# Systematic mapping of nuclear domain-associated transcripts reveals speckles and lamina as hubs of functionally distinct retained introns

Graphical abstract

Authors

A. Rasim Barutcu, Mingkun Wu,
Ulrich Braunschweig, ...,
Philipp G. Maass,
Anne-Claude Gingras,
Benjamin J. Blencowe

Correspondence

b.blencowe@utoronto.ca

In brief

Barutcu, Wu et al. use marker proteins fused to APEX2 to survey coding and non-coding transcripts concentrated in diverse nuclear domains. Speckles and lamina are found to associate with structurally and functionally distinct classes of retained introns implicated in the control of core cellular processes, including the cell cycle.

Highlights

- Transcripts associated with diverse nuclear domains mapped using APEX2-seq

- Speckles and lamina associate with functionally distinct classes of retained introns

- Speckle-retained introns are enriched in RNA processing and cell-cycle factors

- Speckle-associated retained intron splicing is linked to cell-cycle progression

CellPress

Molecular Cell

Resource

Systematic mapping of nuclear domain-associated transcripts reveals speckles and lamina as hubs of functionally distinct retained introns

A. Rasim Barutcu,[1,7] Mingkun Wu,[1,2,7] Ulrich Braunschweig,[1] Boris J.A. Dyakov,[2,3] Zheng Luo,[1] Kyle M. Turner,[1] Tanja Durbic,[1] Zhen-Yuan Lin,[3] Robert J. Weatheritt,[4,5] Philipp G. Maass,[2,6] Anne-Claude Gingras,[2,3] and Benjamin J. Blencowe[1,2,8,*]
[1]Donnelly Centre, University of Toronto, Toronto, ON M5S 3E1, Canada
[2]Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 1A8, Canada
[3]Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, ON M5G 1X5, Canada
[4]EMBL Australia, Garvan Institute of Medical Research, Darlinghurst, NSW 2010, Australia
[5]St. Vincent Clinical School, University of New South Wales, Darlinghurst, NSW 2010, Australia
[6]Genetics and Genome Biology Program, SickKids Research Institute, Toronto, ON M5G 0A4, Canada
[7]These authors contributed equally
[8]Lead contact
*Correspondence: b.blencowe@utoronto.ca
https://doi.org/10.1016/j.molcel.2021.12.010

SUMMARY

The nucleus is highly compartmentalized through the formation of distinct classes of membraneless domains. However, the composition and function of many of these structures are not well understood. Using APEX2-mediated proximity labeling and RNA sequencing, we surveyed human transcripts associated with nuclear speckles, several additional domains, and the lamina. Remarkably, speckles and lamina are associated with distinct classes of retained introns enriched in genes that function in RNA processing, translation, and the cell cycle, among other processes. In contrast to the lamina-proximal introns, retained introns associated with speckles are relatively short, GC-rich, and enriched for functional sites of RNA-binding proteins that are concentrated in these domains. They are also highly differentially regulated across diverse cellular contexts, including the cell cycle. Thus, our study provides a resource of nuclear domain-associated transcripts and further reveals speckles and lamina as hubs of distinct populations of retained introns linked to gene regulation and cell cycle progression.
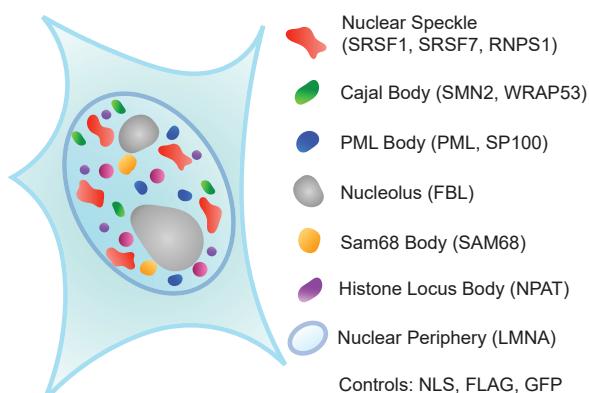
INTRODUCTION

In addition to organizing chromosomal DNA, the cell nucleus contains multiple distinct membraneless nuclear domains or bodies (Staněk and Fox, 2017). These include the following: nucleoli, which form around ribosomal (r)RNA genes and function in the biogenesis of ribosomal subunits (Boisvert et al., 2007; Lafontaine et al., 2021); speckles (also known as interchromatin granule clusters), which concentrate factors that function in transcription, pre-mRNA processing and export (Galganski et al., 2017; Gordon et al., 2021; Spector and Lamond, 2011); paraspeckles, which regulate gene expression through sequestration of specific transcripts and protein components (Fox et al., 2018); Cajal bodies, which function in the maturation of small nuclear and small nucleolar ribonucleoprotein particles (snRNPs and snoRNPs) (Machyna et al., 2013); histone locus bodies (HLBs), which are associated with the biogenesis of histone mRNAs (Marzluff and Koreski, 2017; Nizami et al., 2010); and promyelocytic leukemia (PML) bodies, which have been impli-

cated in the cell cycle, as well as genome repair and maintenance (Corpet et al., 2020; Lallemand-Breitenbach and de The, 2018).
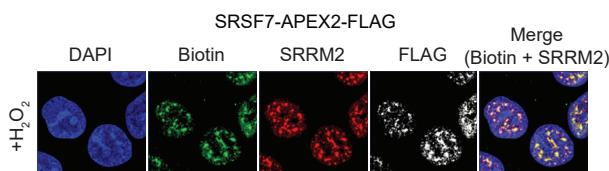
Nuclear domains assemble through the coalescence of nucleic acid and protein components. For example, the long non-coding (lnc)RNA NEAT1 acts as a scaffold to concentrate specific protein components to nucleate formation of paraspeckles (Clemson et al., 2009; Yamazaki et al., 2018). This coalescence may arise through phase separation, whereby RNA facilitates multivalent interactions involving intrinsically disordered domains of bound proteins (Brangwynne et al., 2009). This concept could apply more broadly to nuclear organization, in light of early and more recent evidence for important roles of RNA in shaping nuclear architecture (Barutcu et al., 2019; Creamer et al., 2021; Nickerson et al., 1995; Quinodoz et al., 2021; Rinn and Guttman, 2014; Shin et al., 2018). These studies collectively emphasize the importance of delineating the RNA composition of nuclear domains. Important questions include the following: where and how are specific classes of

(legend on next page)

Molecular Cell
Resource

CellPress
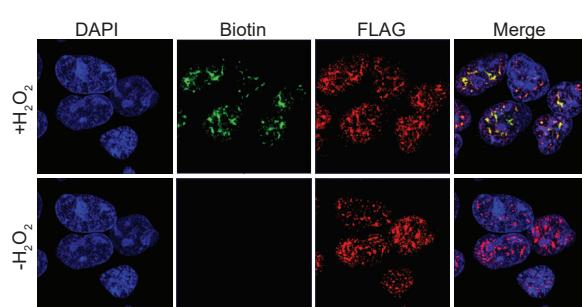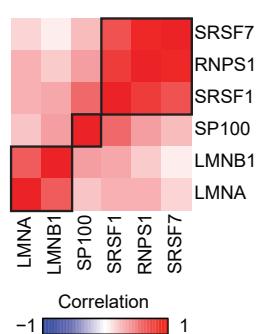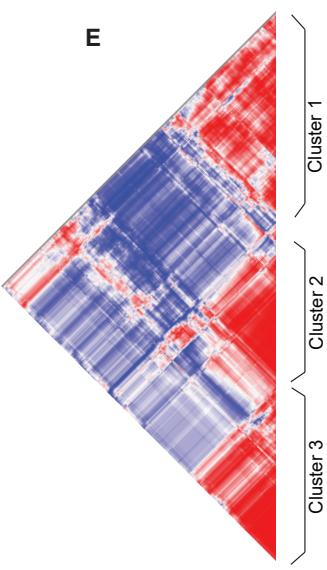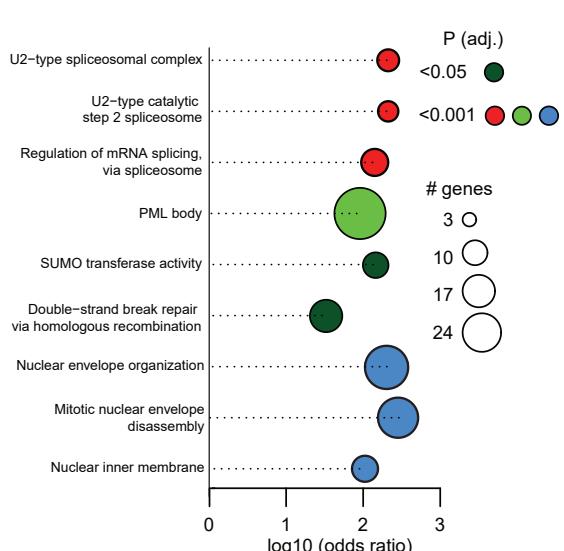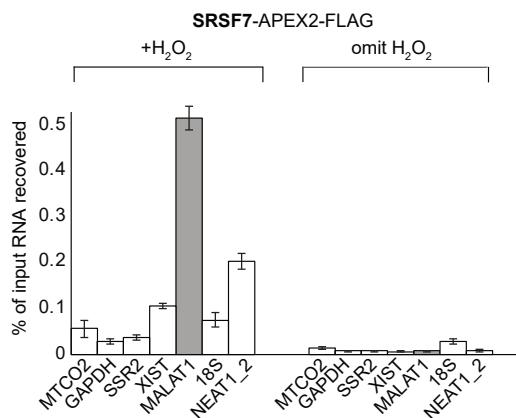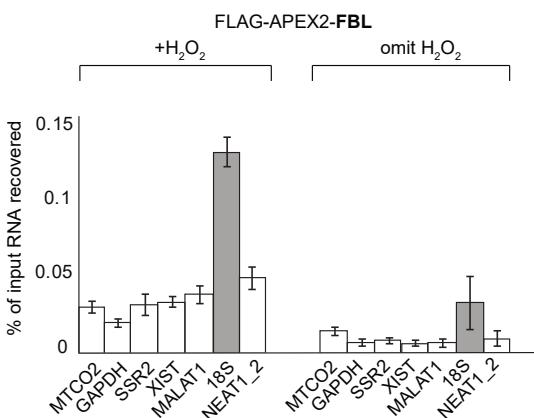
transcripts synthesized and processed in the nucleus, and which RNAs impart critical nuclear organization-function relationships?

These questions particularly apply to the 20–50 irregularly shaped speckles detected in diverse cell types (Galganski et al., 2017; Gordon et al., 2021). Speckles are associated with open chromatin and sites of active transcription (this study; Kim et al., 2020; Su et al., 2020) and have been proposed to function as sites of storage and recruitment of pre-mRNA processing factors, the splicing of specific transcripts (Hall et al., 2006; Spector and Lamond, 2011), and in RNA export (Chen et al., 2019; Dias et al., 2010; Wang et al., 2018). However, while small nuclear (sn)RNAs (Stanek and Neugebauer, 2006), the abundant lncRNA MALAT1 (Hall et al., 2006; Spector and Lamond, 2011; Tripathi et al., 2010), and a few additional transcripts (Duronio and Marzluff, 2017; Yang et al., 2011) have been mapped to speckles, the RNA composition of these and other nuclear domains has not been systematically determined.

A recently described strategy for investigating the composition of cellular compartments is APEX-seq (Fazal et al., 2019; Padrón et al., 2019). APEX2, an engineered ascorbate peroxidase with enhanced activity, is localized to sub-cellular compartments of interest through fusion to marker proteins. Following treatment of cells with biotin-phenol (BP) and hydrogen peroxide, the APEX2-markers biotinylate proteins, DNA, and RNA within an approximate 20 nm radius, which generally is smaller than the dimensions of cellular domains. This enables the recovery and mapping of factors concentrated in diverse cellular structures (Fazal et al., 2019; Padrón et al., 2019). In this study, we used APEX2-seq to generate a resource of nuclear domain-associated transcripts enabling the exploration of nuclear organization-function relationships. Although some domains are associated with relatively low overall levels of RNA that more frequently comprise spliced transcripts, speckles and the lamina are highly enriched for transcripts harboring retained introns. Distinct characteristics of the speckle and lamina-associated retained introns suggest that these nuclear structures coordinate the regulation of large networks of functionally linked genes required for multiple stages of gene expression and the cell cycle.

RESULTS

APEX2-labeling of diverse nuclear domains

We generated HEK293 cell lines expressing N- or C-terminal FLAG epitope and APEX2-tagged fusions representing protein markers of speckles (SRSF7, SRSF1, and RNPS1), paraspeckles (NONO, PSPC1, and EWSR1), Cajal bodies (COILIN, SMN2, and WRAP53), PML bodies (PML, SP100), the SAM68 body (KHDRBS1/SAM68) (Mannen et al., 2016), HLBs (NPAT), lamina (LMNA), and the nucleolus (FBL) (final set of markers analyzed shown in Figure 1A; STAR Methods and further on for details). To facilitate uniform expression and control levels, each APEX2-marker protein was stably expressed using the Flp-In system under doxycycline induction (Figure S1A). To control for non-specific labeling, APEX2 fusions containing FLAG epitope, green fluorescent protein (GFP), or an SV40 nuclear localization signal (NLS) were also analyzed.

The APEX2-markers were assessed for their localization and biotin-labeling specificity by immunofluorescence microscopy with anti-FLAG antibody and a modified streptavidin-dye conjugate, respectively. In general, the FLAG and biotin staining patterns concentrated in the expected nuclear domains (Figures 1A, 1B, and S1B). For example, the speckle marker SRSF7-APEX2 specifically biotin-labeled components in speckles, as determined by co-localization with anti-FLAG antibody and SC35 antibody, which specifically immunolabels these domains (Figure 1B and 1C, top row) (Fu and Maniatis, 1990; Ilik et al., 2020). However, biotinylation of domain components was not detected when H_2O_2 was omitted (Figure 1C, bottom row). Specific co-localization between biotinylation and immunofluorescence signals was also observed for most other tested APEX2-markers. Exceptions were those representing paraspeckles and Coilin (STAR Methods for details). Therefore, these markers were excluded from further analysis. It should also be noted that WRAP53, in addition to concentrating in Cajal bodies, displays cytoplasmic localization (Mahmoudi et al., 2010). This could account for significant transcript labeling differences between WRAP53 and SMN2 (see further on). Analysis of the negative controls revealed that APEX2-NLS has a relatively specific and uniform nucleoplasmic distribution, whereas APEX2-GFP and

Figure 1. APEX2-labeling of nuclear domains

(A) Overview of nuclear domain markers analyzed by APEX2-seq. Each indicated marker was N- or C-terminally fused to APEX2 and FLAG epitope (see STAR Methods for details).

(B) Immunofluorescence microscopy localization of SRSF7-APEX2-FLAG and its associated biotin-labeling signal, as detected by anti-FLAG or anti-SRRM2 (SC35) antibodies, and a NeutrAvidin dye conjugate, respectively. Localization of speckles is indicated by detection of endogenous SRRM2. Yellow color indicates overlapping biotin and SC35 signal in the merged images. Scale bars, 5 μm.

(C) Localization of SRSF7-APEX2-FLAG protein and associated biotin labeling in the presence and absence of H_2O_2. Scale bars, 5 μm.

(D) Heatmap showing degree of correlation between peptides detected by mass spectroscopy (MS) following recovery from selected cell lines expressing APEX2 markers concentrated in speckle domains (SRSF7, SRSF1, and RNPS1), PML bodies (SP100), and in lamina (LMNA and LMNB1). Scale, Pearson's correlation coefficient.

(E) Heatmap showing correlation between prey proteins detected by MS following recovery using the APEX2-nuclear marker cell lines shown in (D). Color scale, correlation value as in (D). Correlating prey proteins labeled by speckle, PML-, and lamina-marker-expressing cell lines form three major clusters that are enriched in the gene ontology (GO) terms indicated. The top three most significantly enriched terms are shown. Circle size, numbers of proteins enriched in a GO term; circle color, adjusted p value of the enriched GO term.

(F) qRT-PCR analysis of representative cellular RNAs following recovery from the SRSF7-APEX2-FLAG-expressing cell line in the presence of biotin-phenol (BP), with and without H_2O_2 treatment. RNAs analyzed include MALAT1 (localized to speckles), NEAT1 (localized to paraspeckles), 18S (localized to nucleoli), SSR2 (localized to the endoplasmic reticulum membrane), MTCO2 (localized to mitochondria), and GAPDH (localized to the cytoplasm). Error bars, mean of three biological replicates ±1 standard deviation.

(G) qRT-PCR analysis of representative RNAs as in (F) after recovery from the FLAG-APEX2-FBL-expressing cell line.
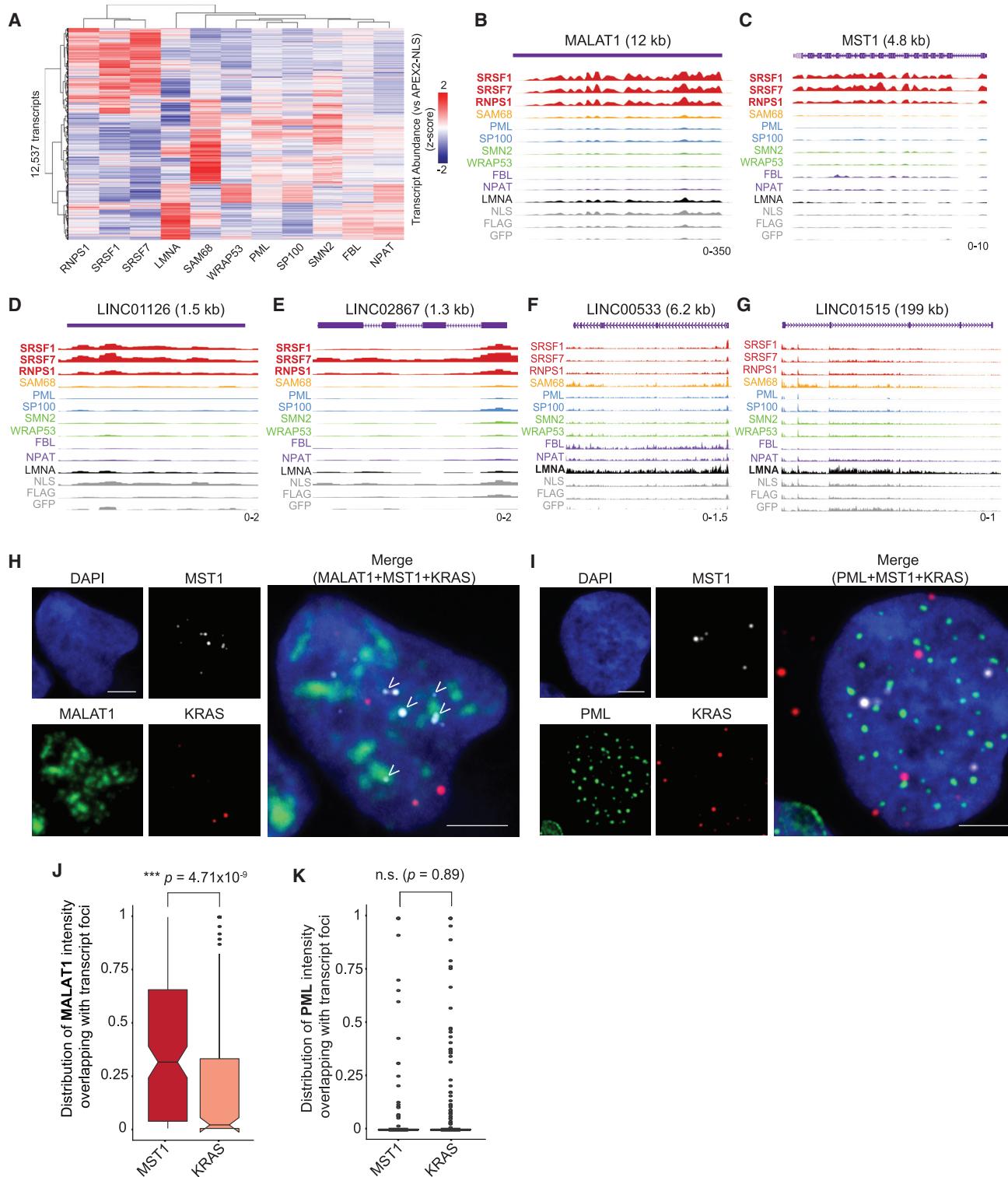
Figure 2. APEX2 RNA-seq analysis of nuclear domain marker-associated RNAs

(A) Heatmap showing unsupervised hierarchical clustering of transcripts recovered from APEX2-nuclear marker-expressing cell lines, over levels recovered from the APEX2-NLS control line, and after normalizing for expression changes in the respective input samples. Scale, relative transcript abundance as z-score (STAR Methods for details).

Molecular Cell
Resource

CellPress

APEX2-FLAG display less intense nuclear localization and substantial cytoplasmic signal (data not shown). Based on these differences and the RNA and protein labeling profiles of the control APEX2 fusions (STAR Methods), and unless stated otherwise, analyses described further on focus on comparisons with the APEX2-NLS control.

To further assess the specificity of biotin labeling, pull-downs followed by mass spectrometry (MS) were performed for a subset of the APEX2-markers, including the speckle proteins SRSF7, SRSF1 and RNPS1, PML body marker SP100, and lamina protein LMNA, with LMNB1 added for further comparison (Table S1). Biotinylated proteins enriched in these pull-downs significantly correlate with each other when the APEX2-markers localize to the same versus different domains (Figures 1D and S1C, p < 7.79×10^{-121}, hypergeometric test). Moreover, they show gene ontology (GO) enrichment for expected functional categories (Figure 1E; Table S2); proteins labeled by speckle markers are enriched in RNA-processing-related GO terms; proteins labeled by SP100 are enriched in PML body, SUMO transferase activity and DNA-damage-related terms; and Lamin-labeled proteins are enriched in terms related to the nuclear periphery (Figure 1E; Table S2). Moreover, the speckle marker-enriched proteins significantly overlap those identified in previous MS analyses of speckle composition (Dopie et al., 2020; Mintz et al., 1999) (Figure S1C; p < 1.64×10^{-41}, hypergeometric test).

To assess the specificity of biotin labeling of RNA components, qRT-PCR assays were performed on several non-coding and protein-coding transcripts (Fazal et al., 2019) recovered from a subset of the APEX2-fusion-expressing cell lines (Figures 1F and 1G). The speckle-localized lncRNA MALAT1 (Hutchinson et al., 2007) is the most highly enriched transcript from cells expressing SRSF7-APEX2 (Figure 1F). Conversely, 18S rRNA is the most enriched transcript recovered from cells expressing the nucleolar marker APEX2-FBL (Figure 1G). Consistent with the microscopy experiments, omission of H_2O_2 greatly reduces the recovery of all analyzed RNAs (Figures 1F and 1G).

Collectively, these and additional results described further on indicate that the APEX2-marker cell lines enable the specific labeling of components in diverse nuclear domains. To systematically investigate the RNA composition of these domains, we performed RNA-seq analysis of biotinylated transcripts recovered from each cell line.

APEX2-seq analysis of nuclear domain associated RNAs
RNA-seq was performed following rRNA depletion to analyze biotin-labeled polyadenylated and non-polyadenylated transcripts recovered from each APEX2-marker-expressing cell line, and also on rRNA-depleted RNA from the respective input

samples, focusing on transcripts longer than ~200 nucleotides. On average, 85% of the RNA-seq reads from each sample were uniquely mapped and biological replicates showed a high degree of expression correlation (Pearson's correlation r > ~0.9), as well as proximal clustering by T-distributed stochastic neighbor embedding (tSNE) analysis (Figures S2A–S2C; data not shown). Consistent with the lack of appreciable changes in cell morphology upon treatment with BP and H_2O_2, transcripts from only four loci showed significant changes following these labeling conditions (data not shown). Figure 2A depicts unsupervised hierarchical clustering of transcripts enriched in pull downs from the APEX2-marker-expressing cell lines, over the level detected in the pull down from the APEX2-NLS cell line, and relative to levels detected in the corresponding input samples (Figures 2A, S3A, and S3B; Table S3). Each APEX2-marker labels a distinct—although in some cases substantially overlapping—subpopulation of transcripts (Figures 2A, S2D, S3A, and S3B). The strongest relative enrichment of transcripts is observed for the three APEX2-speckle markers and SAM68, consistent with the known RNA-binding activities of these proteins, and previous evidence indicating that RNA concentrates in speckles (Dias et al., 2010; Hall et al., 2006; Molenaar et al., 2004). Interestingly, APEX2-LMNA also results in a strong relative enrichment of a subpopulation of transcripts (Figures 2A and S3A).

Consistent with APEX2-SRSF7, APEX2-SRSF1, and APEX2-RNPS1 labeling a common set of speckle-associated RNAs, transcripts enriched from these marker-expressing cell lines significantly overlap with each other but not transcripts recovered from the other cell lines (i.e., 21%–28.5% overlap versus 1%–4.5% overlap; Jaccard similarity index, p < 7.42×10^{-37}, Fisher's exact test) (Figures S2D and S3A). Similarly, the PML body markers PML and SP100 labeled transcripts that are more similar to each other than those labeled by other markers (30% versus 11.7% mean overlap) (p = 1.08×10^{-12}, Fisher's exact test; Figures 2A, S2D, and S3A). Smaller subsets of transcripts are commonly labeled by two or more of the PML (SP100), Cajal body (WRAP53), and histone locus body (NPAT) markers. It is possible that these overlapping patterns may reflect functional interactions between these domains (Boisvert et al., 2007; Imada et al., 2021; Nizami et al., 2010), although we cannot exclude that they also arise through variable cytoplasmic localization or less-specific RNA associations with these APEX2-marker proteins.

To identify candidate transcripts uniquely associated with each nuclear domain, we established a dual-index scoring method, which ranks a transcript based on its relative enrichment in a given APEX2-marker pull-down over all other

---

(B–E) Genome tracks showing raw RNA-seq reads recovered from the APEX2-speckle marker and control cell lines, mapped to loci encoding lncRNAs MALAT1, LINC01126, and LINC02867, and the protein-coding gene MST1. Transcript expression range (read counts per million) is indicated at the bottom left of each panel.

(F and G) Genome tracks showing RNA-seq reads recovered from the FLAG-APEX2-LMNA cell line, mapped to loci encoding lncRNAs LINC00533 and LINC01515. Transcript expression range indicated as in B–E.

(H) RNA-FISH analysis using probes targeting the APEX2-speckle marker-enriched transcript MST1, KRAS (enriched from the FLAG-APEX2-SAM68 cell line), and probes targeting MALAT1 for comparison. Arrowheads, representative RNA foci overlapping MALAT1 signal in merged images. Scale bars, 5 μm.

(I) RNA-FISH analysis using probes targeting MST1 and KRAS as in (H), in comparison with immunolocalization of PML bodies. Scale bars, 5 μm.

(J and K) Boxplots quantifying distributions of MALAT1 (J) or PML intensities (K) (normalized to a 0–1 scale) overlapping RNA-FISH foci detected using probes targeting MST1 and KRAS transcripts. The data represent quantification of 172 MST1 and 239 KRAS foci from 72 cells localized for MALAT1, and 123 MST1 and 234 KRAS foci from 60 cells localized for PML (***p < 0.001, Wilcoxon rank-sum test).

pull-downs, including the three negative control APEX2 samples (index 1), and based on the extent to which the expression of a transcript is altered in response to induced expression of the APEX2-marker in the corresponding input sample (index 2) (Table S4; see STAR Methods for details). Thus, transcripts with high ranks for both indices are considered strongly associated with a nuclear domain independent of any possible expression change in input samples, which in any case is infrequently observed (Figure S3B). A full list of dual-index-ranked transcripts labeled by each nuclear domain marker is provided in Table S4.

As expected, MALAT1 has a high dual-index score (0.96:0.85) across the three nuclear speckle markers. Interestingly, ~150 additional lncRNAs and transcripts from ~850 protein-coding genes have comparable dual-index scores as MALAT1. Representative examples of transcripts preferentially labeled by speckle and LMNA markers are highlighted by genome tracks of reads mapped from the APEX2-RNA-seq data (Figures 2B–2G; Table S4). RNA fluorescence *in situ* hybridization (FISH) analysis further indicates that transcripts from the protein-coding gene MST1, which are preferentially labeled by all three APEX2-speckle markers (Figure 2C), show a significantly greater degree of overlap with these domains, as revealed by MALAT1 detection, compared with transcripts from the KRAS locus, which are preferentially labeled by the APEX2-SAM68 marker (p = 4.7 × 10^{-9}, Wilcoxon rank-sum test; Figures 2H–2K and S3C–S3G). Similar results were obtained when analyzing the overlap between MST1 transcript foci and speckle domains detected using anti-SRRM2 antibody (Blencowe et al., 2000) (Figure S3E). Moreover, MST1 transcripts do not show significantly greater overlap with PML bodies compared with KRAS transcripts (Figures 2I–2K). Additional RNA-FISH data supporting the expected localization specificity of the nuclear marker-associated transcripts detected by APEX2-seq are presented later (see Figures 5 and S5). Collectively, the results indicate that markers for different nuclear domains are associated with distinct and in some cases partially overlapping subpopulations of transcripts.

Transcriptomic and genomic features associated with nuclear domain markers

Next, we investigated features of the transcripts specifically labeled by the APEX2-nuclear domain markers. Speckle-associated coding transcripts have a significantly higher GC content compared with coding transcripts associated with the other nuclear domain markers, as well as the average percent GC of all nuclear-expressed transcripts (Figure 3A; p < 3.22 × 10^{-24}, Wilcoxon rank-sum test). On average, the speckle marker-associated transcripts are also significantly shorter than those labeled by the other nuclear markers (Figure 3B; p < 4.47 × 10^{-17}, Wilcoxon rank-sum test). LncRNAs associated with the speckle markers also show significantly increased GC-content compared with lncRNAs associated with the other markers, and the average GC content of all detected nuclear lncRNAs (p < 1.21 × 10^{-6}, Wilcoxon rank-sum test), whereas their transcript length is comparable (Figures S4A and S4B). Binning transcripts associated with each domain marker based on the magnitude of their index 1 scores reveals that percent GC-con-

tent and length relate to the degree of relative enrichment in the pull-downs (Figures S4C–S4H).

Next, we investigated characteristics of the genomic loci encoding transcripts detected in association with the domain markers. Previously, the tyramide signal amplification sequencing (TSA-seq) procedure was used to label DNA proximal to the speckle protein SON, providing an approximate measure of intranuclear distances of individual gene loci from speckles (Chen et al., 2018). Reanalyzing these data, we quantified overlaps between domain-associated transcripts detected by APEX2-seq labeling and positions of their corresponding genomic loci, binned according to their TSA-seq-mapped relative distance to the SON protein (Figure 3C). Transcripts preferentially labeled by the three APEX2-speckle markers show the highest degree of overlap with corresponding loci that, from TSA-seq mapping data, are the most proximal to speckles; ~69% of speckle-associated transcript-coding loci overlap with the top three most speckle-proximal bins as measured by TSA-seq, whereas ~24% of LMNA transcript loci overlap with these bins (Figure 3C; p < 9.77 × 10^{-72}, Fisher's exact test). Conversely, ~26% of the LMNA-associated loci overlap with the three most speckle-distal bins assigned by TSA-seq, compared with a ~6% overlap between loci encoding the speckle-associated transcripts and these bins (Figure 3C; p < 4.45 × 10^{-159}, Fisher's exact test, Table S5).

We also investigated the genomic compartments overlapping loci encoding transcripts associated with each nuclear domain marker. The chromosome conformation capture method Hi–C has been used to map open and closed chromatin regions, referred to as A-type and B-type genomic compartments, respectively (Lieberman-Aiden et al., 2009). A-type compartments have been further subdivided into A1- and A2-types, representing candidate nuclear speckle-associated and other active chromatin regions, respectively. In contrast, B-type compartments have been sub-divided into B1–B3 types, representing loci associated with polycomb bodies (B1), nucleolar and lamin-associated domains (B2), as well as HP1a heterochromatin (B3), which is associated with the nuclear periphery in some cell types (Hildebrand and Dekker, 2020; Rao et al., 2014; Xiong and Ma, 2019). We observe a higher degree (~49%) of overlap between speckle marker-associated transcripts and their corresponding coding loci when encompassed within A1 compartments, compared with a ~12% overlap between A1 compartment loci corresponding to LMNA-associated transcripts (Figure 3D; p < 2.06 × 10^{-128}, Fisher's exact test). Moreover, transcripts labeled by APEX2-LMNA show a 32% overlap with loci within B2 and B3 sub-compartments compared with an 11% overlap for speckle transcript loci within these sub-compartments (Figure 3D; p < 1.96 × 10^{-83}, Fisher's exact test, Table S6).

Together, these analyses support a close interrelationship between the RNA composition of nuclear domains and the conformation and gene content of proximal chromatin, i.e., speckle-associated transcripts primarily derive from proximal "open" chromatin, whereas transcripts associated with various other nuclear domains are more often associated with condensed chromatin that is distal to speckles (Kim et al., 2020; Su et al., 2020). Furthermore, the significant overlapbetween the speckle marker-associated transcripts and TSA-seq-mapped

Molecular Cell
Resource

CellPress
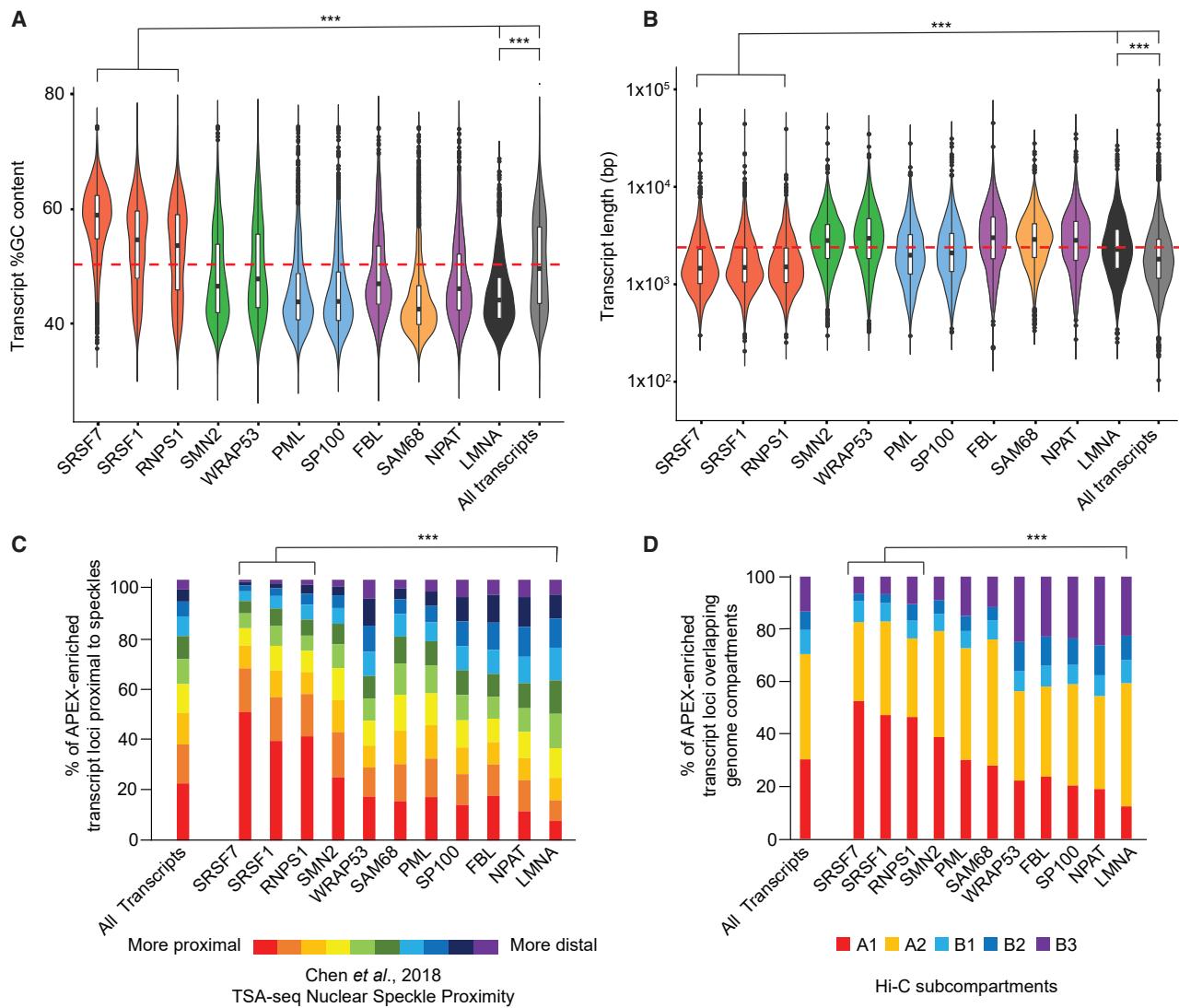
Figure 3. Transcriptomic and genomic features associated with nuclear domain markers

(A) Violin plots comparing distributions of GC content of coding transcripts recovered from APEX2-marker-expressing cell line. Dashed line, average %GC of all analyzed coding transcripts (***p < 0.001, Wilcoxon rank-sum test).

(B) Violin plots comparing distributions of lengths of coding transcripts recovered from APEX2- marker-expressing cell lines. Dashed line, average length of all analyzed coding transcripts (***p < 0.001, Wilcoxon rank-sum test).

(C) Bar plot showing distribution of coding DNA loci of APEX-detected transcripts with respect to their proximity to nuclear speckles, as assessed by tyramide signal amplification sequencing (TSA-seq) (Chen and Belmont, 2019) (***p < 0.001, Fisher's exact test).

(D) Bar plot showing distribution of coding DNA loci of APEX2-seq-detected transcripts with respect to their overlap with mapped Hi–C genomic sub-compartments (Xiong and Ma, 2019) (***p < 0.001, Fisher's exact test).

speckle-proximal loci encoding these transcripts provides additional evidence that the APEX2-seq mapping data represent the specific RNA content of these nuclear domains.

Nuclear domains associated with intron retention

Next, we used the APEX2-seq data to investigate splice variants associated with nuclear domains. Transcripts enriched in each APEX2 marker pull-down were analyzed for differential usage of cassette alternative exons (CE), alternative 3′ or 5′ splice sites (Alt3, Alt5), and intron retention (IR) events by comparing PSI

(percentage of transcripts from a gene with an exon sequence spliced in) and percent intron retention (PIR) (percentage of transcripts from a gene with the intron retained) values between the APEX2-marker and APEX2-NLS pull downs, after normalizing for possible PSI or PIR changes between the corresponding input samples (Figures 4A, S5A, and S5B).

Unsupervised hierarchical clustering of the data based on the magnitude of the PSI/PIR differences reveals particularly striking patterns of differential IR (Figures 4A, and S5A–S5C; Table S7). The highest frequencies of IR are detected in transcripts
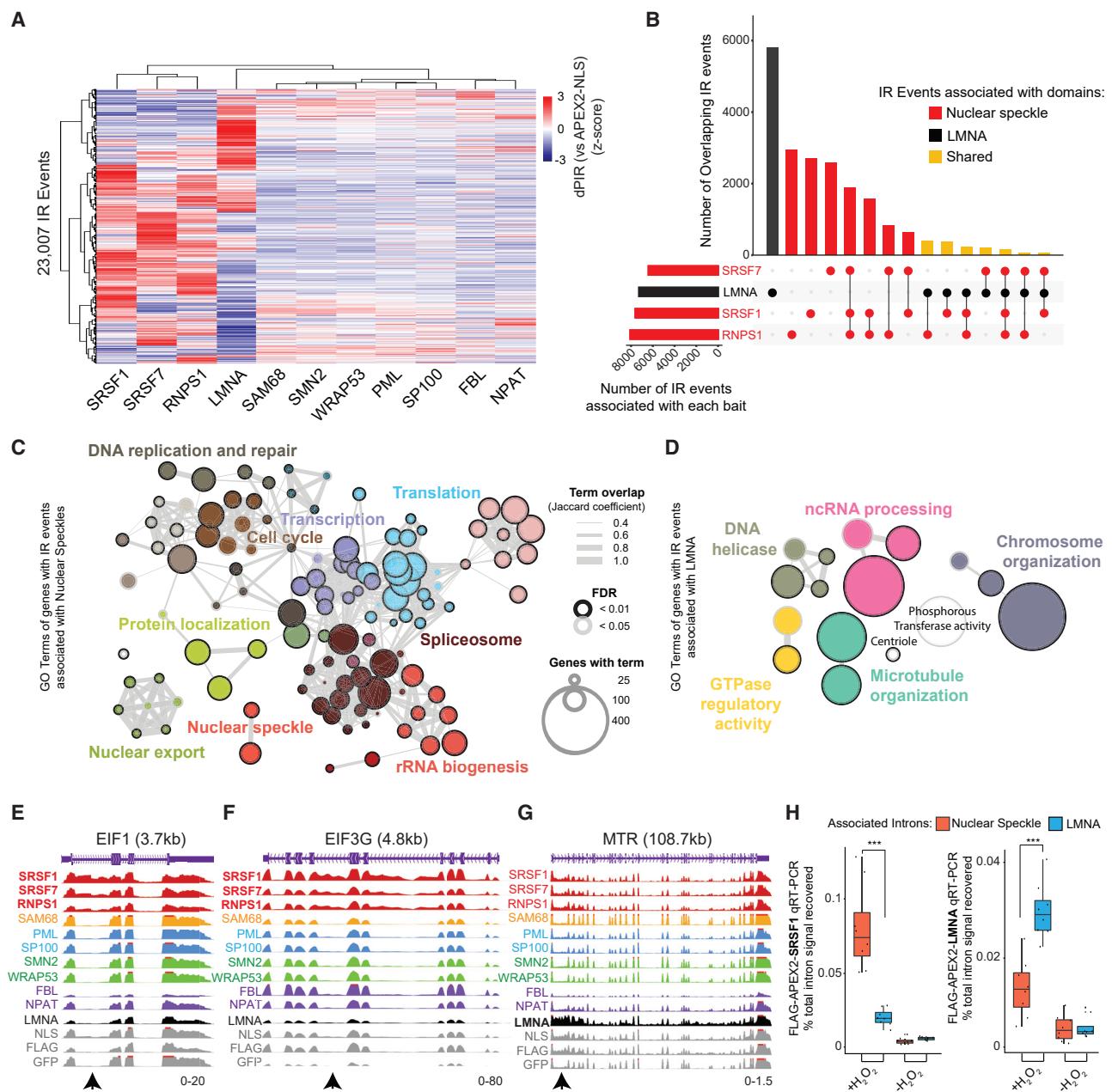
Figure 4. Analysis of intron retention (IR) events associated with nuclear domains

(A) Heatmap showing unsupervised hierarchical clustering of differential IR patterns between transcripts recovered from APEX2-marker-expressing cell lines (main text and STAR Methods for details). Scale, difference in percent intron retention (dPIR, the percentage of total transcript with the intron retained) expressed as a z-score.

(B) Upset plot showing overlaps between IR events detected in transcripts enriched from cells expressing APEX2-speckle and -LMNA markers.

(C and D) GO terms enriched in genes with IR events detected in association with the speckle (C) and LMNA (D) markers. Circle sizes and borders indicate gene numbers associated with a GO term and false discovery rate (FDR) cutoff, respectively. Edge thickness indicates degree of overlap between genes annotated with different terms.

(E–G) Representative genome tracks showing RNA-seq reads recovered from the APEX2-domain marker and control cell lines, highlighting retained introns enriched with the speckle markers (from EIF1 and EIF3G translation initiation factor genes), and LMNA (MTR gene). Arrowheads, introns detected as retained and analyzed by RNA-FISH in Figures 5A, 5B, and S5E–S5I.

(H) Boxplots showing qRT-PCR quantification of IR events detected in biotinylated transcripts enriched from FLAG-APEX2-SRSF1 (left) and FLAG-APEX2-LMNA (right)-expressing cell lines. Each plot shows the mean of enrichment of eight analyzed IR events, ±1 standard deviation (***$p < 0.001$, Student's t test).
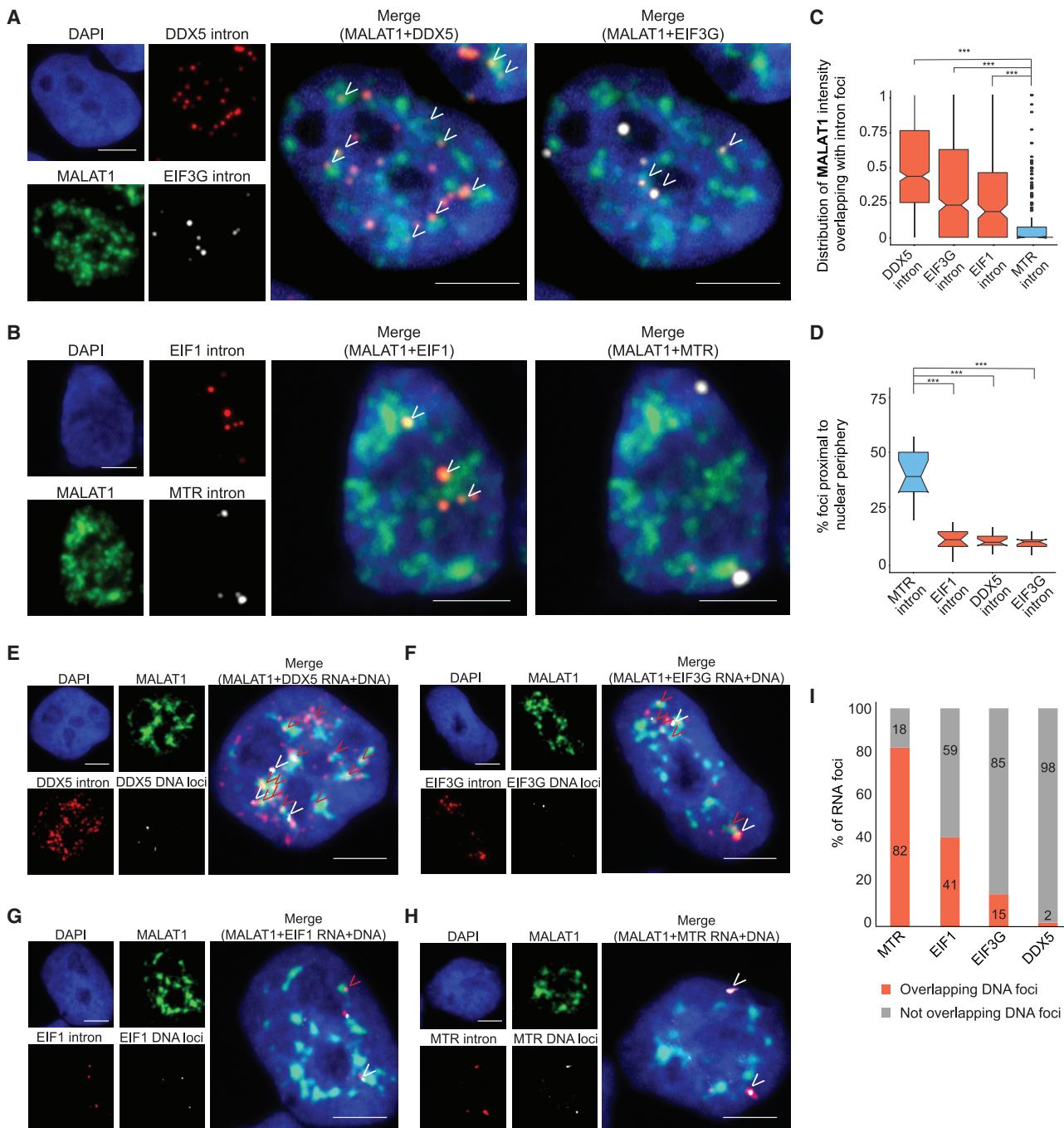
Figure 5. FISH analysis of speckle and lamin-associated retained introns and their corresponding gene loci

(A) RNA-FISH analysis (using RNAScope probes) of APEX2-speckle marker-associated retained introns in DDX5 and EIF3G transcripts. MALAT1 is also probed to mark speckles. Arrowheads, representative overlaps. Scale bars, 5 μm.

(B) RNA-FISH detection as in (A) of speckle marker-associated retained intron in EIF1 transcripts, and LMNA-associated retained intron in MTR transcripts. Arrowheads, representative overlaps with MALAT1 signal. Scale bars, 5 μm.

(C) Boxplots quantifying the distribution of MALAT1 intensities (normalized to a 0–1 scale) overlapping RNA-FISH foci corresponding to retained introns in EIF1, EIF3G, DDX5, and MTR transcripts. Data represent quantification of 215 EIF1, 519 EIF3G, 627 DDX5, and 204 MTR intron foci from 71–73 cells also probed for MALAT1. (***p value: Wilcoxon rank-sum test).

(D) Boxplots quantifying percentages of MTR, EIF1, EIF3G, and DDX5 intron foci associated with the nuclear periphery (see STAR Methods) (***p < 0.001, Wilcoxon rank-sum test).

(legend continued on next page)

associated with the speckle markers and LMNA. These differences in PIR are not a consequence of expression of the APEX2 nuclear markers, since only minor (i.e., 0.6%–5.6%) changes in retention of the corresponding introns are observed between the input samples (Figure S5C). Importantly, IR events associated with the speckle markers show a significantly higher degree of overlap with each other (24.1%–32.4%, Jaccard similarity index), compared with their overlap with IR events associated with LMNA or the other nuclear domain marker pull downs (0%–5.5%; Figures 4B and S5D; p < 2.25 × 10^{-308}, Fisher's exact test).

Interestingly, IR events associated with LMNA are significantly enriched in genes that function in microtubule organization, chromosome organization, and ncRNA processing, whereas speckle marker-associated retained introns are enriched in genes that function in RNA processing, translation, and the cell cycle (Figures 4C and 4D; Table S8). Strikingly, the speckle-associated retained introns are detected in numerous factors that function in translation initiation, such as EIF1, EIF2B1, EIF3A, EIF3G, EIF4G1, and EIF5 (p = 0.029, Fisher's exact test; Table S8). Genome browser tracks highlighting the specificity of IR read mapping are shown for representative examples of the speckle and LMNA-associated retained intron genes (Figures 4E–4G), and qRT-PCR assays further validated the specificity of the APEX2-seq mapping results for retained introns (Figure 4H).

Localization of speckle and lamin-associated retained introns and their corresponding gene loci
Confirming the localization of retained introns detected by APEX2-seq, RNA-FISH probes targeting speckle-associated retained introns in DDX5, EIF1, and EIF3G transcripts show a significantly greater degree of co-localization with speckles, as detected by FISH of MALAT1 or immunostaining with anti-SRRM2 antibody, than does a probe targeting the LMNA-associated retained MTR intron (p < 3.09 × 10^{-16}, Wilcoxon rank-sum test; Figures 5A–5C, S3E, S5E, and S5F). Conversely, foci corresponding to the lamina-associated MTR intron show a significantly higher degree of co-localization with the nuclear periphery than foci representing the speckle-associated retained introns (Figure 5D, p < 1.84 × 10^{-19}, Fisher's exact test; see STAR Methods). Further supporting the specificity of these localization patterns, the EIF1, EIF3G, and MTR intron probe signals do not show significant overlap with PML bodies (Figures S5G–S5I). Moreover, although relatively large numbers of DDX5 intron foci are detected, likely reflecting the higher abundance of this intron, as detected by read counts from input RNA-seq data, these foci also display a significantly higher degree of overlap with MALAT1-speckles as compared with PML bodies (p = 0.01, Wilcoxon rank-sum test; Figures S5E–S5I, and data not shown).

To further characterize the MTR, EIF1, EIF3G, and DDX5 intronic foci, we next investigated their spatial relationship with

the corresponding gene loci, using combined RNA-FISH and DNA-FISH analysis (Figures 5E–5H). Consistent with the results described earlier, and an extensive body of literature showing that relatively long introns are typically co-transcriptionally spliced (Gordon et al., 2021), we observe that 81% of the MTR retained intron foci, of which there are generally two to four per nucleus, overlap MTR gene loci, which accordingly are also concentrated at the nuclear periphery (Figures 5H and 5I). In contrast, 41%, 15%, and 2% of EIF1, EIF3G, and DDX5 intronic foci, respectively, overlap with their gene loci (Figures 5E–5I). These loci are frequently observed adjacent to speckles, consistent with evidence that active transcription and pre-mRNA processing, possibly involving nascent RNA polymerase II transcripts, may often occur at the periphery of these domains (Chen and Belmont, 2019; Spector and Lamond, 2011). However, additional intronic RNA foci that do not coincide with their DNA loci, in particular from the DDX5 gene, overlap or are at the edge of speckle domains (Figures 5E–5H). Therefore, these foci likely derive from retained introns within transcriptionally released but incompletely processed RNA and possibly also from populations of relatively stable excised introns.

Collectively, the results provide evidence that speckles and lamina are preferentially associated with transcripts harboring functionally distinct sets of retained introns, and moreover that different subsets of RNA foci corresponding to retained intron transcripts concentrate at these nuclear structures through mechanisms that are dependent and independent of proximal gene loci.

Compositional and regulatory features of nuclear domain-associated retained introns
Retained introns comprise different subtypes, including the following: those that are (1) associated with otherwise spliced and polyadenylated transcripts exported to the cytoplasm, where the retained intron may elicit nonsense-mediated decay (NMD) or control cellular targeting of a transcript; (2) associated with nuclear or cytoplasmic turnover through other degradation pathways to remove spurious transcripts from cells; and (3) associated with relatively stable nuclear transcripts but which may be spliced out to facilitate export and translation in response to specific signaling events (Boutz et al., 2015; Braunschweig et al., 2014; Monteuuis et al., 2019). Introns comprising the latter class have been referred to as "detained introns" (Boutz et al., 2015; Braun et al., 2017). These tend to be relatively short, conserved, and exhibit elevated GC content and PIR levels. Next, we investigated which of these subtypes of retained introns are associated with the speckle markers and LMNA.

Retained introns associated with the speckle markers are significantly shorter, have a higher GC content, and display elevated PIR levels, when compared with the LMNA-associated retained introns and a set of retained introns randomly selected from those detected in the APEX2-NLS pull down control

---

(E–H) Combined RNA- and DNA FISH localization of retained introns and corresponding gene loci for speckle-associated retained introns from the DDX5 (E), EIF3G (F), and EIF1 (G) genes and the LMNA-associated MTR intron (H). MALAT1 is detected to mark speckles. White arrows, representative overlapping RNA and DNA foci; red arrows, representative RNA foci overlapping MALAT1-speckles. Scale bars, 5 μm.

(I) Bar graph showing percentages of RNA foci for each indicated gene that do or do not overlap with their corresponding gene loci. The data represent quantification of 416 MTR, 375 EIF1, 1054 EIF3G, and 6921 DDX5 RNA foci from 85–112 cells.
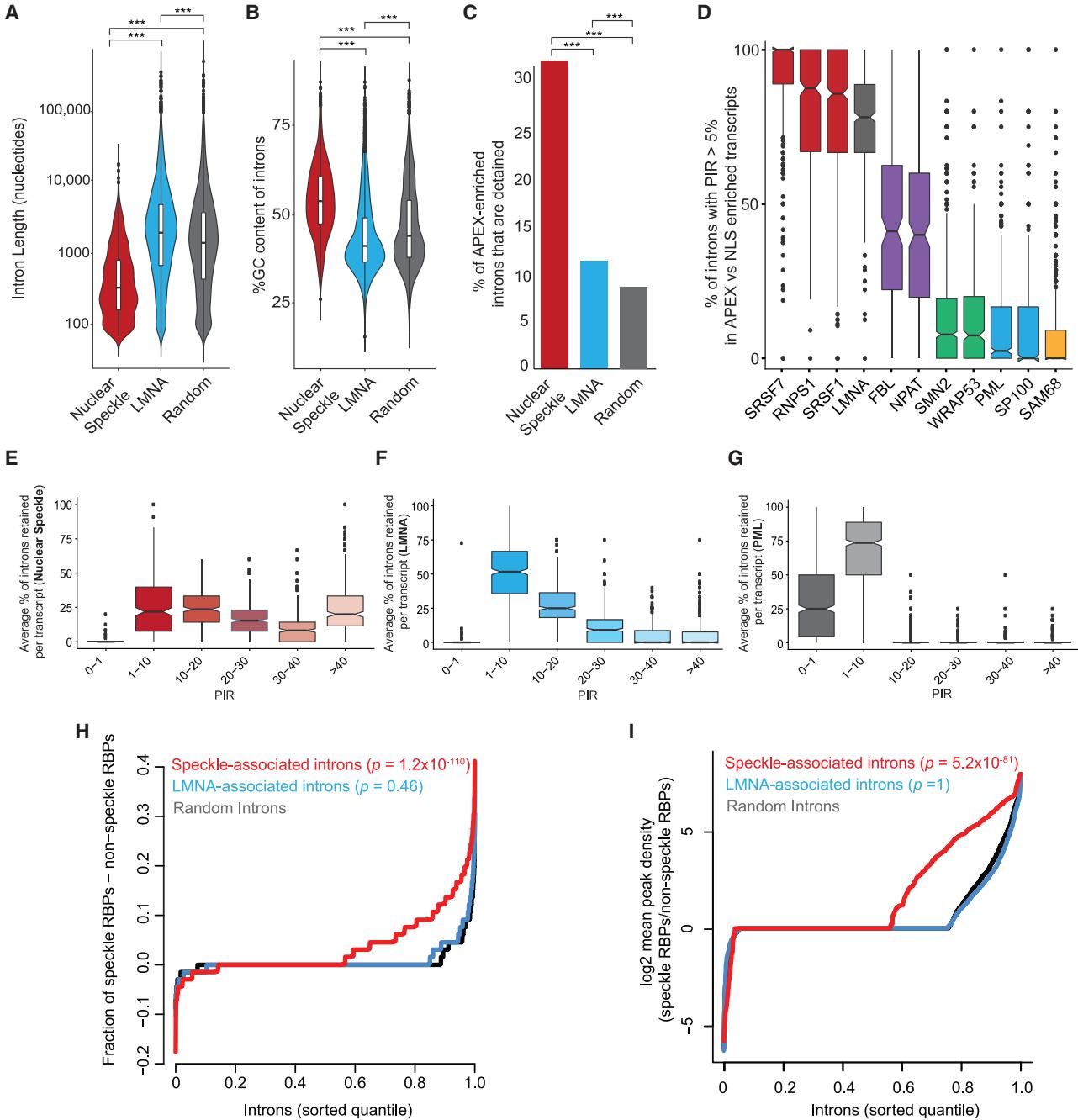
Figure 6. Composition and features of nuclear domain-associated retained introns

(A) Violin plot showing length distributions ($\log_{10}$ scale) of retained introns associated with APEX2-speckle markers, APEX2-LMNA, and a random set of IR events (***$p < 0.001$, Wilcoxon rank-sum test).

(B) Violin plots showing distribution of GC content of retained introns associated with APEX2-speckle markers, APEX2-LMNA, and a random set of IR events (***$p < 0.001$, Wilcoxon rank-sum test).

(C) Bar plot showing percentage of annotated detained introns (Boutz et al., 2015) that overlap with speckle marker-associated retained introns, LMNA-associated retained introns, and a random set of IR events. (***$p < 0.001$, Fisher's exact test.)

(D) Boxplot showing percentage of introns with retention (PIR > 5) detected in transcripts enriched from each APEX2-marker-expressing cell line.

(E–G) Boxplots showing percentage of introns with different IR levels within transcripts associated with nuclear speckle markers (E), LMNA (F), and PML (G).

samples (Figures 6A–6C, p < 1.45 × 10^{-182}, Wilcoxon rank-sum test). In contrast, the LMNA-associated retained introns are significantly longer and have a reduced GC content when compared with the randomly selected retained introns (Figures 6A and 6B, p < 9.77 × 10^{-41}, Wilcoxon rank-sum test). Given that the speckle-associated retained introns bear hallmark features of detained introns, we assessed the overlap between these two groups. Notably, approximately 31% of the retained introns associated with all three speckle markers overlap annotated detained introns (Boutz et al., 2015), as compared with ~12% and 8% overlaps with the LMNA-associated and randomly selected retained introns, respectively (Figure 6C; p = 3.93 × 10^{-62}, Fisher's exact test).

Next, we assessed the extent to which the speckle and LMNA-associated transcripts are processed. When calculating the percentages of total introns per gene that show retention in each pull-down, as expected, the highest frequencies of IR are detected in transcripts associated with the three speckle markers (79.7%–91.6%) and LMNA (76.0%) versus 6.2%–10.4% associated with the other markers (p < 2.48 × 10^{-158}, Wilcoxon rank-sum test), which have correspondingly higher frequencies of spliced junctions (Figures 6D and S6A). Moreover, introns associated with the speckle markers generally display more variable degrees of retention compared with those belonging to LMNA-associated transcripts (Figures 6E–6G and S6B).

The extent of processing of the speckle-associated transcripts appears to depend in part on the relative distances of their corresponding loci from these domains. Specifically, speckle-associated transcripts originating from DNA loci that, by TSA-mapping, are the most speckle-proximal (i.e., the first decile bin in Figure 3C), have higher frequencies of IR than speckle-associated transcripts from loci that are the most distal from speckles (the last decile bin in Figure 3C) (Figures S6C–S6F). Taken together, these observations provide evidence that speckles and lamina are associated with structurally and functionally distinct subpopulations of retained introns, and that the frequency of IR depends on the distance between transcripts accumulating at speckles and their sites of transcription. Moreover, consistent with the DNA-FISH and RNA-FISH analyses in Figure 5, and additional data from polyA+ sequencing of transcripts enriched from the APEX2-SRSF7-expressing cell line in which a high frequency of IR is also observed (data not shown), the results suggest that a substantial fraction of partially retained intron-containing RNAs associated with speckles are unlikely derived from nascent transcripts.

Next, we investigated functional binding associations between RNA-binding proteins (RBPs) and the speckle- and LMNA-associated retained introns. From analyzing ENCODE knockdown RNA-seq, CLIP-seq, and localization data for 227 RBPs (Van Nostrand et al., 2020), we observe that speckle-associated introns, compared with LMNA-associated introns, are

significantly enriched for splicing changes upon knockdown of RBPs that localize to speckles, as compared with RBPs that are nuclear but that do not overlap with speckles (Figures 6H and S6G; p < 1.12 × 10^{-110}, Wilcoxon rank-sum test). Consistently, speckle-associated retained introns, compared with LMNA-associated introns, are significantly enriched for CLIP-seq binding peaks of speckle-localized RBPs (Figures 6I and S6H, p < 2.8 × 10^{-16}, Wilcoxon rank-sum test). For example, the U2 snRNP auxiliary factor (U2AF1), components of the U2 snRNP SF3 complex (e.g., SF3B4 and SF3A3), and U5 snRNP and core spliceosomal protein PRPF8, are among the RBPs with the highest relative enrichment of binding peaks over speckle-associated introns (Table S9). These findings suggest that speckle-associated retained introns may concentrate within these nuclear domains at least in part through their propensity to bind RBPs that have increased residency in speckles. In contrast, LMNA-associated retained introns are not significantly enriched for CLIP-seq binding peaks of splicing factors and other RBPs that are concentrated in speckles (Figures 6H, 6I, S6G, and S6H).

Functions of speckle-associated retained introns

To investigate functional attributes of the nuclear domain-associated retained introns, we compared their regulatory characteristics across diverse cellular contexts. Consistent with their broader range of PIR levels (Figure 6E), and overlap with detained introns (Figure 6C), the speckle-associated retained introns generally display higher and larger differential inclusion levels across diverse human cell and tissue types, as compared with the LMNA-associated and a randomly selected control set of introns (Figure 7A; see legend for details). In contrast, the oocyte, early embryo, and ES cells show low overall levels of IR across all three groups of introns. It is interesting to consider that these observations may relate to previous findings indicating that the formation of speckles is associated with mammalian cell differentiation, as well as overall differences in the transcriptional outputs of cells (Gordon et al., 2021; Spector and Lamond, 2011).

To explore a possible relationship between the integrity of nuclear speckles and the splicing of their associated retained introns, we analyzed the RNA-seq data generated following knockdown of SON (Lu et al., 2013), which functions in the formation of speckles (Ilik et al., 2020; Lu et al., 2013; Sharma et al., 2010). Strikingly, knockdown of SON results in frequent increases in IR (Lu et al., 2013), and these SON knockdown-dependent events significantly overlap those we have detected in association with speckles but not the IR events associated with LMNA (Figure 7B) (5.13%–9.87% versus 1.82%; p < 7.97 × 10^{-23}, Fisher's exact test).

Speckles disassemble during mitosis to form mitotic interchromatin granules during the cell cycle (Ferreira et al., 1994;

(H) Cumulative distribution plot comparing fractions of speckle-localized RNA-binding proteins (RBPs) versus non-speckle localized nuclear RBPs whose depletion causes increased retention of introns associated with APEX2-speckle markers (red), LMNA (blue), and a random set of retained introns (gray). Knockdown RNA-seq data from HepG2 cells from Van Nostrand et al. (2020) (p -value: Wilcoxon rank-sum test).

(I) Cumulative distribution plot comparing ratios of eCLIP-seq peaks from speckle-localized RNA-binding proteins (RBPs) versus non-speckle-localized nuclear RBPs, within retained introns detected in association with the APEX2-nuclear speckle markers (red), LMNA (blue), and a random set of retained introns (gray). CLIP-seq data from HepG2 cells from (Van Nostrand et al., 2020) (p-value: Wilcoxon rank-sum test).
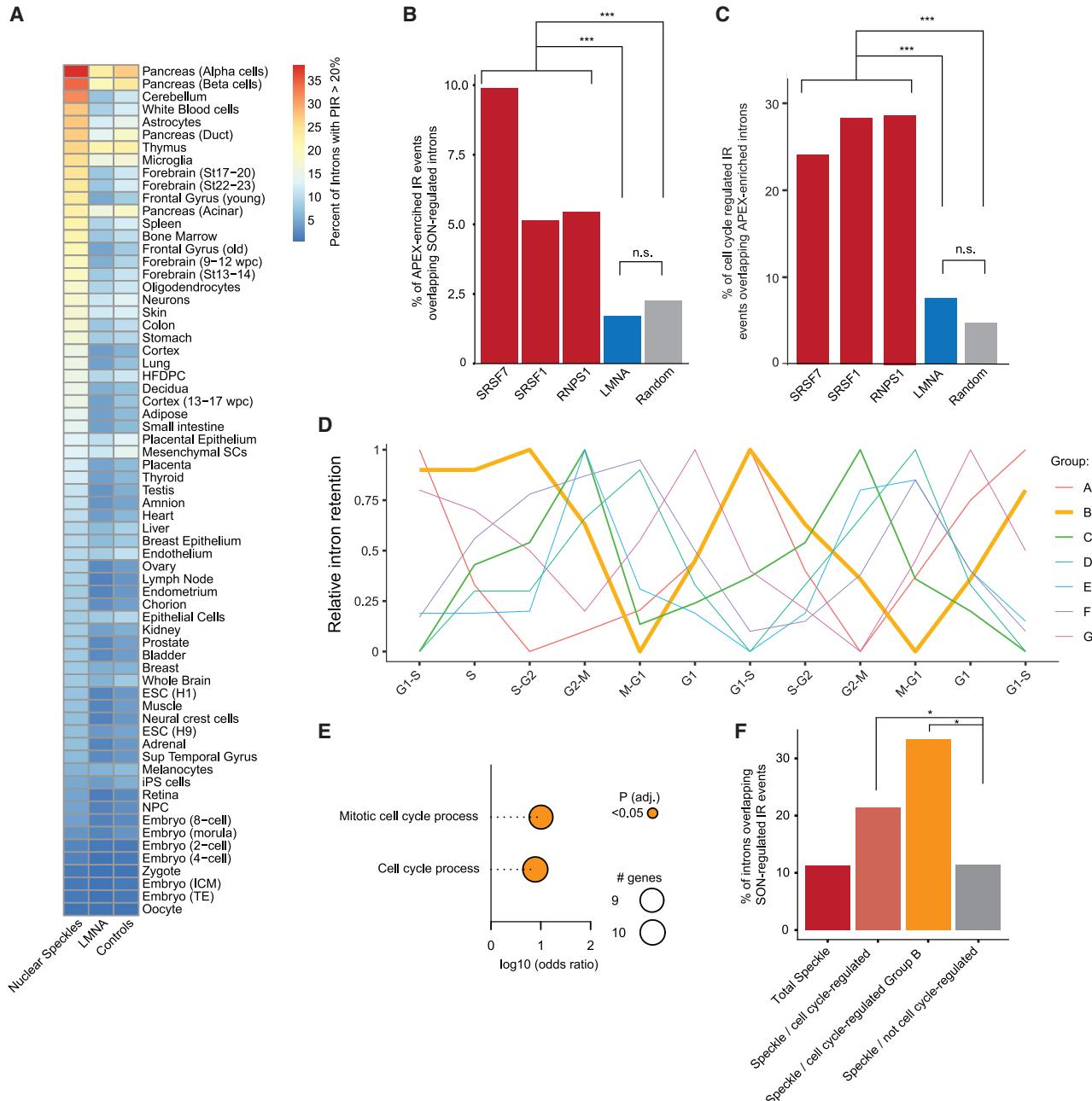
Figure 7. Regulation and function of speckle-associated retained introns

(A) Heatmap showing percentages of introns with PIR > 20% detected in various cell and tissue types (data from VASTDB [Tapial et al., 2017]) that correspond to speckle- or LMNA-associated introns, as well as a randomly selected set of control retained introns matched to have a similar distribution of median PIR in VASTDB as the union of speckle- and LMNA-associated introns for comparison. St, Carnegie stage of brain development; wpc, weeks post conception; (E) SC, (embryonic) stem cell; iPS cells, induced pluripotent cells; NPC, neural progenitor cells; ICM, inner cell mass; TE, trophectoderm.

(B) Bar graph showing percentage of retained introns associated with the APEX2-markers SRSF1, SRSF7, RNPS1, LMNA, and a randomly selected control set of retained introns, which overlap retained introns regulated by SON ([Lu et al., 2013]). (***p < 0.001, Fisher's exact test)

(C) Bar graph showing percentage of cell-cycle-regulated IR events ([Dominguez et al., 2016]) associated with different APEX2-markers. (***, p < 0.001, Fisher's exact test)

(D) Line plot showing clusters of introns that display distinct patterns of periodic change in IR splicing throughout the cell cycle. Line thickness represents the fraction of overlap between IR events in each cluster and speckle marker-associated IR events: thin lines, 4% (lowest) overlap; thick lines; 92%, (highest) overlap.

Tripathi and Parnaik, 2008). Therefore, we investigated whether the cell cycle is linked to the differential regulation of speckle-associated retained introns. Remarkably, retained introns associated with the speckle markers, but not LMNA, significantly overlap with retained introns that display differential regulation across the cell cycle (Figure 7C, $p < 1.41 \times 10^{-20}$, Fisher's exact test) (Dominguez et al., 2016). Of note, a subgroup (Group B) of the speckle-associated introns shows a marked decrease in PIR during the G2-M and M-G1 transitions, phases that coincide with the disassembly of speckles during mitosis (Figure 7D; Ferreira et al., 1994; Tripathi and Parnaik, 2008). These speckle-associated retained introns represent genes that are enriched in functional terms related to cell-cycle control (Figure 7E, $p < 0.021$, Fisher's exact test). Furthermore, these retained introns are also significantly enriched among those that show increased PIR upon knockdown of SON ($p = 0.013$, Fisher's exact test), which, in addition to perturbing speckle integrity, causes cell-cycle arrest (Ahn et al., 2011; Lu et al., 2013; Sharma et al., 2010). Taken together, the results suggest that the formation of speckles, the splicing of their associated retained introns, and cell-cycle progression, are coupled to one another.

DISCUSSION

Our results collectively support a role for speckles and lamina as hubs of distinct classes of retained introns associated with the coordinated regulation of multiple steps in gene expression and the cell cycle. Although a subset of retained introns associated with speckles co-localize with or are proximal to their corresponding genes, we also detect speckle-localized retained introns that are distal from their gene loci, consistent with growing evidence for speckle-proximal co-transcriptional splicing and speckle-coincident post-transcriptional splicing (Gordon et al., 2021). It is possible that gene loci and corresponding transcripts harboring retained introns are positioned proximal to speckles to ensure efficient and regulated processing and as a quality control step to ensure complete processing prior to mRNA export. Moreover, it is interesting to consider that the speckle-enriched retained introns may contribute to the formation of these structures, potentially by promoting coalescence with bound RBPs.

The distinct spatial relationships between nuclear domains, gene loci, and their corresponding transcripts, may have an important quantitative and qualitative impact on the splicing of proximal introns. Notably, we observe that speckles are enriched for relatively short and high-GC content introns with increased retention levels, which presumably are more often spliced by intron definition mechanisms (Amit et al., 2012). In contrast, genes representing lamin-proximal transcripts have relatively long and low-GC content introns and are likely to be more often spliced by exon definition-type mechanisms. Consistently, results from a concurrent study reveal that the peripheral and cen-

tral regions of the nucleus harbor genes and corresponding transcripts with similar differential intron length, GC content, and predicted splicing outcomes (Tammer et al., 2022). The strong enrichment of speckle-associated retained introns in genes with functions in transcription, RNA processing, and translation initiation, further suggests that they could contribute to establishing cell fate and maintenance by coordinating multiple regulatory networks in diverse contexts. Moreover, the extensive overlap of these retained introns with "detained" introns suggests that many of them may be spliced in rapid response to changes in cell growth conditions (Boutz et al., 2015; Braun et al., 2017).

Further supporting the importance of the speckle-associated retained introns is that a subset are differentially regulated and function during the cell cycle. Examples are in transcripts encoding the mitotic checkpoint kinase AURKB and the CDC2-like dual specificity kinases CLK1 and CLK4, which control phosphorylation levels of SR proteins, thereby impacting their associations with speckles, as well as cell growth and proliferation (Colwill et al., 1996; Dominguez et al., 2016; Sacco-Bubulya and Spector, 2002). Based on these observations, we propose that the speckle-associated retained introns are intimately connected with cell-cycle regulation. Notably, knockdown of SON, which perturbs speckle integrity and inhibits cell-cycle progression (Ahn et al., 2011; Lu et al., 2013; Sharma et al., 2010), increases retention of many of the same introns that show increased splicing during the M-G1 cell-cycle transition, which coincides with speckle reassembly following metaphase. Thus, speckle formation appears to be coupled to the regulation of speckle-associated retained introns that are linked to control of the cell cycle.

Limitations of the study

Consistent with evidence that APEX2-fusions label cellular components within an approximate 20 nm radius (Fazal et al., 2019; Hung et al., 2016), we observe discrete labeling of targeted nuclear domains by microscopy, and through detection of significantly overlapping profiles of RNA and proteins labeled by markers residing within the same domains versus different domains. However, the resolution limits of our data are unclear as they depend on undefined parameters, such as the relative concentrations of each APEX2-marker protein, and diffusion rates of biotin phenol within each domain relative to the surrounding nucleoplasm. Moreover, although our dual indexing method identifies transcripts associated with each nuclear domain, it does not distinguish whether a transcript is concentrated within or else proximal but within the APEX2-labeling radius of a domain. As we have shown, differentiating these possibilities requires follow-up RNA-FISH analyses. Future improvements in mapping resolution may be achieved by the analysis of increased numbers of markers specific for the same domains, employment of labeling using "split" enzymes that can be reconstituted through fusion with interacting proteins that concentrate

(E) Plot showing the GO terms enriched in genes with retained introns detected in Group B genes (D) and that are associated with all three APEX2-speckle markers.

(F) Bar plot showing percentage of SON-regulated retained introns that overlap with speckle-associated retained introns ("total speckle"), speckle-associated retained introns that are also cell-cycle-regulated, speckle-associated, retained introns that are cell-cycle regulated in Group B (Figure 6D), or speckle-associated retained introns that are not cell-cycle regulated (*$p < 0.05$, Fisher's exact test).

Molecular Cell
Resource

CellPress

in the same domain, and methods enabling reduced distances or increased efficiencies of substrate labeling (Benhalevy et al., 2018; Cho et al., 2020).

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Experimental Model and Subject Details
  - Data and code availability
- METHODS DETAILS
  - Generation of APEX fusion constructs
  - Generation of Stable Flp-In-HEK293 Cell Lines
  - APEX labeling in living cells
  - Immunofluorescence staining
  - Fluorescence microscopy
  - APEX-MS
  - Mass spectrometry acquisition
  - RNA extraction for RT-qPCR or RNA-seq
  - Streptavidin pull-down and RT-qPCR of APEX-labeled RNA
  - APEX Labeling of cells expressing paraspeckle marker proteins and coilin
  - RNA-Sequencing
  - Immunoblotting
  - RNA fluorescence *in situ* hybridization (RNA-FISH)
  - Dual RNA and DNA fluorescence in situ hybridization (RNA-DNA FISH)
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Data-dependent acquisition mass spectrometry analysis
  - SAINT analysis
  - RNA-seq mapping and visualization
  - Differential gene expression analysis using DESeq2
  - Evaluation of sequence features of APEX2-marker enriched transcripts
  - Dual-index scoring method to identify transcripts uniquely associated with each nuclear domain
  - Post-processing and fluorescence microscopy analysis
  - RNAScope fluorescence in situ hybridization quantification
  - Differential alternative splicing analysis
  - Extended analysis of intron retention levels of nuclear speckle and LMNA associated transcripts
  - Regulation of nuclear speckle and LMNA associated introns across tissues
  - Overlap of nuclear speckle and LMNA associated introns with RNA binding protein interaction sites and knockdown dependent introns
  - Gene set enrichment analysis
  - Cell cycle associated alternative intron retention
  - Statistical Analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.molcel.2021.12.010.

AUTHOR CONTRIBUTIONS

B.J.B., A.R.B., and M.W. conceived the project. A.R.B., B.J.B., U.B., and M.W. designed experiments and analyses, with input from the other authors; M.W. and A.R.B. performed APEX2-seq; M.W. and A.R.B. performed immunofluorescence microscopy; M.W. performed RNA- and DNA-FISH; A.R.B. and M.W. performed image analyses, with input from P.G.M. A.-C.G., B.J.A.D., and Z.Y.L. performed APEX2-MS experiments and analyses. K.M.T. and T.D. performed library preparations and RNA-seq. R.W. performed analysis of cell-cycle data; M.W. A.R.B., and U.B. performed all other bioinformatic analyses, with contributions from Z.L. B.J.B., A.R.B., and M.W. wrote the manuscript with input from the other authors. B.J.B., A.-C.G., and P.G.M. supervised the study.

DECLARATION OF INTERESTS

B.J.B. is a member of the *Molecular Cell* advisory board. The authors declare no other competing interests.

REFERENCES

Ahn, E.Y., DeKelver, R.C., Lo, M.C., Nguyen, T.A., Matsuura, S., Boyapati, A., Pandit, S., Fu, X.D., and Zhang, D.E. (2011). SON controls cell-cycle progression by coordinated regulation of RNA splicing. Mol. Cell *42*, 185–198.

Amit, M., Donyo, M., Hollander, D., Goren, A., Kim, E., Gelfman, S., Lev-Maor, G., Burstein, D., Schwartz, S., Postolsky, B., et al. (2012). Differential GC content between exons and introns establishes distinct strategies of splice-site recognition. Cell Rep *1*, 543–556.

Barutcu, A.R., Blencowe, B.J., and Rinn, J.L. (2019). Differential contribution of steady-state RNA and active transcription in chromatin organization. EMBO Rep 20, e48068.

Benhalevy, D., Anastasakis, D.G., and Hafner, M. (2018). Proximity-CLIP provides a snapshot of protein-occupied RNA elements in subcellular compartments. Nat. Methods 15, 1074–1082.

Berriz, G.F., Beaver, J.E., Cenik, C., Tasan, M., and Roth, F.P. (2009). Next generation software for functional trend analysis. Bioinformatics 25, 3043–3044.

Blencowe, B.J., Baurén, G., Eldridge, A.G., Issner, R., Nickerson, J.A., Rosonina, E., and Sharp, P.A. (2000). The SRm160/300 splicing coactivator subunits. RNA 6, 111–120.

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

Boutz, P.L., Bhutkar, A., and Sharp, P.A. (2015). Detained introns are a novel, widespread class of post-transcriptionally spliced introns. Genes Dev 29, 63–80.

Brangwynne, C.P., Eckmann, C.R., Courson, D.S., Rybarska, A., Hoege, C., Gharakhani, J., Jülicher, F., and Hyman, A.A. (2009). Germline P granules are liquid droplets that localize by controlled dissolution/condensation. Science 324, 1729–1732.

Braun, C.J., Stanciu, M., Boutz, P.L., Patterson, J.C., Calligaris, D., Higuchi, F., Neupane, R., Fenoglio, S., Cahill, D.P., Wakimoto, H., et al. (2017). Coordinated splicing of regulatory detained introns within oncogenic transcripts creates an exploitable vulnerability in malignant glioma. Cancer Cell 32, 411–426.e11.

Braunschweig, U., Barbosa-Morais, N.L., Pan, Q., Nachman, E.N., Alipanahi, B., Gonatopoulos-Pournatzis, T., Frey, B., Irimia, M., and Blencowe, B.J. (2014). Widespread intron retention in mammals functionally tunes transcriptomes. Genome Res 24, 1774–1786.

Chen, S., Wang, R., Zheng, D., Zhang, H., Chang, X., Wang, K., Li, W., Fan, J., Tian, B., and Cheng, H. (2019). The mRNA export receptor NXF1 coordinates transcriptional dynamics, alternative polyadenylation, and mRNA export. Mol. Cell 74, 118–131.e7.

Chen, Y., and Belmont, A.S. (2019). Genome organization around nuclear speckles. Curr. Opin. Genet. Dev. 55, 91–99.

Chen, Y., Zhang, Y., Wang, Y., Zhang, L., Brinkman, E.K., Adam, S.A., Goldman, R., van Steensel, B., Ma, J., and Belmont, A.S. (2018). Mapping 3D genome organization relative to nuclear compartments using TSA-seq as a cytological ruler. J. Cell Biol. 217, 4025–4048.

Cho, K.F., Branon, T.C., Udeshi, N.D., Myers, S.A., Carr, S.A., and Ting, A.Y. (2020). Proximity labeling in mammalian cells with TurboID and split-TurboID. Nat. Protoc. 15, 3971–3999.

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

Colwill, K., Pawson, T., Andrews, B., Prasad, J., Manley, J.L., Bell, J.C., and Duncan, P.I. (1996). The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution. EMBO J 15, 265–275.

Corpet, A., Kleijwegt, C., Roubille, S., Juillard, F., Jacquet, K., Texier, P., and Lomonte, P. (2020). PML nuclear bodies and chromatin dynamics: catch me if you can. Nucleic Acids Res 48, 11890–11912.

Creamer, K.M., Kolpa, H.J., and Lawrence, J.B. (2021). Nascent RNA scaffolds contribute to chromosome territory architecture and counter chromatin compaction. Mol. Cell 81, 3509–3525.e5.

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

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21.

Dominguez, D., Tsai, Y.H., Weatheritt, R., Wang, Y., Blencowe, B.J., and Wang, Z. (2016). An extensive program of periodic alternative splicing linked to cell cycle progression. Elife 5, e10288.

Dopie, J., Sweredoski, M.J., Moradian, A., and Belmont, A.S. (2020). Tyramide signal amplification mass spectrometry (TSA-MS) ratio identifies nuclear speckle proteins. J. Cell Biol. 219, e201910207.

Duronio, R.J., and Marzluff, W.F. (2017). Coordinating cell cycle-regulated histone gene expression through assembly and function of the histone locus body. RNA Biol 14, 726–738.

Eng, J.K., Jahan, T.A., and Hoopmann, M.R. (2013). Comet: an open-source MS/MS sequence database search tool. Proteomics 13, 22–24.

Fazal, F.M., Han, S., Parker, K.R., Kaewsapsak, P., Xu, J., Boettiger, A.N., Chang, H.Y., and Ting, A.Y. (2019). Atlas of subcellular RNA localization revealed by APEX-seq. Cell 178, 473–490.e26.

Ferreira, J.A., Carmo-Fonseca, M., and Lamond, A.I. (1994). Differential interaction of splicing snRNPs with coiled bodies and interchromatin granules during mitosis and assembly of daughter cell nuclei. J. Cell Biol. 126, 11–23.

Fox, A.H., Nakagawa, S., Hirose, T., and Bond, C.S. (2018). Paraspeckles: where long noncoding RNA meets phase separation. Trends Biochem. Sci. 43, 124–135.

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

Galganski, L., Urbanek, M.O., and Krzyzosiak, W.J. (2017). Nuclear speckles: molecular organization, biological function and role in disease. Nucleic Acids Res 45, 10350–10368.

García-Alcalde, F., Okonechnikov, K., Carbonell, J., Cruz, L.M., Götz, S., Tarazona, S., Dopazo, J., Meyer, T.F., and Conesa, A. (2012). Qualimap: evaluating next-generation sequencing alignment data. Bioinformatics 28, 2678–2679.

Gordon, J.M., Phizicky, D.V., and Neugebauer, K.M. (2021). Nuclear mechanisms of gene expression control: pre-mRNA splicing as a life or death decision. Curr. Opin. Genet. Dev. 67, 67–76.

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

Hildebrand, E.M., and Dekker, J. (2020). Mechanisms and functions of chromosome compartmentalization. Trends Biochem. Sci. 45, 385–396.

Hung, V., Udeshi, N.D., Lam, S.S., Loh, K.H., Cox, K.J., Pedram, K., Carr, S.A., and Ting, A.Y. (2016). Spatially resolved proteomic mapping in living cells with the engineered peroxidase APEX2. Nat. Protoc. 11, 456–475.

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

Ilik, İ.A., Malszycki, M., Lübke, A.K., Schade, C., Meierhofer, D., and Aktaş, T. (2020). SON and SRRM2 are essential for nuclear speckle formation. eLife 9, e60579.

Imada, T., Shimi, T., Kaiho, A., Saeki, Y., and Kimura, H. (2021). RNA polymerase II condensate formation and association with Cajal and histone locus bodies in living human cells. Genes Cells 26, 298–312.

Kim, J., Venkata, N.C., Hernandez Gonzalez, G.A., Khanna, N., and Belmont, A.S. (2020). Gene expression amplification by nuclear speckle association. J. Cell Biol. 219, e201904046.

Knight, J.D.R., Choi, H., Gupta, G.D., Pelletier, L., Raught, B., Nesvizhskii, A.I., and Gingras, A.C. (2017). ProHits-viz: a suite of web tools for visualizing interaction proteomics data. Nat. Methods 14, 645–646.

Kucukural, A., Yukselen, O., Ozata, D.M., Moore, M.J., and Garber, M. (2019). DEBrowser: interactive differential expression analysis and visualization tool for count data. BMC Genomics 20, 6.

Lafontaine, D.L.J., Riback, J.A., Bascetin, R., and Brangwynne, C.P. (2021). The nucleolus as a multiphase liquid condensate. Nat. Rev. Mol. Cell Biol. 22, 165–182.

Lallemand-Breitenbach, V., and de Thé, H. (2018). PML nuclear bodies: from architecture to function. Curr. Opin. Cell Biol. *52*, 154–161.

Lambert, J.P., Tucholska, M., Go, C., Knight, J.D., and Gingras, A.C. (2015). Proximity biotinylation and affinity purification are complementary approaches for the interactome mapping of chromatin-associated protein complexes. J. Proteomics *118*, 81–94.

Li, D., Hsu, S., Purushotham, D., Sears, R.L., and Wang, T. (2019). WashU epigenome Browser update 2019. Nucleic Acids Res *47*, W158–W165.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289–293.

Liu, G., Knight, J.D., Zhang, J.P., Tsou, C.C., Wang, J., Lambert, J.P., Larsen, B., Tyers, M., Raught, B., Bandeira, N., et al. (2016). Data independent acquisition analysis in ProHits 4.0. J. Proteomics *149*, 64–68.

Liu, G., Zhang, J., Larsen, B., Stark, C., Breitkreutz, A., Lin, Z.Y., Breitkreutz, B.J., Ding, Y., Colwill, K., Pasculescu, A., et al. (2010). ProHits: integrated software for mass spectrometry-based interaction proteomics. Nat. Biotechnol. *28*, 1015–1017.

Liu, Y., Nanni, L., Sungalee, S., Zufferey, M., Tavernari, D., Mina, M., Ceri, S., Oricchio, E., and Ciriello, G. (2021). Systematic inference and comparison of multi-scale chromatin sub-compartments connects spatial organization to cell phenotypes. Nat. Commun. *12*, 2439.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.

Lu, X., Göke, J., Sachs, F., Jacques, P.É., Liang, H., Feng, B., Bourque, G., Bubulya, P.A., and Ng, H.H. (2013). SON connects the splicing-regulatory network with pluripotency in human embryonic stem cells. Nat. Cell Biol. *15*, 1141–1152.

Machyna, M., Heyn, P., and Neugebauer, K.M. (2013). Cajal bodies: where form meets function. Wiley Interdiscip. Rev. RNA *4*, 17–34.

Mahmoudi, S., Henriksson, S., Weibrecht, I., Smith, S., Söderberg, O., Strömblad, S., Wiman, K.G., and Farnebo, M. (2010). WRAP53 is essential for Cajal body formation and for targeting the survival of motor neuron complex to Cajal bodies. PLOS Biol *8*, e1000521.

Mannen, T., Yamashita, S., Tomita, K., Goshima, N., and Hirose, T. (2016). The Sam68 nuclear body is composed of two RNase-sensitive substructures joined by the adaptor HNRNPL. J. Cell Biol. *214*, 45–59.

Martin, R.M., Rino, J., Carvalho, C., Kirchhausen, T., and Carmo-Fonseca, M. (2013). Live-cell visualization of pre-mRNA splicing with single-molecule sensitivity. Cell Rep *4*, 1144–1155.

Marzluff, W.F., and Koreski, K.P. (2017). Birth and death of histone mRNAs. Trends Genet *33*, 745–759.

Merico, D., Isserlin, R., Stueker, O., Emili, A., and Bader, G.D. (2010). Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. PLoS One *5*, e13984.

Mintz, P.J., Patterson, S.D., Neuwald, A.F., Spahr, C.S., and Spector, D.L. (1999). Purification and biochemical characterization of interchromatin granule clusters. EMBO J *18*, 4308–4320.

Molenaar, C., Abdulle, A., Gena, A., Tanke, H.J., and Dirks, R.W. (2004). Poly(A)+ RNAs roam the cell nucleus and pass through speckle domains in transcriptionally active and inactive cells. J. Cell Biol. *165*, 191–202.

Monteuuis, G., Wong, J.J.L., Bailey, C.G., Schmitz, U., and Rasko, J.E.J. (2019). The changing paradigm of intron retention: regulation, ramifications and recipes. Nucleic Acids Res *47*, 11497–11513.

Nickerson, J.A., Blencowe, B.J., and Penman, S. (1995). The architectural organization of nuclear metabolism. Int. Rev. Cytol. *162A*, 67–123.

Nizami, Z., Deryusheva, S., and Gall, J.G. (2010). The Cajal body and histone locus body. Cold Spring Harb. Perspect. Biol. *2*, a000653.

Osoegawa, K., Mammoser, A.G., Wu, C., Frengen, E., Zeng, C., Catanese, J.J., and de Jong, P.J. (2001). A bacterial artificial chromosome library for sequencing the complete human genome. Genome Res *11*, 483–496.

Padrón, A., Iwasaki, S., and Ingolia, N.T. (2019). Proximity RNA labeling by APEX-seq reveals the organization of translation initiation complexes and repressive RNA granules. Mol. Cell *75*, 875–887.e5.

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods *14*, 417–419.

Perkins, D.N., Pappin, D.J., Creasy, D.M., and Cottrell, J.S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis *20*, 3551–3567.

Quinodoz, S.A., Jachowicz, J.W., Bhat, P., Ollikainen, N., Banerjee, A.K., Goronzy, I.N., Blanco, M.R., Chovanec, P., Chow, A., Markaki, Y., et al. (2021). RNA promotes the formation of spatial compartments in the nucleus. Cell *184*, 5775–5790.e30.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res *44*, W160–W165.

Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680.

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., and Vilo, J. (2019). g:profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res *47*, W191–W198.

Rinn, J., and Guttman, M. (2014). RNA function. RNA and dynamic nuclear organization. Science *345*, 1240–1241.

Sacco-Bubulya, P., and Spector, D.L. (2002). Disassembly of interchromatin granule clusters alters the coordination of transcription and pre-mRNA splicing. J. Cell Biol. *156*, 425–436.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682.

Schindelin, J., Rueden, C.T., Hiner, M.C., and Eliceiri, K.W. (2015). The ImageJ ecosystem: an open platform for biomedical image analysis. Mol. Reprod. Dev. *82*, 518–529.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res *13*, 2498–2504.

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

Shin, Y., Chang, Y.C., Lee, D.S.W., Berry, J., Sanders, D.W., Ronceray, P., Wingreen, N.S., Haataja, M., and Brangwynne, C.P. (2018). Liquid nuclear condensates mechanically sense and restructure the genome. Cell *175*, 1481–1491.e13.

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

Staněk, D., and Fox, A.H. (2017). Nuclear bodies: news insights into structure and function. Curr. Opin. Cell Biol. *46*, 94–101.

Stanek, D., and Neugebauer, K.M. (2006). The Cajal body: a meeting place for spliceosomal snRNPs in the nuclear maze. Chromosoma *115*, 343–354.

Su, J.H., Zheng, P., Kinrot, S.S., Bintu, B., and Zhuang, X. (2020). Genome-scale imaging of the 3D organization and transcriptional activity of chromatin. Cell *182*, 1641–1659.e26.

Tammer, L., Hameiri, O., Keydar, I., Roy, V.R., Ashkenazy-Titelman, A., Custódio, N., Sason, I., Shayevitch, R., Vaello, V.R., Rino, J., et al. (2022). Gene architecture directs splicing outcome in separate nuclear spatial regions. Molecular Cell 82, In this issue.

Tapial, J., Ha, K.C.H., Sterne-Weiler, T., Gohr, A., Braunschweig, U., Hermoso-Pulido, A., Quesnel-Vallières, M., Permanyer, J., Sodaei, R., Marquez, Y., et al. (2017). An atlas of alternative splicing profiles and functional associations reveals new regulatory programs and genes that simultaneously express multiple major isoforms. Genome Res 27, 1759–1768.

Teo, G., Koh, H., Fermin, D., Lambert, J.P., Knight, J.D., Gingras, A.C., and Choi, H. (2016). SAINTq: scoring protein-protein interactions in affinity purification - mass spectrometry experiments with fragment or peptide intensity data. Proteomics 16, 2238–2245.

Thrash, A., Arick, M., 2nd, and Peterson, D.G. (2018). Quack: A quality assurance tool for high throughput sequence data. Anal. Biochem. 548, 38–43.

Tripathi, K., and Parnaik, V.K. (2008). Differential dynamics of splicing factor SC35 during the cell cycle. J. Biosci. 33, 345–354.

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

Van Nostrand, E.L., Freese, P., Pratt, G.A., Wang, X., Wei, X., Xiao, R., Blue, S.M., Chen, J.Y., Cody, N.A.L., Dominguez, D., et al. (2020). A large-scale binding and functional map of human RNA-binding proteins. Nature 583, 711–719.

Wang, K., Wang, L., Wang, J., Chen, S., Shi, M., and Cheng, H. (2018). Intronless mRNAs transit through nuclear speckles to gain export competence. J. Cell Biol. 217, 3912–3929.

Xiong, K., and Ma, J. (2019). Revealing Hi-C subcompartments by imputing inter-chromosomal chromatin interactions. Nat. Commun. 10, 5069.

Yamazaki, T., Souquere, S., Chujo, T., Kobelke, S., Chong, Y.S., Fox, A.H., Bond, C.S., Nakagawa, S., Pierron, G., and Hirose, T. (2018). Functional domains of NEAT1 architectural lncRNA induce paraspeckle assembly through phase separation. Mol. Cell 70, 1038–1053.e7.

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

Molecular Cell
Resource

🔗 **CellPress**

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
anti-Flag M2	Sigma-Aldrich	Cat#F1804; RRID: AB_262044
anti-SRRM2	(Blencowe et al., 2000)	N/A
anti-SC35	Sigma-Aldrich	Cat#S4045; RRID: AB_477511
Anti-alpha-tubulin	Sigma-Aldrich	Cat#T6074; RRID: AB_477582
Goat-anti-mouse IgG HRP	GE	Cat#NA931V
Goat-anti-mouse IgG Alexa 488	Invitrogen	Cat#A11001
Goat-anti-mouse IgG Alexa 647	Invitrogen	Cat#A21236
Donkey-anti-rabbit IgG Alexa 555	Invitrogen	Cat#A31572
Donkey-anti-mouse IgG Alexa 555	Invitrogen	Cat#A31570
Alexa Fluor 647 IgG fraction monoclonal mouse anti-digoxin	Jackson Immunoresearch	Cat#200602156
Bacterial and virus strains		
Subcloning Efficiency DH5α Competent Cells	Thermo Fisher Scientific	18265017
Chemicals, peptides, and recombinant proteins		
Human plasma fibronectin	EMD Millipore	Cat#FC010
Trolox	Sigma-Aldrich	Cat#238813
Sodium Ascorbate	Sigma-Aldrich	Cat#11140
Sodium Azide	Bioshop	Cat#SAZ001
Lipofectamine 2000	Invitrogen	Cat#11668019
Dynabeads MyOne Streptavidin C1	Thermo Fisher Scientific	Cat#65001
Streptavidin sepharose beads	Millipore Sigma	Cat#GE 17-5113-01
TRI Reagent	Sigma Aldrich	Cat#T9424
Biotin-phenol (Biotin tyramide)	Sigma Aldrich	Cat#SML2135
NeutrAvidin-Oregon Green 488 Conjugate	Thermo Fisher	Cat#A6374
Hoechst 33342	Sigma Aldrich	Cat#B2261
TSA Plus fluorescein	Akoya Bioscience	Cat#NEL741001KT
TSA Plus Cyanine 3	Akoya Bioscience	Cat#NEL744001KT
TSA Plus Cyanine 5	Akoya Bioscience	Cat#NEL745001KT
10X PBS, pH7.4	Invitrogen	Cat#AM9624
5M NaCl	Invitrogen	Cat#AM9759
Ultrapure 1M Tris-HCl pH7.5	Invitrogen	Cat#15567-027
Ultrapure Distilled water	Invitrogen	Cat#10977-015
Ribolock RNase Inhibitor	Thermo Fisher	Cat#EO0381
Buffer RWT	QIAGEN	Cat#1067933
Hydrogen peroxide solution, 30% (w/w)	Sigma-Aldrich	Cat#H1009
ProLong Gold Antifade Mountant	Thermo Fisher	Cat#P36934
Permafluor Mountant	Thermo Fisher	Cat#TA-030-FM
RNase-Free DNase Set	QIAGEN	Cat#79256
Digoxigenin-dUTP, alkali-stable	Enzo Life Sciences Inc.	Cat#ENZ-NUC113-0025
DNA pol I	Thermo Fisher	Cat#EP0041
DNaseI	Sigma-Aldrich	Cat#D7291
Herring sperm DNA	Sigma-Aldrich	Cat#D6898
Human Cot DNA	IDT	Cat#1080769
Deionized Formamide	Bioshop	Cat#For001.500
Dextran sulfate	Sigma-Aldrich	Cat#D8906

(Continued on next page)

CellPress

Molecular Cell
Resource

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical commercial assays		
SensiFAST SYBR® No-ROX Kit	BIOLINE	Cat#BIO-98005
Maxima First Strand cDNA Synthesis Kit	Thermo Fisher	Cat#K1671
SuperScriptIII	Thermo Fisher Scientific	Cat#18080044
RNeasy Mini Kit	QIAGEN	Cat#74106
OneStep RT-PCR Kit	QIAGEN	Cat#210210
RNAscope Multiplex Fluorescent Reagent Kit v2	Advanced Cell Diagnostics	Cat#323135
RNAscope RNA-Protein Co-Detection Ancillary Kit	Advanced Cell Diagnostics	Cat#323180
NEBNext Ultra II Directional RNA	NEB	Cat#E7760S
Illumina NovaSeq, S1 flowcell v1.5, 100-cycle kit	Illumina	Cat#20028319
Illumina NovaSeq 6000 S2 Reagent Kit	Illumina	Cat##20028315
Illumina NextSeq 500/550 High-Output Kit v2.5	Illumina	Cat#20024906
NEBNext Human/Mouse/Rat rRNA Depletion Kit	NEB	Cat#E6310
Agilent RNA 6000 Pico Kit	Agilent	Cat#5067-1513
Deposited data		
APEX-Seq of nuclear domains	this study	GEO: GSE176439
Periodic alternative splicing linked to cell cycle	(Dominguez et al., 2016)	GEO: GSE81485
SON knockdown RNA-seq	(Lu et al., 2013)	ArrayExpress: E-MTAB-1687
ENCODE eCLIP and knockdown RNA-seq	(Van Nostrand et al., 2020)	(https://www.encodeproject.org) ENCSR456FVU (CLIP), ENCSR369TWP (RNA-seq)
APEX-MS data	this study	MassIVE: MSV000087649, PXD026758
Experimental models: Cell lines		
Human: Flp-In-293 Cell Line	Invitrogen	R75007
Experimental models: Organisms/strains		
Oligonucleotides	N/A	N/A
RNAscope Probe - Hs-MALAT1-O7-C3	Advanced Cell Diagnostics	Cat#578171-C3
RNAscope Probe - Hs-EIF3G-intron7-C2	Advanced Cell Diagnostics	Cat#1057981-C2
RNAscope Probe - Hs-KRAS-O1	Advanced Cell Diagnostics	Cat#522961
RNAscope Probe - Hs-MTR-intron1-C2	Advanced Cell Diagnostics	Cat#1060151-C2
RNAscope Probe - Hs-EIF1-intron1-C1	Advanced Cell Diagnostics	Cat#1057961-C1
RNAscope Probe - Hs-DDX5-intron12-C1	Advanced Cell Diagnostics	Cat#1057991-C1
RNAscope Probe - Hs-MST1-O1-C2	Advanced Cell Diagnostics	Cat#1057931-C2
Recombinant DNA		
pDEST-cDNA5-FRT/TO-3*Flag-APEX2 N-term	this study	N/A
pDEST-cDNA5-FRT/TO-3*Flag-APEX2 C-term	this study	N/A
pBACe3.6_RP11-81D7	BACPAC resource center (Osoegawa et al., 2001)	RP11-81D7
pBACe3.6_RP11-156E5	BACPAC resource center (Osoegawa et al., 2001)	RP11-156E5
pBACe3.6_RP11-298C17	BACPAC resource center (Osoegawa et al., 2001)	RP11-298C17
pBACe3.6_RP11-959M8	BACPAC resource center (Osoegawa et al., 2001)	RP11-959M8
Software and algorithms		
vast-tools 2.5.1	(Tapial et al., 2017)	https://github.com/vastgroup/vast-tools
DESeq2	(Love et al., 2014)	https://github.com/mikelove/DESeq2
debrowser	(Kucukural et al., 2019)	https://github.com/UMMS-Biocore/debrowser
STAR v2.7.4.a	(Dobin et al., 2013)	https://github.com/alexdobin/STAR

(Continued on next page)

Molecular Cell
Resource

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
featureCounts	(Dobin et al., 2013)	http://bioinf.wehi.edu.au/featureCounts/
R	v4.0.2	https://www.r-project.org/
pheatmap	N/A	https://github.com/raivokolde/pheatmap
ProHits-LIMS	(Liu et al., 2010)	http://prohitsms.com/Prohits_download/list.php
Fiji	(Schindelin et al., 2012)	https://imagej.net/software/fiji/
Funcassociate 3.0	(Berriz et al., 2009)	http://llama.mshri.on.ca/funcassociate/documentation
SAINTexpress v3.6.1	(Teo et al., 2016)	N/A
FASTQC	(Thrash et al., 2018)	https://github.com/s-andrews/FastQC
qualimap	(Garcia-Alcalde et al., 2012)	http://qualimap.conesalab.org/
ImageJ	(Schindelin et al., 2015)	https://imagej.net/
Zen black	Zeiss	N/A
Zen blue	Zeiss	N/A
Samtools	(Li et al., 2009)	http://www.htslib.org/
Deeptools	(Ramirez et al., 2016)	https://deeptools.readthedocs.io/en/develop/
Cytoscape v3.5.1	(Shannon et al., 2003)	https://cytoscape.org/cy3.html
Enrichment Map Pipeline Collection v1.0.0	(Merico et al., 2010)	https://apps.cytoscape.org/apps/enrichmentmappipelinecollection
g:Profiler	(Raudvere et al., 2019)	https://biit.cs.ut.ee/gprofiler/gost
WashU epigenome Browser	(Li et al., 2019)	https://epigenomegateway.wustl.edu/
Salmon 0.14	(Patro et al., 2017)	https://salmon.readthedocs.io/en/latest/salmon.html

RESOURCE AVAILABILITY

Lead Contact
Further information and requests for resources, reagents, and materials should be directed to and will be fulfilled by the lead contact, Benjamin J. Blencowe (b.blencowe@utoronto.ca).

Materials Availability
All unique/stable reagents generated in this study are available from the lead contact with a completed Materials Transfer Agreement.

Experimental Model and Subject Details
HEK293 Flp-In cells were grown in DMEM (high glucose; Sigma-Aldrich) supplemented with 10% FBS, non-essential amino acids, and penicillin/streptomycin. Cells were maintained at sub-confluent conditions, maintained at 37°C with 5% CO_2 and were passaged every 2-3 days. Parental cell line with no integrated trans-gene was selected with 10μg/mL Blasticidin S, and cell lines with integrated transgenes were selected with 200μg/mL Hygromycin. For APEX RNA-Seq, imaging or RT-qPCR experiments, plates or cover slips were coated with 10 μg/mL fibronectin (EMD Millipore Sigma) for 1 hour at 37°C before cell plating.

Data and code availability
The datasets supporting the conclusions of this article have been deposited to the Gene Expression Omnibus (GEO) repository (GSE176439). Proteomics data has been deposited to the MassIVE repository (https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp) with the accession number MSV000087649. The raw gel and microscopy images have been deposited to Mendeley Data (https://data.mendeley.com/datasets/p3m8776742/1). This paper does not report original code. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

METHODS DETAILS

Generation of APEX fusion constructs
Backbone of APEX2 expressing vector was amplified from the mito-V5-APEX2 plasmid (Addgene # 72480) (Fazal et al., 2019), and inserted into pcDNA5/FRT/TO-FLAG using the Gibson Assembly Cloning kit (NEB). Gateway cloning was used to fuse APEX2 to either to

the 5′ or the 3′ ends of bait proteins of interest using ORF cDNAs from available entry clones. All constructs generated were validated by Sanger sequencing. Fusion constructs with an N-terminal FLAG-APEX2 sequence include RNPS1, SRSF1, LMNA, FBL, SAM68, WRAP53, PSPC1, NONO, and COILIN ORFs; whereas constructs with a C-terminal FLAG-APEX2 sequence include SRSF7, PML, SP100, NPAT, EWSR1, and SMN2 ORFs. Sequences of the oligos used to clone the constructs are provided in Table S10.

Generation of Stable Flp-In-HEK293 Cell Lines
Doxycycline-inducible Flp-In HEK293 cell lines were generated by transfecting 500 ng of each pcDNA5/FRT/TO-based plasmid with 2 μg of plasmid encoding pOG44 recombinase using Lipofectamine 2000 (Invitrogen), as per the manufacturer's recommendations. Cell lines with stably integrated constructs were selected and maintained with 10 μg/mL blasticidin S and 200 μg/mL hygromycin B. Transgene expression was induced by addition of 2 μg/mL doxycycline for 24 hours.

APEX labeling in living cells
The APEX procedure was carried out as described previously (Fazal et al., 2019) with modifications (https://protocolexchange. researchsquare.com/article/pex-71/v1). Briefly, 24 hours after inducing cells expressing the APEX2 fusion construct, APEX labeling was performed by changing to fresh growth medium containing 500 μM biotin-phenol (Sigma-Aldrich), followed by incubation at 37°C under 5% CO_2 for 30 minutes. H_2O_2 (Sigma Aldrich) was then added to each well to a final concentration of 1 mM and the culture dishes were gently agitated for 1 minute. The reaction was quenched by replacing the medium with an equal volume of 2X quench buffer (10 mM Trolox , 20 mM sodium ascorbate, and 20 mM sodium azide in Dulbecco's phosphate-buffered saline). The cells were washed one more time with 2X quench buffer before imaging, RT-qPCR or RNA-seq experiments. The unlabeled controls were processed identically, except that H_2O_2 was omitted.

Immunofluorescence staining
After the second wash with 2X quench buffer in the APEX labeling protocol, the cells were immediately fixed with ice-cold methanol at –20°C for 15 minutes. Cells were washed three times with 1xPBS and blocked with blocking buffer (4% BSA, 1% FBS, and 0.1% Triton in 1X PBS) at room temperature for 1 hour. Cells were then incubated with primary antibodies in blocking buffer overnight at 4°C. Antibody dilutions: Mouse SC35-Sigma S4045 at 1:2000; rabbit anti-SRRM2 at 1:5000; mouse anti-Flag M2-antibody-(Sigma Aldrich #F1804) at 1:1000.) After washing three times with 1xPBS, cells were incubated with secondary detection reagents (Neutravidin-Oregon Green 488, AlexaFluor555 or AlexaFluor647 at 1:1000 dilution) in blocking buffer for 1 hour at room temperature. Subsequently, cells were stained with Hoechst 33258 (Sigma Aldrich) at room temperature for 15 minutes to indicate nuclei, then washed three times with PBS, mounted with Permafluor mountant (Thermo Fisher Scientific), and imaged.

Fluorescence microscopy
The LSM880 with Airyscan Fast (Zeiss) was used for all microscopy experiments at the Imaging Facility at The Hospital for Sick Children (SickKids Research Institute, Toronto, Canada). The ZEN black edition software version 2.3 (Zeiss) was used for acquisition. All images shown were acquired in resolution-versus-sensitivity (R-S) mode with 8 bit, 0.3-0.45 μm z-slices. The oil immersion objective Plan-Apochromat 63X/1.4 oil DIC M27 was used for all acquisitions.

APEX-MS
APEX-MS experiments were performed using methods adapted from Hung et al. (2016). All APEX-MS experiments were performed with biological duplicates, and results were compared to the following control samples: 6 bait samples where no H_2O_2 was added, 2 NLS-APEX2 baits where no H_2O_2 was added, and 2 each of NLS-APEX2, GFP-APEX2, and FLAG-APEX2 where labeling was catalyzed by addition of H_2O_2. Cells were pelleted with at least ∼0.1 g per sample, and snap frozen. For processing, cells were lysed at a 10:1 mL:g ratio in modified RIPA lysis buffer [50 mM Tris (pH 7.5), 150 mM NaCl, 1.5 mM $MgCl_2$, 1 mM EDTA, 0.1% SDS (w/v), 1% IGEPAL CA-630 NP-40 Substitute] with freshly added sodium deoxycholate (final 0.5% w/v), protease inhibitors (Sigma-Aldrich P8340), 5mM Trolox (in DMSO), 10mM Sodium Ascorbate (Vitamin C), and 10mM Sodium Azide. Trolox, Sodium Ascorbate, and Sodium Azide were added from freshly made stock solutions. Pellets weighing under 0.1g were treated as 0.1g. After addition of lysis buffer, samples were solubilized for 20 minutes, gently rotating at 4°C, and all additional steps performed at 4°C incubations were also performed with gentle rotation. Samples were sonicated at 30% amplitude for 5 sec on, 3 sec off cycles, for 3 cycles, using a Qsonica sonicator with a CL-18 probe. 1 μL each of TurboNuclease (BioVision Cat#9207-50KU from CEDARLANE) and RNAse (Sigma-Aldrich) were added to each sample and incubated for 30 minutes at 4°C. Samples were spiked with additional SDS to a final concentration of 0.25% and incubated for 15 minutes at 4°C. Samples were spun at 16,000 x g for 20 minutes and the cleared lysates (supernatants) were transferred to new tubes. At this stage, samples were normalized (by volume) to the cleared lysate of 0.1g pellet samples. Streptavidin Sepharose beads (GE 17-5113-01) were washed 3X with modified RIPA buffer with 0.4% SDS, and 30 μl (bed volume) was added to each sample and incubated rotating for 3 hours, 4°C. Samples were washed once with wash buffer (2% SDS, 50 mM Tris), once with modified RIPA buffer with 0.4% SDS, and 3X with ABC buffer [50 mM ammonium bicarbonate (pH 8.5), freshly made]. Samples were spun, supernatant removed, and on-bead trypsin digest of peptides was performed by incubating with 1 μg trypsin dissolved in 50μL of 50 mM ABC buffer, rotating overnight at 37°C. The next day, beads were spun down and supernatant was transferred to new tubes. Beads were washed twice more with 50μL of HPLC-grade water, with each wash followed by a spin and

collection of supernatant that was combined with the initial supernatant from overnight incubation. 0.5 μg trypsin in 50 mM ABC was added to the combined digested sample and incubated at 37°C for 4 hours. Fresh 50% formic acid was added to samples to a final concentration of 2% prior to drying by vacuum centrifugation and subsequent storage at -80°C.

Mass spectrometry acquisition

Each sample (6 μL in 2 % formic acid; corresponding to 1/4 of a 15 cm tissue culture dish) was directly loaded at 800 nL/min onto an equilibrated HPLC column (pulled and packed in-house). The peptides were eluted from the column over a 90 minutes gradient generated by a Eksigent ekspert™ nanoLC 425 (Eksigent, Dublin CA) nano-pump and analysed on a TripleTOF™ 6600 instrument (AB SCIEX, Concord, Ontario, Canada). The gradient was delivered at 400 nL/min starting from 2% acetonitrile with 0.1% formic acid to 35% acetonitrile with 0.1% formic acid over 90 minutes followed by a 15-minute clean-up at 80% acetonitrile with 0.1% formic acid, and a 15-minute equilibration period back to 2% acetonitrile with 0.1% formic acid, for a total of 120 min. To minimize carryover between each sample, the analytical column was washed for 2 hours by running an alternating sawtooth gradient from 35% aceto-nitrile with 0.1% formic acid to 80% acetonitrile with 0.1% formic acid at a flow rate of 1500 nL/min, holding each gradient concentration for 5 min. Analytical column and instrument performance were verified after each sample by loading 30 fmol bovine serum albumin (BSA) tryptic peptide standard with 60 fmol α-casein tryptic digest and running a short 30-minute gradient. TOF MS mass calibration was performed on BSA reference ions before running the next sample to adjust for mass drift and verify peak intensity. Samples were analyzed with two separate injections with instrument method set to data dependent acquisition (DDA) mode. The DDA method consisted of one 250 milliseconds (ms) MS1 TOF survey scan from 400–1800 Da followed by ten 100 ms MS2 candidate ion scans from 100–1800 Da in high sensitivity mode. Only ions with a charge of 2+ to 5+ that exceeded a threshold of 300 cps were selected for MS2, and former precursors were excluded for 7 seconds after one occurrence.

RNA extraction for RT-qPCR or RNA-seq

After the second wash with 2X quench buffer in the APEX labeling protocol, cells were harvested by scraping them in 2X quench buffer including 4μL Ribolock RNase inhibitor (Thermo Fisher). The cells were centrifuged for 5 minutes at 200xg at 4 °C. After aspirating the quench buffer, cells were lysed and processed using the RNeasy plus mini kit (Qiagen) following the manufacturer's protocol with minor modifications, including addition of ß-mercaptoethanol to the RLT lysis buffer, washing with the RWT Buffer (Qiagen) instead of RW1 buffer, and application of the on-column DNaseI digestion (Qiagen). The RNA was eluted in RNase-free water and quantified using Nanodrop. Equal amounts of RNA were aliquoted prior to streptavidin pull-down of RNA for downstream applications. Sequences of the oligos used for RT-qPCR assays are provided in Table S10.

Streptavidin pull-down and RT-qPCR of APEX-labeled RNA

To enrich biotinylated RNAs, 40 μl each of Dynabeads™ MyOne™ Streptavidin C1 magnetic beads (Thermo Fisher) were used per 100 μg of RNA. The beads were washed 3 times in 500μL B&W buffer (5 mM Tris-HCl, pH = 7.5, 0.5 mM EDTA, 1 M NaCl, 0.1% TWEEN 20 (Sigma Aldrich)), followed by 2 times in 500μL Solution A (0.1 M NaOH and 0.05 M NaCl), and 1 time in 500μL Solution B (0.1 M NaCl). Beads were then incubated with purified RNA (100μg) in 0.05M NaCl with 4μl/400μl Ribolock RNase Inhibitor (Thermo) at final volume of 800μl at 4°C for 2 hours with rotations. Following the incubation step, beads were washed 3 times with B&W buffer, followed by the elution of RNA from the beads by adding TRI Reagent (Sigma-Aldrich T9424) on the beads and extracting according to the manufacturer's protocol. For RT-qPCR experiments, first-strand cDNA was generated by using the Maxima H Minus cDNA Synthesis Master Mix (Thermo Fisher) according to the manufacturer's protocol. qRT-PCR was performed in a volume of 10 μL using 2μL of diluted cDNA, 500 nM primers and 5 μL SensiFAST SYBR No-ROX Kit (Bioline) using a CFX96 Real-Time PCR Detection System (BIO-RAD). The percent recovery quantified by RT-qPCR analysis was calculated as the ratio of RNA/cDNA recovered relative to matched input controls using the $2^{(-\text{delta Ct})}$ method.

APEX Labeling of cells expressing paraspeckle marker proteins and coilin

When labeling cells expressing APEX2-tagged paraspeckle marker proteins (NONO, PSPC1, EWSR1), we observed that the FLAG-APEX2 fusion proteins, as well as endogenous paraspeckle proteins (PSPC1 and SFPQ) showed abnormal patterns of localization (data not shown). The mislocalization of paraspeckle marker proteins was not observed in cells without H_2O_2 treatment, nor in parental HEK293 cells that do not express the APEX2-tagged paraspeckle markers. These indicated that the APEX2-labeling of paraspeckle components disrupts paraspeckle morphology, and therefore we did not further analyze paraspeckles in this study. The expected localization pattern of the APEX2-COILIN fusion protein also was not observed, and therefore this marker was omitted from the study (data not shown).

RNA-Sequencing

After the RNA extraction, RNA-seq libraries were prepared using the NextSeq 500/550 High Output Kit v2.5 (Illumina), NEBNext Ultra II Directional RNA Library Prep Kit (NEB), NovaSeq 6000 S2 Reagent Kit (Ilumina), with NEBNext Human/Mouse/Rat rRNA Depletion Kit (NEB). For input RNA samples the median RIN score was 10. 6.5-10 ng of RNA recovered from APEX2-marker expressing cell lines was amplified for 15 cycles, and input RNA samples (500 ng) were amplified for 10 cycles. Ribosomal RNA-depleted samples were sequenced using Illumina NovaSeq platform with 75-100bp single-end sequencing.

Immunoblotting

Cells were lysed in Tris lysis buffer (10 mM Tris-HCl pH8.0, 150 mM NaCl, 1% NP-40, and 10% glycerol) and sonicated. Lysate were further centrifuged at $4°C$, 17000g for 10 minutes, supernatants were collected, mixed with Laemmli buffer, and heated 5 minutes at $95°C$. Samples were separated by SDS-PAGE using NuPAGE 4-12% Bis-Tris protein gels (Invitrogen) as per the manufacturer's protocol, and transferred to PVDF membrane. Blots were blocked 1hr with 5% milk/PBST, incubated with primary antibodies (anti-Flag M2 at 1:5000, anti-α-Tubulin at 1:10000 in 5% milk/PBST) for 2hrs. HRP-conjugated anti-mouse secondary antibody was used at 1:5000 in 5%milk/PBST for 45 minutes, immunoblots were developed using the ECL chemiluminescence-detection kit (PerkinElmer) as per the manufacturer's instructions.

RNA fluorescence *in situ* hybridization (RNA-FISH)

To localize whole transcripts and introns associated with nuclear speckle and other nuclear domain markers, RNA-FISH was performed using the RNAscope Multiplex Fluorescent V2 kit (Advanced Cell Diagnostics, Newark, CA) on the HEK293 Flp-In 'parental' cell line, or cells with an integrated Flag-APEX-PML transgene (for PML detection), as per the manufacturer's protocol. Briefly, cells were seeded in a Nunc Lab-Tek 4-well Chamber Slide (Thermo-Fisher), coated with fibronectin (Millipore Sigma) 24 hours prior to achieve 70-90% confluency at the time of fixation. Slides were washed and fixed with 4% paraformaldehyde (PFA) for 15 minutes at room temperature, followed by dehydration with a gradient of 50/70/100% EtOH and rehydration of 70%/50% EtOH and 1x PBS. Next, slides were treated with hydrogen peroxide for 10 minutes at room temperature, followed by washing with ddH$_2$O three times. For co-localization experiments using immunofluorescence detection, the slides were incubated with primary antibody (anti-Flag M2 at 1:250 and anti-SRRM2 at 1:1000) overnight at $4°C$, using the RNAscope RNA-Protein Co-Detection Ancillary Kit (Advanced Cell Diagnostics), as recommended by the manufacturer. After washing three times in PBST (PBS+Tween20 0.1%), the slides were fixed again with 4%PFA for 15 minutes at RT, washed three times with PBST, then treated with protease III (diluted 1:10-1:15 in PBS) for 10 minutes at room temperature. Next, after three washes in PBS, slides were hybridized with RNAscope probes designed to detect MALAT1 and MST1 or KRAS transcripts, or retained intron sequences in EIF1, MTR, DDX5, EIF3G transcripts (Advanced Cell Diagnostics). Hybridizations were performed at $42°C$ using the HybEZ II Hybridization System (Advanced Cell Diagnostics) and detected using the RNAscope Multiplex Fluorescent V2 protocol with TSA plus fluorophores (1:1500-1:4500 dilution for TSA-Cy3; 1:1000 for TSA-Cy5, 1:1500 for TSA Fluorescein) as per the manufacturer's recommendations. After applying the RNAscope assay, for immunofluorescence co-detection, slides were further incubated with a secondary antibody (anti-mouse-Alexa488, anti-rabbit-Alexa488 or anti-rabbit-Alexa555 (for SRRM2 detection) at 1:250) for 30 minutes at RT. Finally, after two washes with PBST, slides were counterstained with DAPI and mounted on coverslips using ProLong Gold antifade mounting media (ThermoFisher). Images were acquired using an LSM880 confocal microscope with Airyscan module (Zeiss) within two days post-hybridization. Intensity settings were applied manually in ZEN blue, based on images with single channel-only detection.

Dual RNA and DNA fluorescence in situ hybridization (RNA-DNA FISH)

Simultaneous RNA-DNA FISH was performed using a protocol adapted in part from a previous study (Martin et al., 2013), as well as using conditions recommended for RNAScope FISH. Cells grown in Nunc Lab-Tek 4-well Chamber Slide (Thermo-Fisher) were fixed, pretreated as per the RNAScope FISH protocol (see above) with the exception that a higher concentration of protease digestion was used (1:5 dilution of Protease III) for 10 minutes at room temperature. RNAscope FISH was then performed, as per the manufacturer's recommendations (see above). Immediately after RNAscope FISH, cells were incubated in 2xSSC for 5 minutes at $40°C$, and denatured by incubating in 70% formamide/2x SSC/50mM sodium phosphate pH7.0 for 4 minutes at $73°C$. Cells were immediately immersed in an ice-cold ethanol series (70%/90%/100%) for 5 minutes each, then air dried briefly before DNA hybridization.

DNA-FISH probes were prepared to have a median length of ~200bp and were generated using nick translation from BAC clones (Osoegawa et al., 2001), using DNA pol I (ThermoFisher), DNase I (Sigma-Aldrich), and digoxigenin-dUTP incorporation (Enzo Life Sciences). 2μg of the labelled BAC probes was diluted into 180μL DNA hybridization buffer (50% formamide/2x SSC/50mM sodium phosphate pH7.0/10% Dextran Sulfate) with 50-100μg sonicated herring sperm DNA (Sigma Aldrich) and 100ug of Human Cot-1 DNA (IDT). The probe mix was denatured for 5 minutes at $75°C$ immediately prior to hybridization.

Denatured hybridization mixtures were applied to RNAScope-FISH-processed cells and incubated overnight at $37°C$ in humidified chambers. Post-hybridization washes were performed 3 x 5 minutes each with 50% formamide/2x SSC/0.01% Tween-20 at $45°C$, then 3 x 5 minutes each with 1X SSC/0.01% Tween-20 at $60°C$. Cells were then blocked with RNAscope co-detection antibody diluent (Advanced Cell Diagnostics) for 15 minutes at room temperature, then incubated with Alexa647-conjugated mouse anti-digoxigenin (Jackson ImmunoResearch) at a dilution of 1:150 in RNAscope co-detection antibody diluent for 60 minutes at room temperature. After two PBST washes, slides were counterstained with DAPI and mounted with ProLong Gold antifade mounting media (ThermoFisher). Images were acquired using an LSM880 confocal microscope with Airyscan module (Zeiss) within 1 day post mounting. For background controls, DNA-FISH probes were omitted from the protocol described above (data not shown).

Molecular Cell
Resource

 CellPress

QUANTIFICATION AND STATISTICAL ANALYSIS

Data-dependent acquisition mass spectrometry analysis

Mass spectrometry data generated were stored, searched, and analyzed using ProHits laboratory information management system (LIMS) platform (Lambert et al., 2015). Within ProHits', WIFF files were converted to an MGF format using the WIFF2MGF converter', and to an mzML format using ProteoWizard (V3.0.10702) and the AB SCIEX MS Data Converter (V1.3 beta). The data were then searched using Mascot (V2.3.02) (Liu et al., 2016) and Comet (V2016.01 rev.2) (Perkins et al., 1999). The spectra were searched with the human and adenovirus sequences in the RefSeq database (version 57, January 30th, 2013) acquired from NCBI, supplemented with "common contaminants" from the Max Planck Institute (http://maxquant.org/contaminants.zip) and the Global Proteome Machine (GPM; ftp://ftp.thegpm.org/fasta/cRAP/crap.fasta), forward and reverse sequences (labeled "gi|9999" or "DECOY"), sequence tags (APEX2, BirA, GST26, mCherry and GFP) and streptavidin, for a total of 72,481 entries. Database parameters were set to search for tryptic cleavages, allowing up to 2 missed cleavages sites per peptide with a mass tolerance of 35ppm for precursors with charges of 2+ to 4+ and a tolerance of 0.15amu for fragment ions. Variable modifications were selected for deamidated asparagine and glutamine and oxidized methionine. Results from each search engine were analyzed through TPP (the Trans-Proteomic Pipeline, v.4.7 POLAR VORTEX rev 1) via the iProphet pipeline (Eng et al., 2013).

SAINT analysis

SAINT version 2.5.0 was used as a statistical tool to calculate the probability of potential protein-protein associations compared to background contaminants using default parameters except: bait compression set to 2, control compression set to 2, and minFold set to 0 (Teo et al., 2016). A 95% FDR iProphet filter was used before running SAINT. AvgP scores of $\geq$ 0.9 were considered high-confidence protein interactions. All non-human protein interactors (did not start with "NP" in Prey column) were removed from the SAINT analysis, except for APEX2. Prey-prey and bait-bait correlation heatmaps were generated in ProHits-viz (Knight et al., 2017; Teo et al., 2016) using SAINTexpress file generated from ProHits, with a minimum Spectral count sum (of two replicates) cutoff of 5, Pearson correlation, and Euclidean distance settings.

RNA-seq mapping and visualization

All fastq files were first quality checked with FastQC. The RNA-seq libraries generated were mapped to the human reference genome (GRCh38) with GENCODE v34 gene annotation using the STAR alignment tool (Dobin et al., 2013). The mapped reads were counted using featureCounts (Liao et al., 2014). To quantify both unprocessed and mature transcripts, reads that map to each gene locus were counted using the option "-t gene". The mapped data were visualized using WashU Epigenome Browser (Li et al., 2019). To generate genome tracks, samtools was used to merge data from two biological replicates of each pull-down and convert to bigwig file using bamCoverage in Deeptools with CPM normalizing (Li et al., 2009; Ramirez et al., 2016). The sequencing reads displayed a high percentage (~85%) of uniquely mapped reads (Figure S2A), with an average correlation between sequenced replicates of r=~0.9 (Pearson's Correlation; Figures S2C and S2D, and data not shown). For splicing analyses, reads were trimmed to 50 nucleotide segments for vast-tools analysis.

Differential gene expression analysis using DESeq2

Genes differentially enriched in APEX2 pull-downs were determined with DESeq2 (Kucukural et al., 2019; Love et al., 2014) with count tables from featureCounts and first removing genes with <10 reads in any one sample, and by using a threshold of p.adj < 0.05 and log-fold-change (LFC) > 0.5. This study focused on analyzing transcripts enriched in nuclear domains, therefore unless specified, the nuclear-localizing APEX2-NLS fusion was used as control. The main rationale for applying this control in most of our analyses is that APEX2-NLS has a relatively specific and uniform nucleoplasmic distribution, whereas APEX2-GFP and APEX2-FLAG display less intense nuclear localization and substantial cytoplasmic signal (data not shown). Based on these observations, and also the numbers of transcripts and alternative splicing events enriched in the control pull-down samples, APEX2-NLS was considered to represent the most appropriate background control for the APEX2-nuclear domain marker labelling analyses.

Evaluation of sequence features of APEX2-marker enriched transcripts

For transcripts enriched in the APEX2 pull-downs compared to NLS control, an APEX-bait specific weighted mean for each gene based on the GC content and length of all GENCODE-annotated isoforms of the corresponding genes was calculated where the weights are relative isoform abundances quantified by the Salmon pipeline (Patro et al., 2017). For the "all transcripts" control, all transcripts expressed in the HEK293 APEX-NLS sample (cRPKM > 1) were used to calculate weights as described above. In order to investigate characteristics of the genomic loci encoding transcripts associated with nuclear domain markers, we extracted the genomic positions of associated transcripts and quantified the percentage that overlap with TSA-Seq distance bins (Chen et al., 2018) or Hi-C sub-compartments (Liu et al., 2021).

Dual-index scoring method to identify transcripts uniquely associated with each nuclear domain

In order to identify transcripts that are specifically associated with each targeted nuclear domain, we established a dual index scoring method. Index 1 assesses the enrichment of a transcript with a nuclear domain marker in comparison to other nuclear domain marker

pull down samples, as well as the NLS, FLAG and GFP control pull-down samples. To calculate this, the markers for a nuclear domain of interest were compared to all the pull-down samples from all the other nuclear domain marker proteins combined using DESeq2 after filtering by requiring a threshold of p.adj < 0.05 and LFC > 0. Next, a second DESeq2 comparison using the same thresholds was performed between the markers for a nuclear domain of interest and a combination of NLS, FLAG and GFP pull-down samples. For enriched transcripts that overlap between the two comparisons, index 1 was calculated as the quantile of the LFC value from the first comparison (i.e. vs. all the other nuclear body markers). Index 2 represents the reverse quantile of the DESeq2 LFC value from a comparison between the input sample corresponding to a given APEX2 pull-down vs. the combination of the NLS, FLAG and GFP inputs. Therefore, transcripts with the highest enrichment levels in a nuclear domain of interest over the other nuclear body markers, and the controls, harbor high Index 1 scores. Index 2, on the other hand, measures the extent of increased expression of a transcript due to expression of the APEX2 nuclear marker fusion protein. Therefore, a transcript with a high expression level in the input of an APEX2 pull down sample versus input samples from the NLS, FLAG and GFP controls is assigned a low Index 2 score. Thus, a transcript with high Index 1 and Index 2 scores has strong enrichment in an APEX2 nuclear marker pull-down over the other pull-down samples, without showing increased expression in the input controls. To mitigate skewing of Index scores for genes with relatively low expression, we initially removed genes from the analysis with an expression <10 reads in any given sample (see above). Nevertheless, to assess whether the dual index scores might be biased for relatively lowly or highly expressed genes, we plotted the Index 1 and 2 scores against transcript abundance (cRPKM). This revealed a weak correlation (r = $\sim$0.15) between transcript abundance and index scores corresponding to transcripts labelled by the nuclear speckle, lamina, SAM68 and PML body markers (data not shown). As such, we concluded that the index scores are not strongly influenced by transcript abundance.

Post-processing and fluorescence microscopy analysis
Airyscan super-resolution raw data were deconvoluted and processed using ZEN black edition version 2.3 (release version 11.0, Zeiss). Maximum intensity projections (MIP) were used to merge the z-stacks and intensify signals. Pixel intensity display settings were automatically applied with 'linear best-fit analysis' to the entire field of view for the DAPI channel, and manually for the other channels separately using ZEN blue edition (Zeiss). The same thresholds were used for APEX-labelled and unlabeled samples. Z-stacks for all images were analyzed and confirmed that quantified signals are within the nuclear volume (i.e. overlapping with DAPI signal).

RNAScope fluorescence in situ hybridization quantification
Images acquired with the LSM880 Airyscan Fast instrument provide increased sensitivity when compared to standard confocal microscopy, which can be differentiated during post-acquisition image processing. The analysis flow is described in Figures S3C and S3D. After merging z-stack planes (each 0.40 μm) by MIPs, the 'best fit' model (Zeiss Zen) was used to preliminarily adjust intensity thresholds. Next, marker and gene/intron foci were selected with a set threshold based on the average signal intensity across acquired images using Fiji/ImageJ (Schindelin et al., 2015). Then, gene/intron foci (i.e regions of interest, ROIs) were selected by using ImageJ/Fiji's "analyze particles" tool. The assigned ROIs of the different introns and genes displayed a consistent distribution (Figure S3D). These ROIs were then overlaid with the marker channel (i.e MALAT1 or PML), and the fluorescent intensities of marker channel within each ROI, relative to the min/max fluorescent intensity scale of the channel (minimum intensity being 0, and maximum as 1), were quantified in $\sim$67 cells on average. For analyzing foci located at the nuclear periphery (Figure 5D), boundaries were calculated by using the ROIs of the nuclear DAPI staining and intron foci that either overlapped within 0.5 μm of the nuclear boundaries (considered "peripherally located"), or all the other intron foci (considered as "interior located"). For quantification of KRAS foci in Figures 2J and 2K, only those foci that localize in the nucleus were included in the analysis. All statistical tests used for microscopic quantifications are two-sided unless otherwise indicated.

Differential alternative splicing analysis
Alternative splicing analysis of RNA-Seq was performed with vast-tools version 2 (Tapial et al., 2017). Differential alternative splicing was scored with vast-tools' diff module, requiring |dPSI| $\geq$5% and MV_dPSI_at0.95CI > 0. For Figures 4A, S6A, and S6B, only events that meet the minimum threshold in all pull-down samples (score 1 or 2 of read coverage above OK/LOW/VLOW in vast-tools), and have significant changes in at least one nuclear domain marker pull-down were considered. To normalize the potential effect of a change of alternative splicing caused by nuclear domain marker protein expression, we analyzed RNA-seq data from the corresponding input samples. A linear regression model was used to evaluate the relationship between alternative splicing changes of events in pull-down samples (vs NLS pull-down) and in input samples (vs NLS input) and used the residuals of pull-down dPSI after linear regression (i.e after normalizing for PSI changes in the input comparison) as corrected dPSI in Figure 4A, S6A, and S6B. We did not observe a significant effect of input differential AS on pull-down differential AS levels (Figure S5C).

Nuclear speckle-associated introns were defined as intron retention events that are more retained in all three nuclear speckle marker pull-downs compared to the NLS control pull down (dPSI $\geq$5% and MVdPSI>0). To evaluate features of intron retention events associated with nuclear speckle and LMNA, %GC content and lengths of the introns were plotted and compared to a set of 10,000 introns randomly selected from events that were reliably detected by vast-tools in NLS input samples (score 4 of read counts mapped to exon-exon and exon-intron junctions > 10, score 5 of read balance > 0.05, see https://github.com/vastgroup/vast-tools#combine-output-format).

Molecular Cell
Resource

⟲ CellPress

Extended analysis of intron retention levels of nuclear speckle and LMNA associated transcripts

Transcripts differentially enriched in APEX2 pull-downs were called using a threshold of p.adj < 0.05 and LFC > 0.5 as specified above. Only genes with more than three introns and which have at least 75% of the introns quantified by vast-tools (with less than 25% of NAs in vast-tools *combine* table) were considered in the analysis. Percentages of introns per gene, where the introns have PIR>5%, were quantified and plotted, after removing intron retention events with insufficient coverage for quantification.

Regulation of nuclear speckle and LMNA associated introns across tissues

Intron retention across cell and tissue types in VastDB (https://vastdb.crg.eu) (Tapial et al., 2017) was analyzed for introns that were enriched in APEX2 pull-downs for speckle markers, LMNA, or a set of random introns as defined above. VastDB PIR values were first filtered by quality scores (score 4 > 10, score 5 > 0.05), and subsamples of the same sample type according to the VastDB sample annotation were averaged. In each subsample, the percentage of introns among introns associated with speckles, LMNA or random introns with a PIR > 20 was scored.

Overlap of nuclear speckle and LMNA associated introns with RNA binding protein interaction sites and knockdown dependent introns

ENCODE knockdown RNA-seq data was processed with vast-tools v2 as described above requiring $|dPSI| \geq 5\%$ and MV_dPSI_at0.95CI > 0. Only intron retention events that are reliably detected in APEX2 pull-down samples and HepG2 RNA-Seq data (score 4 > 10, score 5 > 0.05), and whose host genes are expressed both in HEK293 and HepG2 cells (ENCODE, FPKM>3) were considered. RNA binding proteins with nuclear speckle localization/association were defined based on a BioID proximity map of the HEK293 nuclear body proteome generated by B.J.A.D. and A.-C. G (manuscript in preparation). Nuclear-localizing non-speckle proteins were defined as proteins annotated as nuclear (Van Nostrand et al., 2020), and not annotated as speckle-localizing by either ENCODE or by BioID proximity mapping data. For each intron, the fraction of nuclear speckle proteins whose knock-down caused its increased retention, minus the fraction of non-speckle nuclear factors causing increased retention, was used to test for the intron being preferentially regulated by nuclear speckle proteins.

eCLIP data from ENCODE was analyzed by downloading IDR-filtered narrow-peak BED files of eCLIP-seq experiments performed in HepG2 from the ENCODE data portal. For each intron in the three groups mentioned above, the difference in the fraction of RBPs from among annotated speckle proteins and nuclear-localizing non-speckle proteins that have a peak, as well as the difference in mean peak density (peaks/kb) between the two groups of RBPs, were scored. Figure 6H and 6I, respectively, show the cumulative distributions of these metrics, while Figures S6G and S6H show the fraction of bound RBPs and peak density for both groups separately.

Gene set enrichment analysis

To assess functional enrichment of genes with introns preferentially enriched with nuclear speckle markers or LMNA, we first obtained enriched terms using the G:GOSt module of g:Profiler with a custom gene space comprising all genes with introns within VastDB that survived quality filtering described above and the 'custom over all known genes' option, with the three GO domains as annotation sources. Only terms comprising at least 5 and at most 1000 genes, and with an intersection with the input of at least 3 genes, were considered. Results were uploaded to the Cytoscape app Enrichment Map together with the gene space GMT file from g:Profiler. A network was constructed using the default parameters. Node size was mapped to intersection size, node border to FDR, and edge width was mapped to overlap represented as the Jaccard coefficient between pairs of nodes. Nodes above an FDR cutoff of 0.05 and edges with a Jaccard coefficient above 3.75 were displayed. The AutoAnnotate app in Cytoscape was then used to cluster nodes with default settings, and cluster names were manually curated from automatically generated three-word cluster labels.

Cell cycle associated alternative intron retention

Analyzed cell cycle-associated intron retention events are from Dominguez et al. (2016). Overlapping intron retention (IR) events detected in the APEX2-nuclear marker pull down data (with cutoffs of dPIR > 5% and MV_dPSI_at0.95CI > 0) were included in the analysis. Clusters of IR events with similar splicing patterns were identified by EuclideanDistanceED (Dominguez et al., 2016) and assigned to distinct clusters (Figure 7D, cluster A-G). Genes enriched in Cluster B overlapping speckle-associated IR events were analysed for functional enrichment compared to a background of genes assessed for IR events using FuncAssociate 3.0 (Berriz et al., 2009). The background set represents genes with intron retention events reliably detected using vast-tools (score 4 > 10, score 5 > 0.05) in HEK293 cells with FDR multiple test correction.

Statistical Analysis

All statistical tests were two-sided, unless stated otherwise. If multiple tests were carried out on the same data, p-values were corrected for multiple testing by Bonferroni correction or as stated in the Results.