

The periphery of nuclear speckles defines a spatially and temporally regulated compartment of long-lived intron-retained RNAs that resolves during mitosis

Received: 13 October 2025

Accepted: 10 July 2026

Published online: 9 September 2026

 Check for updates

Josep Biayna ^{1,2,3,8}, Artem Baranovskii^{4,7,8}, Anusha Chaudhuri ^{1,2,3}, Marta Paladin ^{1,2,3}, Berkay Erdem^{1,2,3}, Luise-Elektra Keller^{1,2,3}, A. Rasim Barutcu⁵, Stefanie Dimmeler ^{1,2,3}, Annalisa Marsico⁴ & Gabrijela Dumbović ^{1,2,3,6} 

RNA localization adds a fundamental layer to gene expression by determining when and where translation-ready mRNAs become available, yet how this timing is coordinated with nuclear architecture and cell-cycle progression remains unclear. Here we identify a subnuclear RNA niche at the nuclear speckle periphery that couples intron retention to cell-cycle-timed RNA release. Using compartment-resolved transcriptional inhibition, sequence-based deep learning and single-molecule and super-resolution RNA imaging in human pluripotent stem cells, we define a class of nuclear RNAs with long-lived retained introns that persist for hours and are enriched in transcripts encoding regulators of genome maintenance and mitosis, including centromere and kinetochore assembly, DNA repair and telomere maintenance. Long-lived retained introns exhibit elevated GC content, predicted structural stability and enrichment for nuclear speckle-associated RNA-binding proteins. In interphase, these RNAs localize to a distinct nuclear speckle-peripheral RNA niche in a spatial arrangement conserved across cell types. During mitotic remodelling, they undergo coordinated, kinase-dependent splicing and are released into the cytoplasm of early G1 daughter cells. Together, these findings link *cis*-encoded intronic features, subnuclear organization and mitotic remodelling to temporal control of RNA fate.

Spatial and temporal RNA control governs when and where mRNAs become accessible for translation, shaping development, cell polarization and stress responses, yet many transcriptomic approaches lack the resolution needed to determine where RNAs reside, how long they persist there or how localization relates to function^{1–9}. RNA mislocalization contributes to disease development, making these principles clinically relevant^{10–14}.

Persistence of an intron in otherwise mature RNA, known as intron retention (IR), is linked to nuclear RNA retention, transcriptome

diversification and rapid gene activation in response to external cues, with implications in differentiation, stress adaptation and disease development^{8,15–25}. Retained introns (RIs) are functionally diverse and can occur in cytoplasmic transcripts that are translated or degraded, or whose RIs direct RNA localization; in unstable transcripts cleared through RNA-surveillance pathways; and in nuclear transcripts whose RIs can be removed in response to specific cues, a class termed detained introns^{8,15,17,18,21,26–35}. Detained introns established nuclear IR as a regulated RNA state rather than inefficient splicing. However, it remains

A full list of affiliations appears at the end of the paper. ✉ e-mail: dumbovic@med.uni-frankfurt.de

unclear whether nuclear IR comprises kinetically distinct populations with different residence times and regulatory behaviours, how these populations are spatially organized, whether specific microenvironments coordinate their timed release, and how physiological transitions involving large-scale nuclear reorganization, such as mitosis, couple to their regulation.

Among nuclear membraneless organelles³⁶, nuclear speckles are enriched in splicing and mRNA processing factors and have been proposed to act as hubs for RNA processing, nuclear mRNA retention and export licensing^{37–42}. They contain concentrated serine and arginine-rich (SR) proteins and other splicing regulators whose phosphorylation controls their partitioning to nuclear speckles and splicing activity⁴³. CDC-like kinases (CLKs) and dual-specificity tyrosine-phosphorylation-regulated kinase 3 (DYRK3) regulate phosphorylation-dependent remodelling of nuclear speckles and their associated splicing factors^{44–47}. Thus, nuclear speckles represent dynamic, kinase-regulated environments in which RNA processing, retention and release are spatially coordinated.

Splicing can occur co- and post-transcriptionally^{48–57}. Nascent transcripts from strongly induced genes localize near nuclear speckles⁵⁸, and highly active transcription sites can generate enlarged nuclear speckle-like domains, supporting coupling between transcriptional activity and nuclear speckles⁵⁹. Recent proximity-labelling studies showed that intron-retained (IR-) RNAs frequently associate with nuclear lamina or nuclear speckle markers⁶⁰; however, these approaches lack the resolution to define the precise spatial relationship of IR-RNAs with nuclear speckles or to distinguish condensate association from signals arising from transcription sites, often near nuclear speckles^{61–64}.

Here, we combine compartment-resolved transcriptional inhibition, sequence-based deep learning and RNA imaging in human induced pluripotent stem (hiPS) cells and primary human umbilical vein endothelial cells (HUVECs) to define the kinetic and spatial organization of IR-RNAs. We identify a class of nuclear IR-RNAs with long-lived IR (LL-IR-RNAs) in hiPS cells that exhibits elevated IR levels and enrichment in genome maintenance and mitosis-linked pathways. High-resolution single-molecule RNA (smRNA) imaging revealed that poly(A)⁺ IR-RNAs occupy a distinct spatial niche compared with bulk poly(A)⁺ RNAs: whereas global poly(A)⁺ RNAs accumulate within nuclear speckles^{38,41,42}, IR-RNAs are enriched at the nuclear speckle periphery, with LL-IR-RNAs more proximal to and distributed around multiple nuclear speckles, defining a juxta-speckle RNA niche conserved across cell types.

Sequence-based deep learning distinguished retained from efficiently spliced introns and long-lived from short-lived RIs, revealing that intron longevity is associated with elevated GC content, structured RNA elements, nuclear speckle-associated RNA-binding proteins (RBPs) and proximity, indicating that IR-RNAs reside in a splicing-permissive environment where *cis*-inhibitory features may delay intron removal. LL-IR-RNAs undergo coordinated, kinase-dependent intron removal in mitosis and populate early G1 daughter cells as cytoplasmic spliced transcripts, consistent with prolonged intron detention followed by

a major mitotic splicing wave. Together, our data link *cis*-encoded intronic features, nuclear organization and mitotic remodelling to the temporal control of RNA fate, suggesting that nuclear storage and mitotic release of RNAs encoding genome maintenance and cell cycle regulators can promote their inheritance into early G1 daughter cells.

Results

IR associates with distinct nuclear localization patterns in hiPS cells

To examine the association of IR with RNA compartmentalization in hiPS cells, we biochemically fractionated cells into nuclear and cytoplasmic fractions, performed poly(A)⁺ RNA sequencing (RNA-seq) and quantified IR (Methods; Fig. 1a). As expected, long noncoding RNAs (lncRNAs) exhibited higher nuclear percentage of IR (PIR) than mRNAs^{16,17,23,65} (Extended Data Fig. 1a). We applied a stringent PIR threshold ($\geq 30\%$) to define IR to minimize nascent RNA background. Gene Ontology (GO) analysis showed IR-RNA enrichment in mRNA processing, chromatin regulation, splicing, cell cycle and DNA replication pathways (Fig. 1b), similar to studies in mouse and quiescent stem cells^{17,21}. RIs were shorter than spliced introns (mean 1.8 kb versus 4.0 kb), had higher GC content (mean 52% versus 46%) and their host RNA showed higher nuclear enrichment (mean $\log_2$ fold change (FC) 1.3 versus -0.05 ; all $P < 2.2 \times 10^{-16}$, two-sided Welch's *t*-test; Fig. 1c), consistent with previous findings in other cell types^{16,24,35,66,67} and suggesting sequence composition biases introns towards retention.

To investigate the localization of IR-RNAs at the single-molecule level, we applied smRNA fluorescence in situ hybridization (smRNA FISH) on 21 candidates (18 mRNAs and 3 lncRNAs) spanning diverse functional categories identified in Fig. 1b, including chromatin regulation, RNA metabolism, cell cycle, DNA damage response, genome maintenance and vesicle trafficking (Fig. 1d,e and Extended Data Fig. 1b). IR for some candidates was previously reported, although not necessarily the same intron^{23,68–76}. Quantitative distribution analysis revealed three major localization patterns (Fig. 1e,f and Extended Data Fig. 1b). Class I transcripts (33%) were predominantly nuclear ($\geq 70\%$ of RNA molecules) in their IR form with broad distribution, including TERT, CENPT, FANCA, BRD8, RAD52, LINC00106 and MEG3. Class II (38%) showed intermediate nuclear localization (55–70% of RNA molecules), most displaying a broad distribution, consistent with substantial coexistence of IR- and spliced isoforms, including METTL3, EZH2, HDAC7, COG4, DDX17, LAMA5, TUG1 and TEO2. Class III (29%) exhibited lower nuclear abundance ($< 55\%$ of RNA molecules) and predominant cytoplasmic localization of spliced isoform, with RI signals frequently, but not necessarily, at or near transcription sites, including TP53BP1, SFPQ, WEE1, DNMT3B, SON and PTBP1.

Together, this underscores two fates of IR-RNAs, a broadly mobile nuclear pool and a locally constrained population, which is indistinguishable from PIR values in compartment-resolved RNA-seq alone, suggesting alternative routes of post-transcriptional control. Interestingly, most nuclear-enriched IR-RNAs encode regulators of essential pathways, including centromere and kinetochore assembly (CENPT⁷⁷), DNA repair (FANCA⁷⁸), chromatin regulation (BRD8⁷⁹),

Fig. 1 | IR is associated with distinct nuclear RNA localization patterns in hiPS cells. **a**, Schematic of the experimental design. **b**, GO analysis of genes with IR (nuclear PIR $\geq 30\%$). **c**, Violin plots showing length (left) and GC content (middle) of RIs (IR; $n = 12,019$ introns) and spliced introns (no IR; $n = 73,704$ introns). Intron length is displayed on a $\log_{10}$ scale. Right: nuclear enrichment ($\log_2(\text{TPM}_{\text{nuc}}/\text{TPM}_{\text{cyto}})$) of host RNAs containing at least one RI (IR-RNA; $n = 5,686$ transcripts) or only spliced introns (no IR-RNA; $n = 4,468$ transcripts). Midline, median; box limits, first and third quartiles; whiskers, most extreme values within 1.5 $\times$ interquartile range (IQR) of the hinges. *** $P \leq 0.001$, two-sided Welch's *t*-test. All comparisons, $P < 2.2 \times 10^{-16}$. **d**, Schematic of the smRNA FISH experimental and analysis workflow. Co-localized exon and intron signals were classified as IR-transcripts, whereas exon-only signals were classified as spliced transcripts.

TS, transcription site. **e**, Left: dot plot showing nuclear PIR of individual introns for the indicated genes, calculated using vast-tools from nuclear poly(A)⁺ RNA-seq ($n = 2$). Introns selected for smRNA FISH are marked with an asterisk. Middle: representative smRNA FISH maximum intensity projections. Exon, white; intron, magenta; nucleus, blue; scale bar, 2 μm . Right: smRNA FISH quantification of nuclear transcript fraction (%) ($n_{\text{exon,nuc}}/n_{\text{exon,total}}$), nuclear exon spots and cytoplasmic exon spots ($n = 50$ cells per gene, except TERT ($n = 55$), two independent stainings). Dots, individual cells; horizontal lines, means. Additional representative images are shown in Extended Data Fig. 1b. **f**, Schematic of observed spatial distribution patterns of IR-RNAs. Arrow indicates increasing cytoplasmic transcript fraction.

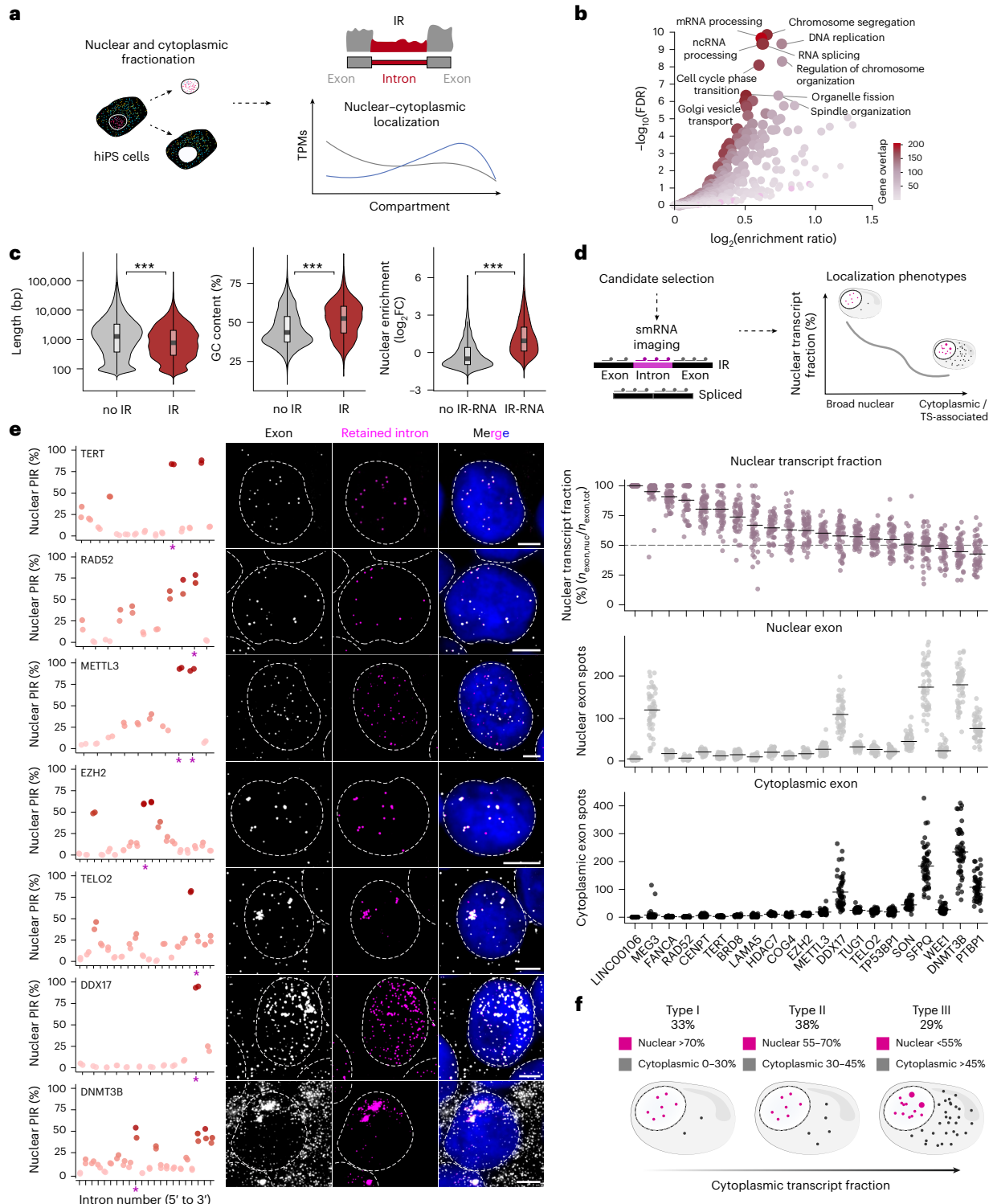
telomere maintenance (TERT⁸⁰) and m⁶A RNA methylation (METTL3⁸¹), indicating IR as a spatial checkpoint.

Transcriptional inhibition time course reveals long-lived IR-RNAs

To determine how long RIs persist on subcellular timescales, we performed poly(A)⁺ RNA-seq on nuclear and cytoplasmic fractions after transcriptional inhibition with Actinomycin D (ActD) at 0.5, 2 and 4.5 h (Fig. 2a). The majority of IR events were enriched in nuclear RNA, as expected^{16–18,35}, with a substantial fraction resolved between 0.5 h and 2 h, and a subset persisting throughout (Fig. 2b). RIs were classified as

unstable at 0.5, 2 or 4.5 h, if PIR dropped below 50% of baseline at the respective timepoint, or stable if maintained throughout (Fig. 2b–d). Among 9,111 RIs, 1,031 (11.3%) resolved within 0.5 h, 3,312 (36.3%) by 2 h and 899 (9.9%) by 4.5 h, while 3,869 (42.5%) remained stable (Fig. 2c and Extended Data Fig. 2a). Stratification by localization class (Fig. 1e,f) revealed a correspondence between spatial pattern and kinetic stability: class I introns were predominantly stable (85.7%), class II mixed (37.5% stable) and class III were predominantly unstable (83.3%).

Stable RIs had higher GC content (53.7% versus 51.5–51.9% in unstable groups, $P_{\text{adj}} < 1.3 \times 10^{-4}$, Tukey's honestly significant difference (HSD) test stable versus unstable groups), whereas intron length



showed no substantial differences (Fig. 2e), indicating that sequence composition rather than length correlates with persistence. Baseline PIR and nuclear RNA enrichment both increased progressively with stability (Fig. 2e). We termed this stable class, defined by higher GC content, elevated baseline PIR and nuclear retention, long-lived RIs (LL-RIs), and the unstable group short-lived RIs (SL-RIs). GO analysis showed that LL-RIs are enriched in DNA repair, replication, cell cycle and mitosis-related pathways, whereas SL-RIs are enriched in RNA processing, metabolism and protein localization pathways (Fig. 2f), linking IR kinetics to gene function and indicating that genome maintenance and cell cycle pathways associate with the longest-lived IR.

Analysis of mean relative RNA levels across compartments showed that LL-IR-RNAs were, on average, more stable in the nucleus, with a decrease at 4.5 h (Fig. 2g). By contrast, groups with shorter IR persistence showed progressively lower average nuclear RNA stability. Cytoplasmic RNA levels declined faster than nuclear levels across all groups, most prominently for the unstable 0.5 h group, indicating partial uncoupling between nuclear and cytoplasmic RNA stability. These data indicate that temporal IR stability is associated with increased nuclear RNA persistence, while cytoplasmic RNA is subject to higher turnover.

Because the 30% PIR cutoff can include transcripts in which RIs represent a minority of the pool, we repeated the analysis at 50% PIR to focus on IR-dominant transcript populations. Trends in basal PIR, GC content, nuclear enrichment and nuclear decay were preserved, with GC content differences becoming sharper and increasing progressively across stability groups (mean 50.5%, 51.6%, 52.8%, 54.6%, $P_{\text{adj}} = 1.93 \times 10^{-6}$, Tukey's HSD test stable 4.5 h versus unstable 0.5 h; Extended Data Fig. 2b), reinforcing the link between GC content and IR stability, while nuclear decay of the unstable 0.5 h group became stronger (Fig. 2g). Increasing the PIR threshold therefore filtered out marginal events and sharpened the distinction between rapidly changing IR events and a kinetically distinct, long-lived IR population. We validated these results with smRNA FISH on 18 RIs (9 stable), confirming the persistence of LL-RIs after transcriptional inhibition, and rapid depletion of SL-RIs. Furthermore, stable RIs were predominantly found in broadly distributed RNAs, with high basal PIR and high nuclear transcript fraction (Fig. 2h and Extended Data Fig. 2c,d).

Because detained introns represent the best-characterized reference of nuclear polyadenylated IR-RNAs that can undergo post-transcriptional splicing, we compared the stability classes to detained introns reported by Boutz et al.¹⁷ and Braun et al.¹⁸. Annotated detained introns distributed across multiple stability groups (Extended Data Fig. 3a,c), with 42–56% having lower baseline PIR in our

dataset and mapping to the no IR category, probably reflecting differences in both cell type and event definition. Among those that could be assigned to our IR stability groups (46–58%), 59–61% mapped to unstable groups, whereas 39–41% to the stable group. Only a minority of our stable introns overlapped annotated detained introns, corresponding to ~15% in the Braun dataset and ~30% in the Boutz dataset (Extended Data Fig. 3a,c).

In the work of Boutz et al., half-life (HL) was measured for six detained introns and averaged ~29 min (ref. 17). Consistent with this, annotated detained introns remained shorter-lived than the stable class (mean HL 3.2–3.4 h versus 9.99 h), had lower basal PIR (mean ~35–42% versus ~66%), greater intron length (mean $\log_{10}(\text{length})$ ~3.07–3.21 versus ~2.86) and lower GC content (mean ~46–47% versus ~54%; all $P_{\text{adj}} < 2.2 \times 10^{-16}$, Tukey's HSD test; Extended Data Fig. 3b,d). Detained introns in the work of Boutz et al. were enriched in RNA processing, stress response, DNA damage and cell cycle regulation pathways¹⁷, consistent with the broad functional enrichment in Fig. 1b, whereas the long-lived class shows a more specific enrichment in mitosis, DNA repair and genome maintenance pathways. These comparisons place stable IR within the broader intron detention framework as a long-lived population.

Collectively, LL-RIs (42% of the IR pool) are GC-rich, nuclear-enriched, stable and linked to cell cycle, mitosis and DNA repair regulators, while SL-RIs predominate among RNA metabolism genes, suggesting that IR dynamics contribute to post-transcriptional control in a context-dependent manner.

RIs are predictable from sequence features linked to RBP binding and are enriched for nuclear speckle-associated RBPs

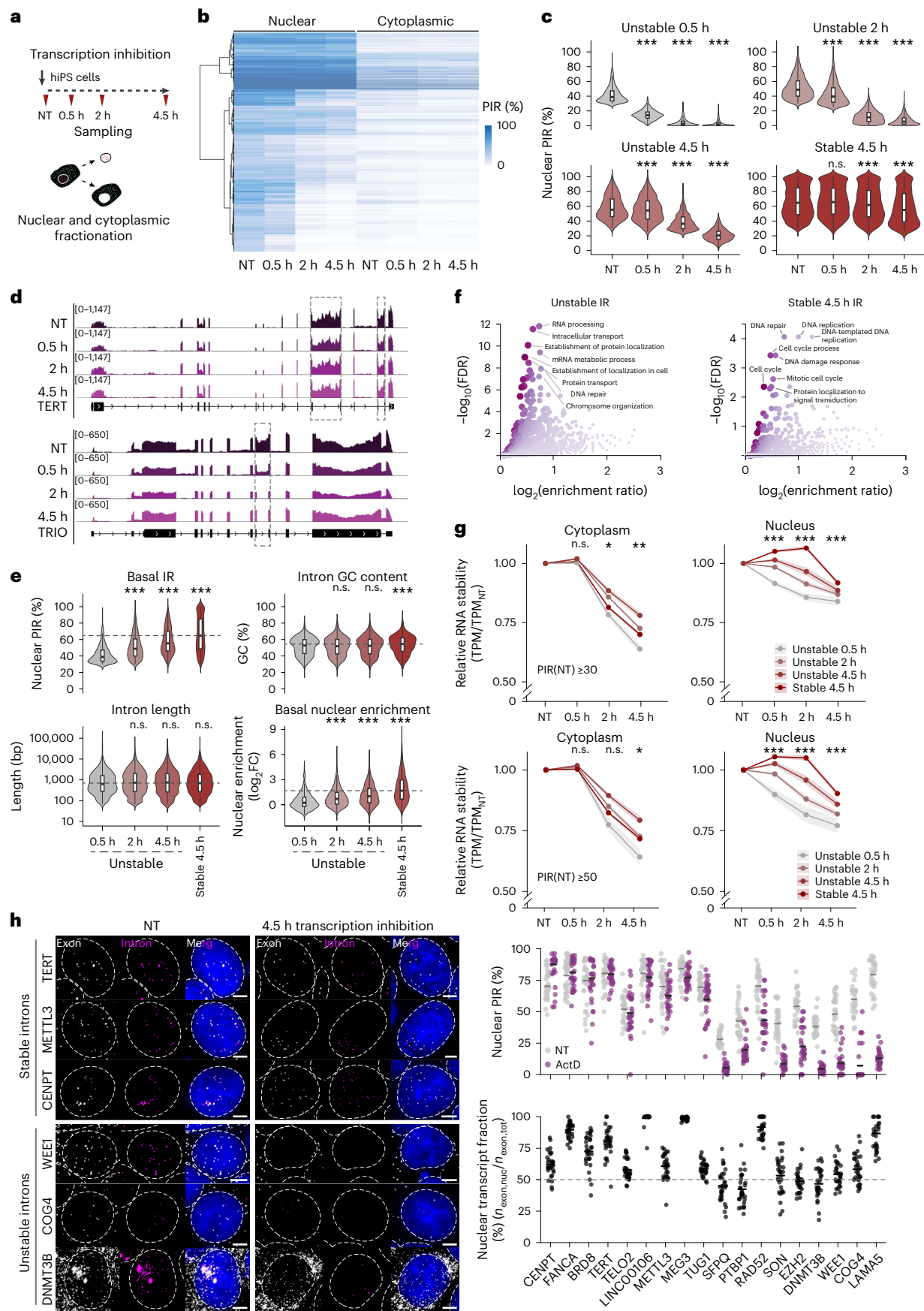
We next sought to unbiasedly characterize sequence features of RIs and LL-RIs using Parnet⁸², an experimentally grounded in-house RNA foundation model trained on the full ENCORE RBP eCLIP dataset. We first fine-tuned Parnet to classify retained versus non-retained introns (Methods), independent of HL (Fig. 3a and Extended Data Fig. 4a). The model distinguished retained from non-retained introns with high accuracy (area under receiver operating characteristic curve and average precision 86% and 83%, respectively; henceforth, AUROC and AP, respectively; Fig. 3b) and the two classes separated in embedding space (Fig. 3c), indicating that IR is encoded in distinct sequence features, capturing RNA–RBP interaction potential and local sequence information.

Although LL-RIs show higher basal retention than unstable introns (Fig. 2e and Extended Data Fig. 4b), we next fine-tuned Parnet to distinguish LL-RIs from SL-RIs (Methods), defined by the lower $\log(\text{IR HL}) < -1.34$ and upper $\log(\text{IR HL}) \geq -0.36$ quartiles of the HL distribution (Methods; Fig. 3d). Despite the smaller, more complex training set, Parnet achieved clear separation between SL- and LL-RIs (Fig. 3e) and

Fig. 2 | Transcriptional inhibition time course reveals a stable class of IR-RNAs.

a, Schematic of the experimental approach. **b**, Heatmap of PIR across four timepoints after transcriptional inhibition in nuclear and cytoplasmic fractions. **c**, Violin plots of nuclear PIR across timepoints for introns grouped by IR stability ($n = 1,031, 3,312, 899$ and $3,869$ introns; unstable 0.5 h, unstable 2 h, unstable 4.5 h, stable 4.5 h groups, respectively). n.s., not significant, *** $P \leq 0.001$, evaluated by one-way ANOVA followed by Tukey's HSD test versus NT. P_{adj} for 0.5 h, 2 h and 4.5 h versus NT, respectively: unstable 0.5 h, 3.19×10^{-8} , 3.19×10^{-8} , 3.19×10^{-8} ; unstable 2 h, $P_{\text{adj}} = 3.20 \times 10^{-10}$, 3.20×10^{-10} , 3.20×10^{-10} ; unstable 4.5 h, $P_{\text{adj}} = 1.65 \times 10^{-4}$, 5.65×10^{-9} , 5.65×10^{-9} ; stable 4.5 h, $P_{\text{adj}} = 3.26 \times 10^{-7}$, 1.60×10^{-8} . **d**, Poly(A)⁺ RNA-seq coverage across genes with stable IR (top) or unstable IR (bottom) across four timepoints. **e**, Violin plots of basal PIR, GC content and length of RIs grouped by IR stability as in **c** (with the same intron n), and of host RNA nuclear enrichment ($\log_2(\text{TPM}_{\text{nuc}}/\text{TPM}_{\text{cyto}})$) for the corresponding transcript groups ($n = 405, 1,391, 478$ and $2,438$ transcripts across increasing stability groups). n.s., not significant, *** $P \leq 0.001$, evaluated by one-way ANOVA followed by Tukey's HSD test versus unstable 0.5 h. Basal PIR, $P_{\text{adj}} < 2.2 \times 10^{-16}$ for unstable 2 h, unstable 4.5 h and stable 4.5 h; GC content, $P_{\text{adj}} = 1.71 \times 10^{-6}$ for stable 4.5 h; host RNA nuclear enrichment, $P_{\text{adj}} = 4.98 \times 10^{-8}$, 3.51×10^{-8} and 3.51×10^{-8} for unstable 2 h, unstable 4.5 h and stable 4.5 h, respectively. **f**, GO analysis of

stable and unstable IR-RNAs (circle size, gene count). **g**, Relative RNA stability compared with non-treated cells ($\text{TPM}/\text{TPM}_{\text{NT}}$) in cytoplasmic (left) and nuclear (right) fractions for IR-RNAs grouped by IR stability and stratified by baseline PIR $\geq 30\%$ (top; $n = 368, 1,232, 422$ and $1,675$) or PIR $\geq 50\%$ (bottom; $n = 110, 749, 295$ and $1,233$), with group sizes reported in the same order as in **c**. * $P < 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, evaluated by two-sided Welch's t -test stable 4.5 h versus unstable 0.5 h. PIR $\geq 30\%$, cytoplasm, $P_{\text{adj}} = 4.28 \times 10^{-2}$, 3.41×10^{-3} at 2 h and 4.5 h, respectively; nucleus, $P_{\text{adj}} = 6.17 \times 10^{-29}$, 2.46×10^{-25} , 5.20×10^{-4} at 0.5 h, 2 h and 4.5 h, respectively. PIR $\geq 50\%$, cytoplasm, $P_{\text{adj}} = 4.31 \times 10^{-2}$ at 4.5 h; nucleus, $P_{\text{adj}} = 7.71 \times 10^{-11}$, 6.00×10^{-10} , 1.78×10^{-4} at 0.5 h, 2 h and 4.5 h, respectively. Dots, mean; shadows, standard error of the mean. **h**, Representative smRNA FISH maximum intensity projections of stable and unstable IR-RNAs in non-treated cells and 4.5 h after transcriptional inhibition. Exon, white; intron, magenta; nucleus, blue; scale bar, 2 μm . Right: nuclear PIR and nuclear transcript fraction (%) ($n_{\text{exon,nuc}}/n_{\text{exon,total}}$) from smRNA FISH quantification. Dots, individual cells; horizontal lines, means; $n = 30$ cells, two independent stainings. Additional representative images and quantifications are shown in Extended Data Fig. 2c,d. For all panels, NT, non-treated cells. For **c** and **e**, midline, median; box limits, first and third quartiles; whiskers, most extreme values within $1.5 \times \text{IQR}$ of the hinges.



performed well above random (AUROC 71%, AP 73%; Fig. 3b), indicating that sequence features contribute to IR stability.

Independent analysis revealed an enrichment of nuclear speckle-associated RBP binding sites as a hallmark of RIs overall, particularly of LL-RIs (Fig. 3f and Extended Data Fig. 4c), which had increased ENCODE eCLIP RBP peaks compared with efficiently spliced introns and SL-RIs (Methods; Extended Data Fig. 4d). We interpreted the learned embeddings to assess whether the fine-tuned foundation model captures these features. Projecting SL- and LL-RI representations and colouring by nuclear speckle-associated RBP peaks revealed a clear gradient: LL-RIs clustered where ENCODE nuclear speckle RBP binding was higher (Extended Data Fig. 4e). Thus, the model encodes intronic features linked to nuclear speckle association and discriminates SL- from LL-RIs accordingly. Consistently, assay of reverse transcription-based RBP binding-site sequencing (ARTR-seq) data⁷⁴, approximating RNA enrichment by tethering reverse transcriptase to nuclear speckle markers, showed a similar LL-RI-to-SL-RI gradient and RI enrichment in nuclear speckle-associated fractions (Fig. 3g,h). Together, these results reinforce the association between nuclear speckles and intron fate.

In line with this, Parnet embeddings revealed a strong link between RIs, LL-RIs and GC content (Fig. 3i and Extended Data Fig. 4f), confirming their GC richness (Fig. 2e). This is consistent with proximity- and antibody-guided RNA-labelling studies showing that RIs enriched with nuclear speckle markers tend to be shorter and GC-rich^{60,74} (Extended Data Fig. 4f). Because GC-rich sequences can promote R-loop formation and transcriptional pausing⁸³, we analysed single-strand DNA ligation-based DNA:RNA immunoprecipitation sequencing (ssDRIP-seq, a method for genome-wide detection of R-loops) data from the same cell line⁸⁴. While R-loop peaks correlated with GC content, nuclear speckle enrichment and localization, we found no direct relationship between R-loops and IR or IR HL (Extended Data Fig. 4f,g). Bisulfite sequencing⁸⁴ showed that most introns were highly methylated, except for a small group of short, conserved, demethylated first introns overlapping R-loops⁸⁵ (Extended Data Fig. 4h–j). Intron length and repeat content showed negative correlation with basal PIR, whereas association with IR HL approached non-significance (Extended Data Fig. 4f,i,k). As expected, both basal PIR and IR HL correlated with the nuclear enrichment of the host gene (Extended Data Fig. 4f,l). Parnet embeddings revealed that annotated retained introns^{17,18} mostly overlapped with SL-RIs and had lower ARTR-seq nuclear speckle enrichment⁷⁴ (mean log₂FC 0.8 versus 1.6, respectively, $P < 2.2 \times 10^{-16}$, Tukey's HSD test; Extended Data Fig. 4m,n).

Fig. 3 | IR-RNAs are associated with nuclear speckles and enriched in G-quadruplexes.

a, Schematic illustrating the process of Parnet's fine-tuning to predict IR or IR HL. One-hot encoded sequences of intron termini are passed through the Parnet model to generate embeddings, which are then decomposed by the CAM modules assigning per-position importance score and a probability of a given intron being retained/stable. Two probabilities are then fused to generate the final prediction. **b**, Precision–recall (PRC; left) and receiver operating characteristic (ROC; right) curves across five folds of Parnet fine-tuning for IR (purple) and IR HL (pink) classification. Best-fold average precision (AP) and AUROC values are shown. **c**, UMAP projection of IR fine-tuned Parnet intron embeddings, where each point represents an intron and distances reflect learned feature similarity. Contour lines indicate kernel density estimates for the two IR classes. In the legends, dim stands for UMAP dimension. **d**, Distribution of gene-level normalized IR HL, with LL-RI and SL-RI cutoffs indicated. **e**, As in **c**, using IR HL fine-tuned embeddings; contours represent densities of IR HL classes defined in **d**. **f**, Boxplots showing the distribution of bootstrapped average deltas between nuclear speckle and non-nuclear speckle RBP peaks in the four intron classes defined in **d** and Extended Data Fig. 4a ($n = 2,000$ in all four cases). Midline, median; box limits, first and third quartiles; whiskers, most extreme values within $1.5 \times$ IQR of the hinges. LL-RIs versus SL-RIs, $P_{\text{adj}} = 3.57 \times 10^{-22}$; retained versus spliced introns, $P_{\text{adj}} = 9.96 \times 10^{-43}$, as estimated by two-way ANOVA. *** $P \leq 0.001$. **g**, Boxplots showing the distribution of nuclear

We next asked whether the GC richness of LL-IRs reflects enhanced RNA structure. LL-RIs displayed significantly lower length-normalized predicted minimum free energy (MFE) than unstable or spliced introns, indicating greater structural stability (Fig. 3j). Among structural elements, G-quadruplexes (G4s), which modulate pre-mRNA processing^{86,87} and accumulate near weak splice sites⁸⁶, were ~3-fold enriched in RIs (Fig. 3k). For all motif categories (Methods), meta-profiles revealed a preferential G4 localization near the 5' of RIs, declining towards the intron body and 3' end (Fig. 3l). Two G4 motifs showed a secondary enrichment near the 3' end, in proximity to branch site. G4 motifs were modestly enriched in LL-RIs (~1.5-fold) while preserving positional bias (Extended Data Fig. 5a,b).

Attribution analyses (Methods) identified additional predictive motifs: canonical 5' splice site and 3' polypyrimidine tract motifs, hallmarks of efficient splicing^{88–90} (Extended Data Fig. 5c), were depleted in RIs, consistent with previous studies⁹¹, suggesting that weakened splice signals and local structures jointly promote IR. Attribution maps also highlighted RBP binding sites⁹²: positive attributions in RIs included RBFOX1/3, ILF2, EWSR1, TAF15, RBM23, PCBP2, HNRNPF and SRSF1; the latter four localized to nuclear speckles⁹³ (Extended Data Fig. 5d, bottom). Negative attributions, enriched in efficiently spliced introns, identified six significant matches: CPEB3/4, ELAVL1/2, NUPL2 and RBM15 (Extended Data Fig. 5d, top). All except NUPL2 are established regulators of alternative splicing^{94,95}; NUPL2, a nucleoporin-like factor, links late splicing to mRNP export⁹⁶.

Compared with SL-RIs, LL-RIs showed weaker canonical 5' splice sites, stronger GC-rich enrichment at the 5' end, and overrepresented motifs for nuclear speckle-localized RBPs: RBM4/RBM4B, which bind G-rich sequences such as G4s, and PCBP2, ARGLU1 and HNRNPF, known alternative splicing regulators^{97–105} (Extended Data Fig. 5e,f).

Together, these analyses show that IR and stability are encoded in distinct sequence and structural features. RIs display weakened splice sites, G4 motif enrichment and splice silencing elements characterized by RBP motifs for SRSF1 or HNRNPF, while LL-IRs are further defined by GC-rich, structure-prone sequences and nuclear speckle-associated RBP interactions, suggesting that intron longevity arises from the interplay of *cis*-encoded features, RBP associations and spatial environment.

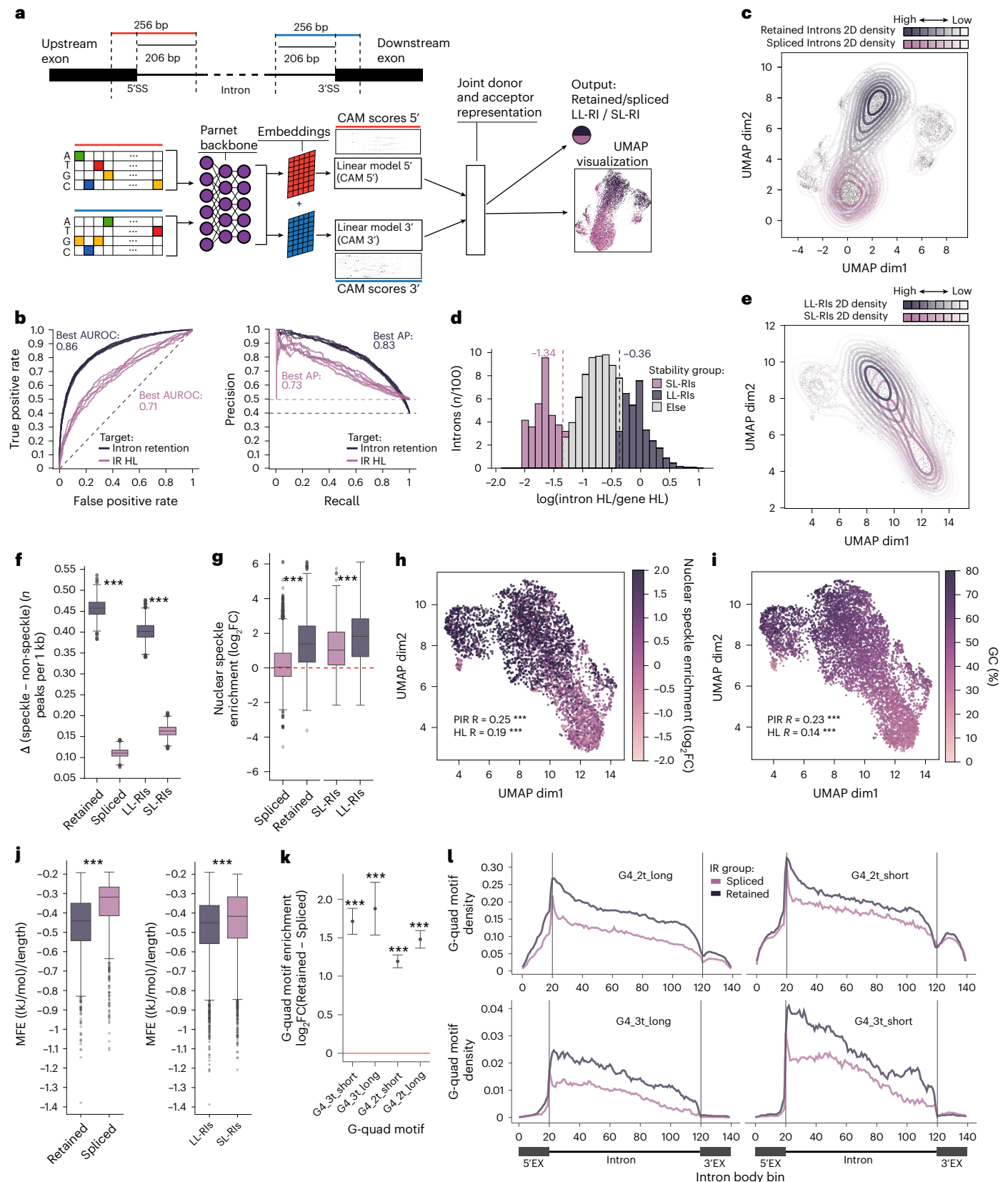
IR-RNAs localize to a distinct niche at the periphery of nuclear speckles

Nuclear speckles have a role in nuclear RNA retention^{38,41,42}, and proximity labelling showed that IR-RNAs associate with nuclear speckle markers⁶⁰. However, the limited spatial resolution of proximity-based

speckle enrichment across the four classes defined in **d** and Extended Data Fig. 4a. Box bounds are defined as in **f**. *** $P \leq 0.001$, as estimated by Welch's two-sided *t*-test. LL-RIs ($n = 2,140$) versus SL-RIs ($n = 2,322$), $P_{\text{adj}} = 4.30 \times 10^{-57}$; retained ($n = 5,867$) versus spliced ($n = 7,618$) introns, $P_{\text{adj}} = 2.2 \times 10^{-308}$. Nuclear speckle enrichment data were derived from ref. 74. **h**, Same as **e**, but coloured by nuclear speckle enrichment of the intron hosting gene. Pearson's *r* values between IR and IR HL, and colour-coded variables are reported on the plot. **i**, Same as in **e**, but introns coloured by GC content. **j**, Boxplots showing the distributions of length-normalized MFE estimates for introns subsampled from four classes defined in **d** and Extended Data Fig. 4a. *** $P \leq 0.001$, Welch's two-sided *t*-test. LL-RIs ($n = 2,982$) versus SL-RIs ($n = 2,988$), $P_{\text{adj}} = 2.42 \times 10^{-11}$; retained ($n = 734$) versus spliced ($n = 966$), $P_{\text{adj}} = 3.97 \times 10^{-40}$. Box bounds are defined as in **f**. **k**, Dot plot showing the enrichment of four G4 motif counts. Points, log₂ fold enrichment of motif-hit rates in retained ($n = 9,111$) relative to spliced ($n = 11,131$) introns; error bars, 95% confidence intervals for the log₂ fold enrichment. *** $P \leq 0.001$. Significance was estimated from the corresponding Poisson generalized linear model class coefficient (see 'G-quadruplex motif enrichment' section in the Methods). *G4_3t_short*, $P_{\text{adj}} = 3.4 \times 10^{-80}$; *G4_3t_long*, $P_{\text{adj}} = 2.11 \times 10^{-22}$; *G4_2t_short*, $P_{\text{adj}} = 1.26 \times 10^{-190}$; *G4_2t_long*, $P_{\text{adj}} = 3.13 \times 10^{-149}$. **l**, Meta-profiles showing four G4 motif distributions along the intron body and flanking exons of retained (purple) and spliced (pink) introns as defined in Extended Data Fig. 4a. EX stands for exon and denotes the borders between the intron body and the flanking exons.

biotinylation, constrained by diffusion of reactive intermediates and the nanoscale organization of nuclear subdomains, leaves unclear how IR-RNAs are organized relative to nuclear speckles, how spatially constrained these associations are, and how their association relates to the internal core and shell architecture of nuclear speckles^{106,107} (Fig. 4a). To address this, we employed high-resolution microscopy combining

immunofluorescence (IF) for SC35 (SRSF2) to detect nuclear speckle core and U2 small nuclear RNA (snRNA) smRNA FISH to label the shell, and computed shortest distances of each IR-RNA spot across an IR-RNA panel spanning different stability categories to the segmented surfaces of the SC35-positive nuclear speckle core and the U2-positive nuclear speckle shell (Methods).



Oligo(dT) smRNA FISH and SC35 IF showed global poly(A)⁺ RNAs within nuclear speckles, as expected^{41,42} (Fig. 4b). By contrast, IR-RNAs were predominantly excluded from the SC35-positive nuclear speckle core and positioned close to its outer surface, with median nearest distances to its outer surface ranging from -0.0470 to $0.261\ \mu\text{m}$ (median across RNAs, $0.123\ \mu\text{m}$; Fig. 4c and Extended Data Fig. 6a). This positioning was non-random; all IR-RNAs were significantly shifted towards the core surface compared with randomly generated nuclear spots (one-sided Kolmogorov–Smirnov test, hereafter KS test; $P_{\text{adj}} < 2.2 \times 10^{-16}$; Extended Data Fig. 6b), with median distance shifts from -0.565 to $-0.257\ \mu\text{m}$ and strong enrichment at $\leq 0.1\ \mu\text{m}$ from the core surface relative to random spots (median across RNAs, 16.4-fold; range, 7.9–32.4-fold; Supplementary Table 1). This indicates that IR-RNAs predominantly localize at the nuclear speckle periphery outside the SC35-positive core. To determine whether this distribution is conserved across cell types, we analysed HUVECs, which preserve native nuclear architecture. IR-RNAs showed the same peripheral positioning relative to the SC35-positive core (median distances from -0.0420 to $0.243\ \mu\text{m}$; median across RNAs, $0.0885\ \mu\text{m}$; Fig. 4d and Extended Data Fig. 6a), and were significantly shifted towards the core surface compared with random spots (one-sided KS test; all $P_{\text{adj}} < 4.3 \times 10^{-9}$; Extended Data Fig. 6b), with median shifts from -0.370 to $-0.085\ \mu\text{m}$ and robust enrichment at $\leq 0.1\ \mu\text{m}$ from the core surface (median across RNAs, 9.30-fold; range, 4.53–17.3-fold; Supplementary Table 1).

Relative to the U2 shell, IR-RNAs predominantly localized close to the shell–nucleoplasm interface in both hiPS cells (median distances to the outer surface of the U2 shell ranging from -0.117 to $0.124\ \mu\text{m}$, median across RNAs, $-0.002\ \mu\text{m}$) and HUVECs (median distances to the outer surface of the U2 shell ranging from -0.0980 to $0.0640\ \mu\text{m}$, median across RNAs, $-0.0185\ \mu\text{m}$; Fig. 4e,f and Extended Data Fig. 7a). IR-RNAs were significantly shifted towards the U2 shell compared with random spots in hiPS cells (one-sided KS test; $P_{\text{adj}} < 2.2 \times 10^{-16}$), with median shifts ranging from -0.394 to $-0.154\ \mu\text{m}$ and a median 4.05-fold enrichment at $\leq 0.1\ \mu\text{m}$ from the shell surface across RNAs (range, 2.39–5.07-fold; Extended Data Fig. 6b and Supplementary Table 1). HUVECs showed the same pattern, with a significant shift towards the U2 shell compared with random spots and an enrichment at $\leq 0.1\ \mu\text{m}$ from the shell surface (one-sided KS test; $P_{\text{adj}} < 1.6 \times 10^{-11}$; Extended Data Fig. 6b and Supplementary Table 1). Although IR-RNAs were closer to nuclear speckles than expected under a random null model, the effect size was heterogeneous, consistent with a graded spatial relationship. IR-RNAs with the largest proximity shifts clustered tightly at the nuclear speckle periphery, whereas more distal transcripts showed correspondingly smaller shifts relative to random. Importantly, nuclear speckle proximity was strongly correlated between hiPS cells and HUVECs (Pearson's $r = 0.90$, $P = 2.82 \times 10^{-5}$; Fig. 4g), supporting conserved spatial organization across cell types.

To determine the spatial organization beyond the diffraction limit, we employed stimulated emission depletion (STED) super-resolution microscopy on TERT and RAD52 RIs, which showed with greater precision that they localize immediately adjacent to and largely outside the nuclear speckle core, reinforcing juxta-speckle localization (Fig. 4h).

IR-RNAs frequently localized around multiple nuclear speckles, consistent with diffusion-based distribution. Transcriptional inhibition, which induces poly(A)⁺ RNA storage within nuclear speckles⁴² and rounded nuclear speckle morphology¹⁰⁸ (Extended Data Fig. 6c), did not disrupt the peripheral association of most IR-RNAs, indicating that their peripheral localization is stable and independent of transcription or intact nuclear speckle morphology, further distinguishing poly(A)⁺ IR-RNAs from global poly(A)⁺ RNAs.

Collectively, IR-RNAs localize at the nuclear speckle periphery, largely outside the SC35-positive core at the U2 shell–nucleoplasm interface, and frequently along the boundaries of multiple nuclear speckles, in a pattern that is conserved across cell types, stable upon transcriptional inhibition, and distinct from the interior accumulation of global poly(A)⁺ RNA, indicating that IR-RNAs have distinct retention mechanisms and a different role of nuclear speckles in their regulation.

LL-IR-RNAs and their transcription sites are proximal to nuclear speckles

To relate nuclear speckle enrichment measured by ARTR-seq⁷⁴ to physical proximity, we correlated ARTR-seq $\log_2\text{FC}$ of IR-RNAs with their median distances to nuclear speckles measured by smRNA FISH. SON enrichment was strongly inversely correlated with measured distances. Exon-derived SON $\log_2\text{FC}$ correlated with smRNA FISH proximity measurements (Pearson's $r = -0.75$, $P_{\text{adj}} = 2.15 \times 10^{-4}$; Extended Data Fig. 7b). Intron-derived and total RNA (exon and intron) SON $\log_2\text{FC}$ showed stronger correlations (total RNA, Pearson's $r = -0.89$, $P_{\text{adj}} = 9.97 \times 10^{-7}$; intron, Pearson's $r = -0.86$, $P_{\text{adj}} = 3.98 \times 10^{-6}$), consistent with these measurements more directly reflecting IR-RNA positions. These results show that $\log_2\text{FC}$ SON enrichment can provide a transcriptome-wide readout of RNA positions relative to nuclear speckles. Accordingly, ARTR-seq showed that stable RIs are enriched in SON-proximal fractions and are among the IR classes with higher SON enrichment (Fig. 3g,h). Thus, LL-IR-RNAs are associated with a nuclear speckle-proximal environment. Importantly, this enrichment does not imply accumulation within the nuclear speckle core; IR-RNAs measured by smRNA FISH localized preferentially to the periphery rather than within the core.

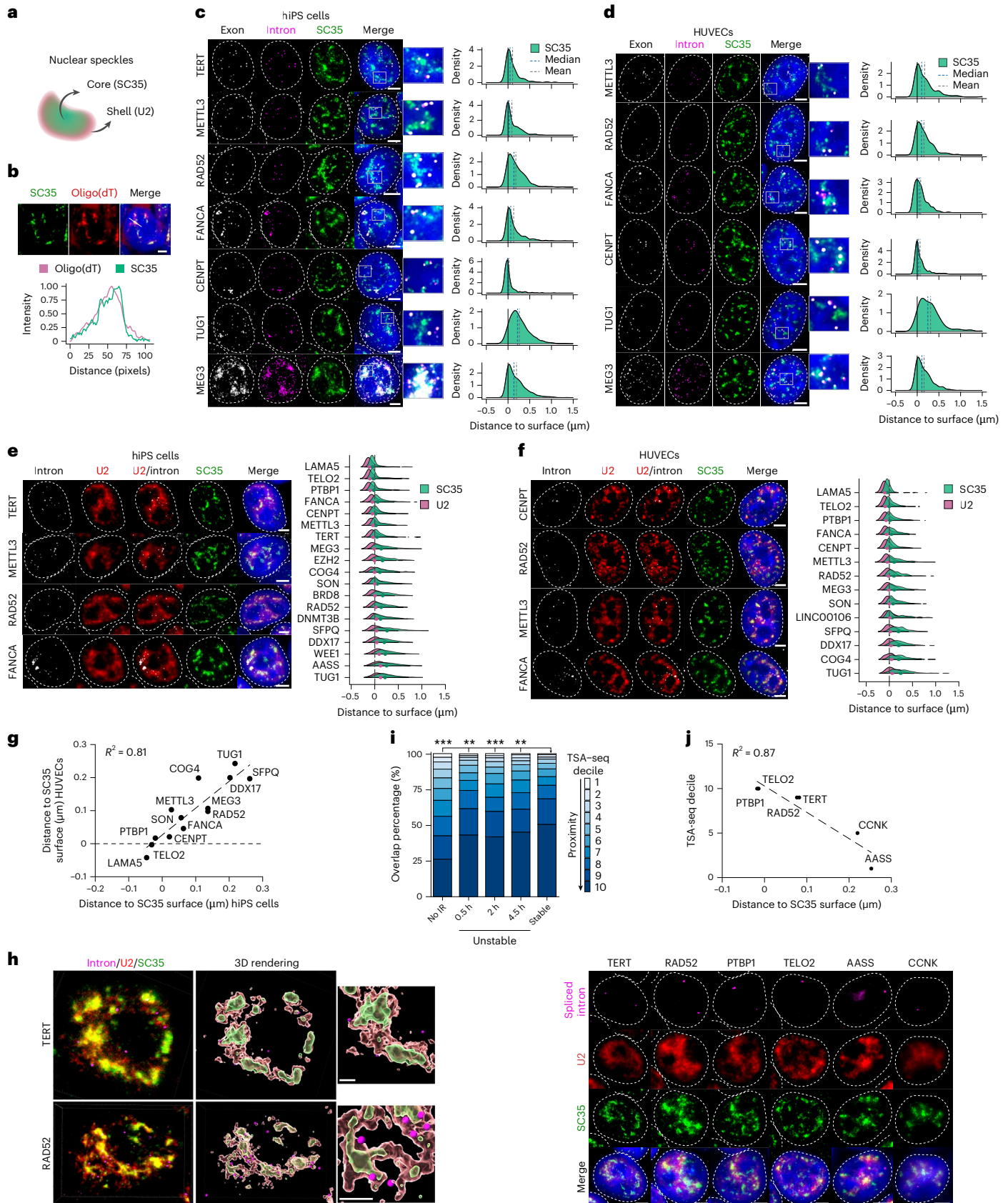
To determine whether gene loci with stable or unstable IR differ in their proximity to nuclear speckles, we analysed SON tyramide signal amplification sequencing (TSA-seq)¹⁰⁹, which estimates distances from gene loci to nuclear speckles and the resulting spatial patterns have been shown to be conserved across cell types¹¹⁰. We assigned genes stratified by IR and IR stability to ten TSA-seq proximity deciles, with

Fig. 4 | IR-RNAs localize at the nuclear speckle periphery. **a**, Schematic illustrating nuclear speckle architecture. **b**, Top: oligo(dT) smRNA FISH (red) and SC35 IF (green) maximum intensity projections. Nucleus, blue; scale bar, 2 μm . Bottom: co-localization plot of oligo(dT) and SC35 fluorescence intensity along the line indicated in the maximum intensity projection. Four independent stainings. **c,d**, Left: maximum intensity projections of exon (white) and intron (magenta) smRNA FISH and SC35 IF (green) in hiPS cells (**c**) and HUVECs (**d**). Nucleus, blue; scale bar, 2 μm . Right: density plots of distances (μm) to the surface of the SC35-positive nuclear speckle core for individual IR-RNA molecules. Positive values indicate molecules outside the segmented surface, negative values indicate the inward distance of RNA molecules located within the segmented surface. Median, dashed blue line; mean, dashed grey line. Three independent stainings. **e,f**, Left: maximum intensity projections of intron (white) and U2 (red) smRNA FISH combined with SC35 IF (green) in hiPS cells (**e**) and HUVECs (**f**). Nucleus, blue; scale bar, 2 μm . Right: density plots of IR-RNA distances (μm) to the nuclear speckle surface. Distances were computed to the segmented SC35 or U2 surface; positive values indicate molecules outside the segmented surface,

negative values indicate the distances of RNA molecules inside the segmented surface relative to the boundary. Dots, median. Three independent stainings. Additional representative images are provided in Extended Data Figs. 6a and 7a. **g**, Correlation between median distances (μm) of IR-RNAs to nuclear speckles in hiPS cells and HUVECs. $n = 13$ IR-RNAs. Pearson's $r = 0.90$, $R^2 = 0.81$, $P = 2.82 \times 10^{-5}$. **h**, Maximum intensity projections (left) and 3D rendering (right) of smRNA FISH and SC35 IF using STED microscopy in hiPS cells. Intron, magenta; U2, red; SC35, green; scale bar, 1 μm . **i**, Stacked bar plot showing distribution of DNA loci of transcripts grouped by IR status and IR stability with respect to their proximity to nuclear speckles assessed by TSA-seq¹⁰⁹. $^{***}P \leq 0.01$, $^{***}P \leq 0.001$, Fisher's exact test. P_{adj} versus stable IR-RNAs: unstable 0.5 h, 7.63×10^{-3} ; unstable 2 h, 4.63×10^{-5} ; unstable 4.5 h, 7.58×10^{-3} ; no IR, 3.28×10^{-91} . **j**, Top: correlation between TSA-seq score and the median distances of transcription sites to the nuclear speckle surface obtained by smRNA FISH ($n = 6$ genes). Pearson's $r = -0.94$, $R^2 = 0.87$, $P = 6.2 \times 10^{-3}$. Bottom: maximum intensity projections of the efficiently spliced intron (magenta) and U2 (red) smRNA FISH combined with SC35 IF (green). Nucleus, blue; scale bar, 2 μm . Three independent stainings.

groups 9–10 being the most nuclear speckle-proximal, which revealed that IR-containing genes were more proximal than genes without IR (–69% versus –43% of stable IR and no IR genes, respectively, groups 9–10; $P_{\text{adj}} < 2.2 \times 10^{-16}$, Fisher's exact test), and that genes with stable IR were more proximal than those with unstable IR (–69% versus –60–62%

overlap with groups 9–10 of stable and unstable IR; all $P_{\text{adj}} < 7.7 \times 10^{-3}$, Fisher's exact test stable versus unstable; Fig. 4i). As a validation, we performed smRNA FISH targeting exons and efficiently spliced introns of six representative genes to mark transcription sites, which were identified as one to two punctate nuclear foci with overlapping exon



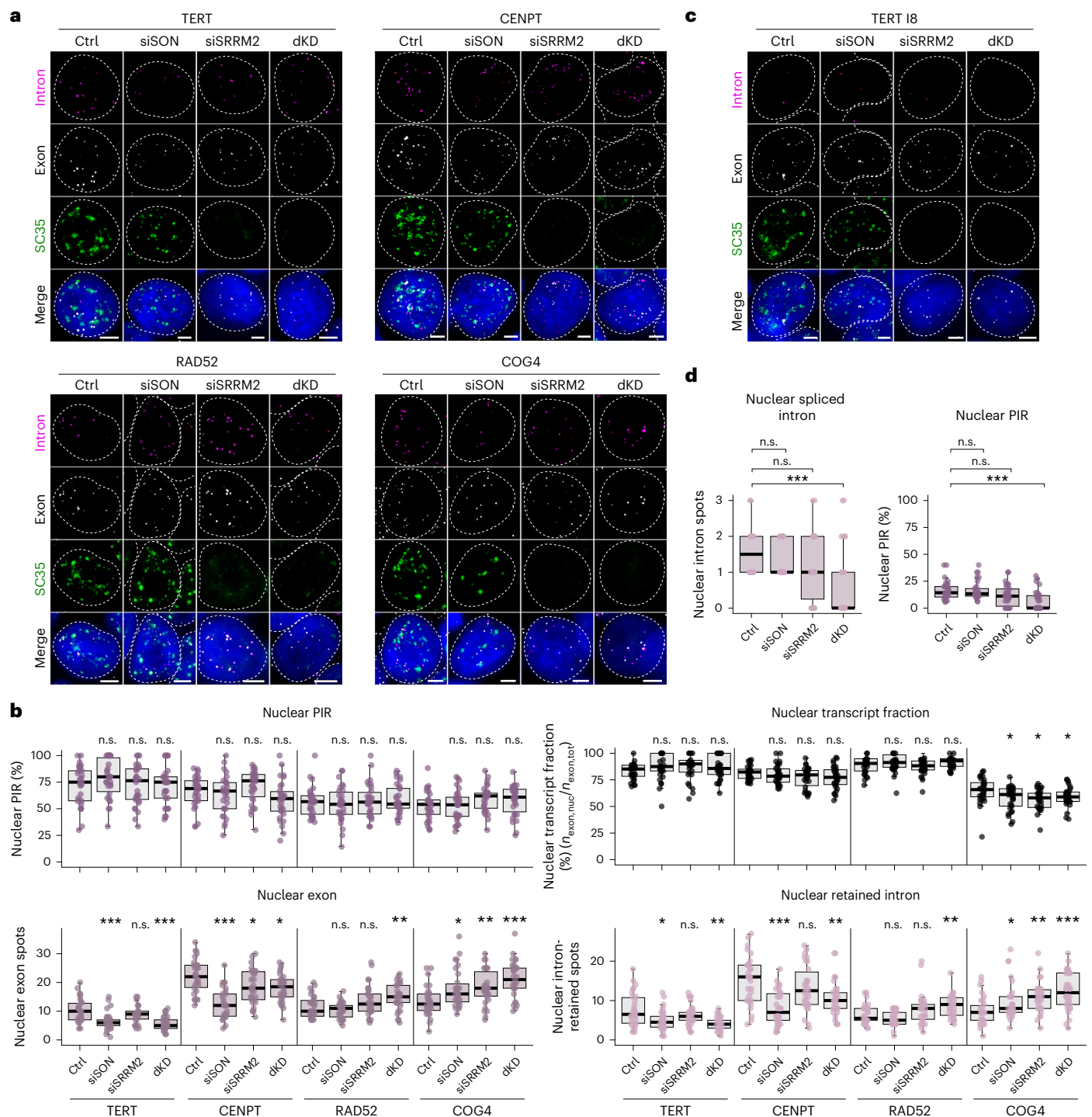


Fig. 5 | Intact nuclear speckle structure is not required for maintaining IR and nuclear retention of IR-RNAs. a, Maximum intensity projections of exon (white) and intron (magenta) smRNA FISH and SC35 IF (green) following knockdown of SON, SRM2 or both (double knockdown, dKD). Nucleus, blue; scale bar, 2 μm . **b**, Quantification of nuclear PIR, nuclear exon spots, nuclear transcript fraction (%) ($n_{\text{exon, nuc}}/n_{\text{exon, tot}}$), and nuclear intron-retained RNA spots for the indicated gene and condition. Each dot represents one cell. Midline, median; box limits, first and third quartiles; whiskers, most extreme values within 1.5 $\times$ IQR of the hinges; points, individual observations. $n = 30$ cells, three independent stainings. n.s., not significant, $*P < 0.05$, $**P \leq 0.01$, $***P \leq 0.001$, as evaluated by two-sided Wilcoxon rank-sum test versus control siRNA (Ctrl). Nuclear transcript fraction, COG4, $P_{\text{adj}} = 4.63 \times 10^{-2}$, 1.46×10^{-2} and 4.63×10^{-2} , for siSON, siSRM2 and dKD, respectively. Nuclear exon spots, TERT, $P_{\text{adj}} = 4.06 \times 10^{-4}$ and 4.65×10^{-5} , for siSON and dKD, respectively; CENPT, $P_{\text{adj}} = 1.70 \times 10^{-6}$, 2.82×10^{-2} and 1.49×10^{-2} , for siSON, siSRM2 and dKD, respectively; RAD52, $P_{\text{adj}} = 9.63 \times 10^{-3}$, for dKD;

COG4, $P_{\text{adj}} = 1.11 \times 10^{-2}$, 4.54×10^{-3} and 1.15×10^{-4} , for siSON, siSRM2 and dKD, respectively. Nuclear retained-intron spots: TERT, $P_{\text{adj}} = 2.38 \times 10^{-2}$ and 1.22×10^{-3} , for siSON and dKD, respectively; CENPT, $P_{\text{adj}} = 1.67 \times 10^{-4}$ and 4.41×10^{-3} , for siSON and dKD, respectively; RAD52, $P_{\text{adj}} = 7.32 \times 10^{-3}$, dKD; COG4, $P_{\text{adj}} = 3.88 \times 10^{-2}$, 1.34×10^{-3} and 3.18×10^{-4} , for siSON, siSRM2 and dKD, respectively. **c**, Maximum intensity projections of TERT exon (white) and efficiently spliced intron 8 (magenta) smRNA FISH and SC35 IF (green) following knockdown of SON and/or SRM2. Nucleus, blue; scale bar, 2 μm . **d**, Quantification of nuclear PIR and nuclear intron spots of TERT efficiently spliced intron 8, marking the active transcription site. Each dot represents one cell. Midline, median; box limits, first and third quartiles; whiskers, most extreme values within 1.5 $\times$ IQR of the hinges; points, individual observations. $n = 30$ cells, two independent stainings. n.s., not significant, $*P < 0.05$, $**P \leq 0.01$, $***P \leq 0.001$, as evaluated by two-sided Wilcoxon rank-sum test versus control siRNA (Ctrl). Nuclear spliced intron, $P_{\text{adj}} = 7.95 \times 10^{-4}$ for dKD; nuclear PIR, $P_{\text{adj}} = 4.34 \times 10^{-4}$ for dKD.

and spliced intron signals (Methods; Extended Data Fig. 8a). Measured distances of transcription sites to nuclear speckles correlated strongly with TSA-seq-based proximity estimates (Pearson's $r = -0.94$, $P = 6.2 \times 10^{-3}$), supporting the robustness of the analysis (Fig. 4j).

We then compared the positions of transcription sites with those of the corresponding IR-RNAs. Median transcription-site distances correlated strongly with median IR-RNA distances to the surfaces of the SC35-positive nuclear speckle core (Pearson's $r = 0.92$, $P = 3.0 \times 10^{-3}$) and the U2 shell (Pearson's $r = 0.91$, $P = 4.7 \times 10^{-3}$; Extended Data Fig. 8b). Thus, genes more proximal to nuclear speckles tended to produce IR-RNAs that also remained relatively nuclear speckle-proximal after release from transcription sites. IR-RNAs generally showed broader positioning than transcription sites, consistent with redistribution after release, but we found no universal pattern. DNMT3B, PTBP1 and RAD52 IR-RNAs were slightly more distal to nuclear speckles than their transcription sites, whereas for TERT and AASS the two distributions were comparable. Overall, these data indicate that between-gene differences in nuclear speckle proximity are largely preserved after RNA release, supporting gene-level spatial determinants as the dominant source of positional variation. Consistent with this, gene proximity to nuclear speckles is associated with features such as GC content and intron length¹¹¹.

Together, IR-RNAs occupy a niche at the nuclear speckle-nucleoplasm interface, linking IR stability and gene-level nuclear speckle proximity.

IR and nuclear retention of IR-RNAs are maintained independently of nuclear speckle integrity

We next asked whether nuclear speckle structure is required for maintaining IR or nuclear retention of IR-RNAs. To this end, we depleted the core nuclear speckle scaffold proteins SON and SRRM2, known to disrupt nuclear speckle structure¹¹². We knocked down SON and SRRM2 individually and in combination, validated the disruption of nuclear speckles by SC35, RBM25 and SON IF (Fig. 5a; Extended Data Fig. 8c) and performed smRNA FISH on four IR-RNAs. Expression levels of individual genes responded variably; TERT nuclear and cytoplasmic RNA decreased, CENPT nuclear RNA decreased with minimal cytoplasmic changes, RAD52 was largely unchanged, and COG4 increased in both compartments across all knockdowns (Fig. 5a,b and Extended Data Fig. 8d).

Critically, despite changes in expression, PIR of examined IR-RNAs remained high and their nuclear transcript fraction was unchanged or changed only minimally across conditions (Fig. 5a,b). To rule out global export defects, we examined cytoplasmic GAPDH, whose distribution was unaffected (Extended Data Fig. 8e). To determine if these effects could be attributed to general splicing defects, we examined the splicing efficiency of *TERT* intron 8, an efficiently spliced intron. Although SRRM2 and SON/SRRM2 knockdown reduced the number of active *TERT* transcription sites, consistent with the observed reduction of *TERT* RNA molecules, splicing efficiency remained unchanged, indicating that disruption of SON, SRRM2 or nuclear speckle structure does not cause accumulation of unspliced transcripts (Fig. 5c,d).

Collectively, SON, SRRM2 and nuclear speckle integrity are not required to maintain IR, efficient splicing of constitutive introns, or nuclear retention of IR-RNAs examined here; instead, they contribute to the transcriptional output of examined nuclear speckle-associated genes.

Mitosis drives coordinated LL-IR-RNA resolution

Building on our previous findings that *TERT* RI is spliced in mitosis²³, we asked whether mitosis broadly functions as a temporal window for resolving nuclear IR-RNAs from genes carrying LL-RIs. We treated hiPS cells with nocodazole and performed smRNA FISH on six genes harbouring LL-RIs (*BRD8*, *METTL3*, *TERT*, *RAD52*, *FANCA* and *CENPT*), defined as having at least one LL-RI and selected for high basal PIR and nuclear

RNA enrichment (Extended Data Fig. 9a,b). Most nocodazole-treated hiPS cells were in mitosis, and approximately 20% were in interphase, as shown previously¹¹³.

In mitosis, all six LL-IR-RNAs underwent extensive splicing and were released into the cytosol. IR-RNA numbers decreased sharply, while spliced RNAs (Δ RI) increased reciprocally (Extended Data Fig. 9a,b). The number of newly spliced transcripts closely matched that of IR-RNAs in interphase, suggesting that the spliced pool arises from the interphase IR-RNAs. Consequently, total PIR dropped sharply, while free intron signal, which was nearly absent in interphase, strongly increased, indicating active intron excision and transient stabilization of excised introns (Extended Data Fig. 9b).

To monitor LL-IR-RNAs across the cell cycle and immediately after mitosis, we synchronized hiPS cells using a double thymidine block and collected cells at defined intervals after release spanning early-to-late S and G2 phase, mitosis and early G1 for smRNA FISH with single-cell-level staging using PCNA immunostaining, cell size and absent or early-onset transcription in early G1, as defined by exon-intron smRNA FISH (Methods; Fig. 6a and Extended Data Fig. 9c).

All six LL-IR-RNAs remained enriched in the nucleus throughout early and late S phase and persisted in G2, with stable PIR and IR-RNA numbers, indicating sustained IR and nuclear retention before mitosis (Fig. 6b,c and Extended Data Fig. 9d). By contrast, mitosis was accompanied by a rapid splicing transition. Total PIR dropped sharply across all candidates (Fig. 6b,c), paralleled by a marked reduction in IR-RNA count (Extended Data Fig. 9d). This shift was accompanied by a reciprocal increase in fully spliced transcripts (Δ RI) and an accumulation of free intron signal. Importantly, total RNA abundance per cell did not change significantly between G2 and mitosis, supporting mitosis-triggered processing rather than degradation (Extended Data Fig. 9d).

Upon mitotic exit, early G1 cells were dominated by cytoplasmic spliced transcripts and showed a marked increase in cytoplasmic RNA fraction, despite a lack of detectable transcription or its early-onset (Fig. 6b,c). The cytoplasmic RNA pool in early G1 corresponded to approximately half of the RNA in mitosis, consistent with partitioning into daughter cells, and exceeded cytoplasmic RNA levels measured in G2 (Fig. 6b). A fraction of free intron signal persisted in early G1 cytoplasm (Extended Data Fig. 9d), consistent with reports of cytoplasmic introns in vertebrates¹¹⁴.

Together, these data demonstrate that LL-IR-RNAs undergo coordinated splicing during mitosis, converting nuclear IR-RNAs into spliced mRNAs without total RNA loss. The resulting spliced transcripts are inherited by daughter cells in early G1 as dominant RNA species prior to transcriptional reactivation. We propose that this mechanism provides early G1 daughter cells with mature transcripts encoding critical regulators, without requiring new transcription, splicing and export, which collectively can take tens of minutes to more than an hour^{115–118}. These findings place LL-IR-RNAs operationally within the intron-detention class as a long-lived subtype and identify mitosis as the physiological cue for their splicing.

Kinase-dependent regulation of mitotic IR resolution

Entry into mitosis is accompanied by a broad phospho-remodelling programme driven by a kinase-dominant state established by cyclin-dependent kinase 1 (CDK1) activation and reduced protein phosphatase 1 and 2 (PP1/PP2) activity^{119–121}. CLKs, which localize to nuclear speckles and peak at G2/M phase, phosphorylate SR proteins and other RS-domain-containing splicing factors, weakening their multivalent interactions, promoting release into the nucleoplasm and nuclear speckle dissolution in mitosis^{44–46,122}, while DYRK3 drives a broader phosphorylation programme promoting dissolution of multiple condensates, including nuclear speckles⁴⁷.

To determine whether mitotic LL-IR-RNA splicing depends on kinase-driven remodelling, we inhibited CLKs and DYRK3 (CLKi/DYRK3i) to block broad phosphorylation-dependent remodelling of

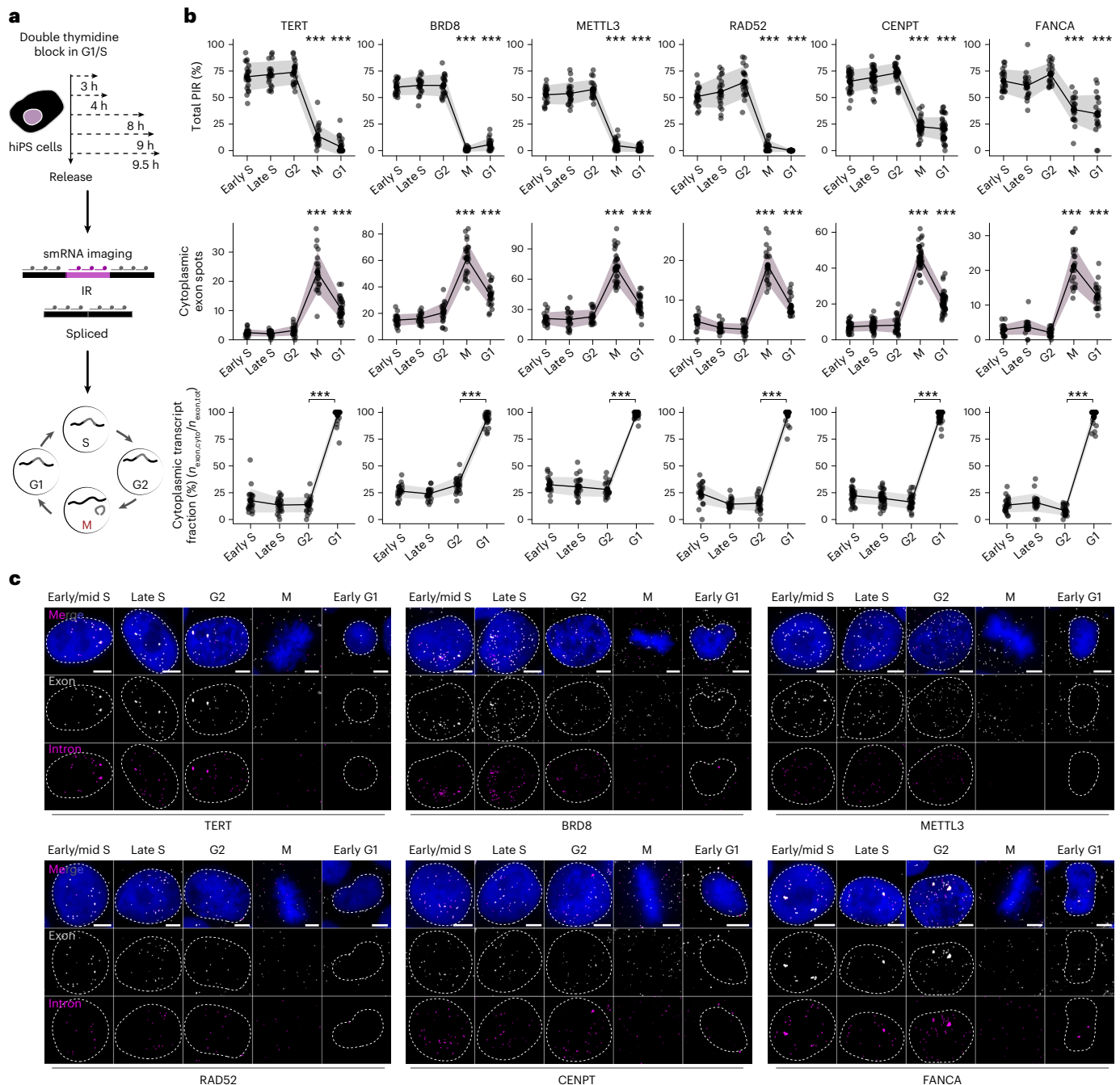


Fig. 6 | Mitosis drives splicing of stable IR-RNAs and cytoplasmic accumulation of the resulting mRNAs. **a**, Schematic of the experimental design. hiPS cells were synchronized by double thymidine block at the G1/S boundary, released and analysed by smRNA FISH at the indicated timepoints corresponding to early/mid S, late S, G2, M and early G1 phases. **b**, Quantification of total PIR (top), cytoplasmic exon spots (middle) and cytoplasmic transcript fraction (%) ($n_{\text{exon, cyto}}/n_{\text{exon, total}}$) (bottom) for the indicated RNAs across synchronized cell-cycle stages. Each dot represents one cell. Black line, mean value; shadow, standard deviation; n values are given for early/mid S, late S, G2, mitosis and early G1, respectively: TERT 21, 20, 21, 21, 29; BRD8 20, 20, 20, 23, 21; METTL3 20, 23, 20, 32, 27; RAD52 20, 20, 20, 20, 21; CENPT 30, 38, 29, 36, 44; FANCA 22, 20, 20, 21, 22; three independent experiments. Cytoplasmic transcript fraction is not shown for mitotic cells. *** $P \leq 0.001$, as evaluated by two-sided Wilcoxon rank-sum test

versus G2. Total PIR, M and G1, respectively: TERT, $P_{\text{adj}} = 4.66 \times 10^{-8}$, 1.59×10^{-9} ; BRD8, $P_{\text{adj}} = 3.96 \times 10^{-8}$, 5.49×10^{-8} ; METTL3, $P_{\text{adj}} = 5.48 \times 10^{-9}$, 6.89×10^{-9} ; RAD52, $P_{\text{adj}} = 5.73 \times 10^{-8}$, 6.89×10^{-9} ; CENPT, $P_{\text{adj}} = 3.51 \times 10^{-11}$, 3.91×10^{-12} ; FANCA, $P_{\text{adj}} = 5.73 \times 10^{-8}$, 1.94×10^{-7} . Cytoplasmic exon spots, M and G1, respectively: TERT, $P_{\text{adj}} = 4.36 \times 10^{-8}$, 1.61×10^{-8} ; BRD8, $P_{\text{adj}} = 4.36 \times 10^{-8}$, 3.56×10^{-5} ; METTL3, $P_{\text{adj}} = 5.48 \times 10^{-9}$, 6.13×10^{-5} ; RAD52, $P_{\text{adj}} = 6.11 \times 10^{-8}$, 5.59×10^{-8} ; CENPT, $P_{\text{adj}} = 3.37 \times 10^{-11}$, 9.18×10^{-11} ; FANCA, $P_{\text{adj}} = 4.87 \times 10^{-8}$, 5.51×10^{-8} . Cytoplasmic transcript fraction, G1 versus G2: TERT, $P_{\text{adj}} = 7.09 \times 10^{-10}$; BRD8, $P_{\text{adj}} = 4.58 \times 10^{-8}$; METTL3, $P_{\text{adj}} = 9.21 \times 10^{-9}$; RAD52, $P_{\text{adj}} = 1.71 \times 10^{-8}$; CENPT, $P_{\text{adj}} = 1.14 \times 10^{-12}$; FANCA, $P_{\text{adj}} = 1.71 \times 10^{-8}$. **c**, Representative maximum intensity projections of exon (white) and intron (magenta) smRNA FISH in synchronized hiPS cells at the indicated cell-cycle stages. Hoechst, blue; scale bar, 2 μm .

splicing and condensate environment in mitosis, and CDK1 (CDK1i) to block upstream mitotic phospho-remodelling. As previously reported, CLKi/DYRK3i promoted poly(A)⁺ RNA accumulation in nuclear speckles in interphase, and disrupted nuclear speckle disassembly in mitosis⁴¹

(Fig. 7a and Extended Data Fig. 10a). CDK1i also prevented nuclear speckle dissolution in mitosis, indicating that CDK1 contributes to this process. Undissolved mitotic nuclear speckles maintained poly(A)⁺ RNA retention capacity (Fig. 7a).

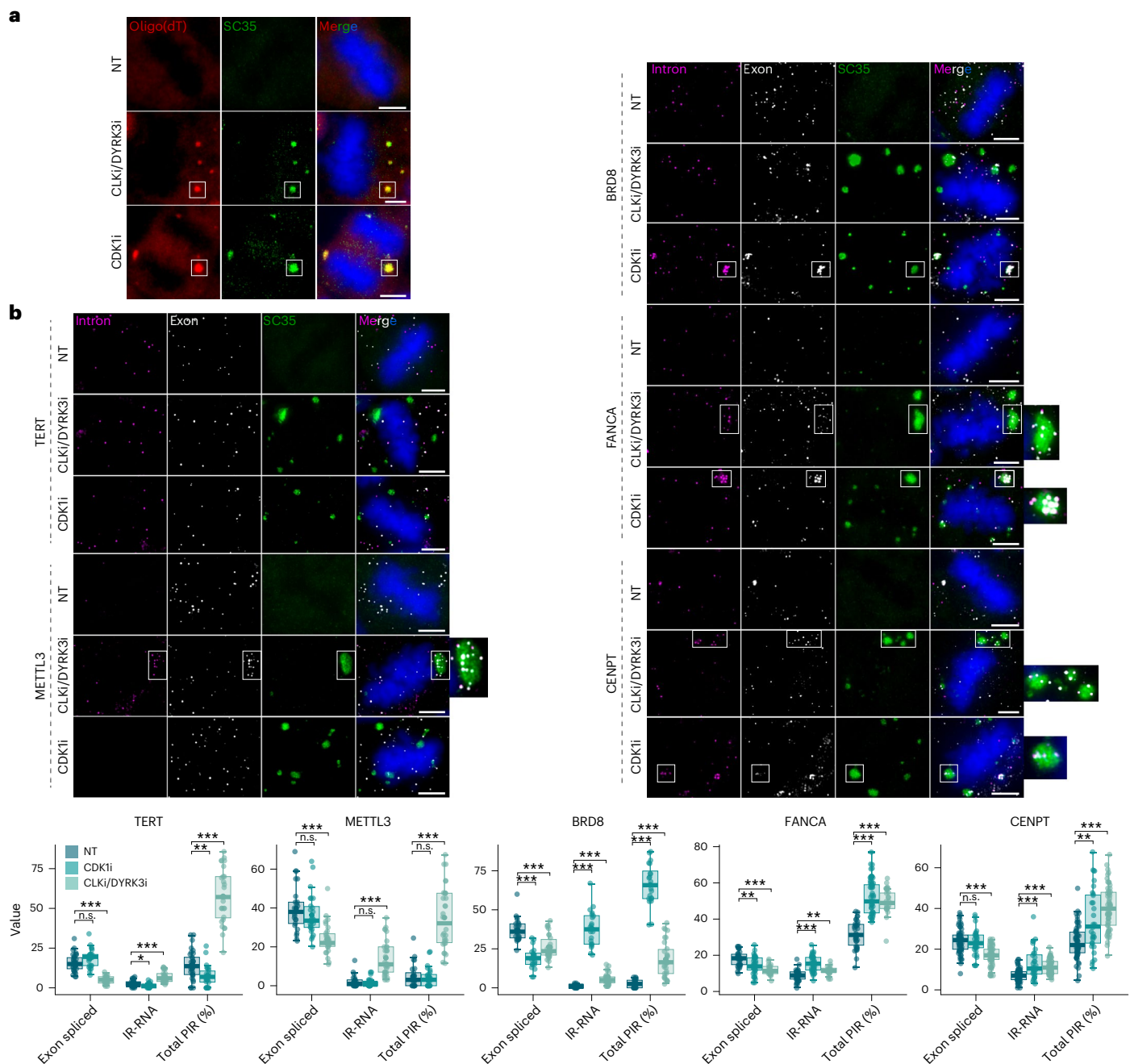


Fig. 7 | Splicing of RIs in mitosis is kinase dependent. a, Maximum intensity projections of oligo(dT) smRNA FISH (red) and SC35 IF (green). Hoechst, blue; scale bar, 2 μ m. **b**, Top: maximum intensity projections of exon (white) and intron (magenta) smRNA FISH and SC35 IF (green) in non-treated (NT) hiPS cells or after CLKi/DYRK3i or CDK1i. Hoechst, blue; scale bar, 2 μ m. Bottom: smRNA FISH quantification of spliced exon spots (Δ RI), retained-intron RNA spots and total PIR in mitosis. Midline, median; box limits, first and third quartiles; whiskers, most extreme values within 1.5 $\times$ IQR of the hinges; points, individual cells. n.s., not significant, $^*P < 0.05$, $^{**}P \leq 0.01$, $^{***}P \leq 0.001$ as evaluated by two-sided Wilcoxon rank-sum test versus non-treated cells (NT). Spliced exon spots: CDK1i, $P_{\text{adj}} = 2.86 \times 10^{-7}$ and 1.07×10^{-3} , for BRD8 and FANCA, respectively;

CLKi/DYRK3i, $P_{\text{adj}} = 3.41 \times 10^{-9}$, 1.34×10^{-6} , 4.16×10^{-5} , 1.91×10^{-6} and 1.23×10^{-8} , for TERT, METTL3, BRD8, FANCA and CENPT, respectively. Retained-intron RNA spots: CDK1i, $P_{\text{adj}} = 1.05 \times 10^{-2}$, 6.53×10^{-8} , 4.84×10^{-7} and 6.31×10^{-4} , for TERT, BRD8, FANCA and CENPT, respectively; CLKi/DYRK3i, $P_{\text{adj}} = 1.84 \times 10^{-6}$, 7.69×10^{-8} , 2.86×10^{-7} , 2.13×10^{-3} and 3.59×10^{-7} , for TERT, METTL3, BRD8, FANCA and CENPT, respectively. Total PIR: CDK1i, $P_{\text{adj}} = 3.61 \times 10^{-3}$, 7.01×10^{-8} , 1.07×10^{-8} and 1.94×10^{-3} for TERT, BRD8, FANCA and CENPT, respectively; CLKi/DYRK3i, $P_{\text{adj}} = 2.88 \times 10^{-9}$, 1.27×10^{-8} , 7.69×10^{-8} , 1.55×10^{-6} and 6.27×10^{-11} for TERT, METTL3, BRD8, FANCA and CENPT, respectively. *n*, mitotic cells from four independent experiments, for NT, CDK1i and CLKi-DYRK3i, respectively: CENPT 55, 32, 55; METTL3 25, 26, 30; TERT 32, 25, 27; BRD8 25, 20, 30; FANCA 34, 38, 20.

CLKi/DYRK3i increased mitotic PIR across all candidate LL-IR-RNAs, with the strongest effects observed for TERT and METTL3, followed by FANCA, CENPT and BRD8 (+43.1%, +30.1%, +17.3%, +17% and +15.6%, respectively; all $P_{\text{adj}} < 1.6 \times 10^{-6}$, two-sided Wilcoxon rank-sum test), reflecting reduced spliced RNA and increased IR-RNA isoforms (Fig. 7b), indicating splicing inhibition instead of decay.

CDK1i affected mitotic PIR in a transcript-specific manner, with the strongest increase in BRD8, followed by FANCA and CENPT (+63.3%, +20.3% and +10.8%, respectively; all $P_{\text{adj}} < 2.0 \times 10^{-3}$, two-sided Wilcoxon rank-sum test; Fig. 7b). METTL3 was not significantly changed, whereas TERT showed a slight reduction (-7% ; $P_{\text{adj}} = 3.6 \times 10^{-3}$, two-sided Wilcoxon rank-sum test; Fig. 7b). Consistent with these changes, IR-RNA

increased for BRD8, FANCA and CENPT, decreased for TERT, whereas spliced RNA was reduced for BRD8 and FANCA. Unspliced IR-RNAs tended to co-localize with undissolved mitotic nuclear speckles, except for TERT, whereas exon-only or excised intron signals did not, indicating that exon–intron architecture promotes association with nuclear speckles. Thus, while both CDK1 and CLK1/DYRK3 prevented nuclear speckle dissolution, their impact on RI splicing is transcript specific.

To compare this behaviour with a canonical CLK-responsive detained intron, we examined SRSF7, a previously described CLK-responsive target in interphase¹⁷ that falls within the short-lived class in our dataset. As expected, CLK1 strongly reduced SRSF7 detained-intron RNA, accompanied by brighter transcription site foci (Extended Data Fig. 10b). In interphase, CLK1/DYRK3 strongly increased METTL3 PIR and only modestly increased BRD8 PIR (+29.4% and +7.8%; $P_{\text{adj}} = 2.9 \times 10^{-9}$ and $P_{\text{adj}} = 3.4 \times 10^{-2}$, respectively, two-sided Wilcoxon rank-sum test; Extended Data Fig. 10c), with the remaining candidates unaffected, suggesting that RS-domain phosphorylation tunes METTL3 IR across the cell cycle. CDK1 increased BRD8 PIR and modestly decreased METTL3 PIR in interphase (+30.8% and -7.3%; $P_{\text{adj}} = 3.4 \times 10^{-9}$ and $P_{\text{adj}} = 1.3 \times 10^{-2}$, respectively, two-sided Wilcoxon rank-sum test; Extended Data Fig. 10c), with other candidates unchanged. Thus, unlike the broad mitotic response, LL-IR-RNAs showed only selective sensitivity to kinase perturbation in interphase, supporting mitosis as the major window for coordinated kinase-dependent IR resolution.

Together, these data indicate that kinase activity is required for full LL-IR-RNA resolution in mitosis: CLK/DYRK3 inhibition broadly impairs splicing across candidates, while CDK1 contributes in a transcript-specific manner.

Discussion

RI in polyadenylated nuclear transcripts can be resolved through post-transcriptional splicing as a regulated layer of gene expression control^{8,17,18,21}. Here, integrating transcription inhibition with deep-learning and high-resolution imaging, we identify a long-lived population of RIs in hiPS cells that persists for hours. Annotated detained introns distribute across multiple kinetic groups, indicating that LL-IR-RNAs represent a distinct subset within the broader detained-intron framework.

We identify a spatially defined niche for IR-RNAs at the nuclear speckle periphery, largely excluded from the SC35-positive core and concentrated at the U2 shell-nucleoplasm interface, resolving ambiguities left by proximity-labelling studies and distinguishing IR-RNAs from bulk poly(A)⁺ RNAs enriched within nuclear speckles^{38,41,42}. LL-IR-RNAs and their transcription sites occupy more nuclear speckle-proximal positions than SL-IR-RNAs, linking IR stability to subnuclear architecture. IR-RNA localization described here is distinct from paraspeckles, which are typically absent in hiPS cells, lack the SON scaffold and would not be captured with SON-based proximity approaches^{123–125}.

This spatial organization is mechanistically informative. Proximity to nuclear speckles has been associated with efficient splicing⁶⁴, and transcription sites localize adjacent to nuclear speckles^{61–64,126}. Recently, nuclear speckles were linked to RNA processing of GC-rich isochores¹¹¹. Yet, LL-IR-RNAs remain unspliced even while positioned at the nuclear speckle periphery, indicating the presence of *cis*-encoded inhibitory features that delay splicing progression despite a splicing-permissive environment. Consistent with this, RIs and LL-RIs are enriched for weak 5' splice sites, elevated GC content, stronger predicted secondary structures, and sequence motifs capable of forming non-canonical conformations. In particular, G-quadruplex-forming sequences are enriched near splice-regulatory elements (donor site, branch point), where, together with elevated GC content and short intron length, they may impede spliceosome progression until specific cues trigger resolution. Motif analysis identifies candidate *trans*-acting regulators, including HNRNPF and SRSF1, known to influence IR^{127,128}, and EWSR1

and RBM4-family proteins^{129,130}, the latter localizing to nuclear speckles, binding G-quadruplexes, and linked to splicing inhibition, making them plausible candidates for coupling GC-rich, structure-prone *cis*-features to splicing control^{98,101–103}. These observations suggest that IR longevity reflects an interplay between RNA-intrinsic features, RBP interactions and the local nuclear environment.

Nuclear speckle perturbation experiments did not lead to LL-IR-RNA splicing or export, indicating that their periphery may function as a permissive interface where transcripts remain poised for resolution. IR-RNAs distribute around multiple nuclear speckles, consistent with diffusion during extended nuclear residence. Hundreds of IR-RNAs concentrated at the periphery may enable noncoding functions, such as protein or DNA binding^{76,131}, or collectively define a distributed nuclear speckle network through multivalent interactions.

Finally, we identify mitotic reprogramming as a regulatory, coordinated window for synchronous resolution of LL-IR-RNAs, generating spliced transcripts that populate the cytoplasm of mitotic and early G1 daughter cells, consistent with an inheritance mechanism that reduces reliance on *de novo* transcription, splicing and export, which collectively can take tens of minutes to over an hour^{115–118}. We emphasize that our data do not argue that these introns are spliced exclusively during mitosis, but rather that mitotic remodelling is a window in which they are resolved in a coordinated burst. This transition depends on kinase-driven remodelling of the splicing and condensate environment, coupling mitotic reorganization to coordinated intron resolution. CLK-dependent IR regulation has been described in interphase contexts, including promoting or suppressing detained intron splicing, CLK-dependent SR protein phosphorylation in nuclear stress bodies, CLK2 condensates linked to increased IR, and autoregulation of CLKs through IR-Clk1/4 transcripts^{17,34,75,132}. Notably, this pathway is not uniform. METTL3 showed increased interphase PIR upon CLK1/DYRK3i, suggesting flexible tuning to meet ongoing cellular demands, whereas TERT IR showed tight coupling to mitosis, consistent with telomerase restriction to proliferative contexts, as we hypothesized earlier²³. CLK and DYRK3 individual contributions remain to be determined. This underscores dynamic regulation of LL-IR-RNAs, with some transcripts requiring flexible availability and others being tightly gated to mitosis.

A long-lived IR programme was previously described in male germ cells, where IR-RNAs persist throughout differentiation and are spliced post-meiotically when *de novo* mRNA synthesis is suppressed¹³³, suggesting that large-scale nuclear reorganization events during mitosis or meiosis may create regulatory windows for coordinated intron resolution to replenish cytoplasmic mRNAs when transcription is transiently repressed. These findings extend observations that alternative splicing oscillates across the cell cycle^{60,122}, revealing a spatial interphase niche and a mitosis-linked intron resolution programme.

Functionally, this timing may ensure the rapid availability of mature transcripts as cells exit mitosis. Many LL-IR-RNAs encode regulators of DNA repair, genome maintenance and cell cycle. CENP-T is a core kinetochore component whose centromeric incorporation and outer kinetochore assembly are cell-cycle regulated^{77,134}. FANCA maintains centrosome and spindle organization, is required for replication-associated DNA damage repair and supports cytokinesis during mitosis; its depletion induces G1/S arrest^{135–138}. METTL3, the catalytic subunit of the m⁶A writer complex, regulates cell cycle progression and facilitates homologous recombination repair^{81,139–141}. BRD8, a bromodomain-containing chromatin regulator, regulates the expression of prereplicative complex subunits in G1 phase, and its depletion induces G1 arrest¹⁴². TERT, the catalytic dose-limiting component of the telomerase complex, is needed for telomere maintenance in mitotically active cells^{80,143}. Together, these examples support a model in which mitotic splicing generates translation-competent mRNAs precisely when transcription is globally repressed and pre-existing mRNA and protein pools have been halved by cell division, delivering a supply of

mRNAs as daughter cells reform to meet early G1 demands that prime S phase and replication-coupled genome maintenance.

Our study identifies a nuclear speckle–peripheral reservoir of LL-IR-RNAs, defines its sequence and spatial determinants, and links mitotic remodelling to their coordinated release. Together, these findings show how intronic sequence, subnuclear organization and cell cycle transitions converge to gate mRNA availability in time and space.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41556-026-02040-5>.

References

- Chin, A. & Lécuyer, E. RNA localization: making its way to the center stage. *Biochim. Biophys. Acta Gen. Subj.* **1861**, 2956–2970 (2017).
- Wilk, R., Hu, J., Blotsky, D. & Krause, H. M. Diverse and pervasive subcellular distributions for both coding and long noncoding RNAs. *Genes Dev.* **30**, 594–609 (2016).
- Fazal, F. M. et al. Atlas of subcellular RNA localization revealed by APEX-Seq. *Cell* **178**, 473–490 (2019).
- Lécuyer, E. et al. Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. *Cell* **131**, 174–187 (2007).
- Martin, K. C. & Ephrussi, A. mRNA localization: gene expression in the spatial dimension. *Cell* **136**, 719–730 (2009).
- Jung, H., Yoon, B. C. & Holt, C. E. Axonal mRNA localization and local protein synthesis in nervous system assembly, maintenance and repair. *Nat. Rev. Neurosci.* **13**, 308–324 (2012).
- Buxbaum, A. R., Haimovich, G. & Singer, R. H. In the right place at the right time: visualizing and understanding mRNA localization. *Nat. Rev. Mol. Cell Biol.* **16**, 95–109 (2015).
- Mauger, O., Lemoine, F. & Scheiffele, P. Targeted intron retention and excision for rapid gene regulation in response to neuronal activity. *Neuron* **92**, 1266–1278 (2016).
- Benoit Bouvrette, L. P. et al. CeFra-seq reveals broad asymmetric mRNA and noncoding RNA distribution profiles in Drosophila and human cells. *RNA* **24**, 98–113 (2018).
- Davis, B. M., McCurrach, M. E., Taneja, K. L., Singer, R. H. & Housman, D. E. Expansion of a CUG trinucleotide repeat in the 3' untranslated region of myotonic dystrophy protein kinase transcripts results in nuclear retention of transcripts. *Proc. Natl Acad. Sci. USA* **94**, 7388–7393 (1997).
- Fallini, C. et al. The survival of motor neuron (SMN) protein interacts with the mRNA-binding protein HuD and regulates localization of poly(A) mRNA in primary motor neuron axons. *J. Neurosci.* **31**, 3914–3925 (2011).
- Alami, N. H. et al. Axonal transport of TDP-43 mRNA granules is impaired by ALS-causing mutations. *Neuron* **81**, 536–543 (2014).
- Goering, R. et al. FMRP promotes RNA localization to neuronal projections through interactions between its RGG domain and G-quadruplex RNA sequences. *eLife* **9**, e52621 (2020).
- Fernandopulle, M. S., Lippincott-Schwartz, J. & Ward, M. E. RNA transport and local translation in neurodevelopmental and neurodegenerative disease. *Nat. Neurosci.* **24**, 622–632 (2021).
- Yap, K., Lim, Z. Q., Khandelia, P., Friedman, B. & Makeyev, E. V. Coordinated regulation of neuronal mRNA steady-state levels through developmentally controlled intron retention. *Genes Dev.* **26**, 1209–1223 (2012).
- Braunschweig, U. et al. Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res.* **24**, 1774–1786 (2014).
- Boutz, P. L., Bhutkar, A. & Sharp, P. A. Detained introns are a novel, widespread class of post-transcriptionally spliced introns. *Genes Dev.* **29**, 63–80 (2015).
- Braun, C. J. et al. Coordinated splicing of regulatory detained introns within oncogenic transcripts creates an exploitable vulnerability in malignant glioma. *Cancer Cell* **32**, 411–426 (2017).
- Jacob, A. G. & Smith, C. W. J. Intron retention as a component of regulated gene expression programs. *Hum. Genet.* **136**, 1043–1057 (2017).
- Luisier, R. et al. Intron retention and nuclear loss of SFPQ are molecular hallmarks of ALS. *Nat. Commun.* **9**, 2010 (2018).
- Yue, L., Wan, R., Luan, S., Zeng, W. & Cheung, T. H. Dek modulates global intron retention during muscle stem cells quiescence exit. *Dev. Cell* **53**, 661–676 (2020).
- Guo, C.-J. et al. Distinct processing of lncRNAs contributes to non-conserved functions in stem cells. *Cell* **181**, 621–636 (2020).
- Dumbović, G. et al. Nuclear compartmentalization of TERT mRNA and TUG1 lncRNA is driven by intron retention. *Nat. Commun.* **12**, 1–19 (2021).
- Mazille, M., Buczak, K., Scheiffele, P. & Mauger, O. Stimulus-specific remodeling of the neuronal transcriptome through nuclear intron-retaining transcripts. *EMBO J.* **41**, e110192 (2022).
- González-Iglesias, A. et al. Intron detention tightly regulates the stemness/differentiation switch in the adult neurogenic niche. *Nat. Commun.* **15**, 2837 (2024).
- Li, Y. et al. An intron with a constitutive transport element is retained in a Tap messenger RNA. *Nature* **443**, 234–237 (2006).
- Buckley, P. T. et al. Cytoplasmic intron sequence-retaining transcripts can be dendritically targeted via ID element retrotransposons. *Neuron* **69**, 877–884 (2011).
- Khaladkar, M. et al. Subcellular RNA sequencing reveals broad presence of cytoplasmic intron-sequence retaining transcripts in mouse and rat neurons. *PLoS ONE* **8**, e76194 (2013).
- Sterne-Weiler, T. et al. Frac-seq reveals isoform-specific recruitment to polyribosomes. *Genome Res.* **23**, 1615–1623 (2013).
- Floor, S. N. & Doudna, J. A. Tunable protein synthesis by transcript isoforms in human cells. *eLife* **5**, e10921 (2016).
- Sharangdhar, T. et al. A retained intron in the 3'-UTR of Calm3 mRNA mediates its Staufen2- and activity-dependent localization to neuronal dendrites. *EMBO Rep.* **18**, 1762–1774 (2017).
- Ortiz, R. et al. Recruitment of Staufen2 enhances dendritic localization of an intron-containing CaMKII α mRNA. *Cell Rep.* **20**, 13–20 (2017).
- Monteuuis, G., Wong, J. J. L., Bailey, C. G., Schmitz, U. & Rasko, J. E. J. The changing paradigm of intron retention: regulation, ramifications and recipes. *Nucleic Acids Res.* **47**, 11497–11513 (2019).
- Ninomiya, K., Kataoka, N. & Hagiwara, M. Stress-responsive maturation of Clk1/4 pre-mRNAs promotes phosphorylation of SR splicing factor. *J. Cell Biol.* **195**, 27–40 (2011).
- Yeom, K.-H. et al. Tracking pre-mRNA maturation across subcellular compartments identifies developmental gene regulation through intron retention and nuclear anchoring. *Genome Res.* **31**, 1106–1119 (2021).
- Ripin, N. & Parker, R. Formation, function, and pathology of RNP granules. *Cell* **186**, 4737–4756 (2023).
- Hutchinson, J. N. et al. A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics* **8**, 1–16 (2007).
- Dias, A. P., Dufu, K., Lei, H. & Reed, R. A role for TREX components in the release of spliced mRNA from nuclear speckle domains. *Nat. Commun.* **1**, 97 (2010).
- Mao, Y. S., Zhang, B. & Spector, D. L. Biogenesis and function of nuclear bodies. *Trends Genet.* **27**, 295–306 (2011).

40. Spector, D. L. & Lamond, A. I. Nuclear speckles. *Cold Spring Harb. Perspect. Biol.* **3**, a000646–a000646 (2011).
41. McIntyre, A. B. R. et al. Phosphorylation of a nuclear condensate regulates cohesion and mRNA retention. *Nat. Commun.* **16**, 390 (2025).
42. Williams, T. D. et al. mRNA export factors store nascent transcripts within nuclear speckles as an adaptive response to transient global inhibition of transcription. *Mol. Cell* **85**, 117–131 (2025).
43. Galganski, L., Urbanek, M. O. & Krzyzosiak, W. J. Nuclear speckles: molecular organization, biological function and role in disease. *Nucleic Acids Res.* **45**, 10350–10368 (2017).
44. Colwill, K. et al. The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution. *EMBO J.* **15**, 265–275 (1996).
45. Araki, S. et al. Inhibitors of CLK protein kinases suppress cell growth and induce apoptosis by modulating pre-mRNA splicing. *PLoS ONE* **10**, e0116929 (2015).
46. Aubol, B. E., Keshwani, M. M., Fattet, L. & Adams, J. A. Mobilization of a splicing factor through a nuclear kinase-kinase complex. *Biochem. J.* **475**, 677–690 (2018).
47. Rai, A. K., Chen, J.-X., Selbach, M. & Pelkmans, L. Kinase-controlled phase transition of membraneless organelles in mitosis. *Nature* **559**, 211–216 (2018).
48. Tsai, M.-J., Ting, A. C., Nordstrom, J. L., Zimmer, W. & O'Malley, B. W. Processing of high molecular weight ovalbumin and ovomucoid precursor RNAs to messenger RNA. *Cell* **22**, 219–230 (1980).
49. Beyer, A. L., Bouton, A. H. & Miller, O. L. Correlation of hnRNP structure and nascent transcript cleavage. *Cell* **26**, 155–165 (1981).
50. Khodor, Y. L. et al. Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in *Drosophila*. *Genes Dev.* **25**, 2502–2512 (2011).
51. Khodor, Y. L., Menet, J. S., Tolan, M. & Rosbash, M. Cotranscriptional splicing efficiency differs dramatically between *Drosophila* and mouse. *RNA* **18**, 2174–2186 (2012).
52. Coulon, A. et al. Kinetic competition during the transcription cycle results in stochastic RNA processing. *eLife* **3**, e03939 (2014).
53. Carrillo Oesterreich, F. et al. Splicing of nascent RNA coincides with intron exit from RNA Polymerase II. *Cell* **165**, 372–381 (2016).
54. Drexler, H. L., Choquet, K. & Churchman, L. S. Splicing kinetics and coordination revealed by direct nascent RNA sequencing through nanopores. *Mol. Cell* **77**, 985–998 (2020).
55. Reimer, K. A., Mimoso, C. A., Adelman, K. & Neugebauer, K. M. Co-transcriptional splicing regulates 3' end cleavage during mammalian erythropoiesis. *Mol. Cell* **81**, 998–1012 (2021).
56. Choquet, K. et al. Pre-mRNA splicing order is predetermined and maintains splicing fidelity across multi-intronic transcripts. *Nat. Struct. Mol. Biol.* **30**, 1064–1076 (2023).
57. Coté, A. et al. Post-transcriptional splicing can occur in a slow-moving zone around the gene. *eLife* **12**, RP91357 (2024).
58. Huang, S. & Spector, D. L. Nascent pre-mRNA transcripts are associated with nuclear regions enriched in splicing factors. *Genes Dev.* **5**, 2288–2302 (1991).
59. Huang, S. & Spector, D. L. Intron-dependent recruitment of pre-mRNA splicing factors to sites of transcription. *J. Cell Biol.* **133**, 719–732 (1996).
60. Barutcu, A. R. et al. Systematic mapping of nuclear domain-associated transcripts reveals speckles and lamina as hubs of functionally distinct retained introns. *Mol. Cell* **82**, 1035–1052 (2022).
61. Brown, J. M. et al. Association between active genes occurs at nuclear speckles and is modulated by chromatin environment. *J. Cell Biol.* **182**, 1083–1097 (2008).
62. Lawrence, J. B. & Clemson, C. M. Gene associations: true romance or chance meeting in a nuclear neighborhood?. *J. Cell Biol.* **182**, 1035–1038 (2008).
63. Khanna, N., Hu, Y. & Belmont, A. S. HSP70 transgene directed motion to nuclear speckles facilitates heat shock activation. *Curr. Biol.* **24**, 1138–1144 (2014).
64. Bhat, P. et al. Genome organization around nuclear speckles drives mRNA splicing efficiency. *Nature* **629**, 1165–1173 (2024).
65. Melé, M. et al. Chromatin environment, transcriptional regulation, and splicing distinguish lincRNAs and mRNAs. *Genome Res.* **27**, 27–37 (2017).
66. Wong, J. J.-L. et al. Orchestrated intron retention regulates normal granulocyte differentiation. *Cell* **154**, 583–595 (2013).
67. Schmitz, U. et al. Intron retention enhances gene regulatory complexity in vertebrates. *Genome Biol.* **18**, 216 (2017).
68. Memon, D. et al. Hypoxia-driven splicing into noncoding isoforms regulates the DNA damage response. *NPJ Genom. Med.* **1**, 16020 (2016).
69. Rowland, T. J., Dumbović, G., Hass, E. P., Rinn, J. L. & Cech, T. R. Single-cell imaging reveals unexpected heterogeneity of telomerase reverse transcriptase expression across human cancer cell lines. *Proc. Natl Acad. Sci. USA* **116**, 18488–18497 (2019).
70. Kim, J.-H. et al. SON drives oncogenic RNA splicing in glioblastoma by regulating PTBP1/PTBP2 switching and RBFOX2 activity. *Nat. Commun.* **12**, 5551 (2021).
71. Diao, Y. et al. SF3B4 promotes ovarian cancer progression by regulating alternative splicing of RAD52. *Cell Death Dis.* **13**, 179 (2022).
72. Iannone, C. et al. PTBP1-activated co-transcriptional splicing controls epigenetic status of pluripotent stem cells. *Mol. Cell* **83**, 203–218 (2023).
73. Lee, S. et al. Intron retention decreases METTL3 expression by inhibiting mRNA export to the cytoplasm. *BMB Rep.* **56**, 514–519 (2023).
74. Wu, J. et al. Dynamics of RNA localization to nuclear speckles are connected to splicing efficiency. *Sci. Adv.* **10**, 1–21 (2024).
75. Wang, B. et al. CLK2 condensates reorganize nuclear speckles and induce intron retention. *Adv. Sci.* **11**, e2309588 (2024).
76. Pereira, I. T., Mamede, I., Amaral, P. P., Franco, G. R. & Rinn, J. L. Widespread specific intron retention events in nuclear RNA complexes identified by sedimentation analysis of pluripotent cellular extracts. *Genome Res.* **35**, 2189–2198 (2025).
77. Nishino, T. et al. CENP-T provides a structural platform for outer kinetochore assembly. *EMBO J.* **32**, 424–436 (2013).
78. Kottmann, M. C. & Smogorzewska, A. Fanconi anaemia and the repair of Watson and Crick DNA crosslinks. *Nature* **493**, 356–363 (2013).
79. Lashgari, A., Fauteux, M., Maréchal, A. & Gaudreau, L. Cellular depletion of BRD8 causes p53-dependent apoptosis and induces a DNA damage response in non-stressed cells. *Sci. Rep.* **8**, 14089 (2018).
80. Morin, G. B. The human telomere terminal transferase enzyme is a ribonucleoprotein that synthesizes TTAGGG repeats. *Cell* **59**, 521–529 (1989).
81. Liu, S. et al. METTL3 plays multiple functions in biological processes. *Am. J. Cancer Res.* **10**, 1631–1646 (2020).
82. Horlacher, M. et al. Towards in silico CLIP-seq: predicting protein–RNA interaction via sequence-to-signal learning. *Genome Biol.* **24**, 180 (2023).
83. Chen, L. et al. R-ChIP using inactive RNase H reveals dynamic coupling of R-loops with transcriptional pausing at gene promoters. *Mol. Cell* **68**, 745–757 (2017).
84. Yan, P. et al. Genome-wide R-loop landscapes during cell differentiation and reprogramming. *Cell Rep.* **32**, 107870 (2020).
85. Anastasiadi, D., Esteve-Codina, A. & Piferrer, F. Consistent inverse correlation between DNA methylation of the first intron and gene expression across tissues and species. *Epigenetics Chromatin* **11**, 37 (2018).

86. Georgakopoulos-Soares, I. et al. Alternative splicing modulation by G-quadruplexes. *Nat. Commun.* **13**, 2404 (2022).
87. Lejault, P. et al. G-quadruplex as a structural modulator of Intron Retention upon viral infection. Preprint at *bioRxiv* <https://doi.org/10.1101/2023.04.12.536615> (2023).
88. Coolidge, C. J., Seely, R. J. & Patton, J. G. Functional analysis of the polypyrimidine tract in pre-mRNA splicing. *Nucleic Acids Res.* **25**, 888–896 (1997).
89. Hauer, C. et al. Exon junction complexes show a distributional bias toward alternatively spliced mRNAs and against mRNAs coding for ribosomal proteins. *Cell Rep.* **16**, 1588–1603 (2016).
90. Sullivan, P. J. et al. Data-driven insights to inform splice-altering variant assessment. *Am. J. Hum. Genet.* **112**, 764–778 (2025).
91. Sakabe, N. J. & de Souza, S. J. Sequence features responsible for intron retention in human. *BMC Genomics* **8**, 59 (2007).
92. Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S. The MEME Suite. *Nucleic Acids Res.* **43**, W39–W49 (2015).
93. Benoit Bouvrette, L. P. et al. RBP Image Database: a resource for the systematic characterization of the subcellular distribution properties of human RNA binding proteins. *Nucleic Acids Res.* **51**, D1549–D1557 (2023).
94. Zhang, L. et al. Cross-talk between PRMT1-mediated methylation and ubiquitylation on RBM15 controls RNA splicing. *eLife* **4**, e07938 (2015).
95. Qu, W. et al. CPEB3 regulates neuron-specific alternative splicing and involves neurogenesis gene expression. *Aging* **13**, 2330–2347 (2021).
96. Hochberg-Laufer, H., Schwed-Gross, A., Neugebauer, K. M. & Shav-Tal, Y. Uncoupling of nucleo-cytoplasmic RNA export and localization during stress. *Nucleic Acids Res.* **47**, 4778–4797 (2019).
97. Dominguez, C., Fiset, J.-F., Chabot, B. & Allain, F. H.-T. Structural basis of G-tract recognition and engaging by hnRNP F quasi-RRMs. *Nat. Struct. Mol. Biol.* **17**, 853–861 (2010).
98. Wang, Y. et al. The splicing factor RBM4 controls apoptosis, proliferation, and migration to suppress tumor progression. *Cancer Cell* **26**, 374–389 (2014).
99. Wang, L. & Yan, F. Molecular insights into the specific recognition between the RNA binding domain qRRM2 of hnRNP F and G-tract RNA: a molecular dynamics study. *Biochem. Biophys. Res. Commun.* **494**, 95–100 (2017).
100. Ghanem, L. R. et al. Poly(C)-binding protein Pcbp2 enables differentiation of definitive erythropoiesis by directing functional splicing of the Runx1 transcript. *Mol. Cell. Biol.* **38**, e00175–18 (2018).
101. Aceituno-Valenzuela, U., Micol-Ponce, R. & Ponce, M. R. Genome-wide analysis of CCHC-type zinc finger (ZCCHC) proteins in yeast, Arabidopsis, and humans. *Cell. Mol. Life Sci.* **77**, 3991–4014 (2020).
102. Niu, K. et al. Identification of LARK as a novel and conserved G-quadruplex binding protein in invertebrates and vertebrates. *Nucleic Acids Res.* **47**, 7306–7320 (2019).
103. Niu, K., Zhang, X., Song, Q. & Feng, Q. G-quadruplex regulation of VEGFA mRNA translation by RBM4. *Int. J. Mol. Sci.* **23**, 743 (2022).
104. Yao, F. et al. Deletion of ARGLU1 causes global defects in alternative splicing in vivo and mouse cortical malformations primarily via apoptosis. *Cell Death Dis.* **14**, 543 (2023).
105. Huang, H. et al. The RNA-binding protein hnRNP F is required for the germinal center B cell response. *Nat. Commun.* **14**, 1731 (2023).
106. Fei, J. et al. Non-random compositional organization of nuclear speckles. *Biophys. J.* **110**, 16a (2016).
107. Fei, J. et al. Quantitative analysis of multilayer organization of proteins and RNA in nuclear speckles at super resolution. *J. Cell Sci.* **130**, 4180–4192 (2017).
108. Ferreira, J. A., Carmo-Fonseca, M. & Lamond, A. I. 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 (1994).
109. Chen, Y. et al. Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. *J. Cell Biol.* **217**, 4025–4048 (2018).
110. Zhang, L. et al. TSA-seq reveals a largely conserved genome organization relative to nuclear speckles with small position changes tightly correlated with gene expression changes. *Genome Res.* **31**, 251–264 (2021).
111. Małszycki, M. et al. Nuclear speckles enable processing of RNA from GC-rich isochores. *Cell* **189**, 2024–2039 (2026).
112. Ilik, İ. A. et al. SON and SRRM2 are essential for nuclear speckle formation. *eLife* **9**, e60579 (2020).
113. Yiangou, L. et al. Method to synchronize cell cycle of human pluripotent stem cells without affecting their fundamental characteristics. *Stem Cell Rep.* **12**, 165–179 (2019).
114. Talhouarne, G. J. S. & Gall, J. G. Lariat intronic RNAs in the cytoplasm of vertebrate cells. *Proc. Natl Acad. Sci. USA* **115**, E7970–E7977 (2018).
115. Singh, J. & Padgett, R. A. Rates of in situ transcription and splicing in large human genes. *Nat. Struct. Mol. Biol.* **16**, 1128–1133 (2009).
116. Bahar Halpern, K. et al. Nuclear retention of mRNA in mammalian tissues. *Cell Rep.* **13**, 2653–2662 (2015).
117. Chen, T. & van Steensel, B. Comprehensive analysis of nucleocytoplasmic dynamics of mRNA in Drosophila cells. *PLoS Genet.* **13**, e1006929 (2017).
118. Steinbrecht, D. et al. Subcellular mRNA kinetic modeling reveals nuclear retention as rate-limiting. *Mol. Syst. Biol.* **20**, 1346–1371 (2024).
119. Olsen, J. V. et al. Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Sci. Signal.* **3**, ra3 (2010).
120. Mochida, S., Rata, S., Hino, H., Nagai, T. & Novák, B. Two bistable switches govern M phase entry. *Curr. Biol.* **26**, 3361–3367 (2016).
121. Holder, J., Poser, E. & Barr, F. A. Getting out of mitosis: spatial and temporal control of mitotic exit and cytokinesis by PP1 and PP2A. *FEBS Lett.* **593**, 2908–2924 (2019).
122. Dominguez, D. et al. An extensive program of periodic alternative splicing linked to cell cycle progression. *eLife* **5**, e10288 (2016).
123. Chen, L.-L. & Carmichael, G. G. Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA. *Mol. Cell* **35**, 467–478 (2009).
124. Modic, M. et al. Cross-regulation between TDP-43 and paraspeckles promotes pluripotency-differentiation transition. *Mol. Cell* **74**, 951–965 (2019).
125. Bond, C. S. & Fox, A. H. Paraspeckles: nuclear bodies built on long noncoding RNA. *J. Cell Biol.* **186**, 637–644 (2009).
126. Kim, J., Venkata, N. C., Hernandez Gonzalez, G. A., Khanna, N. & Belmont, A. S. Gene expression amplification by nuclear speckle association. *J. Cell Biol.* **219**, e201904046 (2020).
127. Tripathi, V. et al. SRSF1 regulates the assembly of pre-mRNA processing factors in nuclear speckles. *Mol. Biol. Cell* **23**, 3694–3706 (2012).
128. Sohail, M. et al. Evolutionarily emerged G tracts between the polypyrimidine tract and 3' AG are splicing silencers enriched in genes involved in cancer. *BMC Genomics* **15**, 1143 (2014).
129. Blechinger, J., Holm, I. E. & Nielsen, A. L. Characterization and expression analysis in the developing embryonic brain of the porcine FET family: FUS, EWS, and TAF15. *Gene* **493**, 27–35 (2012).
130. Lee, J. et al. EWSR1, a multifunctional protein, regulates cellular function and aging via genetic and epigenetic pathways. *Biochim. Biophys. Acta Mol. Basis Dis.* **1865**, 1938–1945 (2019).

131. Biayna, J. & Dumbović, G. Decoding subcellular RNA localization one molecule at a time. *Genome Biol.* **26**, 45 (2025).
132. Ninomiya, K. et al. LncRNA-dependent nuclear stress bodies promote intron retention through SR protein phosphorylation. *EMBO J.* **39**, 1–20 (2020).
133. Naro, C. et al. An orchestrated intron retention program in meiosis controls timely usage of transcripts during germ cell differentiation. *Dev. Cell* **41**, 82–93 (2017).
134. Navarro, A. P. & Cheeseman, I. M. Kinetochore assembly throughout the cell cycle. *Semin. Cell Dev. Biol.* **117**, 62–74 (2021).
135. Nalepa, G. et al. Fanconi anemia signaling network regulates the spindle assembly checkpoint. *J. Clin. Invest.* **123**, 3839–3847 (2013).
136. Abdul-Sater, Z. et al. FANCA safeguards interphase and mitosis during hematopoiesis in vivo. *Exp. Hematol.* **43**, 1031–1046 (2015).
137. Nalepa, G. & Clapp, D. W. Fanconi anaemia and cancer: an intricate relationship. *Nat. Rev. Cancer* **18**, 168–185 (2018).
138. Wang, W. et al. FANCA facilitates G1/S cell cycle advancement, proliferation, migration and invasion in gastric cancer. *Acta Biochim. Biophys. Sin.* **56**, 973–985 (2024).
139. Kobayashi, M. et al. The RNA methyltransferase complex of WTAP, METTL3, and METTL14 regulates mitotic clonal expansion in adipogenesis. *Mol. Cell. Biol.* **38**, e00116–e00118 (2018).
140. Liu, Q. et al. ZFP217 regulates adipogenesis by controlling mitotic clonal expansion in a METTL3-m6A dependent manner. *RNA Biol.* **16**, 1785–1793 (2019).
141. Zhang, C. et al. METTL3 and N6-methyladenosine promote homologous recombination-mediated repair of DSBs by modulating DNA-RNA hybrid accumulation. *Mol. Cell* **79**, 425–442 (2020).
142. Yamaguchi, K. et al. Bromodomain protein BRD8 regulates cell cycle progression in colorectal cancer cells through a TIP60-independent regulation of the pre-RC complex. *iScience* **26**, 106563 (2023).
143. Liu, Y. et al. The telomerase reverse transcriptase is limiting and necessary for telomerase function in vivo. *Curr. Biol.* **10**, 1459–1462 (2000).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2026

¹Goethe University Frankfurt, Center for Molecular Medicine, Institute of Cardiovascular Regeneration, Frankfurt, Germany. ²Cluster of Excellence Cardio-Pulmonary Institute, Goethe University Frankfurt, Frankfurt, Germany. ³German Center of Cardiovascular Research (DZHK), partner site Rhein/Main, Frankfurt, Germany. ⁴Computational Health Center, Helmholtz Center Munich, Munich, Germany. ⁵ScitoVation LLC, Durham, NC, USA. ⁶Cluster of Excellence SubCellular Architecture of Life, Goethe University Frankfurt, Frankfurt, Germany. ⁷Present address: Berlin Institute for Medical Systems Biology, Max Delbrück Center for Molecular Medicine, Berlin, Germany. ⁸These authors contributed equally: Josep Biayna, Artem Baranovskii.

✉ e-mail: dumbovic@med.uni-frankfurt.de

Methods

Cell culture and cell cycle synchronization

hiPS cells WTC-11 (GM25256, male, Coriell Institute) were cultured on Vitronectin (Thermo Fisher Scientific) coated plates or 18-mm round glass coverslips for smRNA FISH and IF in StemFlex basal medium (Thermo Fisher Scientific) with StemFlex supplement (Thermo Fisher Scientific) and 2.5% penicillin–streptomycin. RevitaCell supplement with ROCK inhibitor (Thermo Fisher Scientific) was added when thawing the cells. hiPS cells were passaged with 0.5 mM EDTA in Dulbecco's phosphate-buffered saline (DPBS). HUVECs (c-12203, pool of male/female cells, PromoCell) were freshly isolated from donors and cultured for no more than two passages to preserve their native transcriptional and nuclear architecture. They were cultured on 25 ng ml⁻¹ fibronectin (Sigma-Aldrich)-coated glass coverslips (for smRNA FISH and IF) in endothelial cell basal medium (PromoCell) with endothelial cell growth medium supplement mix (PromoCell).

Actinomycin D (ActD, Sigma-Aldrich) was used at a final concentration of 5 µg ml⁻¹ in complete growth media. Cell pellets and coverslips were collected at 0 h, 0.5 h, 2 h and 4.5 h after adding ActD. Kinase inhibition was performed using DYRK3 inhibitor GSK626616 (2 µM; MedChemExpress), CLK inhibitor CLK-IN-T3 (1.3 µM; MedChemExpress), CDK1 inhibitor RO-3306 (15 µM; Selleckchem) or dimethyl sulfoxide as a control, diluted in StemFlex medium for 5–7 h.

Cell cycle synchronization at the G2/M phase was performed with nocodazole (75 ng ml⁻¹; Sigma-Aldrich) for 16 h (ref. 144). Synchronization at the G1/S phase was achieved with a double thymidine block¹⁴⁴. For the first block, cells were incubated with 2 mM thymidine (Sigma-Aldrich) for 14.5 h. Thymidine-containing medium was then removed, cells were washed with DPBS and fresh growth medium was added to release cells for 9 h. A second thymidine block was subsequently applied for 16 h. To initiate the second release (defined as $t = 0$ h), thymidine-containing medium was removed, cells were washed with DPBS and fresh growth medium was added.

Coverslips were collected at defined time windows after the second release to enrich for specific cell cycle stages. S phase samples were collected at 3 h and 4 h after release. G2, M, and early G1 samples were collected at 8, 9 and 9.5 h after release. Cell cycle stage was assessed by nuclear size and PCNA staining pattern¹⁴⁵, as explained in the 'Wide-field microscopy and image analysis' section.

siRNA transfection

The day before transfection, hiPS cells were seeded on vitronectin-coated coverslips. Small interfering RNA (siRNA) targeting SON and/or SRRM2 (Thermo Fisher Scientific) were transfected at a final concentration of 40 nM using Lipofectamine RNAiMAX reagent (Thermo Fisher Scientific) following the manufacturer's guidelines. After 6 h, the transfection mix was replaced by complete StemFlex growth media, and the cells were incubated for an additional 24 h. smRNA FISH and IF were performed as described below.

RNA fractionation

RNA fractionation was performed as described previously^{146,147}. In brief, hiPS cells were rinsed twice with ice-cold phosphate-buffered saline (PBS), scraped and centrifuged at 500g for 5 min. Cell pellets were lysed with 175 µl/10⁶ cells of ice-cold lysis buffer (50 mM Tris–HCl pH 8.0, 140 mM NaCl, 1.5 mM MgCl₂, 0.5% NP-40, 1:200 dilution of 20 U µl⁻¹ SUPERase RNase inhibitor) and incubated on ice for 10 min. The lysate was centrifuged at 300g for 3 min at 4 °C to pellet the nuclei, and the supernatant was collected as the cytoplasmic fraction. Finally, the nuclear and the cytoplasmic fractions were processed for RNA extraction as described below. The efficiency of RNA fractionation was assessed by quantitative polymerase chain reaction (qPCR) using GAPDH as a marker for the cytoplasmic fraction and 45S pre-rRNA for the nuclear fraction.

RNA extraction

RNA extraction was performed with Maxwell RSC simplyRNA Tissue Kit (Promega) following the manufacturer's instructions with DNaseI treatment. DNA contamination was assessed after each extraction with a qPCR for genomic DNA on RNA samples.

cDNA synthesis and qPCR analysis

Reverse transcription was performed with SuperScript IV First-Strand Synthesis System (Thermo Fisher Scientific) with random hexamers. Relative expression was determined by qPCR using SYBR Green PCR master mix (Thermo Fisher Scientific) according to the manufacturer's instructions using the following amplification conditions: 95 °C 10 min; 45 cycles of 95 °C 15 s, 57.5 °C 20 s and 72 °C 25 s. Primers used in qPCR analyses are summarized in Supplementary Table 2. Relative expression was analysed by the comparative $\Delta\Delta C_t$ method.

RNA-seq

The quantity, quality and integrity of the RNA samples were assessed using Bioanalyzer 5400 (Agilent Technologies). After isolation, RNA was sent to Novogene for sequencing. In brief, poly(A) RNA was enriched with oligo d(T) magnetic beads. Subsequently, mRNA was randomly fragmented, and cDNA synthesis proceeded through the use of random hexamers and the reverse transcriptase enzyme. After purification, the resulting products underwent end-repair, A-tailing and adapter ligation steps. Fragments of the appropriate size were enriched by PCR, where indexed P5 and P7 primers were introduced, and final products were purified. The library was checked with Qubit 2.0 and real-time PCR for quantification and bioanalyzer Agilent 5400 for size distribution detection. Quantified libraries were pooled and sequenced on Illumina NovaSeq 6000/NovaSeq X Plus platforms. The qualified libraries were sequenced using next-generation sequencing based on Illumina's sequencing-by-synthesis technology, detection by fluorescence of the nucleotide added during the synthesis of the complementary chain, and in a parallelized and massive way. The strategy was paired-end 150 bp.

Analysis of IR from RNA-seq

Vast-tools v2.5.1¹⁴⁸ (<https://github.com/vastgroup/vast-tools>) was used to calculate PIR and transcripts per million (TPM) values. Reads mapping to mid-intron sequences and balanced counts of reads aligning to upstream and downstream exon–intron sequences were used to evaluate IR levels. PIR was measured as a percentage of mean retention reads over the sum of retained and spliced intron reads. Raw values were filtered based on reported quality scores, requiring at least 15 total reads per event and absence of a positive result ($P < 0.01$) for the binomial test for upstream/downstream junction read balance. Additional filtering was applied to retain introns with ≥ 5 non-missing values across RNA-seq datasets after quality filtering and a minimum nuclear expression of 5 TPM. For analyses of IR kinetics after transcriptional inhibition, events with both cytoplasmic untreated replicates showing PIR values $\geq 99\%$ were excluded. For the analysis of nuclear and cytoplasmic RNA stability, RNAs expressed at a minimum of 7 TPM in both compartments were kept. This threshold was chosen because, among IR-mRNAs, RAD52 exhibited the lowest cytoplasmic abundance as assessed by smRNA FISH (mean of one molecule in the cytoplasm per cell), which corresponded to 7 TPMs in the cytoplasm. This ensured that RNAs included in the stability analysis were reliably detected in both compartments and avoided bias from low-abundance RNAs where fluctuations in read counts could disproportionately affect stability measurements.

IR stability classes were classified based on IR kinetics at the intron-event level, such that nuclear PIR values at each timepoint were normalized to the nuclear PIR level of the same intron in untreated cells. Introns were considered classifiable if they showed basal nuclear retention in untreated samples (mean nuclear PIR ≥ 30) and replicate support in untreated nuclei (PIR ≥ 25 in both replicates when available). Stable introns were defined as those retaining at least 50% of their

initial nuclear PIR signal at 0.5 h, 2 h and 4.5 h after transcriptional inhibition, with absolute nuclear PIR values of at least 25 at each treatment timepoint. Unstable classes were defined according to the first timepoint at which normalized nuclear PIR dropped below 50% of their initial nuclear PIR. Introns with untreated nuclear PIR below 30 were assigned to the no IR group. The analysis was repeated using an untreated nuclear PIR cutoff of ≥ 50 while retaining the same stability criteria based on normalized PIR kinetics. Transcripts were assigned to an LL-IR-RNA group if they contained at least one RI classified as stable at 4.5 h. For transcripts containing multiple RIs, transcript- or gene-level assignment reflects the most stable RI event. GO analysis was performed using WebGestalt¹⁴⁹.

Genome assembly used in the downstream tasks

GENCODE Human genome release 29 was used in all downstream tasks. Genome primary assembly was used to extract intronic sequences and respective flanks (https://ftp.ebi.ac.uk/pub/databases/gencode/gencode_human/release_29/GRCh38.primary_assembly.genome.fa.gz). Comprehensive annotation GTF was used to look-up boundaries of transcript bodies (https://ftp.ebi.ac.uk/pub/databases/gencode/gencode_human/release_29/gencode.v29.primary_assembly.annotation.gtf.gz).

ENCORE dataset

All eCLIP narrowPeak files for each RBP were retrieved from the ENCODE portal (ENCODE project) (https://www.encodeproject.org/encode-matrix/?type=Experiment&status=released&internal_tags=ENCORE&searchTerm=ENCORE). For downstream analyses, we summarized RBP peaks per intron and normalized by the intron's length. Normalized peak counts were then used to generate heatmaps across spliced/retained and stable/unstable intron classes to check for RBP peak enrichment in one class versus the other.

Estimation of half-lives at intron and gene level

Cells were sampled at 0, 0.5, 2 and 4.5 h after transcription termination. For each cellular compartment (nuclear/cytoplasmic) and replicate (A and B), gene-level abundance (TPMs) and PIR were quantified with vast-tools (see above). HLs were estimated separately at gene level using TPM measurements and at intron level using PIR measurements.

In both cases, the first timepoint zero (T_0) was isolated, requiring both replicates present. Then, all post- T_0 timepoints were normalized by their matching T_0 within the same replicate, yielding normalized $C(t)/C(T_0)$ values (TPMs in case of genes and PIR in case of introns), whereas T_0 was set to 1.

We assume first-order decay after transcription inhibition. Thus, for each compartment separately and for each gene/intron we estimated the decay rate by fitting a log-linear regression with ordinary least squares over both replicates, using `scipy.stats.linregress()`: $\log C_{\text{norm}}(t) = \alpha + \beta t$. The HL was calculated from the decay constant $k = -\beta$ as

$$\text{HL} = \frac{\ln 2}{k}$$

The HL's standard error estimate was propagated from the slope's standard error SE:

$$\text{SE}_{\text{HL}} \approx \frac{\ln 2}{k^2} \text{SE}_{\beta}$$

After computing compartment-specific intron and gene HLs, we applied three filters, requiring that

- (1) $C_{\text{norm}}(T_1) > 0$, to remove cases below our detection capabilities;
- (2) $\text{HL} > 0$;
- (3) HLs are filtered by relative uncertainty, ensuring that the relative HL error

$$r = \frac{|\text{SE}_{\text{HL}}|}{|\text{HL}|}$$

is below the fourth quartile of the r distribution.

HL estimates exceeding 12 h—reflecting a relatively ‘flat’ exponential decay and thus highly stable RIs—were set to the limit's value, consistent with previous studies¹⁵⁰, as HL cannot be reliably determined in such cases. Finally, for machine learning modelling, single-intron HLs are normalized to the HL of the corresponding full-length gene, to account for differences arising from the overall stability of the transcript.

Prediction of IR and intron HL with the Parnet foundation model

Parnet model. Parnet is a multitask sequence-to-signal deep learning model of protein–RNA interactions, extending our original RBPNet framework⁸². In brief, RBPNet is a dilated convolutional network that maps RNA sequences directly to single-nucleotide crosslink profiles from raw CLIP-seq data, trained on all regions with significant crosslink enrichment signal, without requiring peak calling. By modelling both experimental signal and control background for each RBP experiment separately, it disentangles protein-specific binding from technical noise. Trained on hundreds of thousands of regions across and for each of the ~200 RBPs from ENCORE, RBPNet learns binding profiles and encodes a functional representation of RNA sequences in its embedding space. Parnet builds on this foundation by generalizing to multiple RBPs simultaneously, leveraging the learned embeddings as transferable RNA representations. In practice, this means that Parnet encodes the sequence rules of the RBP interactome for a given RNA, enabling downstream tasks such as RNA classification and modelling RNA processes, either through frozen embeddings or with additional fine-tuning layers.

For predicting IR and intron HL from sequence, we used Parnet (release 0.3.0, 21M parameters; <https://github.com/marsico-lab/parnet>).

Parnet fine-tuning and class activation mapping (CAM) feature importance. Coordinates of alternative splicing events for *Homo sapiens* (hg38) were obtained from VastDB (https://vastdb.crg.eu/downloads/hg38/EVENT_INFO-hg38.tab.gz). Parnet was fine-tuned in two separate classification tasks: first, distinguishing retained from non-retained introns, with introns labelled as ‘spliced’ if $\text{PIR} < 0.1$ and as ‘retained’ if $\text{PIR} \geq 0.1$; and second, classifying unstable ($\log(\text{IR HL}) < -1.34$) versus stable ($\log(\text{IR HL}) \geq -0.36$) RIs ($\text{PIR} \geq 10\%$). Thresholds were defined based on the distributions shown in Fig. 3d and Extended Data Fig. 4a.

Training sequences were generated by extracting 256 bases from the 5' and 3' intron ends, including 50 bases of flanking exons to capture potential exonic splice enhancers and silencers, as most splice regulatory elements lie within a few hundred bases of splice sites. Fasta sequences were extracted based on the selected coordinates with the Python library pysam (<https://github.com/pysam-developers/pysam>). Sequences on the negative strand were reverse complemented. The resulting set of intronic sequences was one-hot encoded into four nucleotide channels (A, C, G and T) and forward-passed through the model's body. For short introns that did not span the full 256-base window, sequences were padded to 256 bases, with valid positions marked as 1 and padding as 0.

For this task, we used the core of the original Parnet model—its series of one-dimensional convolutional blocks with residual connections—wrapped in a ParnetBackbone class. This backbone converts one-hot encoded RNA sequences into a feature map that represents each position in the sequence. Valid positions are tracked using the binary mask defined above. On top of the backbone, we added two CAM attention heads, one for each intron end, based on the approach by Zhou et al.¹⁵¹. These attention heads learn to weight each position in the sequence according to its importance for the classification task. Weighted feature vectors are then pooled to create a single sequence-level representation, which is passed through a linear classifier to produce predictions. Finally, per-base attribution scores are computed to quantify the contribution of each nucleotide to the model's decision, with positions outside the valid sequence ignored.

The outputs from both 5' and 3' attention heads are combined in a small multilayer perceptron to produce the final prediction, the probability that an intron is retained/stable. Model training was performed using fivefold cross-validation. The checkpoint from the fold with the best validation performance was selected for downstream analyses.

Attribution score interpretation. To interpret the sequence elements underlying the model's attribution scores we used *tf-modisco lite* (<https://github.com/jmschrei/tfmodisco-lite>), a motif discovery algorithm tailored for deep learning attribution scores, which given the attribution scores computed above for each sequence, aggregates them to highlight patterns or motifs in the different classes, hereafter called attribution motifs. We followed the standard procedures provided in the package documentation, setting *motif_size*, *window_size* and *n_seqs* to 31, 256 and 1,000, respectively. Only motifs with a minimum of 20 seqlets derived at 0.05 false discovery rate (FDR) were retained. To link enriched motifs across intron sequence classes to potential RBP binding sites, we used *tfmodisco*'s implementation of TomTom to query what is, to our knowledge, the largest and most up-to-date database of human RBP motifs¹⁵². In detail, every attribution motif identified by *tf-modisco* was compared with every motif in the database, recording the top three most likely matches. For visualization, we reported only RBP motifs that matched the attribution motif with a *q* value (FDR) less than 0.05.

Visualization of intron embeddings with Parnet. Embeddings from the model's penultimate layer were extracted and average-pooled into three segments, 5' end, middle and 3' end, then concatenated and saved in an h5 file. For visualization, intron embeddings were projected into two dimensions using Uniform Manifold Approximation and Projection (UMAP; Python implementation)¹⁵³. Default parameters were used except for the number of neighbours increased to 50, the distance metric set to cosine, and the minimum distance between neighbours set to 0.15.

Quantification of RBP enrichment

The annotation of nuclear speckle-associated RBPs was obtained from RBP Image Database from the Lecuyer lab⁹³, selecting RBPs with a quality score >1 that were detected in nuclear speckles in either HepG2 or K562 cell lines (Supplementary Table 3). To test whether peaks of nuclear speckle-associated RBPs (from ENCODE NarrowPeaks) are differentially enriched or depleted in spliced versus RIs, we fitted a linear model with three terms: intron class (spliced or retained), RBP group (speckle or non-speckle) and their interaction. A type II analysis of variance (ANOVA) was then used to assess whether the difference in mean peak counts between speckle and non-speckle RBPs depends on intron class—that is, whether this difference changes between retained and spliced introns.

G-quadruplex motif enrichment

We used canonical G-quadruplex motif definitions, that is, (G{3,}[ATGCN]{1,7}){3,}G{3,}, which are widely employed in Potential G-Quadruplex-forming Sequences identification¹⁵⁴. For the variant motifs with longer loops, we extended this definition accordingly, that is (G{3,}[ATGCN]{8,12}){3,}G{3,}. Motifs containing only two G-tetrads have also been observed to form stable G-quadruplex structures, for example, intralocked G4s with G₂ tracts¹⁵⁵. In summary, the following motif regular expressions were used to analyse G-quadruplex motif enrichment in retained/stable versus non-retained/unstable introns:

```
motifs = {
    "G4_3t_short": r'G{3,}[ACGT]{1,7}G{3,}[ACGT]{1,7}G{3,}[ACGT]{1,7}G{3,}',
    "G4_3t_long": r'G{3,}[ACGT]{8,12}G{3,}[ACGT]{8,12}G{3,}[ACGT]{8,12}G{3,}',
    "G4_2t_short": r'G{2,}[ACGT]{1,7}G{2,}[ACGT]{1,7}G{2,}[ACGT]{1,7}G{2,}'
}
```

```
"G4_2t_long": r'G{2,}[ACGT]{8,12}G{2,}[ACGT]{8,12}G{2,}[ACGT]{8,12}G{2,}',
}
```

To test whether G-quadruplex motifs were enriched in different intron classes, we counted motif occurrences using pattern matching and modelled these counts with a Poisson regression. The model compared motif counts between classes while accounting for intron length. From the fitted model, we calculated how much more (or less) frequent the motifs were in each class and expressed these values as log₂ rate ratios for plotting.

ssDRIP-seq and bisulfite sequencing processing

To assess the enrichment of R-loops in retained versus non-retained introns, we downloaded and reprocessed publicly available ssDRIP-seq dataset⁸⁴. The raw sequencing data can be accessed with BioProject ID PRJNA608890 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA608890>). The dataset was processed as in the original publication. Namely, the reads were first trimmed with *trim_galore* (<https://github.com/FelixKrueger/TrimGalore>) and then mapped with STAR (<https://github.com/alexdobin/STAR>). The resulting bam files were further deduplicated with *samtools* (<https://github.com/samtools/samtools>). Finally, the locations of R-loops were called with *macs3* software (<https://github.com/mac3-project/MACS3>).

In addition, we downloaded and reprocessed bisulfite sequencing runs available under the same BioProject ID, namely SRR1118540, following the processing steps taken by the authors of the original publication. In detail, the reads were first trimmed with *trim_galore* then mapped and deduplicated with *Bismark* (<https://github.com/FelixKrueger/Bismark>). Finally, *Bismark*'s command *bismark_methylation_extractor* was used to extract methylation calls.

The shell scripts to execute the processing as well as the exact software versions are available via GitHub (see 'Code availability' section).

MFE calculation

The MFE of RNA secondary structures was calculated using the ViennaRNA Python API, following the provider's standard procedure. For each intronic sequence (including 50-bp flanks), the MFE of the predicted optimal structure was obtained and then normalized by intron length for downstream analysis¹⁵⁶.

Analysis of detained introns

Before downstream analysis, coordinates of detained introns from Boutz et al.¹⁷ were lifted over from hg19 to hg38 version of human genome assembly using *liftOver* tool provided by UCSC (<https://genome.ucsc.edu/cgi-bin/hgLiftOver>). Then, the coordinates of detained introns from Boutz et al.¹⁷ and Braun et al.¹⁸ were intersected with the introns from the vast-tools database, selecting the best possible match between two pairs of intronic coordinates. Of note, the mapping was successful for 11,729 out of 13,290 and 3,750 out of 3,968 introns, respectively. The mapped intron IDs are provided in Supplementary Table 4.

smRNA FISH and IF

smRNA FISH was performed as described previously^{147,157,158}. In brief, up to 48 tiled oligonucleotides targeting intronic and exonic regions of candidate transcripts were custom designed using Stellaris online RNA FISH Probe Designer (LGC Biosearch Technologies). Of note, due to small intron size, we tiled oligonucleotides across two RIs of METTL3 to image the IR-RNA; as a result, the two introns cannot be reliably distinguished as independent species. The imaged RI of RAD52 was selected based on probe-design feasibility and corresponds to unstable 4.5 h RI, whereas RAD52 harbours at least one LL-RI that was too short for reliable imaging. Probe labelling was carried out as previously described¹⁵⁹. For each target, up to 30 oligonucleotides (200 μM in nuclease-free water; Eurofins) were pooled into a single

master mix and labelled with either 5-Propargylamino-ddUTP-ATTO-647N (Jena Biosciences) or 5-propargylamino-ddUTP-ATTO-550 (Jena Biosciences). For labelling, 5 µl of the oligonucleotide master mix was combined with 1 µl ddUTP-ATTO-647N or ddUTP-ATTO-550, 0.3 µl terminal deoxynucleotidyl transferase (TdT, 20 U µl⁻¹), 3 µl TdT buffer (Thermo Fisher Scientific) and 5.7 µl water in a PCR tube. Reactions were incubated at 37 °C for 16 h in a PCR thermocycler with the lid set to 40 °C; an additional 0.3 µl of fresh TdT enzyme was added after 8 h. Labelled oligonucleotides were purified using the Oligo Clean & Concentrator kit (Zymo Research) according to the manufacturer's instructions. Labelling efficiency was assessed by absorbance at 646 nm (ATTO-647N) or 554 nm (ATTO-550). Oligonucleotide concentrations were calculated using Beer's law, and probes were diluted to a final stock concentration of 12.5 µM in water, which was used at a 1:100 dilution for hybridization.

Cells were seeded on vitronectin-coated (for hiPS cells) or fibronectin-coated (for HUVECs) 18-mm round glass coverslips. On the day of smRNA FISH, coverslips were washed twice with PBS, fixed with 3.7% formaldehyde in PBS for 10 min at room temperature, washed twice with PBS and incubated in 70% ethanol at 4 °C for a minimum of 1 h. Coverslips were washed with 1 ml of wash buffer A (nuclease-free 2× SSC, Thermo Fisher Scientific; 10% deionized formamide, Sigma-Aldrich) for 5 min at room temperature. Cells were hybridized with 80 µl of hybridization buffer (10% dextran sulfate in nuclease-free 2× SSC, 10% deionized formamide) containing smRNA FISH probes (1:100) overnight (ON) at 37 °C in a humid chamber.

The following day, coverslips were washed with 1 ml of wash buffer A for 30 min at 37 °C, followed by a second wash with wash buffer A containing Hoechst DNA stain (1:1,000; Thermo Fisher Scientific) for 30 min at 37 °C. Coverslips were washed with 1 ml of 2× SSC for 5 min at room temperature, equilibrated for 5 min in base glucose buffer (2× SSC, 0.4% glucose solution, 20 mM Tris, pH 8.0, in RNase-free water) and then incubated in base glucose buffer supplemented with a 1:100 dilution of glucose oxidase (stock 3.7 mg ml⁻¹, Sigma-Aldrich) and catalase (stock 4 mg ml⁻¹, Sigma-Aldrich) for 15 min at room temperature. Finally, coverslips were mounted with ProLong Glass Antifade Mountant (Thermo Fisher Scientific) on glass slides before image acquisition.

Coverslips for combinatorial IF and smRNA FISH were processed as described above, with primary antibodies added to the hybridization buffer: anti-SC35 antibody (1:1,500, #s4045, Clone SC-35, ascites fluid, Sigma-Aldrich), anti-RBM25 antibody (1:300, #25297-1-AP, Proteintech), anti-SON antibody (1:300, #83787-5-RR, Clone 240845B7, Proteintech) and anti-PCNA antibody (1:200, #HPA030522, Atlas Antibodies). The first wash with wash buffer A following overnight hybridization contained the appropriate Alexa Fluor 488-conjugated secondary antibody: anti-mouse IgG (1:1,000, #ab150113, Abcam) or anti-rabbit IgG (1:1,000, A-11008, Invitrogen).

For STED, high-precision glass coverslips were used (Hecht Assistant, thickness 170 ± 5 µm), DNA stain was omitted, and the secondary antibody was Abberior STAR 460L-conjugated anti-mouse IgG (1:1,000; #ST460L-1001, Abberior).

Wide-field microscopy and image analysis

smRNA FISH with or without IF z-stacks (200 nm step size) were acquired using the Nikon ECLIPSE Ti2 widefield microscope with a Nikon Plan Apo λ ×100/1.45-numerical aperture oil objective lens and a Nikon DS-Qi2 camera. Maximum intensity projections were generated using Nikon software (NIS-Elements version 5.11.03) and deconvolved in ImageJ (version 2.14.0/1.54f) using Deconvolution-Lab2 plugin (version 2.1.2)¹⁶⁰. Quantification of RNA FISH signal was performed using the 'Big-FISH' v0.6.2 Python package¹⁶¹ or manually when the number of RNA spots was low. RNA spots were assigned to nuclear or cytoplasmic compartments based on Hoechst segmentation. Intron-retained RNA spots were identified by spatial overlap of exon and intron smRNA FISH signals, and spliced RNA for the RI by

exon signal without co-localizing intron. Transcription sites were identified based on overlap between efficiently spliced intron signals and exon signals (for Fig. 4j and Extended Data Fig. 8b) or by high-intensity nascent RNA foci characteristic of active transcription (for Extended Data Fig. 8b; DNMT3B and LAMAS). Nuclear transcript fraction was quantified as the fraction of exon smRNA FISH spots detected in the nucleus relative to the total number of exon spots per cell, calculated as $N_{\text{exon,nuc}}/(N_{\text{exon,nuc}} + N_{\text{exon,cyt}})$, where $N_{\text{exon,nuc}}$ and $N_{\text{exon,cyt}}$ denote the numbers of exon spots detected in the nucleus and cytoplasm, respectively. Cytoplasmic transcript fraction was defined analogously as $N_{\text{exon,cyt}}/(N_{\text{exon,nuc}} + N_{\text{exon,cyt}})$. Nuclear PIR was calculated as $N_{\text{RI,nuc}}/N_{\text{exon,nuc}}$ and total PIR as $N_{\text{RI,nuc}}/(N_{\text{exon,nuc}} + N_{\text{exon,cyt}})$, where $N_{\text{RI,nuc}}$ denotes the number of nuclear IR-RNA spots.

The distances of intron smRNA FISH signals to the nearest nuclear speckle were measured in three dimensions in Imaris (Bitplane, version 11.0) using the 'Shortest Distance to Surfaces' function. RNA FISH signals were segmented as spots using the Imaris 'Spots' module. SC35 or U2 signals were segmented independently as three-dimensional (3D) surfaces representing the SC35-positive nuclear speckle core and the U2-positive nuclear speckle shell, respectively, using the Imaris 'Surface' module. For each RNA spot, the shortest distance from its centroid to the nearest segmented surface was computed using the 'Shortest Distance to Surfaces' function. Distances were calculated separately relative to the SC35 surface and to the U2 surface. For each segmented structure, distances were negative when RNA spots were located within the segmented volume and positive when located outside the respective surface. Values close to 0 indicate localization at the surface interface.

To define distances expected under random positioning, random spots of the same size as smRNA FISH spots were generated in Imaris, and their distances to nuclear speckles were extracted using the same workflow. For each RNA molecule, its distance was expressed as its percentile relative to the random mixture reference. Under random positioning, these percentiles are expected to be uniformly distributed. Nuclear speckle proximity was tested using a one-sided Kolmogorov–Smirnov test comparing the RNA percentile distribution to a uniform distribution, and corrected for multiple testing across RNAs using the Benjamini–Hochberg procedure¹⁶² separately for SC35 and U2. Alongside *P* values, effect sizes are reported as the shift in median and mean distance relative to the random mixture expectation and as fold enrichment within 0.1 µm of nuclear speckles in Supplementary Tables 1 and 5.

For cell cycle-resolved smRNA FISH analyses, cell cycle stage was assigned at single-cell resolution using PCNA IF together with nuclear morphology and size. Early, mid and late S-phase cells were identified by characteristic PCNA replication foci¹⁴⁵. G1 and G2 cells, both lacking PCNA foci, were distinguished by nuclear size, consistent with their 2N and 4N DNA content. Mitosis was identified by condensed chromatin morphology. Early G1 cells were identified as recently divided daughter-cell pairs with mirrored nuclear morphology and absent or early-onset transcriptional foci in exon/intron smRNA FISH. These annotations were used for downstream quantification.

STED imaging

STED imaging was performed as previously described with several changes¹⁶³. Image stacks were acquired on a Leica Stellaris 8 STED FALCON system using a 100× HC PL APO CS2 1.4-numerical aperture oil immersion objective. The samples were excited with a second-generation Leica pulsed White Light Laser at 553, 644 and 449 nm (for the long Stokes-shift dye) and detected using HyDS and HyDX detectors. For STED imaging of the three fluorophores of interest, the pulsed 775-nm STED laser was used and the STED laser intensity was set to 10%. The field of view was optimized for the ideal number of cells and the pixel dwell time was kept between 2.8 µs and 3 µs with a pixel size of 35 nm, and the scan speed was set to 200 Hz. The 3D z-stack

acquisition was performed in a line-by-line fashion for each channel in the sequence with a z-stack slice height of 183 nm. The 3D stacks were visualized in Imaris (Bitplane).

Statistics and reproducibility

In figures and figure legends, asterisk notation was used corresponding to: * $P < 0.05$, ** $P \leq 0.01$ and *** $P \leq 0.001$. P values were adjusted using the FDR procedure (also known as Benjamini–Hochberg procedure)¹⁶², unless otherwise stated in the figure legend where no multiple tests were conducted. Test names and statistics, sample sizes and P values are reported in Supplementary Table 5.

In Fig. 3f, to test whether the means of nuclear speckle-associated RBP peaks are differentially enriched (or depleted) in retained/spliced and stable/unstable introns, we fit a simple linear model with three coefficients: intron class, RBP group and the interaction of the two. Then, we applied a type II ANOVA to test for significance of differences between means of the groups. In this formulation, ANOVA answers the question whether the difference in peak count means between speckle and non-speckle RBPs depends on the intron class, that is, whether the said difference varies depending on whether the intron is retained or spliced.

No statistical method was used to predetermine sample size. Sample sizes for smRNA FISH and imaging experiments were chosen based on standard practice in the field and prior studies using similar approaches. Sample sizes for smRNA FISH quantifications were considered sufficient when consistent trends were observed across individual cells, imaging fields and independent experiments. All experiments were performed with a minimum of three independent replicates, unless otherwise stated in the figure legends. smRNA FISH with ActD-treated samples was performed twice and was considered sufficient given the high concordance between replicates, individual cells and consistency with vast-tools results. Validation of SRSF7 by smRNA FISH was performed in two independent experiments, as it was a validation of previously published findings. No data were excluded from the analyses, except as described in the Methods (for example, quality-control filtering, RNA-seq and intron-level filtering). The experiments were not randomized; samples were assigned to experimental groups based on the experimental design and treatment conditions. Data collection and analysis were not performed blind to the conditions of the experiments.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Raw RNA-seq data and processed data that include PIR and TPM values obtained through vast-tools have been uploaded to [GSE327794](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE327794). Publicly available datasets used in this study include TSA-seq data from [GSE81553](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE81553) (ref. 109), $\log_2$ FC of enrichment with nuclear speckle markers from ARTS-seq⁷⁴, ssDRIP-seq and bisulfite sequencing data from BioProject [PRJNA608890](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA608890) (ref. 84), eCLIP narrowPeak files from the ENCODE portal (ENCODE project) (https://www.encodeproject.org/encode-matrix/?type=Experiment&status=released&internal_tags=ENCODE&searchTerm=ENCODE) and detained introns from Boutz et al.¹⁷ and Braun et al.¹⁸. The mapped detained intron IDs are provided in Supplementary Table 4. A list of nuclear speckle-associated RBPs was obtained from RBP Image Database from the Lecuyer lab⁹³ and is provided in Supplementary Table 3. Source data are provided with this paper.

Code availability

Code to fine-tune the Parnet foundation model, compute and interpret model attribution scores, along with the scripts and notebooks to reproduce the results, is available via GitHub at https://github.com/marsico-lab/IR_iPSCs as well as via Zenodo at <https://doi.org/10.5281/zenodo.21135982> (ref. 164).

References

144. Cuevas-Ocaña, S. et al. A cell-based optimised approach for rapid and efficient gene editing of human pluripotent stem cells. *Int. J. Mol. Sci.* **24**, 10266 (2023).
145. Schönenberger, F., Deutzmann, A., Ferrando-May, E. & Merhof, D. Discrimination of cell cycle phases in PCNA-immunolabeled cells. *BMC Bioinformatics* **29**, 180 (2015).
146. Cabianca, D. S. et al. A long ncRNA links copy number variation to a polycomb/trithorax epigenetic switch in FSHD muscular dystrophy. *Cell* **149**, 819–831 (2012).
147. Dumbović, G. et al. A novel long non-coding RNA from NBL2 pericentromeric macrosatellite forms a perinucleolar aggregate structure in colon cancer. *Nucleic Acids Res.* **46**, 5504–5524 (2018).
148. Tapial, J. et al. 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 (2017).
149. Elizarraras, J. M. et al. WebGestalt 2024: faster gene set analysis and new support for metabolomics and multi-omics. *Nucleic Acids Res.* **52**, W415–W421 (2024).
150. Tushev, G. et al. Alternative 3' UTRs modify the localization, regulatory potential, stability, and plasticity of mRNAs in neuronal compartments. *Neuron* **98**, 495–511 (2018).
151. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A. & Torralba, A. Learning deep features for discriminative localization. In *Proc. 2016 IEEE Conf. Comput. Vis. Pattern Recognit.* 2921–2929 (IEEE Computer Society, 2015).
152. Sasse, A. et al. A resource of RNA-binding protein motifs across eukaryotes reveals evolutionary dynamics and gene-regulatory function. *Nat. Biotechnol.* **44**, 1226–1236 (2026).
153. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. Preprint at <https://arxiv.org/abs/1802.03426> (2020).
154. Puig Lombardi, E. et al. Thermodynamically stable and genetically unstable G-quadruplexes are depleted in genomes across species. *Nucleic Acids Res.* **47**, 6098–6113 (2019).
155. Maity, A., Winnerdy, F. R., Chang, W. D., Chen, G. & Phan, A. T. Intra-locked G-quadruplex structures formed by irregular DNA G-rich motifs. *Nucleic Acids Res.* **48**, 3315–3327 (2020).
156. Lorenz, R. et al. ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **6**, 26 (2011).
157. Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5**, 877–879 (2008).
158. Much, C. et al. Evolutionary divergence of Firre localization and expression. *RNA* **28**, 842–853 (2022).
159. Gaspar, I., Wippich, F. & Ephrussi, A. Enzymatic production of single-molecule FISH and RNA capture probes. *RNA* **23**, 1582–1591 (2017).
160. Sage, D. et al. DeconvolutionLab2: an open-source software for deconvolution microscopy. *Methods* **115**, 28–41 (2017).
161. Imbert, A. et al. FISH-quant v2: a scalable and modular tool for smFISH image analysis. *RNA* **28**, 786–795 (2022).
162. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
163. Dumbović, G., Sanjuan, X., Perucho, M. & Forcales, S.-V. Stimulated emission depletion (STED) super resolution imaging of RNA- and protein-containing domains in fixed cells. *Methods* **187**, 68–76 (2021).
164. Dumbovic, G., Marsico, A. & Baranovskii, A. Code for: The periphery of nuclear speckles organizes long-lived intron-retained RNAs into a mitotically resolved distinct nuclear retention compartment. Zenodo <https://doi.org/10.5281/zenodo.21135982> (2026).

Acknowledgements

We thank T. Zimmerman (EMBL Imaging Facility) for helpful discussions and advice regarding the planning of STED experiments. iPS cells were kindly provided by the laboratory of L. Kurian (Goethe University). We thank S. V. Forcales (University of Barcelona), D. Incarnato (University of Groningen) and L. Kurian (Goethe University) for insightful discussions and detailed feedback, and A. Raj (University of Pennsylvania) for sharing the protocol for probe labelling. We acknowledge the access and services provided by the Imaging Centre at the European Molecular Biology Laboratory (EMBL IC), generously supported by the Boehringer Ingelheim Foundation, and in particular D. Walsh for support with STED imaging setup.

Author contributions

Conceptualization: G.D. Experiments and analyses: J.B., A.C., G.D., M.P., B.E. and L.-E.K. Computational analyses: G.D., A.B., A.R.B. and A.M. Parnet: A.B. and A.M. Visualization: G.D., J.B. and A.B. Writing original draft: G.D. Draft editing: G.D., A.M., J.B. and A.B. Manuscript text revision: G.D. Funding and materials: G.D., S.D. and A.M. All authors have read and approved the manuscript.

Funding

A.C., M.P., B.E., J.B. and G.D. are supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through the TRR 267, Project A09 (project ID 403584255), a DFG Basic Module grant to G.D. (project ID 562550297) and Germany's Excellence Strategy through the Cardio-Pulmonary Institute, EXC 2026 (project ID 390649896). J.B. and B.E. are supported by the

Dr. Rolf M. Schwiete Stiftung (project no. 2024-047) to G.D. G.D. is supported by SCALE, EXC 3094 (project ID 533751785). Support to A.B. and A.M. was obtained by the BMBF Cluster4Future programme (Cluster for Nucleic Acid Therapeutics Munich, CNATM, project ID 03ZU1201AA) and SFB TRR 267 (project ID 403584255, project Z02). S.D. is supported by 403584255–TRR 267 (project B03). Open access funding provided by Johann Wolfgang Goethe-Universität.

Competing interests

The authors declare no competing interests.

Additional information

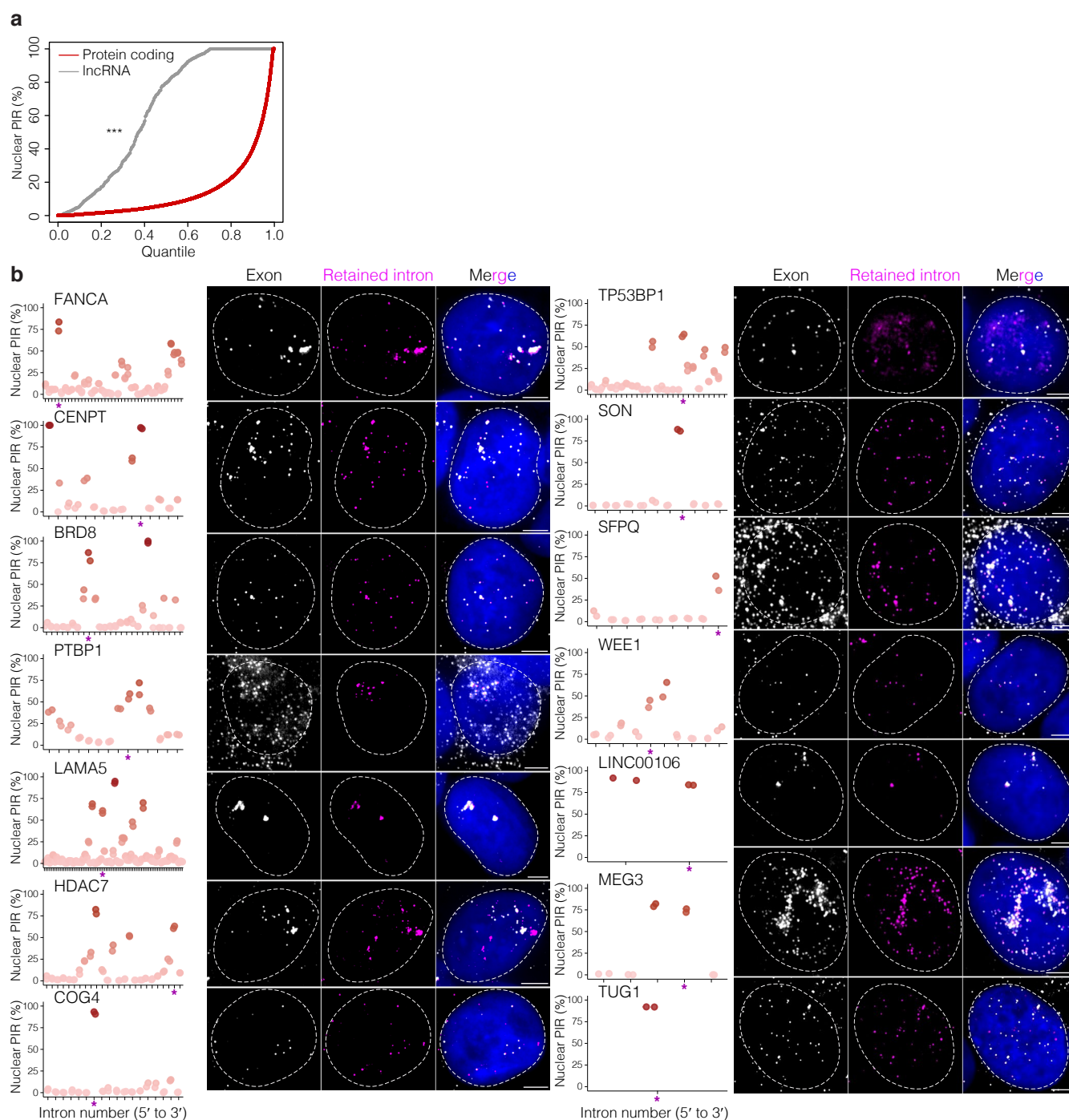
Extended data is available for this paper at <https://doi.org/10.1038/s41556-026-02040-5>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41556-026-02040-5>.

Correspondence and requests for materials should be addressed to Gabrijela Dumbović.

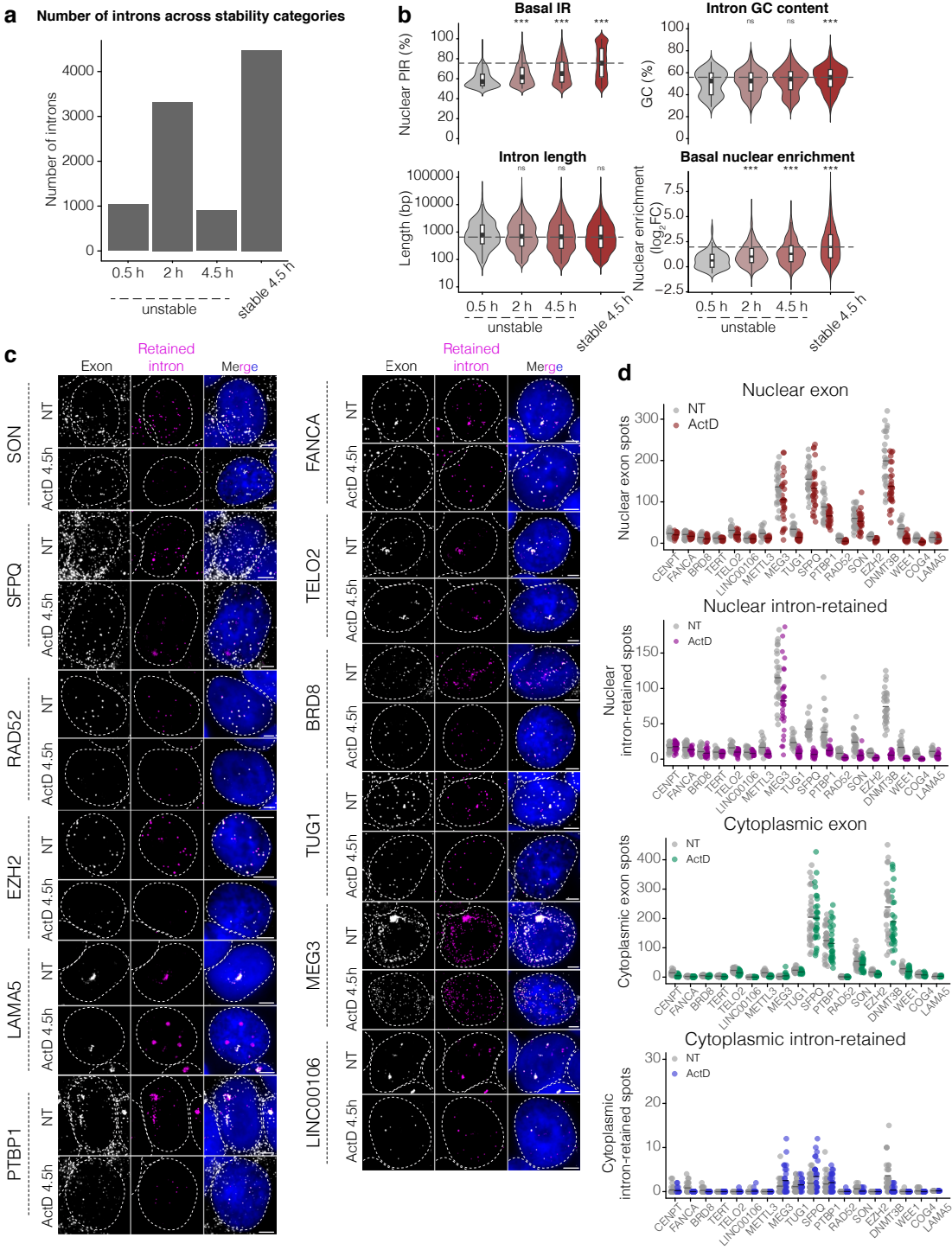
Peer review information *Nature Cell Biology* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.



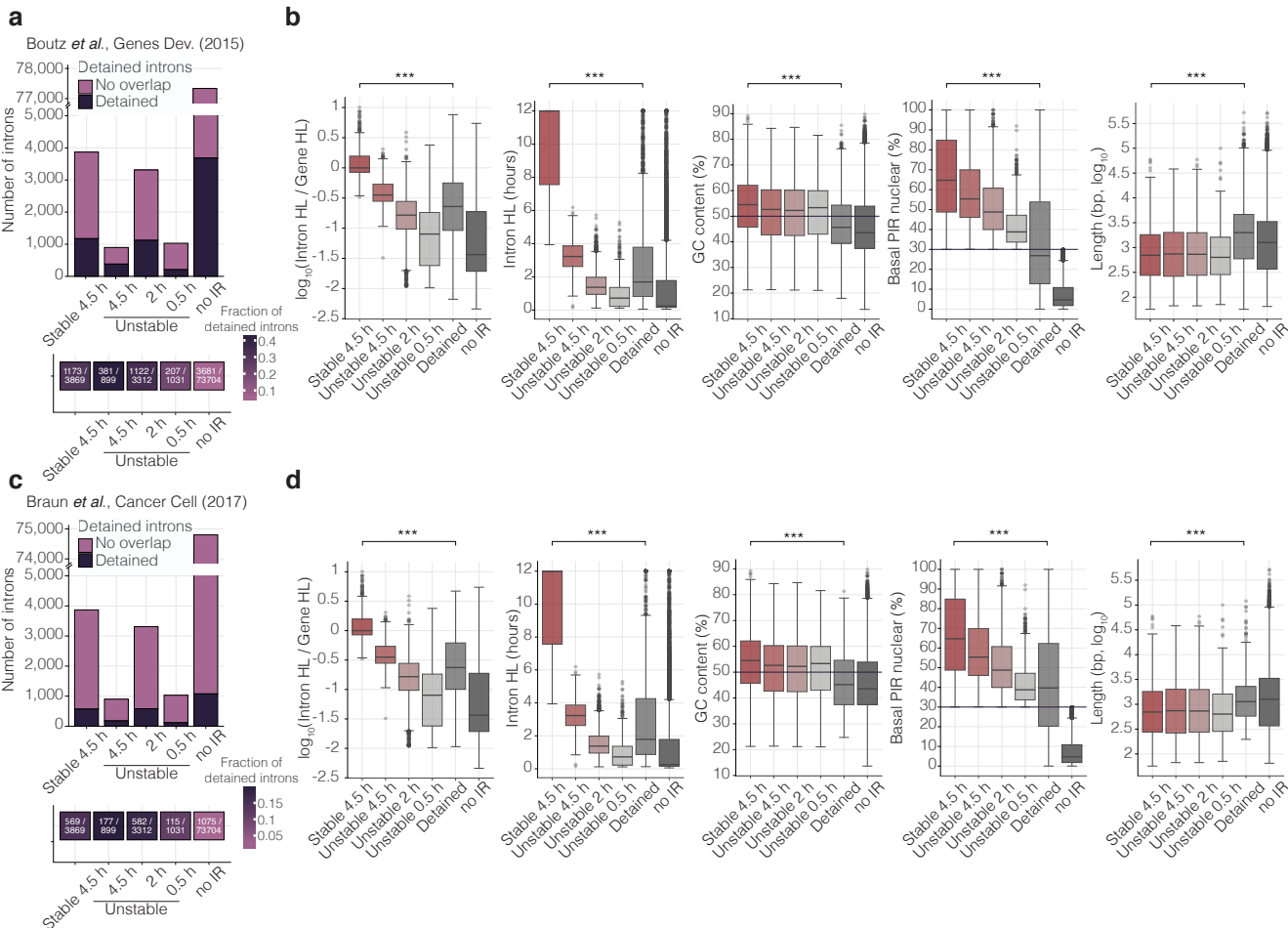
Extended Data Fig. 1 | IR associates with distinct nuclear RNA localization patterns in hiPSCs. a, Cumulative distribution of nuclear PIR for coding (red) and lncRNA (gray) genes. *** $P \leq 0.001$, evaluated by Kolmogorov-Smirnov test. $P < 2.2 \times 10^{-16}$. **b**, Dot plot showing nuclear PIR of individual introns for indicated genes ($n = 2$) obtained by vast-tools from poly(A)⁺ nuclear RNA-seq. Introns selected

for smRNA FISH are indicated with purple asterisks. On the right, representative smRNA FISH maximum intensity projections. Exon, white; intron, magenta; nucleus, blue, outlined with a dashed line; scale bar, 2 μm . Source numerical data are available in source data.



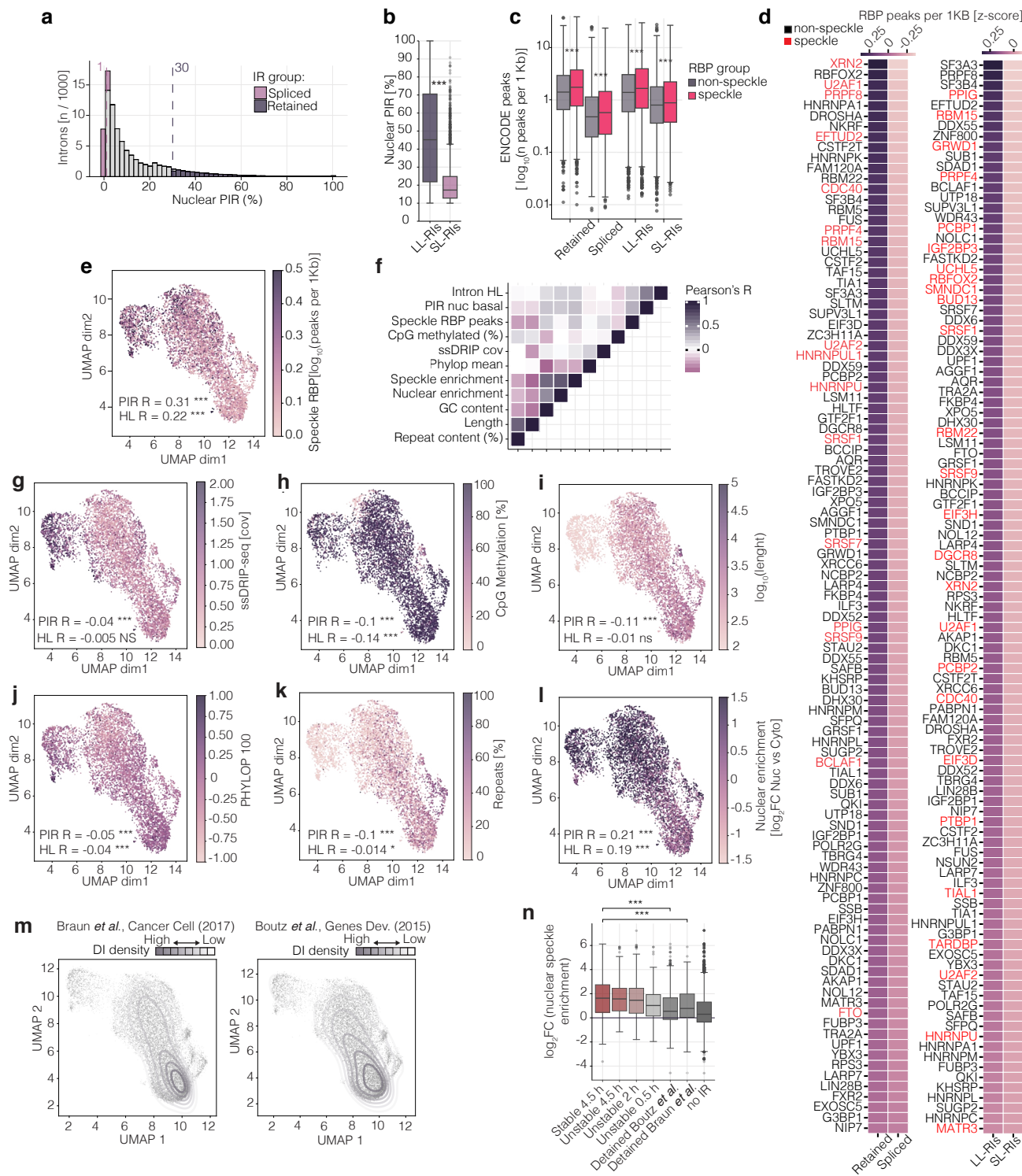
Extended Data Fig. 2 | smRNA FISH analysis of IR stability in hiPSCs following transcriptional inhibition. **a**, Number of introns across different IR stability categories. **b**, Violin plot showing distribution of basal PIR, GC content and length of RIs, as well as host RNA nuclear enrichment [$\log_2(\text{TPM}_{\text{nuc}}/\text{TPM}_{\text{cyto}})$] ($n = 206, 1,547, 568$ and $2,824$ introns for the unstable 0.5 h, unstable 2 h, unstable 4.5 h and stable groups, respectively, and $n = 122, 855, 350$ and $1,919$ transcripts, respectively). Midline, median; box limits, first and third quartiles; whiskers, most extreme values within $1.5 \times \text{IQR}$ of the hinges. ns = not significant, $***P \leq 0.001$, evaluated by Tukey's HSD test versus unstable 0.5 h. Basal PIR,

$P_{\text{adj}} = 2.80 \times 10^{-5}, 2.46 \times 10^{-11}, 4.06 \times 10^{-14}$ for unstable 2 h, unstable 4.5 h, stable 4.5 h, respectively; GC content, $P_{\text{adj}} = 1.93 \times 10^{-6}$ for stable 4.5 h; host RNA nuclear enrichment, $P_{\text{adj}} = 9.27 \times 10^{-4}, 7.05 \times 10^{-6}$ and $< 2.2 \times 10^{-16}$ for unstable 2 h, unstable 4.5 h, stable 4.5 h, respectively. **c**, Representative maximum intensity projections of smRNA FISH images of hiPSC cells treated with ActD for 4.5 h or non-treated cells (NT). Exon, white; intron, magenta; nucleus, blue, outlined with a dashed line; scale bar, 2 μm . **d**, Quantification of smRNA FISH for genes shown in **c** and Fig. 2h. $n = 30$ cells per gene, two independent stainings. Black line, mean values. Source numerical data are available in source data.



Extended Data Fig. 3 | Overlap of IR stability groups with published detained intron datasets. **a**, Stacked bar plots showing the number of introns from the indicated IR stability groups that overlap with detained introns defined by Boutz *et al.* versus introns without overlap. Bottom, fraction of shared introns in each group. **b**, Boxplots showing host RNA nuclear enrichment [$\log_2(\text{TPM}_{\text{nuc}}/\text{TPM}_{\text{cyto}})$], basal PIR, GC content, and intron length for the indicated IR stability groups, introns overlapping the Boutz detained intron set, and introns without IR (no IR).

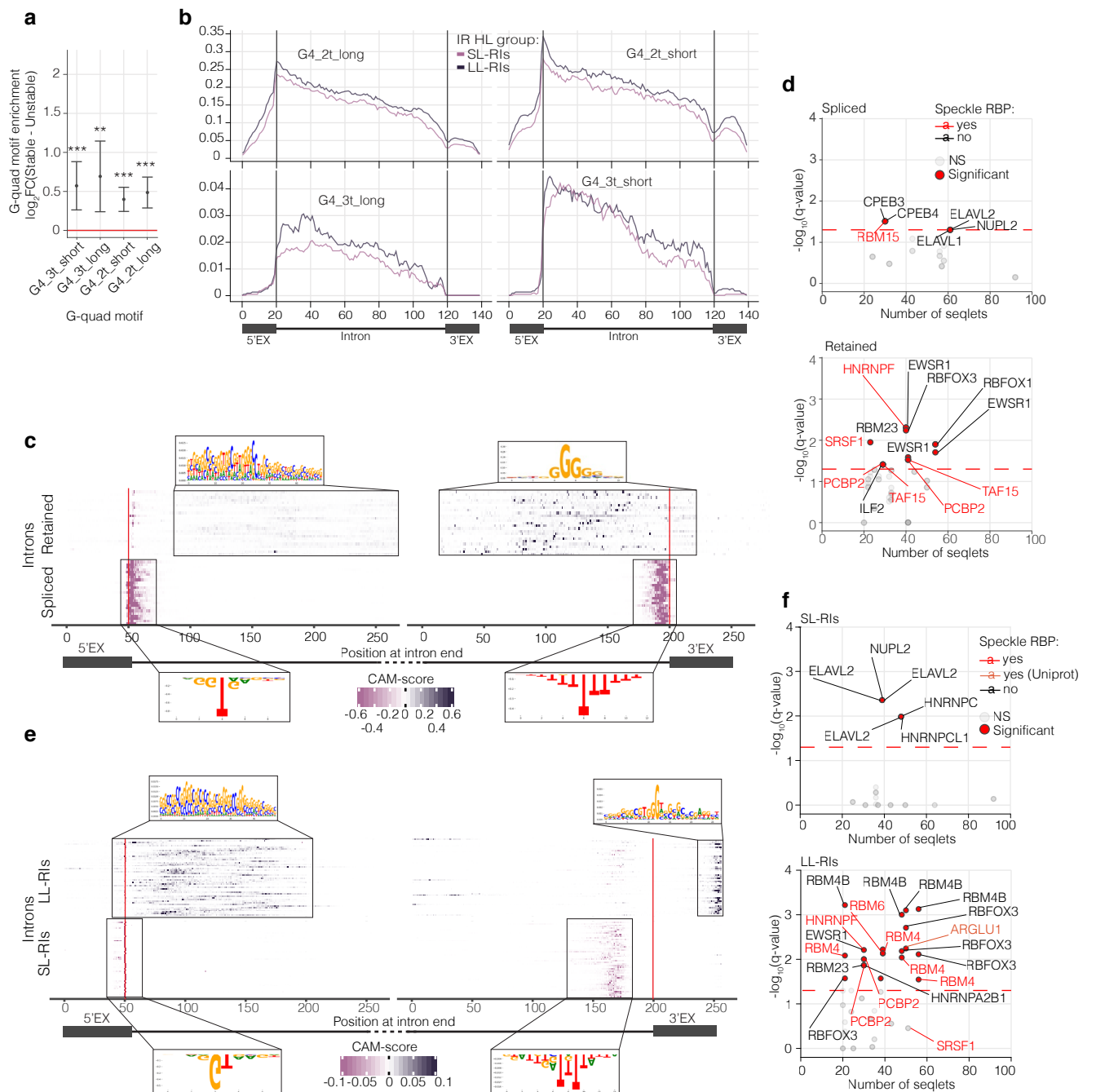
*** $P \leq 0.001$, evaluated by one-way ANOVA followed by Tukey's HSD test stable 4.5 h versus detained introns. **c**, Same as in **a**, but using the detained intron set defined by Braun *et al.* **d**, Same as in **b**, but for introns overlapping the Braun detained intron set. *** $P \leq 0.001$, evaluated by one-way ANOVA followed by Tukey's HSD test stable 4.5 h versus detained introns. For **b** and **d**, $P_{\text{adj}} < 2.2 \times 10^{-16}$ for all plotted features. Source numerical data are available in source data.



Extended Data Fig. 4 | See next page for caption.

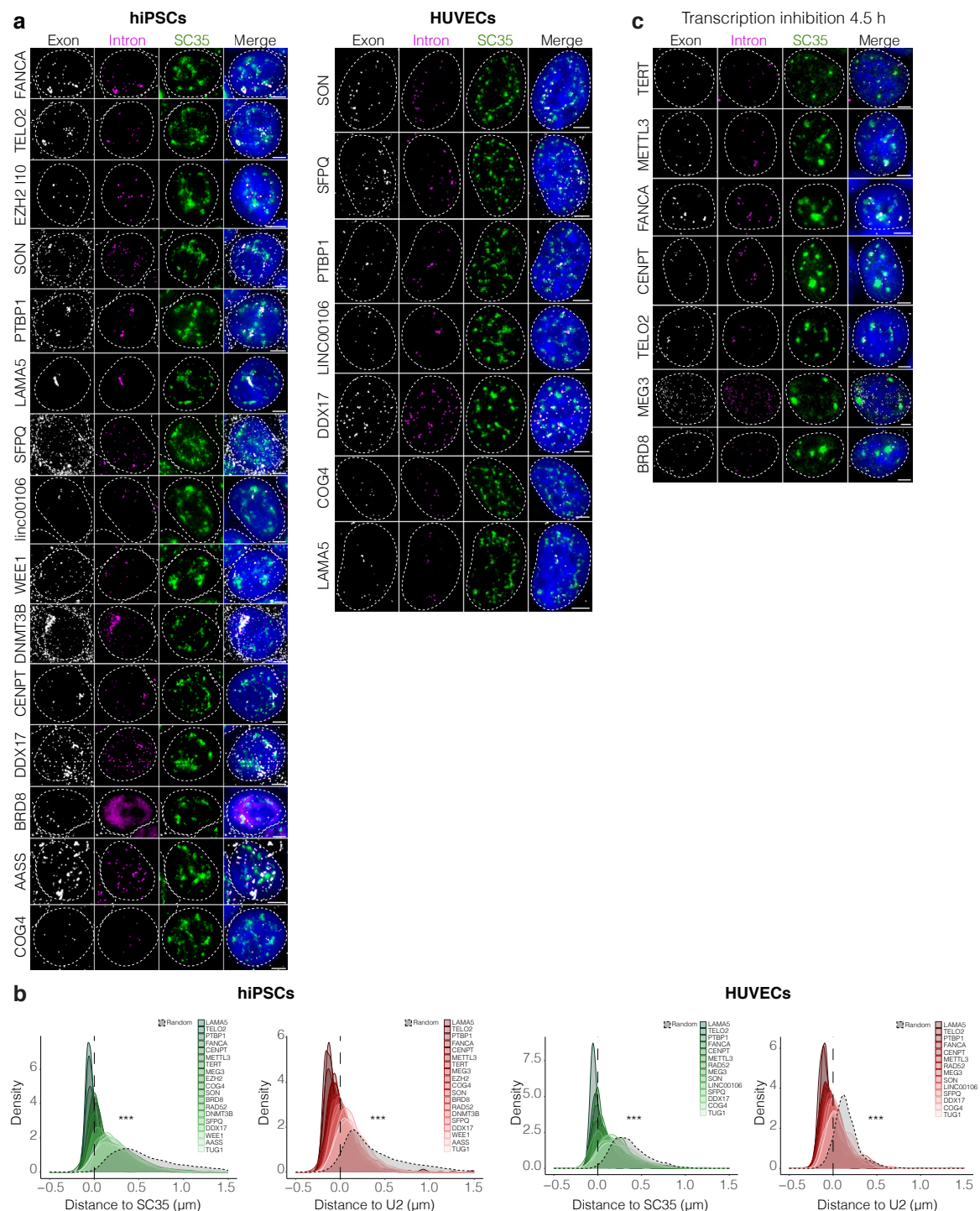
Extended Data Fig. 4 | IR-RNAs associate with nuclear speckles. **a**, Histogram showing the distribution of PIR levels. Cutoffs for retained and spliced classes are labeled on the plot. **b**, Boxplots showing the distribution of PIR levels in IR HL classes as defined in Fig. 3d. Welch's two-sided t-test ($n = 3171$ LL-RIs, $n = 3171$ SL-RIs). $P = 2.2 \times 10^{-308}$. *** $P \leq 0.001$. **c**, Boxplots showing the distributions of nuclear speckle and other RBP peaks. Welch's two-sided t-test ($n = 5000$ per group). Retained $P_{\text{adj}} = 1.13 \times 10^{-14}$; Spliced $P_{\text{adj}} = 6.37 \times 10^{-4}$; LL-RIs $P_{\text{adj}} = 9.56 \times 10^{-25}$; SL-RIs $P_{\text{adj}} = 1.13 \times 10^{-10}$. *** $P \leq 0.001$. **d**, Heatmap showing the enrichment of ENCODE eCLIP RBP peaks in IR HL and IR intron classes. RBPs associated with nuclear speckles are colored in red. **e**, UMAP reduction of Parnet intron embeddings fine-tuned on IR HL. Each point is an intron and is coloured by the average of length-normalized ENCORE speckle RBP peaks. Pearson's two-sided correlation tests between the color-coded covariate and IR $n = 79,436$, $P_{\text{adj}} = 2.2 \times 10^{-308}$; IR HL $n = 25,354$, $P_{\text{adj}} = 9.06 \times 10^{-283}$. *** $P \leq 0.001$. **f**, Correlation plot showing Pearson's R for pairs of variables on X and Y. **g–l**, Same as **e**, but colored by

ssDRIP-seq coverage (**g**; $P_{\text{adj}}(\text{IR}) = 1.85 \times 10^{-30}$); CpG methylation (**h**; $P_{\text{adj}}(\text{IR}) = 6.44 \times 10^{-166}$, $P_{\text{adj}}(\text{HL}) = 9.2 \times 10^{-102}$); intron length (**i**; $P_{\text{adj}}(\text{IR}) = 5.65 \times 10^{-211}$); phyloP (**j**; $P_{\text{adj}}(\text{IR}) = 6.94 \times 10^{-44}$, $P_{\text{adj}}(\text{HL}) = 3.49 \times 10^{-9}$); repeats (**k**; $P_{\text{adj}}(\text{IR}) = 1.03 \times 10^{-173}$, $P_{\text{adj}}(\text{HL}) = 0.0302$); nuclear enrichment (**l**; $P_{\text{adj}}(\text{IR}) = 2.2 \times 10^{-308}$, $P_{\text{adj}}(\text{HL}) = 7.03 \times 10^{-209}$). ns = not significant, * $P < 0.05$, *** $P \leq 0.001$. **m**, Projections of Parnet embeddings of the two sets of detained introns defined by Braun et al. 2017 (left) and Boutz et al. 2015 (right) onto the UMAP of the introns from this study. Contour lines show kernel density estimates for the two sets of detained introns. **n**, Boxplots showing host RNA nuclear speckle enrichment for the IR stability groups, introns overlapping the two detained intron sets, and introns without IR (no IR). Tukey's HSD $P_{\text{adj}} < 2.2 \times 10^{-16}$ for both comparisons; $n(\text{stable } 4.5 \text{ h}) = 2300$, $n(\text{Boutz et al.}) = 4952$, $n(\text{Braun et al.}) = 1972$. *** $P \leq 0.001$. For **b**, **c**, **n**, midline, median; box limits, first and third quartiles; whiskers, most extreme values within $1.5 \times \text{IQR}$ of the hinges. Source numerical data are available in source data.



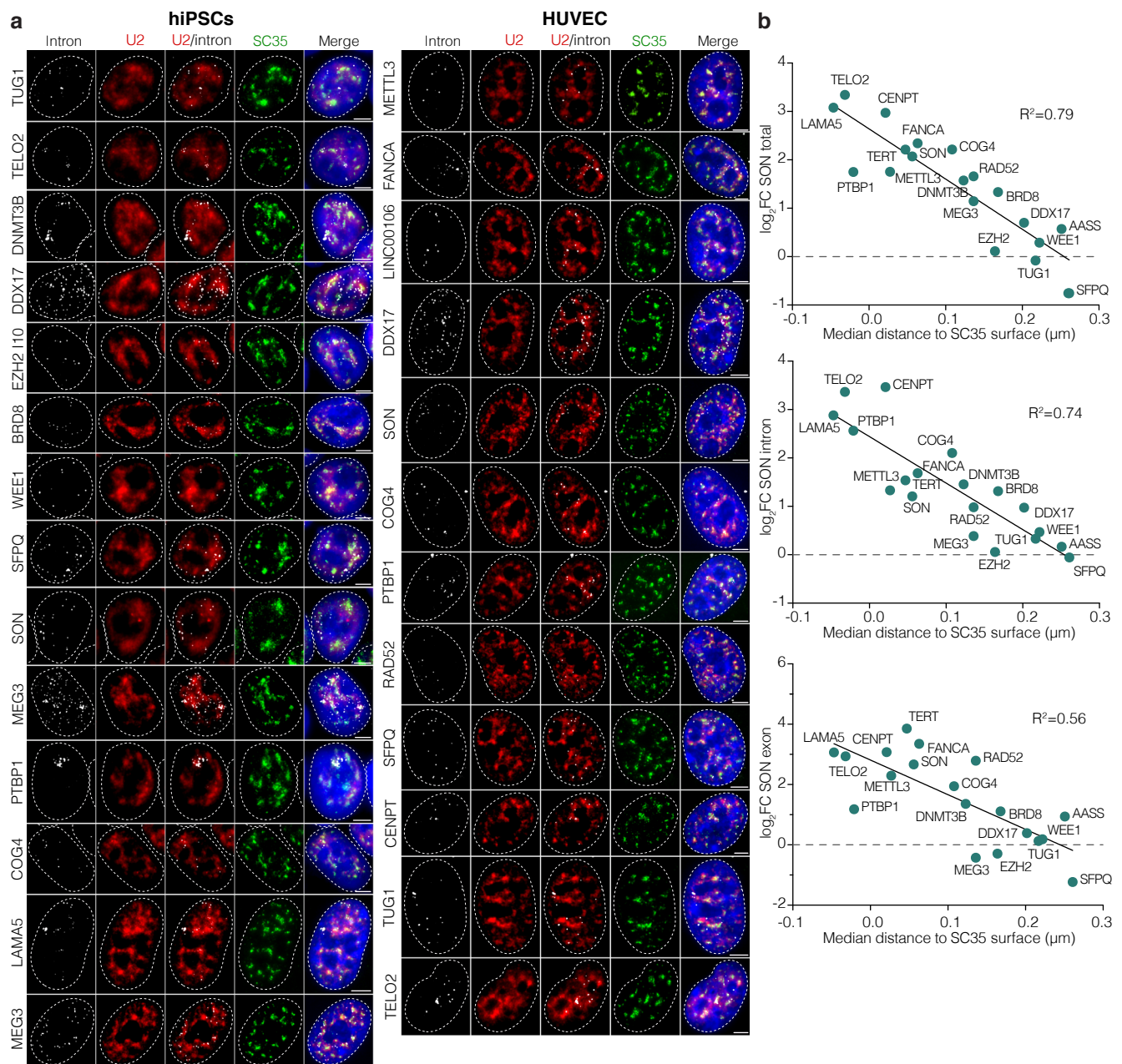
Extended Data Fig. 5 | IR-RNAs are enriched in G-quadruplexes. a, Dot plot showing the enrichment of four G-quadruplex motif counts. Points show $\log_2$ fold enrichment of motif-hit rates in LL- relative to SL-RIs ($n = 3171$ LL-RIs and SL-RIs); error bars indicate 95% confidence intervals for the $\log_2$ fold enrichment. Statistical significance was estimated from the corresponding Poisson GLM class coefficient (*Methods - G-quadruplex motif enrichment*). $G4_3t_short$, $P_{adj} = 3.53 \times 10^{-4}$; $G4_3t_long$, $P_{adj} = 2.6 \times 10^{-3}$; $G4_2t_short$, $P_{adj} = 1.46 \times 10^{-6}$; $G4_2t_long$, $P_{adj} = 2.58 \times 10^{-6}$. * $P < 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$. **b**, Meta profiles showing the distribution of four G-quadruplex motifs along the intron body and flanking exons within LL- (purple) and SL- (pink) RIs. **c**, Class Attribution Maps (CAMs) of ParNet fine-tuned on IR. Per-base class attribution scores for 256 bases on each end of the intron (X) are plotted for top 25 predicted introns in each class (Y).

Top motifs extracted by tf-modisco for each intron's end are reported above (Positive-Retained) and below (Negative-Spliced) (*Methods - ParNet fine-tuning and Class Activation Mapping (CAM) feature importance*). **d**, Results of TomTom analysis of motifs extracted from IR CAMs by tf-modisco. Each dot corresponds to a CAM-derived motif, with TomTom one-sided motif-comparison q-values (probability that a random motif would align to the query motif) shown on Y as $-\log_{10}(q\text{-value})$ and the number of motif seqlets (n) shown on X; q-values and n are reported in the *Source Data*. Up to three RBP motifs can be assigned to each input motif. Nuclear speckle RBPs are coloured in red. **e**, Same as **c**, but for ParNet fine-tuned on IR HLs. **f**, Same as **d**, but the motifs were extracted from the IR HL model CAMs (*Methods - Attribution scores interpretation*). Source numerical data are available in source data.



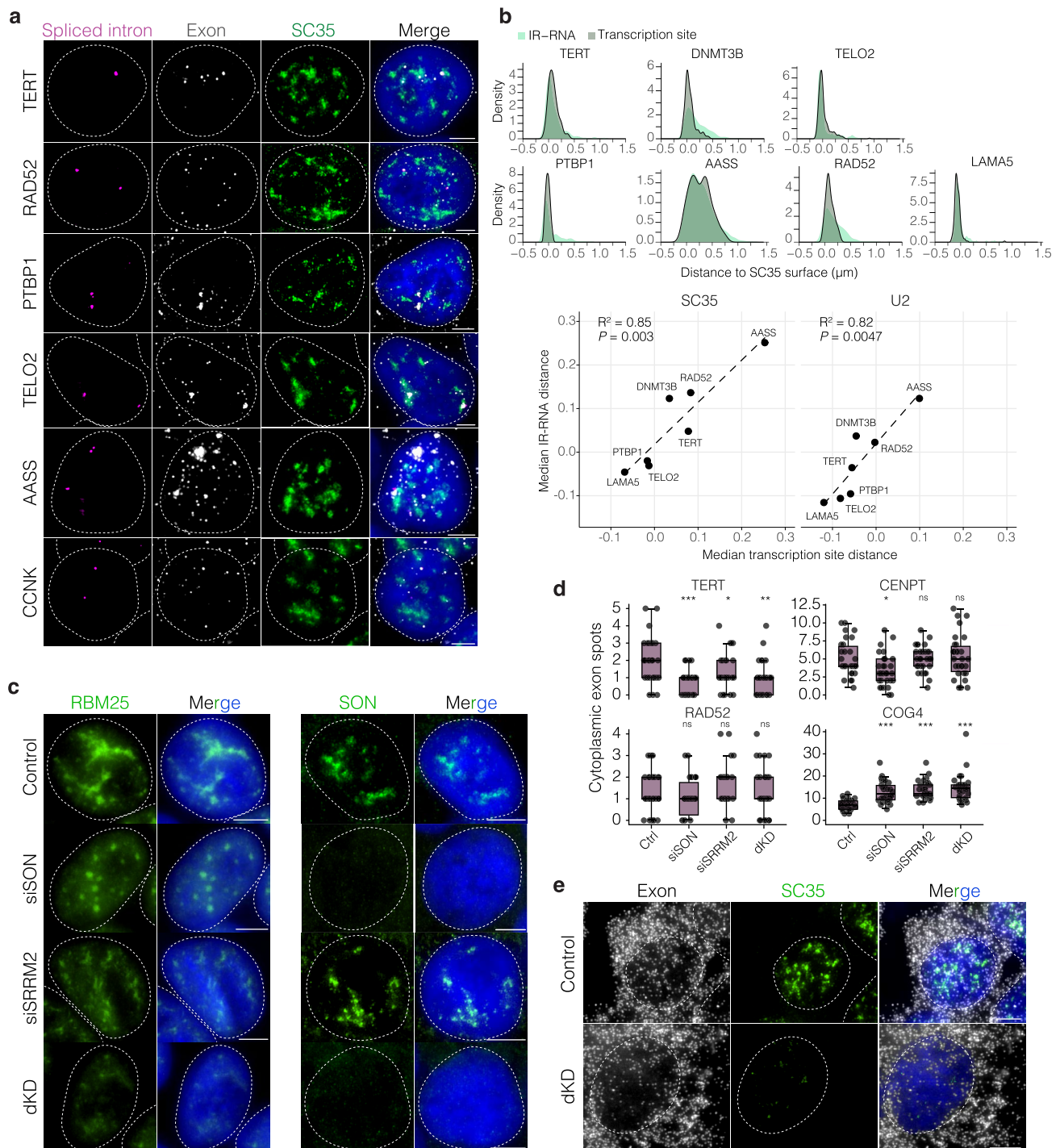
Extended Data Fig. 6 | Subcellular localization of IR-RNAs relative to nuclear speckles in hiPSCs and HUVECs. a, Maximum intensity projections of smRNA FISH and SC35 IF in hiPSCs (left) and HUVECs (right). Exon, white; intron, magenta; SC35, green; nucleus, blue, outlined with a dashed line; scale bar, 2 μm. **b**, Density plots of the distance (μm) of IR-RNA molecules and random spots to nuclear speckle surface. Distances were computed to the segmented U2 and SC35 boundary for hiPSCs (left) and HUVECs (right). *** $P \leq 0.001$, as evaluated by one-sided KS test, random spots versus individual IR-RNAs. In hiPSCs, $P_{\text{adj}} < 2.2 \times 10^{-16}$ for all RNAs except TUG1 ($P_{\text{adj}} = 1.63 \times 10^{-12}$) and COG4 ($P_{\text{adj}} = 2.62 \times 10^{-9}$); distance to U2, $P_{\text{adj}} < 2.2 \times 10^{-16}$ for all RNAs except COG4

($P_{\text{adj}} = 1.60 \times 10^{-11}$). n IR-RNA molecules: hiPSCs, TEL02 718, PTBP1 687, CENPT 1,144, SON 1,308, MEG3 2,579, DDX17 2,977, AASS 2,830, LAMA5 397, FANCA 658, BRD8 993, METTL3 445, DNMT3B 725, WEE1 977, TUG1 824, RAD52 490, EZH2 538, SFPQ 877, TERT 184, COG4 125; HUVECs, MEG3 727, PTBP1 353, LAMA5 137, SON 371, CENPT 176, TEL02 150, FANCA 176, DDX17 879, METTL3 311, LINC00106 151, SFPQ 416, RAD52 117, TUG1 234, COG4 104. Random nulls were generated from 10 nuclei per target (hiPSCs: $n = 9,267$ SC35 and 9,269 U2 random spots; HUVECs: $n = 2,445$ SC35 and 2,445 U2 random spots). **c**, Maximum intensity projections of smRNA FISH in hiPSCs treated with ActD for 4.5 h. Channels are the same as in **a**. Scale bar, 2 μm. Source numerical data are available in source data.



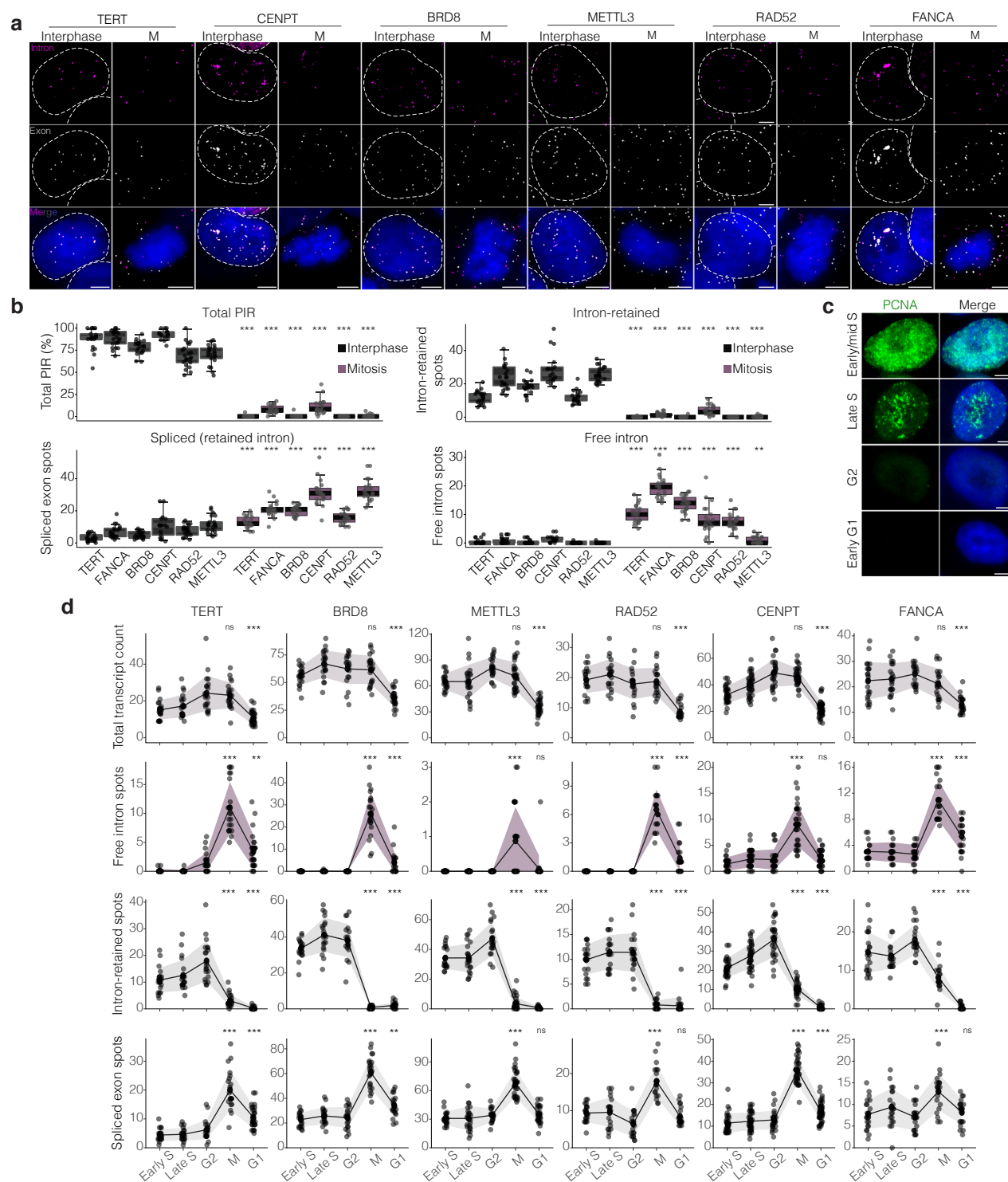
Extended Data Fig. 7 | IR-RNAs localize to the outer shell of nuclear speckles in hiPSCs and HUVECs. a, Maximum intensity projections of smRNA FISH and SC35 IF in hiPSCs. Intron white; U2, red; SC35, green; nucleus, blue, outlined with a dashed line; scale bar, 2 μm . **b,** Correlation of SON $\log_2$ fold change of the indicated RNAs (derived from Wu et al., 2024⁷⁴) and their median distances to the

SC35-positive core surface (μm). Data are shown separately for total RNA (top), intron regions (middle), and exon regions (bottom). For total RNA, Pearson's $r = -0.89$, $R^2 = 0.79$, $P_{\text{adj}} = 9.97 \times 10^{-7}$; for introns, Pearson's $r = -0.86$, $R^2 = 0.74$, $P_{\text{adj}} = 3.98 \times 10^{-6}$; for exon regions, Pearson's $r = -0.75$, $R^2 = 0.56$, $P_{\text{adj}} = 2.15 \times 10^{-4}$. $n = 19$ RNAs. Source numerical data are available in source data.



Extended Data Fig. 8 | Transcription site positioning and perturbation of nuclear speckle integrity. **a**, Maximum intensity projections of exon and intron smRNA FISH combined with SC35 IF. Spliced intron, magenta; exon, white; SC35, green; nucleus in blue, outlined with a dashed line; scale bar, 2 μm . **b**, Top, density plots of the distance (μm) of IR-RNAs (green) and their transcription sites (TSs) (white) to the SC35 surface. n IR-RNA molecules/TS foci, respectively: AASS, 2,830/145; DNMT3B, 725/108; LAMA5, 397/184; PTBP1, 687/93; RAD52, 490/44; TELO2, 718/311; TERT, 184/73. Bottom, correlation between median transcription-site distance and median IR-RNA distance to the SC35 core surface or U2 shell surface. For SC35, Pearson's $r = 0.92$, $P = 3.0 \times 10^{-3}$; for U2, Pearson's $r = 0.91$, $P = 4.7 \times 10^{-3}$. $n = 7$ genes, three independent stainings. **c**, RBM25 and SON IF (green) after knockdown of SON, SRRM2 or both (double knockdown, dKD). Nucleus, blue; outlined with a dashed line; scale bar, 2 μm . **d**, Quantification

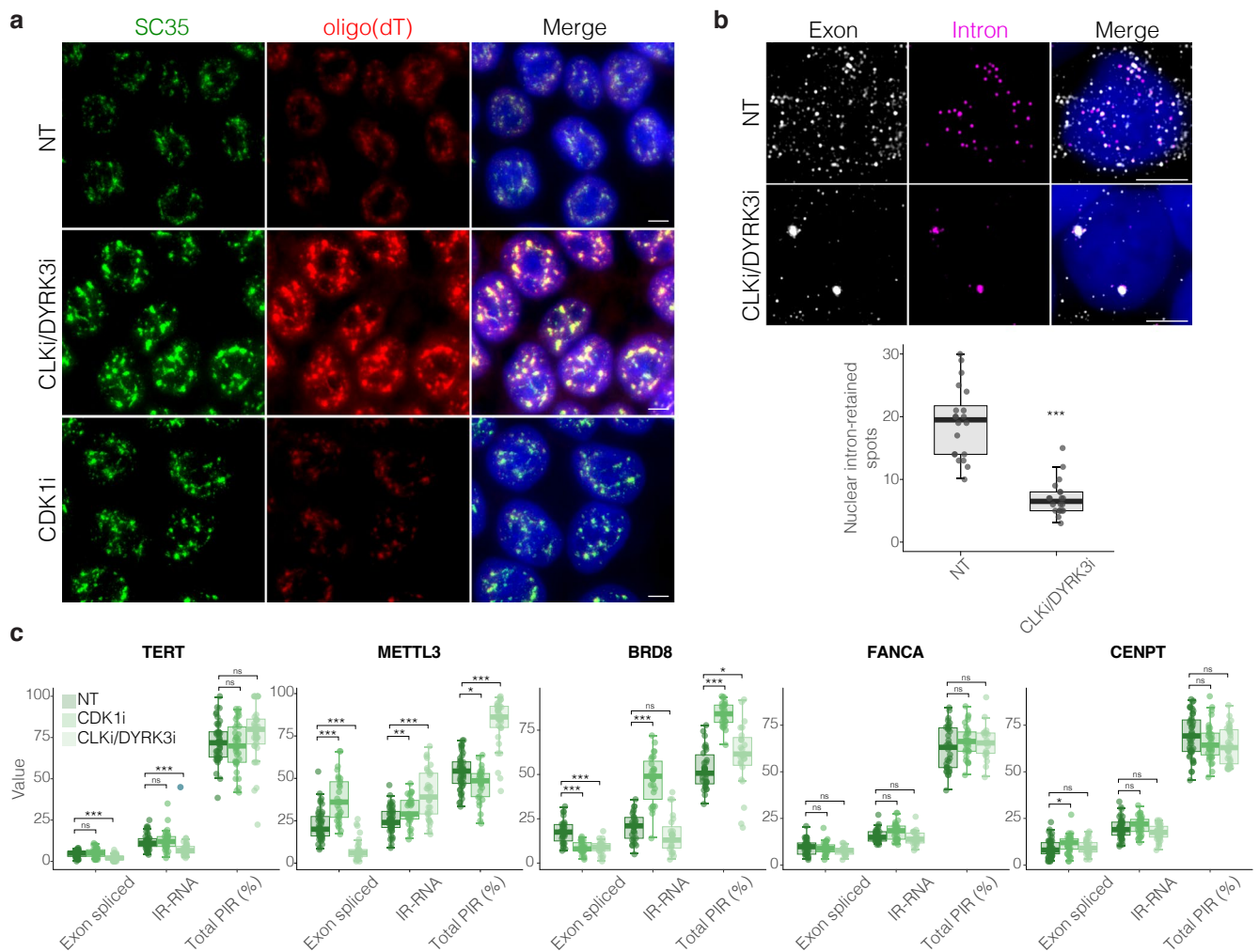
of cytoplasmic exon spots after knockdown of SON, SRRM2, or both (double knockdown, dKD). Each dot represents one cell. Midline, median; box limits, first and third quartiles; whiskers, most extreme values within $1.5 \times \text{IQR}$ of the hinges; points, individual observations. $n = 30$ cells, three independent stainings. ns = not significant, $*P < 0.05$, $**P \leq 0.01$, $***P \leq 0.001$, as evaluated by two-sided Wilcoxon rank-sum test versus control siRNA (Ctrl). Cytoplasmic exon spots: TERT, $P_{\text{adj}} = 3.38 \times 10^{-4}$, 2.23×10^{-2} , 2.39×10^{-3} for siSON, siSRRM2, dKD, respectively; CENPT, $P_{\text{adj}} = 2.63 \times 10^{-2}$ for siSON; COG4, $P_{\text{adj}} = 7.76 \times 10^{-7}$, 1.32×10^{-8} , 2.64×10^{-8} for siSON, siSRRM2, dKD, respectively. **e**, Maximum intensity projections of GAPDH exon smRNA FISH combined with SC35 IF following siRNA-mediated knockdown of SON, SRRM2, or both (double knockdown, dKD). SC35, green; exon, white; nucleus, blue, outlined with a dashed line; scale bar, 2 μm . Source numerical data are available in source data.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Splicing of LL-IR-RNAs in mitosis. a, Representative maximum intensity projections of exon/intron smRNA FISH for the indicated RNAs in hiPSCs in interphase and mitosis (M). Exon, white; intron, magenta; nucleus, blue, outlined with a dashed line; scale bar, 2 μ m. **b**, Quantification of total PIR, intron-retained RNA spots, spliced RNA spots (Δ retained intron) and free intron spots in interphase and mitosis for indicated RNAs. Each dot represents one cell ($n = 20$ cells per gene and cell-cycle stage; $n = 16$ interphase cells at G2 phase for CENPT; three independent experiments). Midline, median; box limits, first and third quartiles; whiskers, most extreme values within 1.5 $\times$ IQR of the hinges; points, individual observations. $**P \leq 0.01$, $***P \leq 0.001$, as evaluated by two-sided Wilcoxon rank-sum test, mitosis versus interphase. Total PIR, $P_{\text{adj}} = 1.25 \times 10^{-8}$, 3.72×10^{-8} , 1.25×10^{-8} , 1.72×10^{-7} , 1.25×10^{-8} , 1.87×10^{-8} for TERT, FANCA, BRD8, CENPT, RAD52, METTL3, respectively; intron-retained RNA spots, $P_{\text{adj}} = 1.23 \times 10^{-8}$, 2.95×10^{-8} , 1.23×10^{-8} , 1.61×10^{-7} , 1.23×10^{-8} , 1.83×10^{-8} , respectively; spliced exon spots, $P_{\text{adj}} = 7.69 \times 10^{-8}$, 9.94×10^{-8} , 7.69×10^{-8} , 9.22×10^{-7} , 7.26×10^{-7} , 7.69×10^{-8} , respectively; free intron spots, $P_{\text{adj}} = 3.38 \times 10^{-8}$, 3.38×10^{-8} , 3.38×10^{-8} , 1.28×10^{-5} , 3.38×10^{-8} , 3.44×10^{-3} , respectively. **c**, Representative PCNA IF patterns used to assign synchronized hiPSCs to early/mid S, late S,

G2, and early G1 phase. PCNA, green; nucleus, blue; scale bar, 2 μ m. **d**, smRNA FISH quantification across synchronized cell-cycle stages; total transcript count, free intron spots, intron-retained RNA spots and total spliced RNA spots for the indicated RNAs. Each dot represents one cell (n given in Fig. 6b, three independent experiments). Black line, mean value; shadow, standard deviation; ns = not significant, $*P < 0.05$, $**P \leq 0.01$, $***P \leq 0.001$, as evaluated by two-sided Wilcoxon rank-sum test versus G2. For gene order TERT, BRD8, METTL3, RAD52, CENPT and FANCA, total transcript count, G1: $P_{\text{adj}} = 1.40 \times 10^{-7}$, 2.95×10^{-6} , 1.98×10^{-8} , 1.80×10^{-6} , 7.20×10^{-12} , 1.51×10^{-7} , respectively; free intron spots, M: $P_{\text{adj}} = 4.83 \times 10^{-8}$, 1.14×10^{-8} , 1.23×10^{-4} , 1.51×10^{-8} , 2.75×10^{-9} , 4.83×10^{-8} , respectively; intron-retained RNA spots, M: $P_{\text{adj}} = 6.12 \times 10^{-8}$, 3.31×10^{-8} , 4.85×10^{-9} , 6.12×10^{-8} , 3.65×10^{-11} , 2.60×10^{-7} , respectively; G1: $P_{\text{adj}} = 6.95 \times 10^{-10}$, 4.11×10^{-8} , 7.98×10^{-9} , 3.01×10^{-8} , 1.03×10^{-12} , 1.83×10^{-8} , respectively; spliced exon spots, M: $P_{\text{adj}} = 2.02 \times 10^{-7}$, 5.23×10^{-8} , 6.16×10^{-9} , 2.38×10^{-7} , 4.08×10^{-11} , 7.73×10^{-5} , respectively; free intron spots, G1: $P_{\text{adj}} = 9.95 \times 10^{-3}$, 1.27×10^{-4} , 3.22×10^{-5} , 5.85×10^{-4} for TERT, BRD8, RAD52, FANCA, respectively; spliced exon spots, G1: $P_{\text{adj}} = 4.22 \times 10^{-4}$, 4.13×10^{-3} , 5.82×10^{-4} for TERT, BRD8, CENPT, respectively. Source numerical data are available in source data.



Extended Data Fig. 10 | Splicing of retained introns in mitosis is kinase-

dependent. **a**, Maximum intensity projections of oligo(dT) smRNA FISH (red), and SC35 IF (green) in non-treated cells (NT) or cells treated with CLKi/DYRK3i or CDK1i. Nucleus, blue; scale bar, 2 μ m. **b**, Maximum intensity projections of SRSF7 exon (white) and intron (magenta) smRNA FISH in non-treated cells (NT) or cells treated with CLKi/DYRK3i. Nucleus, blue; scale bar, 2 μ m. Below, quantification of intron-retained RNA spots of the indicated treatments. Each dot represents one cell. Midline, median; box limits, first and third quartiles; whiskers, most extreme values within $1.5 \times$ IQR of the hinges; points, individual observations. *** $P \leq 0.001$, as evaluated by two-sided Welch's t-test versus non-treated cells (NT), $n = 20$ cells, two independent stainings. $P = 4.47 \times 10^{-9}$. **c**, Bottom, smRNA FISH quantification of spliced exon spots (Δ retained intron), intron-retained RNA spots and total PIR in interphase. Midline, median; box limits, first and third quartiles; whiskers, most extreme values within $1.5 \times$ IQR of the hinges; points, individual

observations. ns = not significant, * $P < 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$ evaluated by two-sided Wilcoxon rank-sum test versus non-treated cells (NT). n values are given for NT, CDK1i and CLKi/DYRK3i, respectively: TERT 42, 36, 36; BRD8 30, 27, 23; METTL3 47, 29, 33; CENPT 45, 35, 34; FANCA 35, 30, 20; four independent experiments. Spliced exon spots, CDK1i, $P_{\text{adj}} = 6.11 \times 10^{-6}$, 1.51×10^{-6} , 2.45×10^{-2} for METTL3, BRD8, CENPT, respectively; CLKi/DYRK3i, $P_{\text{adj}} = 3.96 \times 10^{-5}$, 3.41×10^{-9} , 8.78×10^{-6} for TERT, METTL3, BRD8, respectively. IR-RNA spots, CDK1i, $P_{\text{adj}} = 5.41 \times 10^{-3}$, 4.44×10^{-8} for METTL3, BRD8, respectively; CLKi/DYRK3i, $P_{\text{adj}} = 1.37 \times 10^{-5}$, 4.87×10^{-6} for TERT, METTL3, respectively. Total PIR, CDK1i, $P_{\text{adj}} = 1.32 \times 10^{-2}$, 3.41×10^{-9} for METTL3, BRD8, respectively; CLKi/DYRK3i, $P_{\text{adj}} = 2.88 \times 10^{-9}$, 3.36×10^{-2} for METTL3, BRD8, respectively. n for NT, CDK1i and CLKi/DYRK3i, respectively: TERT 42, 36, 36; BRD8 30, 27, 23; METTL3 47, 29, 33; CENPT 45, 35, 34; FANCA 35, 30, 20; four independent experiments. Source numerical data are available in source data.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒ ☐ A description of all covariates tested
- ☐ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☐ ☒ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Imaging: Nikon ECLIPSE Ti2 widefield microscope with a Nikon Plan Apo $\lambda \times 100/1.45$ -numerical aperture oil objective lens and a Nikon DS-Qi2 camera. Maximum intensity projections were generated using Nikon software (NIS-Elements version 5.11.03) and deconvolved in ImageJ (version 2.14.0/1.54f) using DeconvolutionLab2 plugin (version 2.1.2). Distances between intron smRNA FISH signals and SC35 or U2 surfaces were obtained in Imaris (Bitplane, version 11.0) using the "closest exterior surface distance" function. FISH signals were first detected and segmented as spots using the Imaris "Spots" module, and then Shortest Distance to Surfaces option was applied.

STED: image stacks were acquired on a Leica Stellaris 8 STED FALCON system using a 100x HC PL APO CS2 1.4 NA oil immersion objective.

RNA-seq: Illumina Novaseq 6000/Novaseq X Plus platforms.

The trained Parnet 21M model weights were downloaded from the parnet repository (release 0.3.0, 21M parameters; <https://github.com/marsico-lab/parnet>).

ENCORE dataset: Narrow RBP eCLIP peaks were downloaded from ENCORE portal (https://www.encodeproject.org/encode-matrix/?type=Experiment&status=released&internal_tags=ENCORE)

Data analysis

Image analysis: ImageJ/Fiji version 2.14.0/1.54f. Quantification of RNA FISH signal was performed using "Big-FISH" v0.6.2 Python package or manually when the number of RNA spots was low.

Vast-tools v2.5.1 (<https://github.com/vastgroup/vast-tools>) was used to calculate PIR and TPM values.

Minimum free energy (MFE) of RNA secondary structures was calculated using the ViennaRNA version 2.7.0 Python API, following the provider's standard procedure. For each intronic sequence (including 50 bp flanks), the MFE of the predicted optimal structure was obtained and then normalized by intron length for downstream analysis.

ssDRIP-seq dataset was processed as in the original publication. The reads were trimmed with trim_galore version 0.6.10 (<https://github.com/FelixKrueger/TrimGalore>) and mapped with STAR 2.7.11b (<https://github.com/alexdobin/STAR>). The resulting bam files were deduplicated with samtools 1.23.1 (<https://github.com/samtools/samtools>). Locations of R-loops were called with macs3 software version 3.0.4 (<https://github.com/macs3-project/MACS>).

Bisulfite sequencing reads were trimmed with trim_galore 0.6.10 then mapped and deduplicated with Bismark 0.25.1 (<https://github.com/FelixKrueger/Bismark>). Bismark's command bismark_methylation_extractor was used to extract methylation calls.

Estimation of half-lives were generated by fitting a log-linear regression to the time-series data (Methods – Estimation of half-lives at intron and gene level). The code for that analysis can be found at https://github.com/marsico-lab/IR_iPSCs/blob/main/notebooks/2.hl_revision_workflow.ipynb jupyter notebook.

RBP peaks per intron were summarized and normalized by the intron length. Normalized peak counts were then used to generate heatmaps across spliced/retained and stable/unstable intron classes; to check for RBP peak enrichment in one class versus the other.

Code to fine-tune the Parnet foundation model, compute and interpret model attribution scores, along with the scripts, notebooks, and environments can be found at: https://github.com/marsico-lab/IR_iPSCs and was deposited on Zenodo at <https://zenodo.org/records/21135983>. For interrogation of attribution scores, tf-modisco lite version 2.4.0 (<https://github.com/jmschrei/tfmodisco-lite>) in a python 3.9 conda environment was used.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

ARTR-seq data was retrieved from Wu J et al., 2024. TSA-seq data was retrieved from the GEO accession number GSE81553 (Chen Y et al., 2018).

A list of previously annotated detained introns was retrieved from Boutz et al., 2015 and Braun et al., 2018.

A list of nuclear speckle-associated RBPs was obtained from RBP Image Database from the Lecuyer lab.

Enrichment of R-loops in retained versus non-retained introns was downloaded and reprocessed from the publicly available ssDRIP-seq dataset (Chen L et al., 2021). The raw sequencing data can be accessed with the following BioProject id PRJNA608890 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA608890>).

Bisulphite sequencing runs are available under the same bioproject id (SRR1118540), following the processing steps taken by the authors of the original publication.

All eCLIP narrowPeak files for each RBP were retrieved from the ENCODE portal (ENCODE project) (https://www.encodeproject.org/encode-matrix/?type=Experiment&status=released&internal_tags=ENCODE&searchTerm=ENCODE).

The trained Parnet 21M model weights were downloaded from the parnet repository (release 0.3.0, 21M parameters; <https://github.com/marsico-lab/parnet>).

Raw RNA-seq data and processed data that include PIR and TPM values obtained through vast-tools are available in GEO under accession GSE327794.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design; whether sex and/or gender was determined based on self-reporting or assigned and methods used. Provide in the source data disaggregated sex and gender data, where this information has been collected, and if consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected. Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.

Reporting on race, ethnicity, or other socially relevant groupings

Please specify the socially constructed or socially relevant categorization variable(s) used in your manuscript and explain why they were used. Please note that such variables should not be used as proxies for other socially constructed/relevant variables (for example, race or ethnicity should not be used as a proxy for socioeconomic status). Provide clear definitions of the relevant terms used, how they were provided (by the participants/respondents, the researchers, or third parties), and the method(s) used to classify people into the different categories (e.g. self-report, census or administrative data, social media data, etc.)

Please provide details about how you controlled for confounding variables in your analyses.

Population characteristics

Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."

Recruitment

Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.

Ethics oversight

Identify the organization(s) that approved the study protocol.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

No statistical methods were used to predetermine sample size. Sample sizes for smRNA FISH and imaging experiments were chosen based on standard practice in the field and prior studies using similar approaches. Sample sizes for smRNA FISH quantifications were considered sufficient when consistent trends were observed across individual cells, imaging fields and independent experiments. All experiments were performed with a minimum of three independent biological replicates, unless otherwise stated in the figure legends. smRNA FISH following ActD treatment was performed in two independent experiments and was considered sufficient because of the high concordance between replicates and consistency with the vast-tools results. Validation of SRSF7 by smRNA FISH was performed in two independent experiments, as it was a validation of previously published findings. For sequencing-based and computational analyses, sample sizes correspond to the number of introns, genes, transcripts, passing the predefined filtering criteria; no additional statistical sample-size calculation was performed. Samples were allocated to experimental groups according to the experimental design, treatment condition, time point, cell-cycle stage or computational class definition. Random allocation was not relevant for sequencing-based and computational analyses because groups were defined by predefined biological or computational criteria. Imaging fields and cells were selected randomly. Exact sample numbers for each experiment or analysis are provided in the corresponding figure legends and Source Data.

Data exclusions

No data were excluded from the analyses, except as described in the Methods (for example, quality-control filtering, RNA-seq and intron-level filtering).

Replication

Data reproducibility was verified by performing independent biological and technical measurements.

Randomization

The experiments were not randomized; samples were assigned to experimental groups based on experimental design and treatment conditions. Imaging fields/cells were selected randomly.

Blinding

All experiments in any comparisons used cells grown in same culture conditions. The investigators were not blinded because experiments were designed and performed by the same person.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Monoclonal anti SC-35 antibody (Sigma Aldrich, 1:1500); Alexa Fluor® 488 anti-Mouse IgG (Abcam, 1:1000); Goat anti-Rabbit IgG Alexa Fluor™ 488 (Invitrogen, 1:1000); STAR 460L anti-Mouse IgG (Abberior, 1:1000); RBM25 Polyclonal antibody (Proteintech, 1:300); Anti-PCNA Antibody (Atlas Antibodies, 1:200); SON Recombinant monoclonal antibody (Proteintech, 1:300)
Validation	Monoclonal anti SC-35 antibody (Sigma Aldrich, #s4045, Clone SC-35, ascites fluid): The manufacturer validated the antibody for ELISA, EM, ICC, IF, IP use. The antibody was purchased new and used per manufacturer's instructions in 1:1500 dilution. No further validation was performed. https://www.sigmaaldrich.com/DE/en/product/sigma/s4045 Alexa Fluor® 488 anti-Mouse IgG (Abcam, #ab150113): The manufacturer validated the antibody for ICC/IF, IHC, flow cytometry and ELISA use. The antibody was purchased new and used per manufacturer's instructions in 1:1000 dilution. Further validation was performed by including a condition without the primary antibody. https://www.abcam.com/en-us/products/secondary-antibodies/goat-mouse-igg-h-l-alex-fluor-488-ab150113 Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 488 (Invitrogen, #A-11008): The manufacturer validated the antibody for ICC/IF, IHC, flow cytometry. The antibody was purchased new and used per manufacturer's instructions in 1:1000 dilution. Further validation was performed by including a condition without the primary antibody. https://www.thermofisher.com/antibody/product/Goat-anti-Rabbit-IgG-H-L-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A-11008 STAR 460L anti-Mouse IgG (Abberior, #ST460L-1001-500UG): The manufacturer validated the antibody for STED use. The antibody was purchased new and used per manufacturer's instructions in 1:1000 dilution. No further validation was performed. https://abberior.shop/abberior-STAR-460L-goat-anti-mouse-igg-500-ll-1-mg-ml RBM25 Polyclonal antibody (Proteintech, #25297-1-AP): The manufacturer validated the antibody for WB, IHC, IF/ICC, IP, ELISA. The antibody was purchased new and used per manufacturer's instructions in 1:300 dilution. No further validation was performed. https://www.ptglab.com/products/RBM25-Antibody-25297-1-AP.htm PCNA Antibody (Atlas Antibodies, #HPA030522): The manufacturer validated the antibody for WB, IHC, IF/ICC. The antibody was purchased new and used per manufacturer's instructions in 1:200 dilution. No further validation was performed. https://www.atlasantibodies.com/products/primary-antibodies/triple-a-polyclonals/anti-pcna-antibody-hpa030522/ SON Recombinant monoclonal antibody (Proteintech, #83787-5-RR, Clone 240845B7): The manufacturer validated the antibody for WB, IHC, IF/ICC, flow cytometry and ELISA. The antibody was purchased new and used per manufacturer's instructions 1:300 dilution. No further validation was performed. https://www.ptglab.com/products/SON-Antibody-83787-5-RR.htm

Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)	Human iPSC WTC-11 (GM25256, male, Coriell Institute); HUVECs (c-12203, pool female/male, PromoCell)
Authentication	All cell lines were obtained from repositories with the corresponding certificate of analysis and used without further authentication.
Mycoplasma contamination	All cell lines were tested negative for mycoplasma.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified lines were used.

Plants

Seed stocks	<i>Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.</i>
Novel plant genotypes	<i>Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.</i>
Authentication	<i>Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.</i>