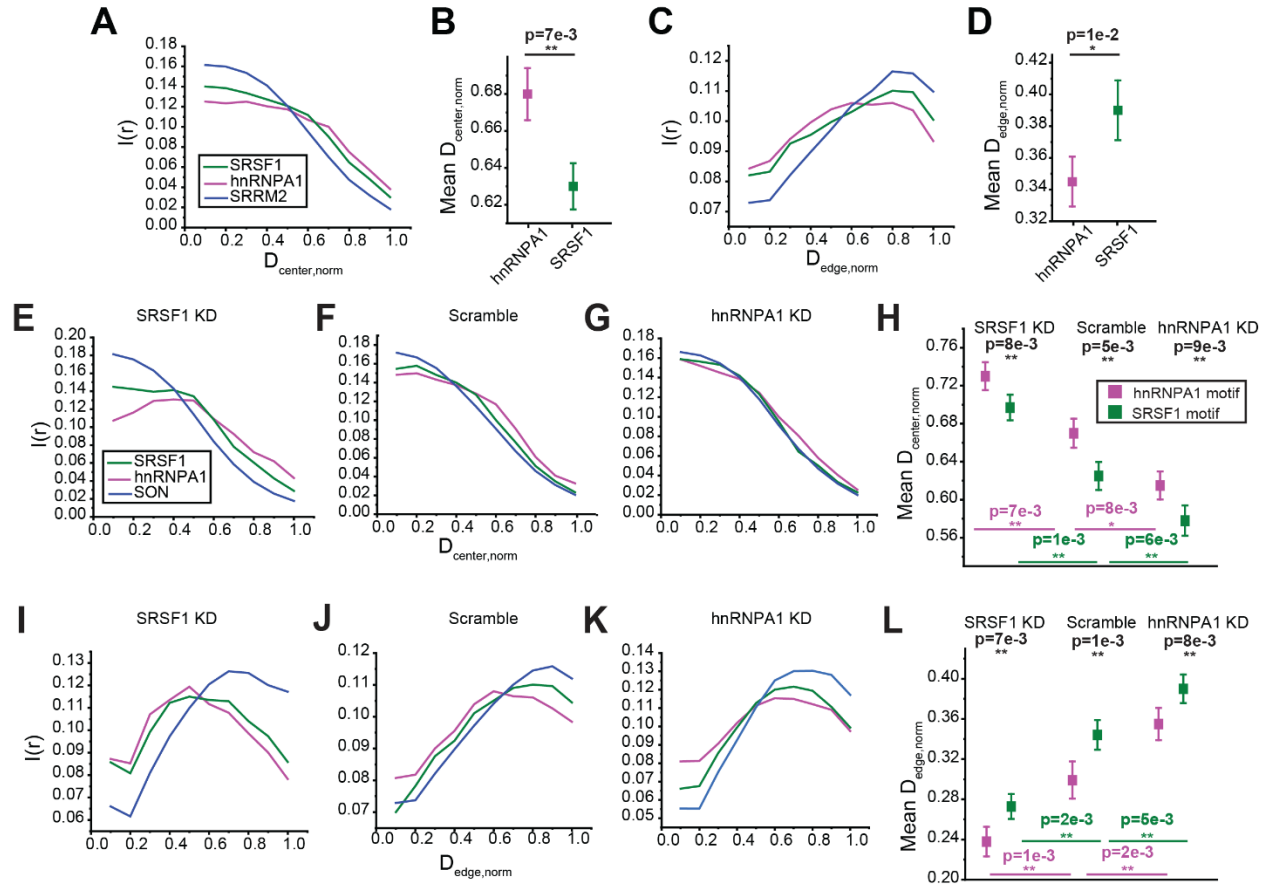


**Figure S5. Subpopulation analysis of MUT<sub>S1-H1</sub> transcripts, related to Figure 2.** (A) Schematic illustration of MUT<sub>S1-H1</sub>. (B) Representative SMLM image of MUT<sub>S1-H1</sub> showing different RNA speckle localization behavior (white arrows represent peripheral behavior; scale bar represents 5  $\mu$ m). (C) Normalized mean distance analysis for different subpopulations of RNA distribution within speckles (mixed, peripheral and combination of mixed and peripheral). Values in scatter plot represent mean  $\pm$  standard error of mean (SEM). p-values in the scatter plots are calculated with paired sample Wilcoxon signed rank test (one-sided), with \* $p < 5e-2$ , \*\* $p < 1e-2$ , \*\*\* $p < 1e-3$ .

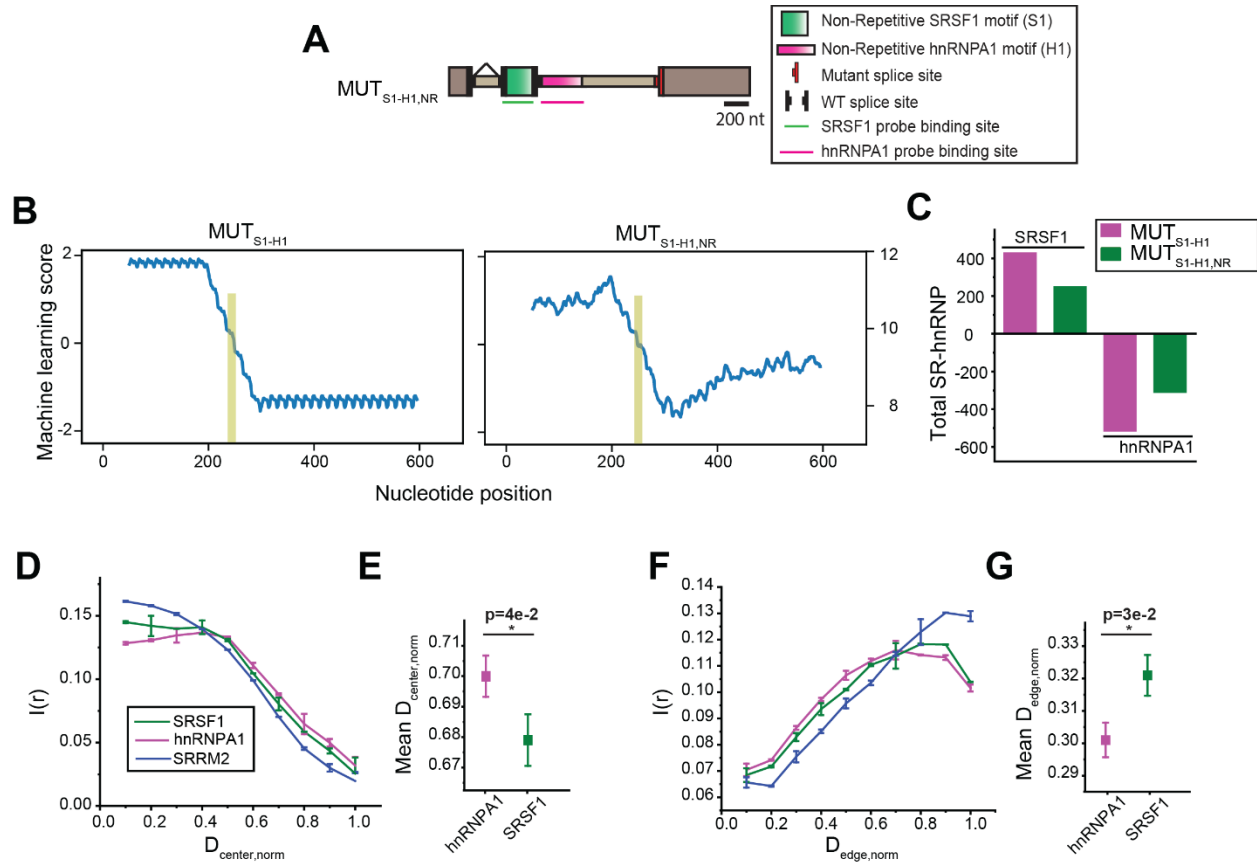


**Figure S6. Validation of mean radial distance difference, related to Figure 2.** (A) Schematic illustration of  $MUT_{S1-H1}$  and  $MUT_{S1-spacer-H1}$  constructs with FISH probe locations.  $**MUT_{S1-H1}$  represents the case where both the AF647 and CF568 labeled RNA FISH probes are on the SRSF1 motifs. Population distribution of SRSF1 motif signals labeled with both AF647 and CF568 for  $MUT_{S1-H1}$  as a function of the normalized distance from the center of the speckle (B) and edge of the speckle (C). Population-weighted mean absolute distance of SRSF1 and hnRNPA1 signal from the center of speckle (D) and edge of the speckle (E) for each speckle for  $MUT_{S1-H1}$  and  $MUT_{S1-spacer-H1}$ . Error bars in the population vs. distance plots report the standard deviation from two replicates, each data set contains at least 80-120 nuclear speckles collected from 8-12 cells. Values in scatter plot represent mean  $\pm$  SEM. p-values in the scatter plots are calculated with paired sample Wilcoxon signed rank test (one-sided), with  $*p<5e-2$ ,  $**p<1e-2$ ,  $***p<1e-3$ . Replicates are biological replicates collected starting from different dishes of cells and measured on different days.

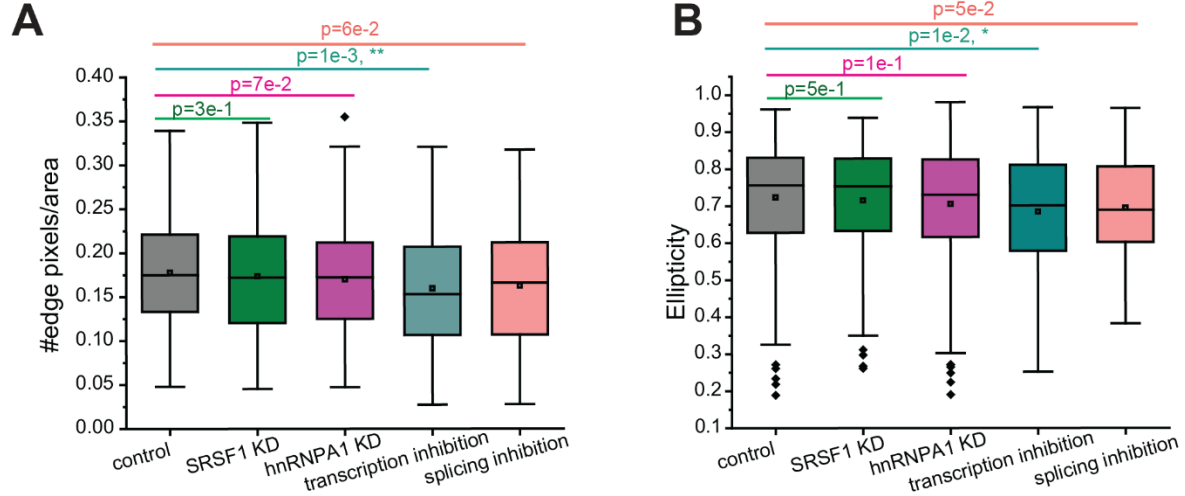




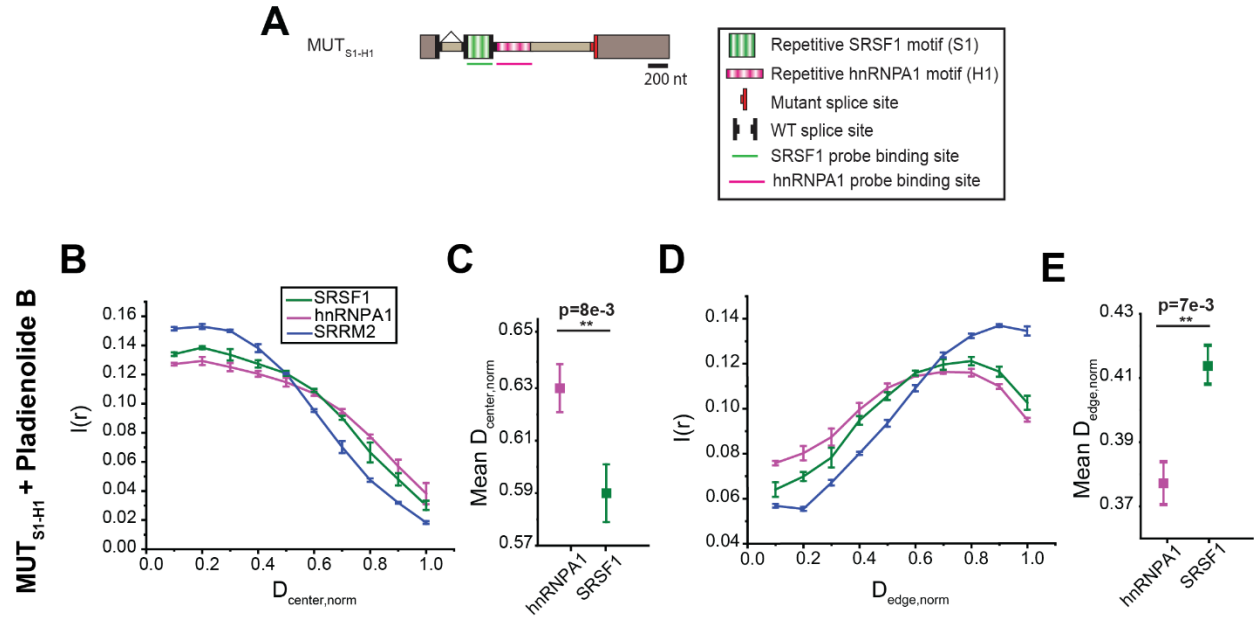
**Figure S7. SMLM imaging and analysis with reversed labeling of probes, related to Figures 2 and 5.** The reversed labeling scheme: hnRNPA1 motifs labeled with CF568 and SRSF1 motifs labeled with AF647. (A-D) Analysis of  $MUT_{S1-H1}$  RNA. Population distribution of SRSF1 and hnRNPA1 motif signals as a function of the normalized distance from the center (A) and from the edge (C) of the speckle. Scatter plot of the population-weighted mean normalized distance of SRSF1 and hnRNPA1 signal from the center (B) and from the edge (D) of speckle for each speckle. (E-L) Analysis of  $MUT_{S1-H1}$  RNA under knockdown conditions. Population distribution of SRSF1 and hnRNPA1 motif signals as a function of the normalized distance from the center (E-G) and from the edge (I-K) of the speckle. Population-weighted mean normalized distance of SRSF1 and hnRNPA1 signal from the center (H) and from the edge (L) of speckle for each speckle. Since the reversed labeling scheme is only used for validation, data was only collected once for each case. Each data set contains 50-60 nuclear speckles collected from 5-6 cells. Values in scatter plot represent mean  $\pm$  SEM. p-values in the scatter plots are calculated with paired sample Wilcoxon signed rank test (black, one-sided) and two sample t-test (magenta and green, one-sided), with  $*p<5e-2$ ,  $**p<1e-2$ ,  $***p<1e-3$ .



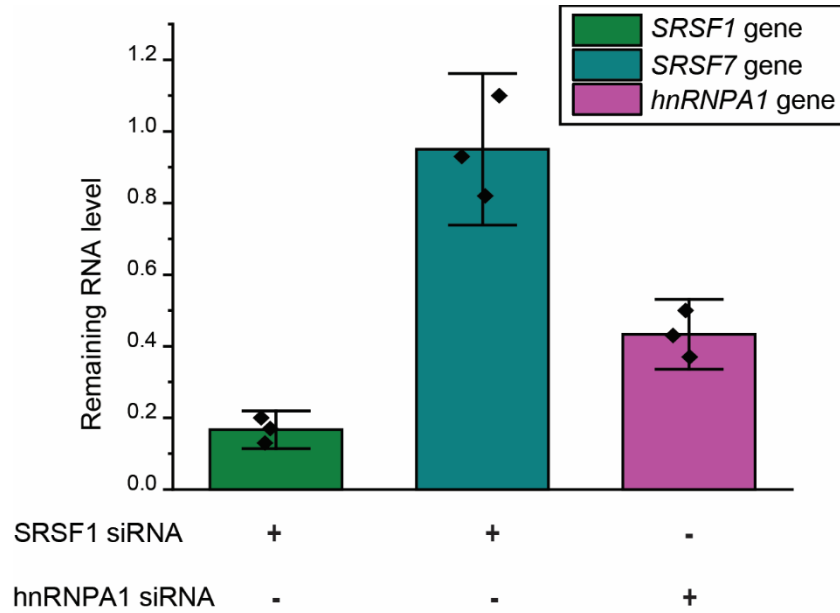
**Figure S8. Intra-speckle positioning of RNA transcripts from constructs devoid of repeats, related to Figure 2.** (A) Schematic illustration of  $MUT_{S1-H1,NR}$  construct. These constructs were obtained from  $MUT_{S1-H1}$  by mutating each nucleotide in the exonic and intronic regions with a probability of 37.5%. (B) Plots showing predicted SR enrichment (positive) or hnRNP enrichment (negative) for  $MUT_{S1-H1}$  and  $MUT_{S1-H1,NR}$  constructs. The prediction is based on a machine learning model trained on splicing data,<sup>1</sup> and smoothed using a 100nt sliding window. Bars show predicted MaxEnt splice site score.<sup>2</sup> (C) Machine learning-predicted total SR enrichment (positive) and hnRNP enrichment (negative) of  $MUT_{S1-H1}$  and  $MUT_{S1-H1,NR}$  constructs. (D-G) Population distribution of SRSF1 and hnRNP1 motif signals as a function of the normalized distance from the center (D) and from the edge (F) of the speckle for  $MUT_{S1-H1,NR}$ . Population-weighted mean normalized distance of SRSF1 and hnRNP1 signal from the center (E) and from the edge (G) of speckle for each speckle in  $MUT_{S1-H1,NR}$ . Error bars in the population vs. distance plots report the standard deviation from two replicates, each containing at least 60-75 nuclear speckles collected from 4-5 cells. Scatter plots are generated by combining all nuclear speckles (120-150) from two replicates. Values in scatter plot represent mean  $\pm$  SEM. p-values in the scatter plots are calculated with paired sample Wilcoxon signed rank test (one-sided), with \* $p < 5e-2$ , \*\* $p < 1e-2$ , \*\*\* $p < 1e-3$ . Replicates here are biological replicates which are collected starting from different dishes of cells and measured on different days.



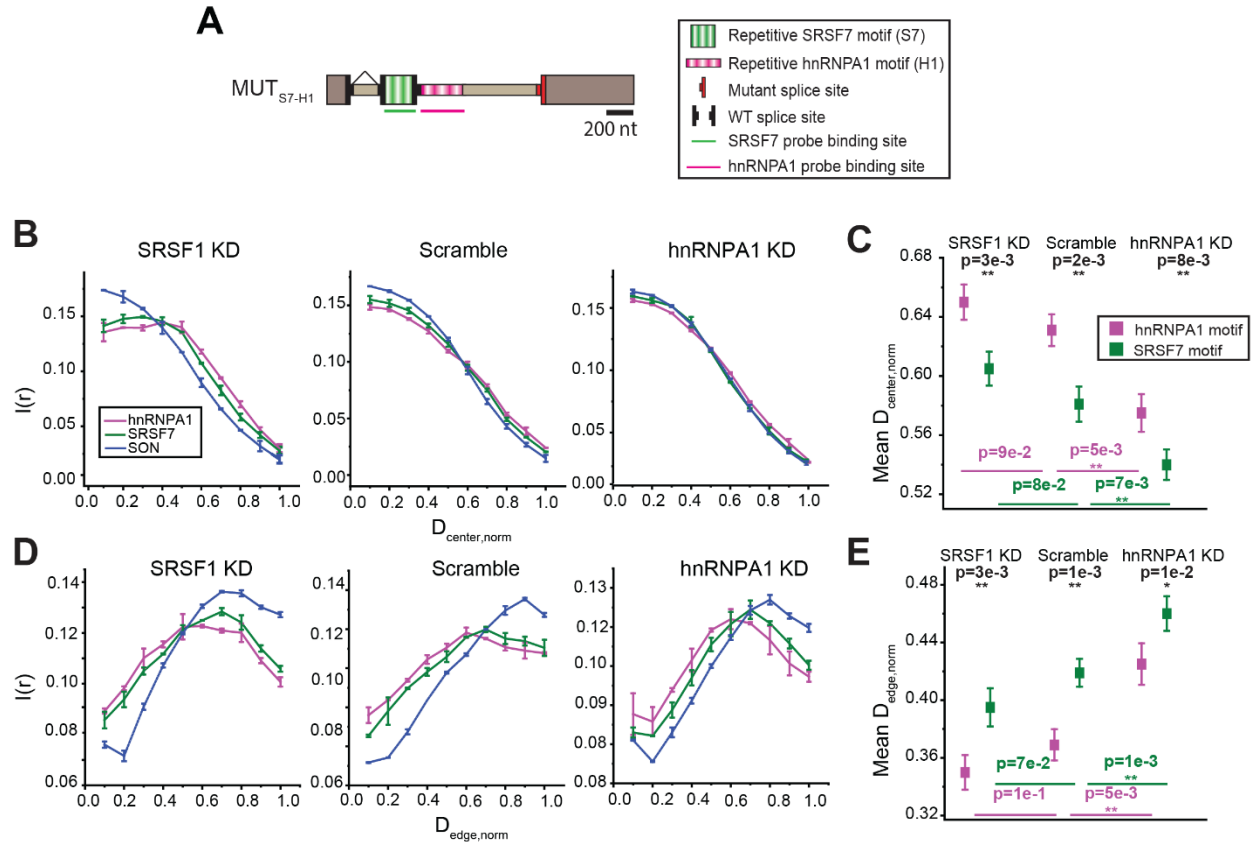
**Figure S9. The effect of treatments on speckle morphology, related to Figures 3 and 5.** (A) The ‘regularity’ or ‘roughness’ of the speckle surface is measured by the number of edge pixels divided by area, where the number of edge pixels is an estimation of the perimeter of the speckle. The lower the value of perimeter-to-area ratio, the more regular or rounded the speckle is. (B) Ellipticity is a measure of the deviation of speckle shape from a perfect sphere and is calculated as,  $\sqrt{((a^2 - c^2)/a^2)}$ , where  $a$  is the equatorial radius and  $c$  is the polar radius assuming nuclear speckles to be an ellipse. An ellipticity value of 0 represents a perfect sphere. Each data set contains 300-400 nuclear speckles collected from 30-40 cells. p-values are calculated with two sample t-tests (two-sided). \* $p < 5e-2$ , \*\* $p < 1e-2$ , \*\*\* $p < 1e-3$ . Description of box plots is same as Figure S1. Control represents MUT<sub>S1-H1</sub> without any treatment.



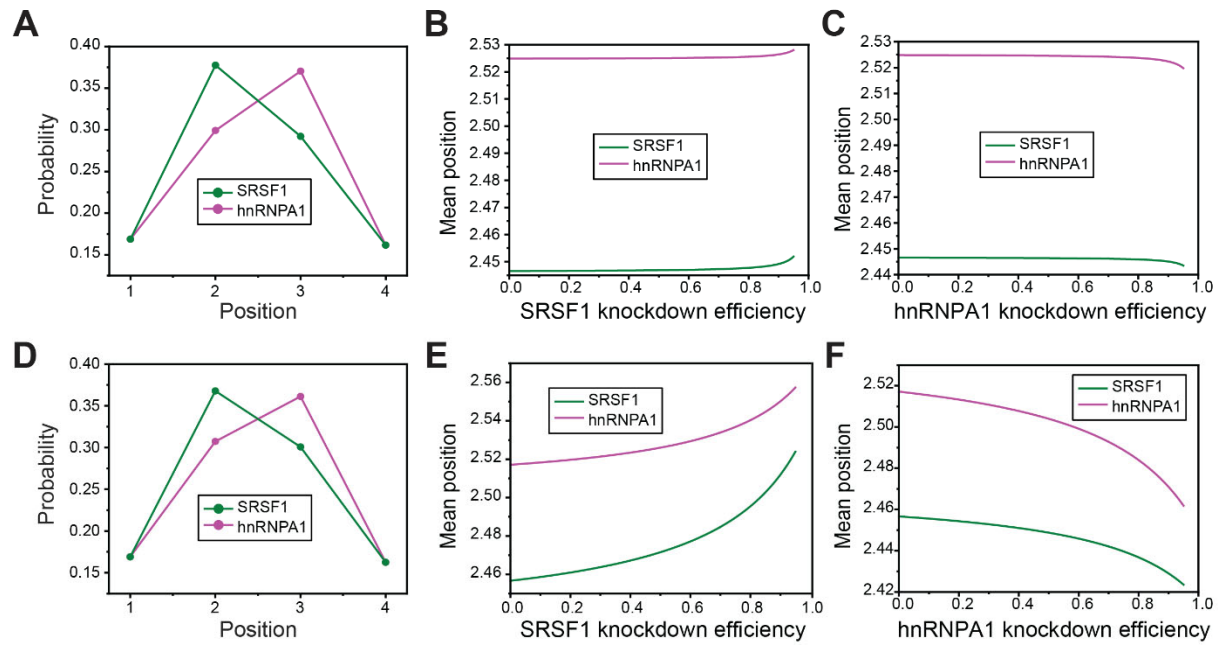
**Figure S10. Intra-speckle organization of MUT<sub>S1-H1</sub> transcripts with splicing inhibition, related to Figure 2.** (A) Schematic illustration of MUT<sub>S1-H1</sub>. (B-E) Population distribution of SRSF1 and hnRNP A1 motif signals for MUT<sub>S1-H1</sub> in the presence of Pladienolide B as a function of the normalized distance from the center of the speckle (B) and edge of the speckle (D). Population-weighted mean normalized distance of SRSF1 and hnRNP A1 signal from the center of speckle (C) and edge of speckle (E) for each speckle for MUT<sub>S1-H1</sub> in the presence of Pladienolide B. Error bars in the population vs. distance plots report the standard deviation from two replicates, each data set contains at least 80-100 nuclear speckles collected from 8-10 cells. Values in scatter plot represent mean  $\pm$  SEM. p-values in the scatter plots are calculated with paired sample Wilcoxon signed rank test (one-sided). \* $p < 5e-2$ , \*\* $p < 1e-2$ , \*\*\* $p < 1e-3$ . Replicates are biological replicates collected starting from different dishes of cells and measured on different days.



**Figure S11. Knockdown efficiencies at the mRNA level measured by qPCR, related to Figure 5.** The Ct value for each target gene under each knockdown condition was normalized with the Ct value of *ACTB* mRNA. The normalized Ct value of each target gene under knockdown conditions was compared to the normalized Ct value of the same gene treated with the scramble siRNA. Error bars report mean  $\pm$  standard deviation from three biological replicates (black points).

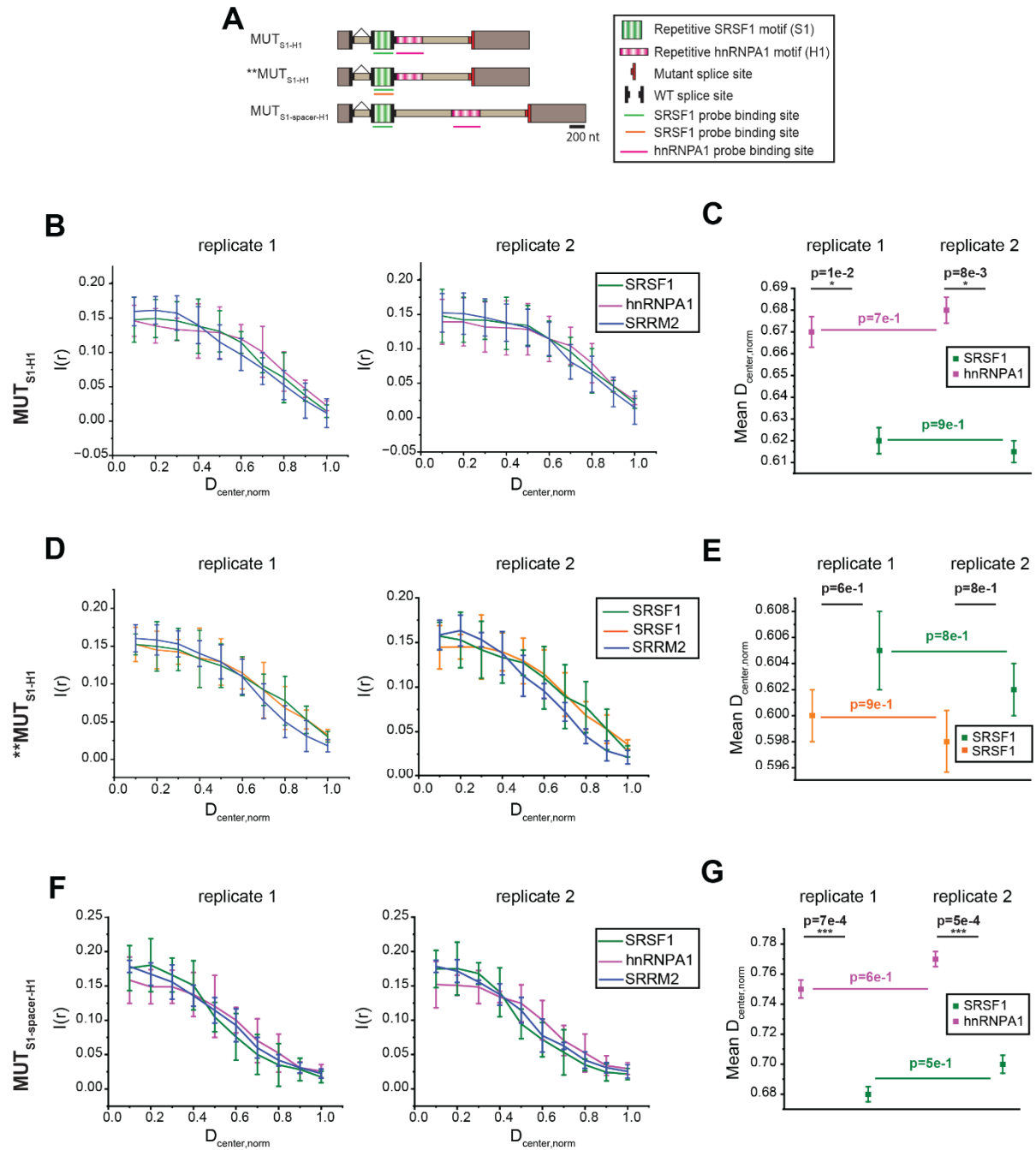


**Figure S12. Effect of SRSF1 and hnRNA1 knockdown on the intra-speckle organization of RNAs containing SRSF7 motifs in exon and hnRNA1 motifs in intron, related to Figure 5.** (A) Schematic illustration of MUT<sub>S7-H1</sub>. (B-E) Population distribution of SRSF7 and hnRNA1 motif signals as a function of the normalized distance from the center (B) and from the edge (D) of the speckle. Population-weighted mean normalized distance of SRSF7 and hnRNA1 signal from the center (C) and from 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 60-90 nuclear speckles collected from 4-6 cells. Scatter plots are generated by combining all nuclear speckles (120-180) from two replicates. Values in scatter plot represent mean  $\pm$  SEM. p-values in the scatter plots are calculated with paired sample Wilcoxon signed rank test (black, one-sided) and two sample t-test (magenta and green, one-sided), with \* $p < 5e-2$ , \*\* $p < 1e-2$ , \*\*\* $p < 1e-3$ . Replicates are biological replicates collected starting from different dishes of cells and measured on different days.



**Figure S13. Toy model for different  $K_d$  values, related to Figure 7.** (A)-(C)  $K_d = 100$  nM: (A) Probability distributions of SRSF1 and hnRNPA1 motif position predicted by the toy model. Mean positions of SRSF1 and hnRNPA1 motif plotted as a function of both SRSF1 (B) and hnRNPA1 (C) knockdown efficiencies. (D)-(F)  $K_d = 10$  uM: (D) Probability distributions of SRSF1 and hnRNPA1 motif position predicted by the toy model. Mean positions of SRSF1 and hnRNPA1 motif plotted as a function of both SRSF1 (E) and hnRNPA1 (F) knockdown efficiencies.

## Supplemental data item



**Data S1. Analysis of intra-speckle organization of RNAs containing SRSF1 motifs in exon and hnRNP A1 motifs in intron for individual biological replicates, related to Figure 2. (A)** Schematic illustration of  $MUT_{S1-H1}$  and  $MUT_{S1-spacer-H1}$  constructs with FISH probe locations. Population distribution of SRSF1 and hnRNP A1 motif signals for  $MUT_{S1-H1}$  (B) and  $MUT_{S1-spacer-H1}$  (F) as a function of the normalized distance from the center of the speckle for two separate biological replicates. Population-weighted mean normalized distance of SRSF1 and hnRNP A1 signal from the center of speckle for  $MUT_{S1-H1}$  (C) and  $MUT_{S1-spacer-H1}$  (G). (D) Population distribution of SRSF1 motif signals labeled with both A647 and CF568 for  $MUT_{S1-H1}$  as



a function of the normalized distance from the center of the speckle for two separate biological replicates. (E) Population-weighted mean absolute distance of SRSF1 motif signals labeled with both A647 and CF568 for MUT<sub>S1-H1</sub> from the center of speckle for each speckle for two biological replicates separately. Error bars in the population vs. distance plots report the standard deviation of all speckles in each biological replicate. Each biological replicate contains at least 40-60 nuclear speckles collected from 4-6 cells. Values in scatter plot represent mean  $\pm$  SEM. For comparison between SRSF1 and hnRNPA1 motifs in each biological replicate, p-values (in black on top of the plots) are calculated with paired sample Wilcoxon signed rank test (one-sided); for comparison of SRSF1 motif or hnRNPA1 motif between two biological replicates, p-values (in green, magenta or orange, labeled within the plots) are calculated using two sample t-test, two-sided), with \* $p < 5e-2$ , \*\* $p < 1e-2$ , \*\*\* $p < 1e-3$ . \*\* MUT<sub>S1-H1</sub> represents the case where both the A647 and CF568 labeled RNA FISH probes are on the SRSF1 motifs. Replicates are biological replicates collected starting from different dishes of cells and measured on different days.

# Supplemental tables

**Table S1. Summary of configurations and states for toy model, related to Figure 7.**

Configuration	State	$x_s$	$x_h$	$\sigma_s$	$\sigma_h$	$E$
		1	2	0	0	0
		1	2	1	0	$\varepsilon_{s1}$
		1	2	0	1	$\varepsilon_{h1}$
		1	2	1	1	$\varepsilon_{s1} + \varepsilon_{h1} + J$
		2	3	0	0	0
		2	3	1	0	$\varepsilon_{s1}$
		2	3	0	1	$\varepsilon_{h2}$
		2	3	1	1	$\varepsilon_{s1} + \varepsilon_{h2} + J$
		3	4	0	0	0
		3	4	1	0	$\varepsilon_{s2}$
		3	4	0	1	$\varepsilon_{h2}$
		3	4	1	1	$\varepsilon_{s2} + \varepsilon_{h2} + J$
		2	1	0	0	0
		2	1	1	0	$\varepsilon_{s1}$
		2	1	0	1	$\varepsilon_{h1}$
		2	1	1	1	$\varepsilon_{s1} + \varepsilon_{h1} + J$
		3	2	0	0	0
		3	2	1	0	$\varepsilon_{s2}$
		3	2	0	1	$\varepsilon_{h1}$
		3	2	1	1	$\varepsilon_{s2} + \varepsilon_{h1} + J$
		4	3	0	0	0
		4	3	1	0	$\varepsilon_{s2}$
		4	3	0	1	$\varepsilon_{h2}$
		4	3	1	1	$\varepsilon_{s2} + \varepsilon_{h2} + J$

**Table S2: Primers, FISH probes and construction oligomers, related to Figures 2-6.**

<b>PCR/qPCR/RT primers</b>	<b>5' - 3'</b>
SRSF1 forward primer	GATTCCTGCCCCAACCAAAC
SRSF1 reverse primer	TCGTGCATCTTAAGGGCTCC
SRSF7 forward primer	GCTGGTTGCGCAGATACCTA
SRSF7 reverse primer	ACTAGCGGTCAAACCTACCGAA
hnRNPA1 forward primer	TGGAGGTGGTGGAAAGCTACA
hnRNPA1 reverse primer	GCCAGAGCTTCTGCCTCCAA
ActB forward primer	CCTTCCAGCAGATGTGGATC
ActB reverse primer	GCCATGCCAATCTCATCTTG
Minigene forward primer	CGACCGCTGAACTGCATCGTCGCCGT
Minigene reverse primer	CAACTATCCAAACCATGACAGCTCCACTTTA
Minigene reverse transcription primer	GAGATGGCCTGGCT
<b>FISH probes</b>	<b>5' - 3'</b>
SRSF1 motifs in WT_SRSF1 and MUT_SRSF1	TCTTGTTTGGCGTTCCTC
hnRNPA1 motifs in WT_SRSF1/7 and MUT_SRSF1/7	CCCTAATTGTTTGGTTCCC
SRSF7 motifs in WT_SRSF7 and MUT_SRSF7	GTCGTCTTTGTTTGGAGTC
SRSF1 motifs in MUT_NR	CGCTATTCCCCTTGTATG
	GCGCTACTCCTGTTTGGT
	ATTTGCCTTCTCCTTGAC
	GCGGTAGTTAAACGTTGC
	CCATCGTCTTGATTGACG
	TTGAAATGCCTCGATGCT
	TTGTCTGGCGTACACGTT
	TCCTCGGTCTTCGACATC
hnRNPA1 motifs in MUT_NR	ACACTAACTGCGTGGGGC
	TCCCGACCCATTGTTTG
	CCTACTTGTTTGTTAGCC
	GGTACCCTAATTAGTAGG
	GCATGGTGCCCAAATTGA
	GCGACTAGGTTGGGTATC
	GGTTCACCTATCCTGAGA
<b>Construction oligomers</b>	
SRSF1-3x-module-F	AAACAAGAGGAACGCCAAACAAGAGGAACGCC AAACAAGAGGAACGCC
SRSF1-3x-module-R	GTTTGGCGTTCCTCTTGTGGCGTTCCTCTT GTTTGGCGTTCCTCTT
SRSF7-3x-module-F	AAACAAAGACGACTCCAAACAAGACGACTCC AAACAAAGACGACTCC
SRSF7-3x-module-R	GTTTGGAGTCGTCTTTGTTTGGAGTCGTCTTT GTTTGGAGTCGTCTTT
hnRNPA1-3x-module-F	AAACAATTAGGGAACCAACAATTAGGGAACC AAACAATTAGGGAACC
hnRNPA1-3x-module-R	GTTTGGTTCCTAATTGTTTGGTTCCTAATTG TTTGGTTCCTAATT
WT_splice_site-F	TCGACTTCAGCAGCCGTCTCCAAACAACACAG GTAAGTCCAAACAGAGACCAGCTGCA
WT_splice_site-R	AGCTTGACAGCTGGTCTCTGTTTGGACTTACCT GTGTTGTTTGGAGACGGCTGCTGAAG
MUT_splice_site-F	TCGACTTCAGCAGCCGTCTCCAAACAACACAG ACAAGTCCAAACAGAGACCAGCTGCA
MUT_splice_site-R	AGCTTGACAGCTGGTCTCTGTTTGGACTTGTCT GTGTTGTTTGGAGACGGCTGCTGAAG

## **Supplemental References**

1. Liao, S.E., Sudarshan, M., and Regev, O. (2023). Deciphering RNA splicing logic with interpretable machine learning. *Proc Natl Acad Sci U S A* *120*, e2221165120. 10.1073/pnas.2221165120.
2. Yeo, G., and Burge, C.B. (2004). Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* *11*, 377–394. 10.1089/1066527041410418.