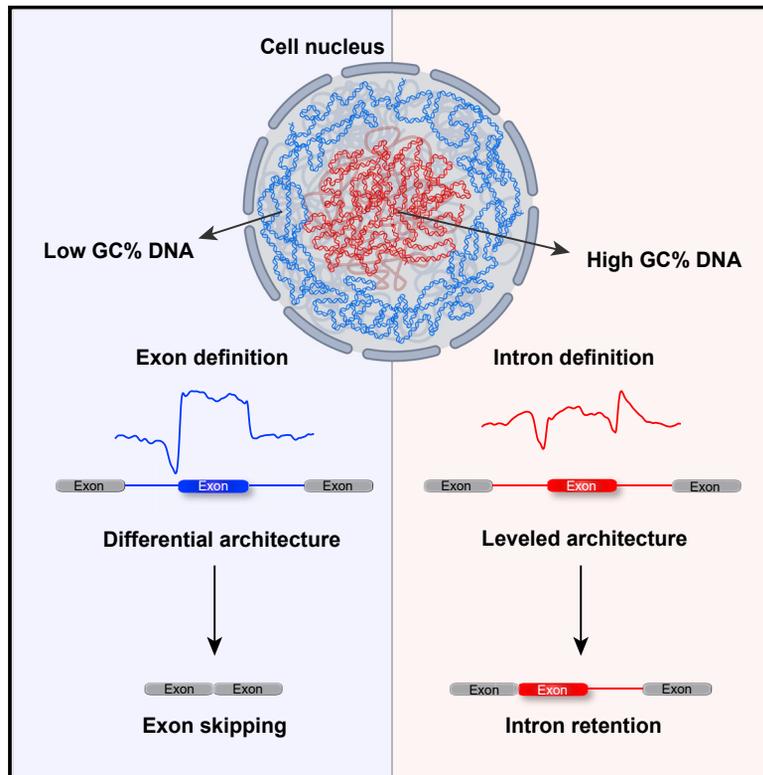


Gene architecture directs splicing outcome in separate nuclear spatial regions

Graphical abstract



Authors

Luna Tammer, Ofir Hameiri, Ifat Keydar, ..., Yaron Shav-Tal, Maria Carmo-Fonseca, Gil Ast

Correspondence

gilast@tauex.tau.ac.il

In brief

Tammer et al. demonstrate that chromatin 3D organization and splicing regulation are coupled. Two main gene architectures occupy different subnuclear regions which are related to two splicing unit recognition modes, exon and intron definition, and to two distinct protein subnetworks. Shifts in the 3D genomic locations lead to changes in splicing.

Highlights

- Two gene architectures and splicing modes colocalize in separate subnuclear regions
- Features related to the two gene architectures regulate exon or intron definition
- Changes in the 3D location of genes affect alternative splicing outcome
- Two protein subnetworks regulate the splicing of genes in distinct subnuclear regions

Article

Gene architecture directs splicing outcome in separate nuclear spatial regions

Luna Tammer,^{1,7} Ofir Hameiri,^{1,7} Ifat Keydar,^{1,7} Vanessa Rachel Roy,¹ Asaf Ashkenazy-Titelman,² Noélia Custódio,³ Itay Sason,⁴ Ronna Shayevitch,¹ Victoria Rodríguez-Vaello,⁵ José Rino,³ Galit Lev Maor,¹ Yodfat Leader,¹ Doha Khair,¹ Erez Lieberman Aiden,⁶ Ran Elkon,¹ Manuel Irimia,⁵ Roded Sharan,⁴ Yaron Shav-Tal,² Maria Carmo-Fonseca,³ and Gil Ast^{1,8,*}

¹Department of Human Molecular Genetics and Biochemistry, Sackler Faculty of Medicine, Tel-Aviv University, Tel Aviv 69978, Israel

²The Mina & Everard Goodman Faculty of Life Sciences and the Institute of Nanotechnology and Advanced Materials, Bar-Ilan University, Ramat Gan 5290002, Israel

³Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, 1649-028 Lisboa, Portugal

⁴Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel

⁵Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona, Spain. Universitat Pompeu Fabra (UPF), Barcelona, Spain, ICREA, Barcelona, Spain

⁶The Center for Genome Architecture, Baylor College of Medicine, Houston, TX 77030, USA

⁷These authors contributed equally

⁸Lead contact

*Correspondence: gilast@tauex.tau.ac.il

<https://doi.org/10.1016/j.molcel.2022.02.001>

SUMMARY

How the splicing machinery defines exons or introns as the spliced unit has remained a puzzle for 30 years. Here, we demonstrate that peripheral and central regions of the nucleus harbor genes with two distinct exon-intron GC content architectures that differ in the splicing outcome. Genes with low GC content exons, flanked by long introns with lower GC content, are localized in the periphery, and the exons are defined as the spliced unit. Alternative splicing of these genes results in exon skipping. In contrast, the nuclear center contains genes with a high GC content in the exons and short flanking introns. Most splicing of these genes occurs via intron definition, and aberrant splicing leads to intron retention. We demonstrate that the nuclear periphery and center generate different environments for the regulation of alternative splicing and that two sets of splicing factors form discrete regulatory subnetworks for the two gene architectures. Our study connects 3D genome organization and splicing, thus demonstrating that exon and intron definition modes of splicing occur in different nuclear regions.

INTRODUCTION

Discrimination of exonic from intronic sequences during pre-mRNA splicing requires a complex interplay between genomic, epigenomic, and transcriptomic features. This recognition process is carried out via exon, or intron, definition mechanisms, in which the initial step of the spliceosomal assembly occurs across exons or introns, respectively (Will and Lührmann, 2011). These two mechanisms are thought to impose length constraints on the recognized unit to limit the distance between the splice sites and thus facilitate their pairing (Fox-Walsh et al., 2005; Gelfman et al., 2013; Sterner et al., 1996). Exon and intron definition mechanisms are widely acknowledged to generate distinct alternative splicing (AS) phenotypes. When the cross-exon complex is inhibited, exon skipping (ES) occurs, whereas disruption of the cross-intron complex leads to intron retention (IR) events (Berget, 1995; De Conti et al., 2013; Li et al., 2019; Wahl et al., 2009). Recent studies have demonstrated that

following the initial splicing unit recognition step, subsequent spliceosomal assembly proceeds in a unified manner, as the cross-exon interactions during the exon definition mode can later be converted to a cross-intron complex (Li et al., 2019; Schneider et al., 2010; Sharma et al., 2008). However, the method by which the splicing machinery initially selects either an exon or intron as the recognized unit is poorly understood.

We have previously demonstrated that with the appearance of homeothermic organisms, high GC content regions underwent a GC content elevation that reshaped exon-intron architectures. This generated two exon-intron architectures from the ancestral architecture, which consisted of low GC content exons flanked by short introns with even lower GC content (referred to as the exon-intron GC differential). Within the GC-poor genomic areas, the exon-intron GC differential was maintained, and introns were elongated (termed the differential architecture). However, in regions with elevated GC content, the exons are flanked by introns that remained short but the GC differential was abolished,

creating the second exon-intron architecture (termed the leveled architecture) (Amit et al., 2012).

GC content has been shown to be associated with chromatin organization and functionality (Eyre-Walker and Hurst, 2001). Notably, chromosome size and GC content yield the highest accuracy in predicting the radial location of individual chromosomes (Girelli et al., 2020). In addition, AT-rich genomic sequences have been found to be located in proximity to the nuclear lamina, structural components of the nuclear envelope (van Steensel and Belmont, 2017). On the other hand, a study that used tyramide signal amplification (TSA) to examine the spatial organization of chromatin (Chen et al., 2018), reported that GC-rich genomic regions are located near nuclear speckles, which are found scattered within the nucleoplasm but are excluded from the nuclear envelope, and are known to harbor splicing factors (SFs) (Spector and Lamond, 2011). These observations suggest that the genome is arranged on a speckle-to-lamina axis that generates a higher-order structure of chromatin in the nucleus (reviewed by Crosetto and Bienko, 2020).

Recent advances in genome-wide chromosome conformation capture technologies have made it possible to study the genomic organization in 3D. For example, the Hi-C method enables the identification of specific genomic loci that interact with each other, making it possible to detect DNA loops and topologically associating domains (TADs), which are regions of the genome that show elevated interaction frequencies (Rao et al., 2014). Comparative analysis across different cell types suggested that although the chromatin organization is largely conserved across different cell types, some changes do exist and are associated with cell-type specific gene expression profiles (Cheng et al., 2020; Rao et al., 2014). Taken together, Hi-C can be used to study the structural and functional components of the 3D genome architecture.

Here, we combined transcriptomic analyses with a 3D study of genome organization to examine the connections between exon and intron definition, the differential or leveled architectures, and the nuclear spatial location. Chromosome conformation capture methods allowed us to identify a periphery-to-center transition in exon-intron GC content architectures. Splicing of pre-mRNAs with differential architecture occurs via exon definition, whereas most splicing in the leveled architecture is carried out via intron definition. The intron definition mode requires exons and introns to be constrained in length. We also demonstrate that the nuclear center and periphery represent two AS regulatory regions, in which two different subnetworks of SFs are associated with pre-mRNAs of the leveled and differential architectures, respectively. Thus, exon and intron definition modes of splicing occur in different nuclear regions.

RESULTS

The differential and leveled exon-intron architectures co-localize with distinct alternative splicing outcomes in different regions of the nucleus

We examined the nuclear spatial distribution of GC content in human cells in order to detect any association between the splicing mechanism and the chromatin spatial organization. For this purpose, we used Chrom3D (Paulsen et al., 2017), a 3D genome

modeling tool that integrates Hi-C data and ChIP-seq data for lamin A/C, to map TADs throughout the nuclear space. We performed ChIP-seq for lamin A/C in GM12878 and K562 cells and used publicly available ChIP-seq data for HeLa cells (Paulsen et al., 2017). We integrated the results using previously published *in situ* Hi-C datasets (Rao et al., 2014) to prepare whole-genome 3D models depicting the GC content distribution of the three cell types (Figures 1A and S1A). Our results demonstrate a transition in the GC content from periphery-to-center, with TADs having low GC in the periphery and those of higher GC in the center of the nucleus. To confirm these results, we used data obtained from three additional chromatin conformation capture methods, namely, GPSeq (Girelli et al., 2020), TSA-seq (Chen et al., 2018), and Dip-C (Tan et al., 2018), and performed Pearson correlation tests to assess the association between the GC content and the spatial location of: genome-wide segments, genes, introns, and exons. Our results indicate high inverse correlation between the distance from nuclear center and GC content, in agreement with (Girelli et al., 2020), in all segments tested, regardless of the method or cell type employed, meaning that the greater the distance from the nuclear center, the lower the GC content (Figure S1B and Table S1). The highest inverse correlation between GC content and spatial localization of genes, introns, and exons are shown for the speckles-to-lamins axis, using TSA-seq ($r = -0.65$, $r = -0.61$, and $r = -0.45$, respectively) (Table S1).

We have previously described two GC content-based exon-intron architectures that we termed differential and leveled (Amit et al., 2012). To examine whether these architectures occupy different regions of the nucleus, we divided the Euclidian distances from the nuclear center of TADs into five equal intervals, referred to as concentric radial scopes 1-to-5 (Figure 1A), and plotted their mean GC content and of exon-intron structures in each concentric radial scope (Figures 1B and 1C, respectively). The results reveal a periphery-to-center transition from differential to leveled architecture, with an overall increase in GC content (Figures 1C and S1C). The decreased GC content observed in the exon-intron junctions is likely correspond to splice site sequences that were not completely removed, as, for instance, the position and length of the polypyrimidine tract can vary (Akerman and Mandel-Gutfreund, 2007; Schwartz et al., 2008). We also observed a periphery-to-center decrease in intron length (Figures 1D and S1D), while exon length remains relatively constant (Figure S1E). Taken together, these results indicate that the differential architecture is predominantly found in the nuclear periphery, whereas the leveled architecture resides primarily in the nuclear center.

To assess whether the spatial location of the exon-intron GC content architectures is associated with a specific splicing mechanism, we studied the spatial distribution of ES and IR events, which are considered the most probable outcomes of the exon and intron definition modes, respectively (De Conti et al., 2013; Fox-Walsh et al., 2005; Li et al., 2019; Penalva et al., 2001). Considering the high prevalence of co-transcriptional splicing (Herzel et al., 2017; Kornblihtt et al., 2013), we assumed that the splicing reaction occurs while the pre-mRNA is close to the DNA. Analysis of previously annotated AS events obtained from the Vertebrate Alternative Splicing and Transcription Data Base (VastDB) (Tapial et al., 2017) revealed that IR

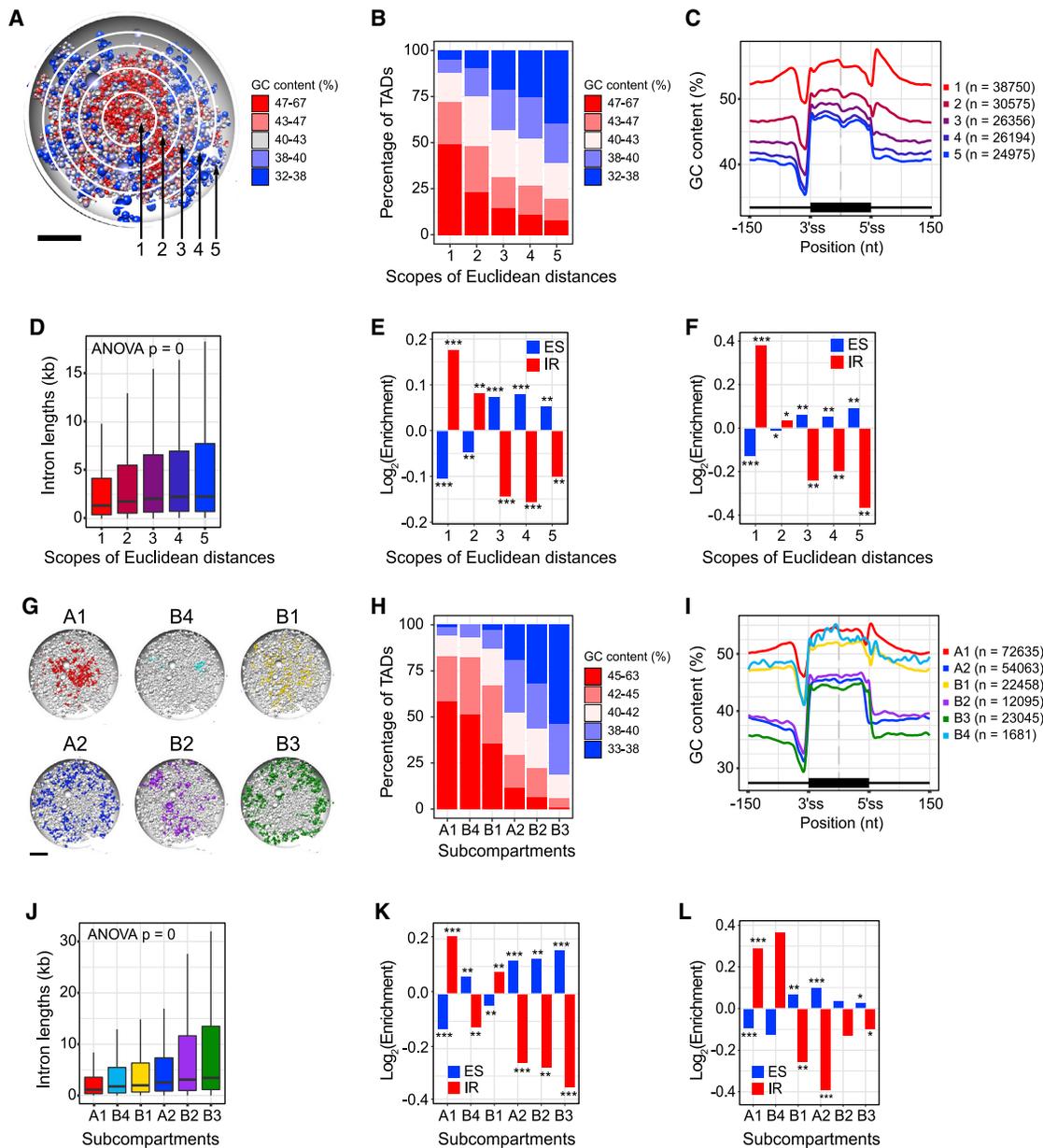


Figure 1. Peripheral and central regions of the nucleus are characterized by different exon-intron gene architectures and alternative splicing outcomes

(A) A nuclear 3D model constructed using Chrom3D depicting the mean GC content (%) of topologically associated domains (TADs, shown as spheres) in GM12878 cells (Paulsen et al., 2017). Black bar represents a scale of 2 μ m. Five concentric radial scopes of Euclidian distance are marked by white circles. Numbers 1-to-5 mark center-to-periphery, respectively.

(B) Stacked bar plot depicting the distribution of TADs according to their mean GC content (%) per scope.

(C-F) Features analyzed with respect to the five concentric radial scopes described in (A): (C) Mean GC content (%) in a sliding window of 75 nt taken from both sides of the exons and the adjacent 150 nt of their flanking introns, excluding splice site regions (20 nt from the 3' end of the upstream introns, the first 6 nt at the 5' end of the downstream intron, and the first 2 and the last 3 nt of the exonic sequences). (D) Intron lengths distribution. P values correspond to one-way ANOVA test. (E) Fold enrichment of exon skipping (ES, $n = 33883$) and intron retention (IR, $n = 18496$) events obtained from VastDB. (F) Fold enrichment of cancer-related 5' SS mutations that lead to ES ($n = 4290$) and IR ($n = 1225$) (Jayasinghe et al., 2018). (E) and (F) hypergeometric tests were applied. * ($p \leq 0.05$), ** ($p \leq 0.01$), *** ($p \leq 1e-10$).

(G) Chrom3D nuclear models of TADs showing the subcompartment annotations in GM12878 cells (Rao et al., 2014). Black bar represents a scale of 2 μ m.

(H-L) Features analyzed with respect to subcompartment as in (B)-(F), respectively.

events are enriched in genomic regions mapped closer to the nuclear center, whereas ES events are enriched in peripheral regions (Figures 1E and S1F). We also analyzed the enrichment patterns of 5' splice site (5'SS) mutations that were shown to cause either ES or IR events (Jayasinghe et al., 2018) and found that 5'SS mutations that lead to ES are more commonly associated with genes located closer to the nuclear periphery, whereas 5'SS mutations leading to IR events are located mostly in the nuclear center (Figures 1F and S1G). Together, these results imply that exon definition is associated with the differential architecture and mainly occurs in the nuclear periphery, while intron definition is associated with the leveled architecture and mainly occurs in the nuclear center.

To further support our results, we analyzed the above features according to the division of the chromatin into six subcompartments (Rao et al., 2014). The Chrom3D models depicting the nuclear spatial locations of the subcompartments and the mean GC content of TADs per subcompartment are shown in Figures 1G and 1H, respectively. Peripherally located subcompartments A2, B2, and B3 exhibit patterns of the differential architecture, whereas subcompartments A1, B4, and B1, which are located closer to the nuclear center, show patterns of the leveled architecture (Figures 1I and 1J), similar to our findings in Figures 1C and 1D. We also used the AS events from VastDB and the 5'SS mutations described above to study their enrichment patterns in each subcompartment. The enrichment patterns are mostly consistent using the two datasets: peripherally located subcompartments A2 and B3 show higher frequency of ES events, and centrally located subcompartment A1 is enriched for IR events. Interestingly, subcompartment B1 shows reversed patterns: VastDB annotations show an enrichment of ES events, but more IR events using the 5'SS mutations (Figures 1K and 1L). This could be due to the intermediate localization of subcompartment B1 between scopes 2 and 3, where the patterns are switched from IR to ES, respectively (Figures 1E and 1F). Subcompartment B4 also shows inconsistency, but the enrichment patterns obtained using the 5'SS mutations are not significant, which is plausible due to the relatively small size of this subcompartment.

We used RNA fluorescence *in situ* hybridization (RNA FISH) (Brody et al., 2011; Custódio et al., 1999; Darzacq et al., 2007; Raj et al., 2008) to validate our predicted locations of the two exon-intron architectures. The results indicate that the active genes from the differential architecture are located in the periphery of the nucleus, whereas those from the leveled are present in the nuclear center (Figures S1H–S1J).

The two exon-intron GC content architectures are associated with either exon definition or intron definition

To examine the interplay between features of the differential and leveled architectures that guide the splicing machinery to define exons or introns, we first cloned four human minigenes from the two architectures, where each is composed of three exons separated by two introns (Figures 2A and S2A). As defined in Gelfman and Ast, (2013), the differential architecture minigenes (*FRG1*, *SR140*) are characterized by relatively long introns and an overall low GC content, with a higher GC content in the middle exons

than the flanking introns. In contrast, the leveled architecture minigenes (*FES*, *CORO1B*) are characterized by relatively short introns and an overall high GC content, which remains constant across the middle exons and the flanking introns. Interestingly, we observed that the differential minigenes display the same splicing pattern as the corresponding endogenous genes when expressed as episomal plasmids or when integrated into the genome (transient and stable transfection, respectively) (Figures S2B–S2C). The leveled minigenes, on the other hand, are not spliced unless stably integrated into the genome (Figures S2D–S2F). This implies that the leveled architecture, unlike the differential, rely on the genomic environment for proper splicing, which could exclude the possibility that the two architectures are spliced by the same mechanism.

To examine the mechanism by which the spliced unit is selected in each group of minigenes, we mutated the 5'SS of the middle exons to generate weaker or inactive sites (see Methods) and observed the splicing outcome by RT-PCR. Minigenes were stably integrated into the same integration site in human embryonic kidney cells (HEK293) and transcribed by the same promoter. The results indicate that weakening or abolishing the 5'SS of the middle exon of the differential minigenes results in ES (Figures 2B and S2G), suggesting that exons are the defined unit for splicing in this group. In contrast, mutating the 5'SS of the leveled minigenes results predominantly in IR of the downstream intron with a low level of ES or alternative 3' splice site (3'SS) (Figures 2C and S2H). This implies that the splicing machinery in the leveled architecture operates mainly via intron definition, although exons can also be identified to some extent. These results also imply that exon-intron GC content architecture has a strong impact on the selection mechanism of the spliced units. The differential architecture is selected by exon definition and the leveled architecture mainly by intron definition.

Until now, intron length has been considered to be the major discriminator between exon and intron definition modes of splicing (De Conti et al., 2013; Fox-Walsh et al., 2005). Therefore, we examined the influence of intron and exon lengths on splicing in the two gene architectures. Changes in intron or exon length were made as far away as possible from the splice sites while maintaining a similar GC content in the introns or exons. We also tried not to disturb known binding sites of splicing regulatory proteins (see Methods). The results show that shortening the long introns of the differential minigenes does not affect the splicing pattern, and intron shortening coupled with mutations at the 5'SS preserves the ES mode (Figures 2D and S2I). This means that exons in the differential architecture are spliced predominantly via exon definition, regardless of intron length. In contrast, lengthening the short downstream intron of the leveled minigenes leads mainly to IR and to a low level of either ES or alternative 3'SS isoforms (Figures 2E, S2J, and S2O). These sets of experiments on the two groups of minigenes demonstrate that intron length influences splicing recognition in the leveled architecture but is not a significant factor in the differential architecture.

Stepwise lengthening of the middle exon in the differential minigenes from 115 nt to over 400 nt leads to a gradual shift from full exon inclusion to ES, or to the generation of alternative 5'SS isoforms (Figures 2F, S2K, and S2P). Full ES occurs when

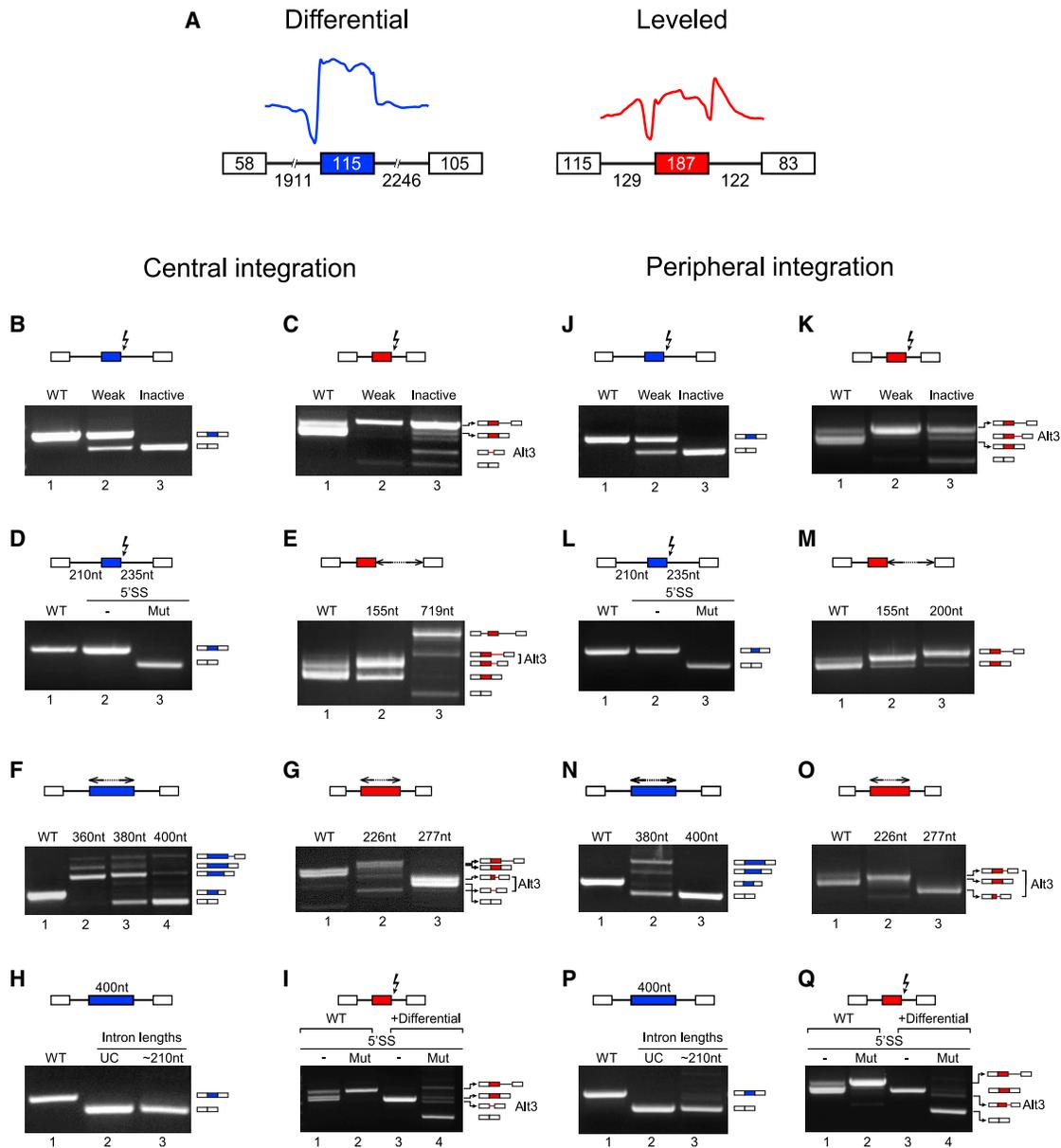


Figure 2. Exon-intron GC content architecture directs exon definition and intron definition

(A) Minigenes composed of three exons and two introns cloned from human genomic regions with differential architecture (*FRG1*) and leveled architecture (*FES*) were stably integrated into the same genomic location in HEK293 cells (FRT-site). For both minigenes, exon-intron GC content profiles are shown in blue or red lines (middle exon and the adjacent 50 nt of the flanking introns, excluding splice site sequences). Boxes indicate exons, lines represent introns; lengths of exons and introns of the wild-type (WT) minigenes are as indicated in nucleotides. Minigenes were inserted into both nuclear center and periphery integration sites (*left* and *right* panels, respectively).

(B and C) The splicing patterns of the differential (B) and the leveled (C) minigenes with the WT 5'SS or the mutated 5'SS sequence of the middle exon (weak or inactive 5'SS as indicated), see Methods for the exact mutations performed.

(D) The splicing patterns of the differential minigene before (lane 1) and after shortening the upstream intron to 210 nt and the downstream intron to 235 nt (lanes 2) and after inserting a 5'SS inactive mutation to the minigene with the shortened introns (lane 3).

(E) The splicing patterns of the leveled minigene before and after lengthening the second intron as indicated.

(F and G) The splicing patterns of differential (F) and leveled (G) minigenes before and after lengthening of the middle exon as indicated.

(H) The splicing patterns of the differential minigene before (lane 1) and after lengthening the middle exon to 400 nt (lanes 2 and 3) and after shortening both introns to indicated lengths in (D) (lane 3). Intron length in the second lane was unchanged (indicated as UC).

(I) The splicing patterns of the leveled minigenes before (lanes 1 and 2) and after replacement of a DNA segment with lower GC content located downstream to the 5'SS of the second intron (lanes 3 and 4), and with or without a 5'SS weak mutation (minus (–) represents the WT 5'SS and Mut represents a weak 5'SS mutation). For each experiment, total RNA was extracted, desired regions were amplified by RT-PCR, and products were separated on 1.5% agarose gel. The RNA products were sequenced and are shown schematically on the right side of each gel image.

(J–Q) Same as (B)–(I) except minigenes were integrated into the peripheral integration site.

the exon is longer than 400 nt, meaning the splicing machinery continues to recognize the exon until it is almost 3-fold the length. In contrast, lengthening the middle exon in the leveled minigenes by a mere 40 nt leads to mainly ES combined with activated alternative 3'SS (Figures 2G and S2L), or mainly IR product, when a different DNA segment was used to lengthen the exon (Figure S2Q). The production of ES isoform suggests that the intron definition mechanism may require the definition of the intron as well as the neighboring exons, as suggested previously (Enculescu et al., 2020). Overall, these findings may indicate the presence of cross-intron as well as cross-exon complexes in the leveled architecture, but only cross-exon complexes are active in the differential architecture.

In light of our findings that the differential architecture exclusively employs the exon definition mode of splicing, we further assessed whether the introns are also identified in this group. Previous reports suggested that long exons may be included in the mature mRNA if the flanking introns are relatively short and thereby spliced via the intron definition mechanism (Fox-Walsh et al., 2005; Sterner et al., 1996). We thus extended the middle exon in the differential minigenes to a length that produces ES (from ~115 nt to ~417 nt), and then shortened the flanking introns (from ~2000 nt to ~210 nt). The results indicated that the middle exon is still skipped, meaning that in the differential minigenes the introns are not recognized as the spliced units (Figures 2H and S2M). This demonstrates that genes with the differential architecture are spliced exclusively via the exon definition mode.

To examine whether the exon-intron differential GC content is the major determinant for exons to be defined as the spliced unit, we modified a minigene with the leveled architecture, by replacing a 100 nt DNA segment downstream of the 5'SS with a sequence of the same length but with lower GC level. This introduces a differential architecture into a minigene with leveled architecture. We observed that for the *FES* clone, the introduction of the differential GC structure leads to elimination of the retained intron product, which is usually found in the wild-type minigene (Figure 2I, compare lane 1 to lane 3). Mutating the 5'SS shifts the splicing outcome from IR to ES (Figure 2I, compare lane 2 to lane 4). Similar results were obtained for the *CORO1B* minigene, except that in this minigene there was a selection of an intronic 3'SS located immediately downstream to the new segment that generated the differential GC content (Figures S2N and S2R, in both lanes 3 and 4). These observations demonstrate that exon-intron differential GC content can shift splicing from intron definition to exon definition. In summary, we can conclude that differential GC content between the exon and the downstream intron directs the recognition of the exon unit, whereas leveled GC content directs intron definition.

To examine the importance of nuclear spatial localization to the mode of splicing, we generated a second recombination-based integration site in a more peripheral genomic locus on chromosome 1. This site was selected from our Chrom3D analysis as being further from the nuclear center (Euclidian distance 3.96) than the first integration site (Euclidian distance 1.93) (Lin et al., 2014, see Methods). We then stably integrated our minigene constructs into the peripheral site. The splicing patterns of the minigenes inserted into the peripheral site were similar to those already observed in the original centrally located integration site

(Figures 2J–2Q [right] compared to Figures 2B–2I [left]), indicating that the exon-intron GC content architecture is the major driving force for the definition of introns or exons as the spliced unit.

To support these results, we analyzed the splicing outcome of exon-intron structures with the leveled and differential architectures throughout the nuclear space, but spatially outside of their main clustering regions (i.e., genes with the differential architecture that are located in the nuclear center and genes with the leveled architecture that are located in the nuclear periphery). We thus used a dataset of ES and IR events (VastDB) and extracted their exon-intron GC content profiles. We sorted the two architectures by their location within five concentric radial scopes and checked the dominant splicing event in each scope. We found that regardless of nuclear localization, the leveled architecture is enriched with IR events, and the differential architecture with ES events (Figure S2S). This analysis supports the notion that gene architecture, rather than nuclear localization, is the primary feature that determines the mode of splicing.

Nuclear localization affects alternative splicing outcome

We next examined whether the nuclear localization affects the levels of AS events. We performed quantitative real-time PCR (RT-qPCR) to calculate the inclusion levels of the middle exon or the downstream retained intron in minigenes with weak 5'SS. The differential minigenes show higher ES levels when inserted into the nuclear periphery than when in the nuclear center (Figure 3A). In contrast, the leveled minigenes inserted close to the nuclear center show higher retention of the downstream intron (Figure 3B), but not exon inclusion (data not shown). These results suggest that the periphery of the nucleus promotes ES events only for the differential architecture, while the center of the nucleus simulates IR for the leveled architecture. When examining the effect of the location on transcription levels of the minigenes, we found that the transcription of the leveled can be affected but not the differential (Figure S3A).

Next, we examined whether gene location within the nucleus is directed by the GC content. For this purpose, we used RNA FISH to detect the active sites of transcription of integrated minigenes from the differential architecture or the leveled architecture when inserted into the same genomic site. The two integration sites selected are within the *lacZ-Zeocin* open reading frame (see Methods), and the site of the *Zeocin* is considered the location of the integration site prior to minigene insertion. Insertion of the *CORO1B* minigene with higher GC content, resulted in a significant movement of the integration site toward the center of the nucleus relative to the original *Zeocin*, but only when integrated in the nuclear center and not in the nuclear periphery (Figures 3C and S3B). In contrast, the differential *FRG1* minigene with a lower GC content moved toward the nuclear periphery when integrated in the nuclear center but remained essentially in the same location when integrated in the nuclear periphery (Figures 3C and S3B). These results imply that the GC content of a gene influences its actual nuclear location only when it is located close to the nuclear center (see Discussion).

These results led us to examine the effect of this observed genomic movement on AS and transcription of endogenous genes adjacent to the integration site. We found that insertion of

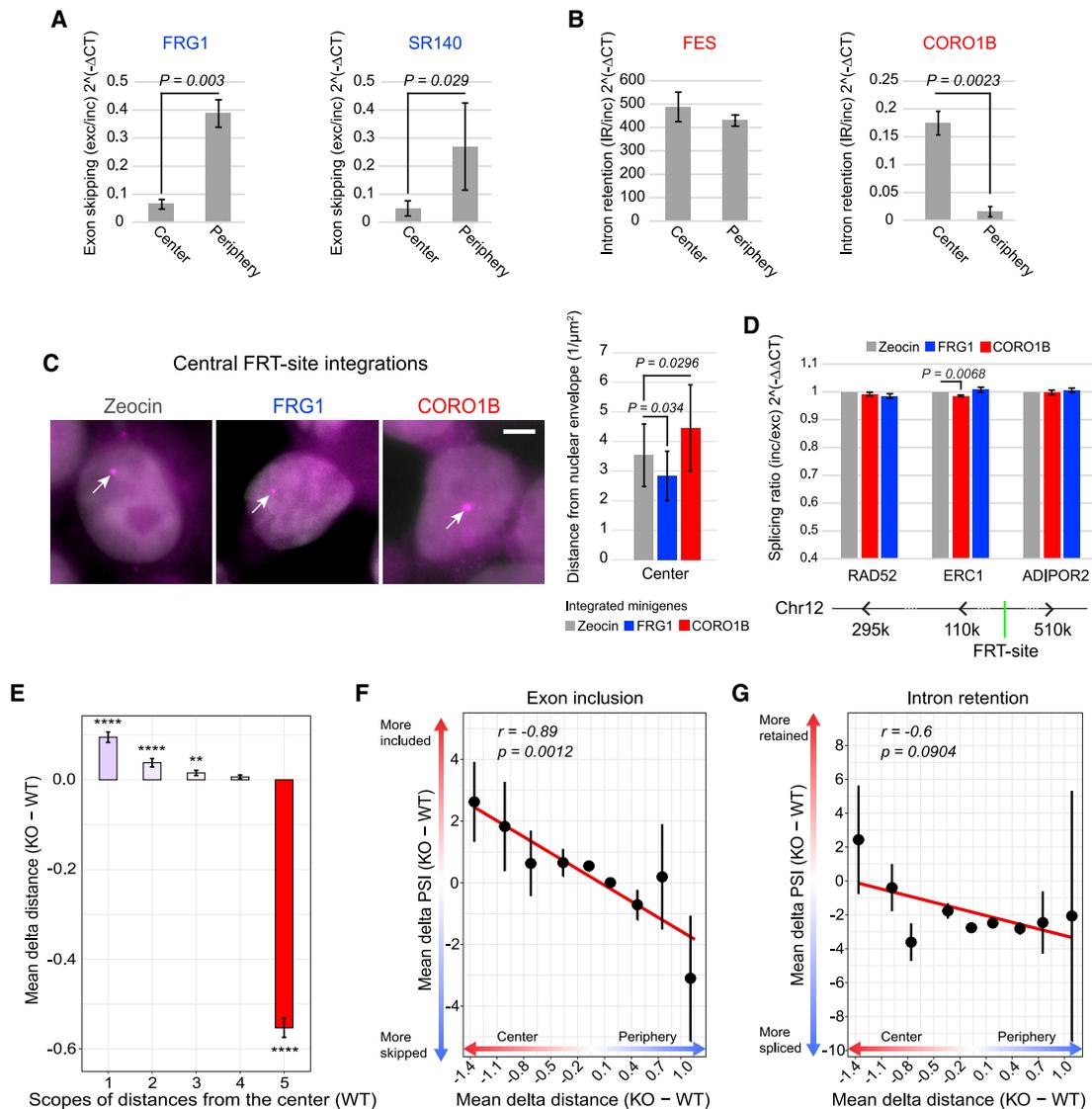


Figure 3. Nuclear spatial localization affects alternative splicing outcome

(A) Bar plot describing the ES levels (exon exclusion/exon inclusion) upon integration of the differential minigenes (*FRG1* and *SR140*) into the center or periphery of the nucleus.

(B) Bar plot portraying the IR levels upon integration of the leveled minigenes (*FES* and *CORO1B*) into the center or periphery of the nucleus (downstream intron inclusion/exon inclusion, as the exon inclusion amounts to the exclusion of the intron).

(C) Left: representative RNA FISH images of fixed HEK293 cells hybridized with probes that detect the Zeocin transcribed gene (control conditions before minigene integration) and the *FRG1* and *CORO1B* minigenes (representing the differential and the leveled architectures, respectively) that were inserted into the central integration site (FRT-site). White bar represents a scale of 5 μm . Right: bar plots depicting the mean distance of the indicated minigenes from the nuclear envelope after insertion into the central integration site. The distance was normalized to the appropriate nuclear volume (see Methods, distance given as $1/\mu\text{m}^2$), nuclei analyzed: Zeocin $n = 59$, *FRG1* $n = 42$, and *CORO1B* $n = 49$.

(D) Bar plots demonstrating the splicing ratios (exon inclusion/exon exclusion) of alternatively spliced exons located in genes (*RAD52*, *ERC1*, *ADIPOR2*) that are proximal to the FRT-site, upon integration of the differential (*FRG1*) or leveled (*CORO1B*) minigene to the central integration site. Splicing ratios of the endogenous adjacent genes are with respect to the splicing ratios prior to the integrations of the minigenes (referred to as Zeocin). The distance of the upstream and downstream genes from the FRT-site is shown at the bottom. (A)–(D) P values correspond to two-tailed t tests.

(E) Bar plot indicating the shifts in nuclear spatial localizations of TADs after LMNB1 KO in MDA-MB-231 cells per scope of Euclidean distance of the WT sample. Scopes 1-to-5 correspond to central-to-peripheral, respectively. P values correspond to two-tailed paired t tests. Significance levels: ****(1e-04), ***(0.001), ***(0.01), *(0.05).

(F) and (G) Error plots with mean PSI values of ES (F) and IR (G) events ($n = 10,181$ and $11,837$, respectively) after LMNB1 KO in MDA-MB-231 cells. Pearson correlation tests were performed. Regression lines are shown in red.

a relatively short high GC content segment (700 nt) 110 kb downstream to the *ERC1*'s alternatively spliced exon, induces a small, yet statistically significant, effect on *ERC1*'s exon inclusion level ~ 0.02 (fold-change) (Figures 3D and S3C), and on transcription levels ~ 0.03 (fold-change) of *ERC1* and *RAD52* genes (Figure S3D), but not on genes that are located farther from the insertion site. Both exon inclusion and gene expression levels change when the high GC content minigene is inserted into the integration site close to the nuclear center, however, inserting either of the minigenes into the nuclear periphery integration site has no effect on splicing (Figures 3D, S3C and S3D). This implies that genomic movement can affect both transcription and AS levels, especially in the center of the nucleus where there is more movement in the genomic locus. These results support previous reports suggesting that greater shifts in genomic localization occur for genes located close to the nuclear speckles compared to genes located close the nuclear periphery (Chen et al., 2018; Jabbari et al., 2019; Lemaire et al., 2019; Wang et al., 2018).

In addition, we asked how genome-wide changes in nuclear spatial organization affect AS. LMNB1 knockout (KO) leads to major chromatin rearrangements (Chang et al., 2020). In contrast to lamin A/C, lamin B1 is found solely in the nuclear periphery and tethers the chromatin to the nuclear matrix (Kolb et al., 2011; Paulsen et al., 2018; Paulsen et al., 2017). To examine the changes in nuclear spatial organization following LMNB1 KO, we generated Chrom3D genomic models of WT and LMNB1 KO by combining the ChIP-seq and Hi-C datasets from Chang et al., (2020). We show, using 3D modeling, that LMNB1 KO causes major shifts in nuclear locations of TADs in scope 5 toward the center, while TADs in scopes 1–4 shift moderately toward the nuclear periphery (Figure 3E).

To examine the effect of the above mentioned chromosomal rearrangements on AS, we analyzed RNA-seq data in WT and LMNB1 KO cells from (Chang et al., 2020), and coupled the percent spliced in (PSI) values of ES and IR events with their Euclidean distances obtained from our Chrom3D models. Our findings indicate that there is a significant correlation between the changes in PSI values of ES events and changes in Euclidean distances from the nuclear center following LMNB1 KO ($r = -0.89$, p value = 0.0072). Genes that move toward the nuclear periphery show lower exon inclusion (exons are more skipped), whereas genes that moved toward the nuclear center show higher exon inclusion (Figure 3F). These results are in line with Figure 3A. For IR events, a mild increment is shown in genes that moved toward the nuclear center (Figure 3G). In addition, this effect on ES events does not seem to be correlated with the elongation rate of RNA polymerase II, as insignificant correlation was observed between changes in transcription level and shifts in nuclear spatial locations of genes (Figure S3E). Taken together, we conclude that shifts in spatial locations of the chromatin due to KO LMNB1 affect the inclusion levels of ES events.

Distinct splicing regulatory subnetworks associate with genes located in the nuclear periphery and center

A previous study showed that the splicing of GC and AT-rich exons is affected by different sets of SFs (Lemaire et al., 2019). For instance, core spliceosomal components like SNRPC and SNRNP70 affect the inclusion levels of GC-rich exons, whereas

AT-rich exons are more likely to be regulated by SF1 or U2AF2 (Lemaire et al., 2019). To examine whether the exon and intron definition modes also involve distinct splicing regulatory proteins, we sought to identify SFs that preferentially bind transcripts encoded by genes located in the periphery or the center of the nucleus. To this end, we combined enhanced cross-linking and immunoprecipitation followed by sequencing (eCLIP-seq) data (Van Nostrand et al., 2016) with our Chrom3D data in K562 cells to determine the nuclear spatial distribution of the RNA binding sites of 55 experimentally validated SFs. Since co-transcriptional splicing is the norm rather than the exception (Herzel et al., 2017; Kornblihtt et al., 2013), we can assume that the genomic sites of genes reflect the location of their splicing activity. Focusing on the crosslinking sites within internal exons and the adjacent 100 nt of their flanking introns, we identified 15 SFs that predominantly bind pre-mRNAs transcribed in the nuclear periphery (referred to as peripherally active SFs or pSFs), and 27 SFs that mostly bind pre-mRNAs transcribed in the nuclear center (referred to as centrally active SFs or cSFs, Figures 4A.I and S4A.I).

To better understand which SFs target the differential and the leveled architectures in the different spatial locations of the nucleus, we examined the spatial distributions of the pre-mRNAs bound by each SF in the five concentric radial scopes (Figure 1A). The results indicate a gradual enrichment of cSFs from scope 3-to-1, and a gradual depletion from scope 4-to-5. Conversely, pSFs are enriched in scopes 4 and 5, and gradually depleted in scopes 3 through 1 (Figures 4A.II and S4A.II). Analysis of the enrichment patterns of the pSFs and cSFs with respect to their locations within the six subcompartments, revealed that the cSFs are enriched for genes residing in the centrally located subcompartment A1, whereas the peripherally located subcompartment A2 has more pSFs. This is in accordance with the enrichment patterns of the two groups of SFs in the five radial scopes. Notably, there is a depletion of all the examined SFs in subcompartments B1–B4 (Figure S4A.III). These subcompartments harbor inactive genes (Rao et al., 2014).

Because our analysis of the exon-intron GC content profile demonstrates that the leveled architecture is mainly found in proximity to speckles (Figure S1C, TSA-seq method), we were interested to relate the SF binding patterns to the distance of genes from speckles. We thus assigned the TSA-seq data (Chen et al., 2018), which describes the lamin-to-speckles axis, to the RNA binding sites obtained from the eCLIP-seq data. Figure S4A.IV demonstrates that cSFs mainly target genes in the top 10th percentile of proximity to the speckles, while other regions are depleted. Conversely, pSFs, are not enriched in the top 10th percentile, and SAFB, HNRNPC, KHSRP, and KHDRBS1 are enriched between the 60th and 80th percentiles (Figure S4A.IV). Notably, all SFs are depleted in the regions furthest from the nuclear speckles (the lowest 30th percentile). These results support the observation that nuclear speckles harbor SFs (Galganski et al., 2017; Spector and Lamond, 2011) and that nuclear speckles are excluded from the periphery (Chen et al., 2018).

Examination of the exon-intron architectures bound by the pSFs and the cSFs, revealed a strong linear correlation

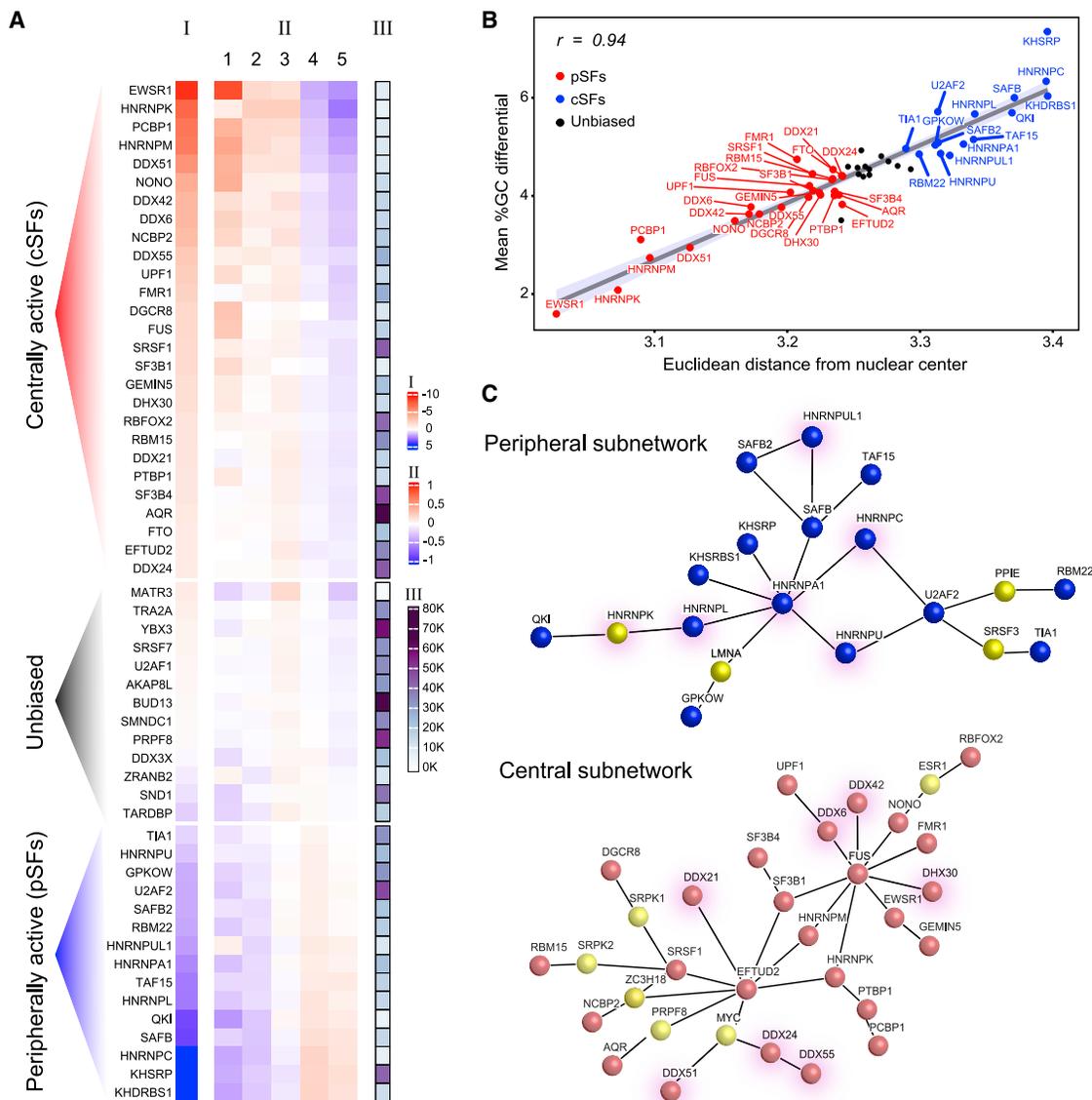


Figure 4. Different sets of SFs target pre-mRNAs transcribed in central and peripheral regions of the nucleus

(A) Heatmaps (I) and (II) depicting the nuclear localization (determined by Chrom3D) of exon-intron structures (a window of 75 nt taken from both sides of the exons and the adjacent 100 nt of the flanking introns) containing cross-linking sites of each SF in K562 cells. (I) Mean relative distances from the nuclear center (see Methods). Two-tailed Wilcoxon rank sum tests followed by FDR corrections were applied. SFs with a significant distance were classified as centrally and peripherally active. (II) Fold enrichment in each concentric radial scope from nuclear center. (I) and (II) See significance levels in Figure S4. (III) Heatmap of the total exon-intron structures bound by each SF.

(B) Pearson correlation analysis between the mean %GC content differential level (exon - intron) and the Euclidean distance of the bound exon-intron structures per SF.

(C) Peripheral (top) and central (bottom) subnetworks. pSFs and cSFs are colored in blue and red, respectively. Proteins found to be associated with the original sets of pSFs and cSFs in the two subnetworks, according to ANAT (Yosef et al., 2011), are colored in yellow. Protein-protein interactions are shown as black lines. HNRNP proteins in the peripheral subnetwork and DEX/D/H RNA helicases in the central subnetwork are highlighted.

between the radial distance and the exon-intron GC content differential levels (exonic - intronic) ($r = 0.94$, p value $< 2.2 \times 10^{-16}$), where pSFs target pre-mRNAs with higher exon-intron differential GC content than cSFs (Figure 4B). Furthermore, pSFs bind exon-intron structures with overall lower GC content and mostly longer introns than the cSFs (Figure S4A.V and VI). Thus, the pSFs preferentially bind the differential ar-

chitecture, whereas the cSFs mostly target the leveled architecture. Because ES events are enriched in peripheral regions and IR events occur mostly in central regions and closer to the nuclear speckles (Figures 1E–1F and S1F–S1G), the binding preferences of the pSFs and the cSFs suggest their involvement in the exon and intron definition modes, respectively.

To examine whether gene architecture or nuclear localization affects the splicing unit recognition, we selected the 10 most extreme SFs (pSFs and cSFs) and categorized their combined sets of targets according to the location in the five concentric radial scopes. We then assessed to which architecture the targets in each scope correspond (architecture definitions given in Gelfman et al., [2013]). We observed that in all scopes, cSFs bind targets that are enriched for the leveled architecture and depleted in the differential. In contrast, targets of pSFs are enriched for the differential architecture, but not for the leveled. Except for scope 1, we find that the enrichment pattern for the unbiased SFs is enriched for the differential architecture (Figure S4B).

From an evolutionary perspective, although pSFs and cSFs are similarly ancient, pSFs have significantly more paralogs (i.e., belong to larger gene families) and ohnologs (i.e., paralog gene pairs retained following whole genome duplication events) than cSFs (Figures S4C–S4E). This suggests that pSFs have a higher degree of neofunctionalization and/or specialization than do cSFs.

The splicing mechanism requires an elaborate interplay between SFs within a complex. Therefore, in order to study the regulatory pathways that might mediate the splicing process of the differential and leveled architectures via the exon and intron definition modes, we used the advanced network analysis tool (Yosef et al., 2011) to construct the physical protein-protein interactions between the cSFs and the pSFs. The results suggest the presence of two separate subnetworks that connect the SFs in each group. Markedly, the proteins with the highest number of interactions, are hnRNPA1 in the pSFs, and EFTUD2 and FUS in the cSFs (Figure 4C). These discrete interactions and proteins in the two networks may reflect different functionalities in terms of splicing modes (see Discussion). In addition, gene ontology analysis reveals significant enrichment of proteins associated with nuclear speckles and Cajal bodies in the central subnetwork (Figures S4F and S4G). Collectively, our results suggest that two distinct regulatory subnetworks mediate the splicing of pre-mRNAs transcribed in the periphery or the center of the nucleus. The peripheral subnetwork preferentially binds transcripts with the differential architecture, and these are spliced by exon definition and associate with the nuclear periphery. In contrast, the central subnetwork mainly binds transcripts with the leveled architecture, and these are spliced by intron definition, and associated with the nuclear center.

DISCUSSION

In this manuscript, we uncovered how exons or introns are selected as the spliced units by demonstrating that distinct gene architectures promote either exon or intron definition. Our finding that these architectures are differentially located in the nucleus is the first to demonstrate the link between the splicing mechanism and the nuclear 3D genome organization. The results demonstrate that: (1) IR and ES are favored in the center or periphery of the nucleus, respectively; (2) the exon-intron GC profile determines whether exons or introns serve as the splicing unit; (3) nuclear localization regulates the level of AS; (4) SF binding segregates between the center and periphery; (5) genomic loci can shift from one nuclear spatial region to another when gene architecture is altered; and (6) changes in

GC content, rather than in exon/intron lengths, can convert the mode of splicing from intron to exon definition. Altogether, our results suggest that the chromatin is organized in the nuclear space in a way that creates different functional zones with respect to the splicing mechanism of exon and intron definition, regulation of AS, and binding of SFs.

We describe the identification of two topologically distinct protein-protein interaction subnetworks of SFs that interact with the differential and leveled architectures. The subnetwork associated with the nuclear periphery is enriched for proteins from the hnRNP family. These SFs were previously shown to be involved in compacting long introns, thus acting as “RNA nucleosomes” (König et al., 2010). hnRNPA1, which is a key SF in the peripheral subnetwork, is a member of the hnRNP A/B subfamily, whose members have been shown to bind splicing elements flanking alternative exons and coordinate their inclusion or exclusion (Ule and Blencowe, 2019). Moreover, hnRNPA1 may favor splicing mediated by exon definition as it has been shown to spread preferentially in a 3' to 5' direction (Okunola and Krainer, 2009). Furthermore, hnRNPA1 and hnRNPC (an additional SF in the peripheral subnetwork) were shown to modulate the interaction of U2AF2 with decoy 3'SSs during exon definition (Howard et al., 2018; Zarnack et al., 2013). An intriguing node in the peripheral subnetwork associates U2AF2, SRSF3, and TIA1, where TIA1 binds U-rich sequences downstream of the 5'SS and promotes the binding of U1 snRNP to the 5'SS, which in turn facilitates exon definition by enhancing U2AF binding to the upstream 3' splice site (Izquierdo et al., 2005; Vivori et al., 2021). It may thus be possible that SRSF3 interacts with U2AF, which binds the polypyrimidine tract and TIA1 at the 5'SS to facilitate cross-exon formation of the splicing complex. Moreover, hnRNPA1 interacts with the nuclear matrix/scaffold attachment factor SAFB2 and with lamin A/C (LMNA). This suggests a role for the nuclear matrix in the splicing of genes transcribed in the nuclear periphery.

In the subnetwork that interacts with the leveled architecture, one-fifth of the SFs belongs to the DEXD/DEXH-box RNA helicase family. These SFs facilitate transcription and splicing by opening secondary structures, particularly at the 5'SS of GC-rich pre-mRNAs (Zhang et al., 2011). Downregulation of FUS, which is a key SF in the central subnetwork, was shown to induce widespread IR events in amyotrophic lateral sclerosis, supporting the notion that FUS promotes intron definition in the central subnetwork (Humphrey et al., 2020). In this context, the central subnetwork also includes two components of the SF3b complex (SF3B1 and SF3B4), which are involved in the U1-U2 interaction network in pre-spliceosomal assembly during intron definition (Shao et al., 2012). FUS is a core component of the paraspeckles, as are NONO and EWSR1 proteins which also participate in the node. The speckles and the paraspeckles are located in the interchromatin spaces and are excluded from the nuclear periphery (Chen et al., 2018). This suggests that the speckles are responsible for pulling high GC content genes toward the center of the nucleus. As lamin proteins bind AT rich sequences (van Steensel and Belmont, 2017), it is possible that the combination of lamin on one hand and speckles on the other are responsible for the GC content transition across the nucleus.

Our results demonstrate that the differential GC content genes in the nuclear periphery are predominantly recognized by exon

definition regardless of intron length, but the maximal length of the exon to be recognized by this mechanism is approximately 400 nt, which is in agreement with previous findings (Amit et al., 2012; De Conti et al., 2013; Enculescu et al., 2020). On the other hand, the lengths of both introns and exons are important for proper recognition by the splicing machinery in the leveled GC content architecture found in the nuclear center. This architecture of genes underwent an elevation of their GC content during homeothermic evolution (Amit et al., 2012). Minigenes generated from this architecture are generally left unspliced when transiently transfected and require genome integration in order to undergo proper splicing. These observations suggest that the leveled architecture requires a higher level of regulation than the differential architecture, and specifically has the need for regulatory factors such as RNA helicases that can open secondary structures (Zhang et al., 2011). We now suggest that the genomic and spatial localization of genes with leveled GC content in the center of the nucleus may bring all these factors into proximity with the nuclear speckles (Chen et al., 2018; Crosetto and Bienko, 2020)

The regulation of the differential architecture, which is the ancestral exon-intron gene structure (Amit et al., 2012), is arguably simpler. As the exons are the only recognized unit in this architecture, while the introns lengthened considerably during evolution, the exons remained short. Exons in the differential architecture are protected by higher nucleosome occupancy than in flanking introns (Amit et al., 2012; Schwartz et al., 2009b; Tilgner et al., 2009) and are marked for splicing by a specific histone modification (H3K36me3) (Lev Maor et al., 2015; Saint-André et al., 2011). Our findings that this architecture is associated with hnRNP proteins that are involved in splicing by compacting long introns (König et al., 2010), suggest that in genes with exon-intron differential GC content, chromatin organization and epigenetic mark may assist in co-transcriptional identification of short exons flanked by long introns (Lev Maor et al., 2015; Saint-André et al., 2011).

In summary, the transition of exon-intron GC-content from the differential to leveled architectures generates two nuclear subregions with different preferences for the mode of splice unit selection and different networks of splicing regulatory proteins. These architectures are a major determinant in directing the splicing machinery to select introns or exons as the spliced unit. Genomes of warm-blooded organisms accumulated mutations that elevated the overall genomic GC-content, which during the course of evolution resulted in the formation of nonhomogeneous nuclei with an increased GC-content along the lamina-to-speckle axis. The positioning of nuclear speckles may drive this higher order of nuclear genomic organization (Chen et al., 2018; Crosetto and Bienko, 2020). This global nucleoplasm spatial genomic organization may also impact other nuclear machineries such as those involved in DNA replication and transcription.

Limitations of the study

One possible limitation lies within the 3D reconstruction of the genomes, and it is possible that a modeling method does not fully capture the radial distribution of chromosomes.

Another possible limitation is that the insertion of segments used for lengthening/shortening exons or introns could contain

splicing regulatory elements (silencers/enhancers), and we cannot rule out this possibility completely despite our efforts.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Cell culture
- METHOD DETAILS
 - Chromatin immunoprecipitation (ChIP)
 - 3D Genomic models
 - LMNB1 KO versus WT 3D Genome models and RNA-seq analysis
 - Exon skipping and intron retention spatial distribution analysis
 - RNA FISH
 - Minigene cloning
 - Site-directed mutagenesis
 - Manipulation of minigene introns and exons
 - Generation of peripheral FRT-site using CRISPR-Cas9
 - Generation of peripheral FRT-site stable expression in HEK293 cell line
 - Transfection of Flp-In-293 HEK cells
 - RNA isolation RT-PCR amplification and RT-qPCR
 - Exon dataset construction for the eCLIP-seq analysis
 - eCLIP-seq analysis
 - Central and peripheral subnetworks construction
 - Subnetworks; separation analysis
 - Gene ontology (GO) enrichment analysis
 - Evolutionary history and gene family size analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.molcel.2022.02.001>.

ACKNOWLEDGMENTS

The research was funded by the Israel Science Foundation [ISF 671/18, ISF 142/13, and ISF 2417/20]; German-Israeli Foundation for Scientific Research and Development [GIF I-1460]; Israel Cancer Association [ICA 20170034]; Israel Cancer Research Fund [ICRF PG-18-105]; and United States – Israel Binational Science Foundation [BSF 2017086]. L.T. was supported by a fellowship from the Planning & Budgeting Committee of the Council of Higher Education of Israel. O.H. and V.R.R. were supported by a fellowship from the Edmond J. Safra Bioinformatics Center at Tel Aviv University. We thank D. Hollander for consultation throughout the process of writing the manuscript.

AUTHOR CONTRIBUTIONS

L.T. designed and performed the biological experiments. O.H. designed and performed the bioinformatics analyses related to the regulation of SFs in the nuclear periphery and center. I.K. constructed the Chrom3D models for

HeLa, K562 and GM12878 cells. I.K. and O.H. performed the bioinformatics analyses related to the chromatin 3D models. V.R.R. constructed the 3D models of the LMNB1 KO and WT in MDA-MB-231 cells and O.H. performed the following RNA-seq analysis. N.C. and J.R. performed the FISH experiment with the endogenous genes. R.S. performed ChIP-seq of lamin A/C and generated the peripheral FRT-site using CRISPR-Cas9. A.A.T. and L.T. performed the FISH experiment with integrated minigenes. I.S. and O.H. performed the analyses of the peripheral and central subnetworks. V.R.-V. performed the evolutionary analyses. G.L.M. helped with constructing additional inserts and minigene cloning. Y.L. and L.T. performed the RT-qPCR. D.K. performed the intron shortening of the constructs. E.L.A and R.E. advised and guided in the design of the bioinformatics analyses. L.T., O.H., V.R.R., I.K., and G.A. wrote the manuscript. All authors mentioned above had the opportunity to edit, comment on, and approve the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We worked to ensure diversity in experimental samples through the selection of the cell lines and genomic datasets. One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in science. One or more of the authors of this paper received support from a program designed to increase minority representation in science. The author list of this paper includes contributors from the location where the research was conducted, and those who participated in the data collection, design, analysis, and/or interpretation of the work.

Received: April 12, 2021

Revised: January 31, 2022

Accepted: January 31, 2022

Published: February 18, 2022

REFERENCES

- Akerman, M., and Mandel-Gutfreund, Y. (2007). Does distance matter? Variations in alternative 3' splicing regulation. *Nucleic Acids Res.* *35*, 5487–5498.
- Almozilino, Y., Atias, N., Silverbush, D., and Sharan, R. (2017). ANAT 2.0: reconstructing functional protein subnetworks. *BMC Bioinformatics* *18*, 495.
- Amit, M., Donyo, M., Hollander, D., Goren, A., Kim, E., Gelfman, S., Lev-Maor, G., Burstein, D., Schwartz, S., Postolsky, B., et al. (2012). Differential GC content between exons and introns establishes distinct strategies of splice-site recognition. *Cell Rep.* *1*, 543–556.
- Ast, G. (2004). How did alternative splicing evolve? *Nat. Rev. Genet.* *5*, 773–782.
- Bae, S., Park, J., and Kim, J.S. (2014). Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* *30*, 1473–1475.
- Bergert, S.M. (1995). Exon recognition in vertebrate splicing. *J. Biol. Chem.* *270*, 2411–2414.
- Braunschweig, U., Barbosa-Morais, N.L., Pan, Q., Nachman, E.N., Alipanahi, B., Gontopoulos-Pournatzis, T., Frey, B., Irimia, M., and Blencowe, B.J. (2014). Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res.* *24*, 1774–1786.
- Brody, Y., Neufeld, N., Bieberstein, N., Causse, S.Z., Böhnlein, E.M., Neugebauer, K.M., Darzacq, X., and Shav-Tal, Y. (2011). The in vivo kinetics of RNA polymerase II elongation during co-transcriptional splicing. *PLoS Biol.* *9*, e1000573.
- Carmel, I., Tal, S., Vig, I., and Ast, G. (2004). Comparative analysis detects dependencies among the 5' splice-site positions. *RNA* *10*, 828–840.
- Chang, L., Li, M., Shao, S., Li, C., Ai, S., Xue, B., Hou, Y., Zhang, Y., Li, R., Fan, X., et al. (2020). Nuclear peripheral chromatin-lamin B1 interaction is required for global integrity of chromatin architecture and dynamics in human cells. *Protein Cell*. Published online November 5, 2020. <https://doi.org/10.1007/s13238-020-00794-8>.
- Chen, Y., Zhang, Y., Wang, Y., Zhang, L., Brinkman, E.K., Adam, S.A., Goldman, R., van Steensel, B., Ma, J., and Belmont, A.S. (2018). Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. *J. Cell Biol.* *217*, 4025–4048.
- Cheng, R.R., Contessoto, V.G., Lieberman Aiden, E., Wolynes, P.G., Di Pierro, M., and Onuchic, J.N. (2020). Exploring chromosomal structural heterogeneity across multiple cell lines. *eLife* *9*, e60312.
- Consortium, E.P.; ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* *489*, 57–74.
- Crosetto, N., and Bienko, M. (2020). Radial Organization in the Mammalian Nucleus. *Front. Genet.* *11*, 33.
- Custódio, N., Carmo-Fonseca, M., Geraghty, F., Pereira, H.S., Grosveld, F., and Antoniou, M. (1999). Inefficient processing impairs release of RNA from the site of transcription. *EMBO J.* *18*, 2855–2866.
- Darzacq, X., Shav-Tal, Y., de Turris, V., Brody, Y., Shenoy, S.M., Phair, R.D., and Singer, R.H. (2007). In vivo dynamics of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* *14*, 796–806.
- De Conti, L., Baralle, M., and Buratti, E. (2013). Exon and intron definition in pre-mRNA splicing. *Wiley Interdiscip. Rev. RNA* *4*, 49–60.
- Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016). Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst.* *3*, 95–98.
- Enculescu, M., Braun, S., Thonta Setty, S., Busch, A., Zarnack, K., König, J., and Legewie, S. (2020). Exon Definition Facilitates Reliable Control of Alternative Splicing in the RON Proto-Oncogene. *Biophys. J.* *118*, 2027–2041.
- Eyre-Walker, A., and Hurst, L.D. (2001). The evolution of isochores. *Nat. Rev. Genet.* *2*, 549–555.
- Fox-Walsh, K.L., Dou, Y., Lam, B.J., Hung, S.P., Baldi, P.F., and Hertel, K.J. (2005). The architecture of pre-mRNAs affects mechanisms of splice-site pairing. *Proc. Natl. Acad. Sci. USA* *102*, 16176–16181.
- Galganski, L., Urbanek, M.O., and Krzyzosiak, W.J. (2017). Nuclear speckles: molecular organization, biological function and role in disease. *Nucleic Acids Res.* *45*, 10350–10368.
- Gelfman, S., and Ast, G. (2013). When epigenetics meets alternative splicing: the roles of DNA methylation and GC architecture. *Epigenomics* *5*, 351–353.
- Gelfman, S., Cohen, N., Yearim, A., and Ast, G. (2013). DNA-methylation effect on cotranscriptional splicing is dependent on GC architecture of the exon-intron structure. *Genome Res.* *23*, 789–799.
- Girelli, G., Custodio, J., Kallas, T., Agostini, F., Wernersson, E., Spanjaard, B., Mota, A., Kolbeinsdottir, S., Gelali, E., Crosetto, N., and Bienko, M. (2020). GPSeq reveals the radial organization of chromatin in the cell nucleus. *Nat. Biotechnol.* *38*, 1184–1193.
- Haeussler, M., Schöning, K., Eckert, H., Eschstruth, A., Mianné, J., Renaud, J.B., Schneider-Maunoury, S., Shkumatava, A., Teboul, L., Kent, J., et al. (2016). Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol.* *17*, 148.
- Han, H., Braunschweig, U., Gontopoulos-Pournatzis, T., Weatheritt, R.J., Hirsch, C.L., Ha, K.C.H., Radovani, E., Nabeel-Shah, S., Sterne-Weiler, T., Wang, J., et al. (2017). Multilayered Control of Alternative Splicing Regulatory Networks by Transcription Factors. *Mol. Cell* *65*, 539–553.e7.
- Harvey, S.E., and Cheng, C. (2016). Methods for Characterization of Alternative RNA Splicing. *Methods Mol. Biol.* *1402*, 229–241.
- Herzel, L., Ottoo, D.S.M., Alpert, T., and Neugebauer, K.M. (2017). Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. *Nat. Rev. Mol. Cell Biol.* *18*, 637–650.
- Howard, J.M., Lin, H., Wallace, A.J., Kim, G., Draper, J.M., Haeussler, M., Katzman, S., Toloue, M., Liu, Y., and Sanford, J.R. (2018). HNRNPA1 promotes recognition of splice site decoys by U2AF2 in vivo. *Genome Res.* *28*, 689–698.

- Hu, M., Deng, K., Selvaraj, S., Qin, Z., Ren, B., and Liu, J.S. (2012). HiCNorm: removing biases in Hi-C data via Poisson regression. *Bioinformatics* 28, 3131–3133.
- Huang, W., Sherman, B.T., and Lempicki, R.A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37, 1–13.
- Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57.
- Huber, D.W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B.S., Bravo, H.C., Davis, S., Gatto, L., Girke, T., et al. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* 12, 115–121.
- Humphrey, J., Birsa, N., Milioto, C., McLaughlin, M., Ule, A.M., Robaldo, D., Eberle, A.B., Kräuchi, R., Bentham, M., Brown, A.L., et al. (2020). FUS ALS-causative mutations impair FUS autoregulation and splicing factor networks through intron retention. *Nucleic Acids Res.* 48, 6889–6905.
- Irimia, M., Weatheritt, R.J., Ellis, J.D., Parikshak, N.N., Gonatopoulos-Pournatzis, T., Babor, M., Quesnel-Vallières, M., Tapial, J., Raj, B., O'Hanlon, D., et al. (2014). A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell* 159, 1511–1523.
- Izquierdo, J.M., Majós, N., Bonnal, S., Martínez, C., Castelo, R., Guigó, R., Bilbao, D., and Valcárcel, J. (2005). Regulation of Fas alternative splicing by antagonistic effects of TIA-1 and PTB on exon definition. *Mol. Cell* 19, 475–484.
- Jabbari, K., Chakraborty, M., and Wiehe, T. (2019). DNA sequence-dependent chromatin architecture and nuclear hubs formation. *Sci. Rep.* 9, 14646.
- Jayasinghe, R.G., Cao, S., Gao, Q., Wendl, M.C., Vo, N.S., Reynolds, S.M., Zhao, Y., Climente-Gonzalez, H., Chai, S., Wang, F., et al. (2018). Systematic Analysis of Splice-Site-Creating Mutations in Cancer. *Cell Rep* 23, 270–281.e3.
- Kolb, T., Maass, K., Hergt, M., Aebi, U., and Herrmann, H. (2011). Lamin A and lamin C form homodimers and coexist in higher complex forms both in the nucleoplasmic fraction and in the lamina of cultured human cells. *Nucleus* 2, 425–433.
- König, J., Zarnack, K., Rot, G., Curk, T., Kayikci, M., Zupan, B., Turner, D.J., Luscombe, N.M., and Ule, J. (2010). iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nat. Struct. Mol. Biol.* 17, 909–915.
- Kombliht, A.R., Schor, I.E., Alló, M., Dujardin, G., Petrillo, E., and Muñoz, M.J. (2013). Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat. Rev. Mol. Cell Biol.* 14, 153–165.
- Labun, K., Montague, T.G., Krause, M., Torres Cleuren, Y.N., Tjeldnes, H., and Valen, E. (2019). CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res.* 47 (W1), W171–W174.
- Leader, Y., Lev Maor, G., Sorek, M., Shayevitch, R., Hussein, M., Hameiri, O., Tammer, L., Zonszain, J., Keydar, I., Hollander, D., et al. (2021). The upstream 5' splice site remains associated to the transcription machinery during intron synthesis. *Nat. Commun.* 12, 4545.
- Lee, F.C.Y., and Ule, J. (2018). Advances in CLIP Technologies for Studies of Protein-RNA Interactions. *Mol. Cell* 69, 354–369.
- Lemaire, S., Fontrodona, N., Aubé, F., Claude, J.B., Polvéche, H., Modolo, L., Bourgeois, C.F., Mortreux, F., and Auboeuf, D. (2019). Characterizing the interplay between gene nucleotide composition bias and splicing. *Genome Biol.* 20, 259.
- Lev Maor, G., Yearim, A., and Ast, G. (2015). The alternative role of DNA methylation in splicing regulation. *Trends Genet.* 31, 274–280.
- Li, X., Liu, S., Zhang, L., Issaian, A., Hill, R.C., Espinosa, S., Shi, S., Cui, Y., Kappel, K., Das, R., et al. (2019). A unified mechanism for intron and exon definition and back-splicing. *Nature* 573, 375–380.
- Liebeskind, B.J., McWhite, C.D., and Marcotte, E.M. (2016). Towards Consensus Gene Ages. *Genome Biol. Evol.* 8, 1812–1823.
- Lin, Y.C., Boone, M., Meuris, L., Lemmens, I., Van Roy, N., Soete, A., Reumers, J., Moisse, M., Plaisance, S., Drmanac, R., et al. (2014). Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nat. Commun.* 5, 4767.
- Litman, T., and Stein, W.D. (2019). Obtaining estimates for the ages of all the protein-coding genes and most of the ontology-identified noncoding genes of the human genome, assigned to 19 phylostrata. *Semin. Oncol.* 46, 3–9.
- Lund, E., Oldenburg, A.R., and Collas, P. (2014). Enriched domain detector: a program for detection of wide genomic enrichment domains robust against local variations. *Nucleic Acids Res.* 42, e92.
- Okunola, H.L., and Krainer, A.R. (2009). Cooperative-binding and splicing-repressive properties of hnRNP A1. *Mol. Cell. Biol.* 29, 5620–5631.
- Paulsen, J., Sekelja, M., Oldenburg, A.R., Barateau, A., Briand, N., Delbarre, E., Shah, A., Sorensen, A.L., Vigouroux, C., Buendia, B., and Collas, P. (2017). Chrom3D: three-dimensional genome modeling from Hi-C and nuclear lamin-genome contacts. *Genome Biol.* 18, 21.
- Paulsen, J., Liyakat Ali, T.M., and Collas, P. (2018). Computational 3D genome modeling using Chrom3D. *Nat. Protoc.* 13, 1137–1152.
- Penalva, L.O., Lallena, M.J., and Valcárcel, J. (2001). Switch in 3' splice site recognition between exon definition and splicing catalysis is important for sex-lethal autoregulation. *Mol. Cell. Biol.* 21, 1986–1996.
- Petersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612.
- Pohl, A., and Beato, M. (2014). bwtool: a tool for bigWig files. *Bioinformatics* 30, 1618–1619.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.
- Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A., and Tyagi, S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* 5, 877–879.
- Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., and Aiden, E.L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–1680.
- Saint-André, V., Batsché, E., Rachez, C., and Muchardt, C. (2011). Histone H3 lysine 9 trimethylation and HP1 γ favor inclusion of alternative exons. *Nat. Struct. Mol. Biol.* 18, 337–344.
- Saito, R., Smoot, M.E., Ono, K., Ruschinski, J., Wang, P.L., Lotia, S., Pico, A.R., Bader, G.D., and Ideker, T. (2012). A travel guide to Cytoscape plugins. *Nat. Methods* 9, 1069–1076.
- Schneider, et al. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, <https://doi.org/10.1038/nmeth.2089>.
- Schneider, M., Will, C.L., Anokhina, M., Tazi, J., Urlaub, H., and Lührmann, R. (2010). Exon definition complexes contain the tri-snRNP and can be directly converted into B-like precatalytic splicing complexes. *Mol. Cell* 38, 223–235.
- Schwartz, S.H., Silva, J., Burstein, D., Pupko, T., Eyraes, E., and Ast, G. (2008). Large-scale comparative analysis of splicing signals and their corresponding splicing factors in eukaryotes. *Genome Res.* 18, 88–103.
- Schwartz, S., Hall, E., and Ast, G. (2009a). SROOGLE: webserver for integrative, user-friendly visualization of splicing signals. *Nucleic Acids Res.* 37, W189–92.
- Schwartz, S., Meshorer, E., and Ast, G. (2009b). Chromatin organization marks exon-intron structure. *Nat. Struct. Mol. Biol.* 16, 990–995.
- Shao, W., Kim, H.S., Cao, Y., Xu, Y.Z., and Query, C.C. (2012). A U1-U2 snRNP interaction network during intron definition. *Mol. Cell. Biol.* 32, 470–478.
- Sharma, S., Kohlstaedt, L.A., Damianov, A., Rio, D.C., and Black, D.L. (2008). Polypyrimidine tract binding protein controls the transition from exon definition to an intron defined spliceosome. *Nat. Struct. Mol. Biol.* 15, 183–191.
- Shin, H., Shi, Y., Dai, C., Tjong, H., Gong, K., Alber, F., and Zhou, X.J. (2016). TopDom: an efficient and deterministic method for identifying topological domains in genomes. *Nucleic Acids Res.* 44, e70.

- Shiraishi, Y., Kataoka, K., Chiba, K., Okada, A., Kogure, Y., Tanaka, H., Ogawa, S., and Miyano, S. (2018). A comprehensive characterization of *cis*-acting splicing-associated variants in human cancer. *Genome Res.* 28, 1111–1125.
- Spector, D.L., and Lamond, A.I. (2011). Nuclear speckles. *Cold Spring Harb. Perspect. Biol.* 3, a000646.
- Stern, D.A., Carlo, T., and Berget, S.M. (1996). Architectural limits on split genes. *Proc. Natl. Acad. Sci. USA* 93, 15081–15085.
- Tan, L., Xing, D., Chang, C.H., Li, H., and Xie, X.S. (2018). Three-dimensional genome structures of single diploid human cells. *Science* 361, 924–928.
- Tapial, J., Ha, K.C.H., Sterne-Weiler, T., Gohr, A., Braunschweig, U., Hermoso-Pulido, A., Quesnel-Vallières, M., Permanyer, J., Sodaei, R., Marquez, Y., et al. (2017). An atlas of alternative splicing profiles and functional associations reveals new regulatory programs and genes that simultaneously express multiple major isoforms. *Genome Res.* 27, 1759–1768.
- Tilgner, H., Nikolaou, C., Althammer, S., Sammeth, M., Beato, M., Valcárcel, J., and Guigó, R. (2009). Nucleosome positioning as a determinant of exon recognition. *Nat. Struct. Mol. Biol.* 16, 996–1001.
- Touceda-Suárez, M., Kita, E.M., Acemel, R.D., Firbas, P.N., Magri, M.S., Naranjo, S., Tena, J.J., Gómez-Skarmeta, J.L., Maeso, I., and Irimia, M. (2020). Ancient genomic regulatory blocks are a source for regulatory gene deserts in vertebrates after whole genome duplications. *Mol. Biol. Evol.* 37, 2857–2864.
- Ule, J., and Blencowe, B.J. (2019). Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution. *Mol. Cell* 76, 329–345.
- Van Nostrand, E.L., Pratt, G.A., Shishkin, A.A., Gelboin-Burkhardt, C., Fang, M.Y., Sundararaman, B., Blue, S.M., Nguyen, T.B., Surka, C., Elkins, K., et al. (2016). Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods* 13, 508–514.
- van Steensel, B., and Belmont, A.S. (2017). Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell* 169, 780–791.
- Vivori, C., Papasaikas, P., Stadhouders, R., Di Stefano, B., Rubio, A.R., Balaguer, C.B., Generoso, S., Mallol, A., Sardina, J.L., Payer, B., et al. (2021). Dynamics of alternative splicing during somatic cell reprogramming reveals functions for RNA-binding proteins CPSF3, hnRNP UL1, and TIA1. *Genome Biol.* 22, 171.
- Wahl, M.C., Will, C.L., and Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell* 136, 701–718.
- Wang, H., Xu, X., Nguyen, C.M., Liu, Y., Gao, Y., Lin, X., Daley, T., Kipriss, N.H., La Russa, M., and Qi, L.S. (2018). CRISPR-Mediated Programmable 3D Genome Positioning and Nuclear Organization. *Cell* 175, 1405–1417.e4.
- Wheeler, E.C., Van Nostrand, E.L., and Yeo, G.W. (2018). Advances and challenges in the detection of transcriptome-wide protein-RNA interactions. *Wiley Interdiscip. Rev. RNA* 9.
- Will, C.L., and Lührmann, R. (2011). Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* 3, a003707.
- Yosef, N., Zalckvar, E., Rubinstein, A.D., Homilius, M., Atias, N., Vardi, L., Berman, I., Zur, H., Kimchi, A., Ruppin, E., and Sharan, R. (2011). ANAT: a tool for constructing and analyzing functional protein networks. *Sci. Signal.* 4, pi1.
- Zarnack, K., König, J., Tajnik, M., Martincorena, I., Eustermann, S., Stévant, I., Reyes, A., Anders, S., Luscombe, N.M., and Ule, J. (2013). Direct competition between hnRNP C and U2AF65 protects the transcriptome from the exonization of Alu elements. *Cell* 152, 453–466.
- Zhang, J., Kuo, C.C., and Chen, L. (2011). GC content around splice sites affects splicing through pre-mRNA secondary structures. *BMC Genomics* 12, 90.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse monoclonal anti-Lamin A/C	Santa-Cruz Biotechnology	sc-7292; RRID: AB_627875
Deposited data		
Raw and analyzed data of Lamin A/C ChIP-seq	This paper	GEO: GSE174396
Original RT-PCR gels	This paper; Mendeley Data	https://doi.org/10.17632/mtsbg6brmw.1
Experimental models: Cell lines		
Human: Flp-In TM -293	Invitrogen TM	Cat# R75007; RRID:CVCL_U421
Human: HEK293 with peripheral FRT integration site	This paper	N/A
Human: GM12878	BioBank - Coriell Institute	RRID:CVCL_7526;
Human: K-562	ATCC	Cat# CCL-243; RRID:CVCL_0004
Oligonucleotides		
Probes used for Stellaris [®] FISH of integrated minigenes analysis, see Table S2	This paper	N/A
Primers used for cloning and constructing the minigenes, see Table S2	This paper	N/A
Primers used for specific integration of FRT-site to the genome using CRISPR-Cas9, see Table S2	This paper	N/A
Primers used for RT-qPCR, see Table S2	This paper	N/A
Recombinant DNA		
pcDNA TM 5/FRT/TO	Invitrogen TM	Cat# V601020
pSpCas9(BB)-2A-GFP	Addgene	Cat# 48138; RRID:Addgene_48138
pX458-GFP-Cas9-2sgRNA (with 2 expression cassette sgRNA)	This paper	N/A
pX458-2gRNA_mCherry-A2 (with 2 expression cassette sgRNA)	This paper	N/A
pFRT/LacZeo	Invitrogen TM	Cat# V601520
Software and algorithms		
EDD version 1.1.19	Lund et al., 2014	RRID:SCR_001693; https://github.com/CollasLab/edd
Chrom3D version: 1.0.2	Paulsen et al., 2017	https://github.com/Chrom3D
bwtool version 1.0	Pohl and Beato 2014	RRID:SCR_003035; https://github.com/CRG-Barcelona/bwtool
HiCNorm	Hu et al., 2012	https://github.com/ren-lab/HiCNorm
TopDom version 0.0.2	Shin et al., 2016	RRID:SCR_016964; https://github.com/HenrikBengtsson/TopDom
BEDtools version 2.29.1	Quinlan and Hall, 2010	RRID:SCR_006646; https://bedtools.readthedocs.io/en/latest/
UCSF Chimera version 1.13.1	Pettersen et al., 2004	RRID:SCR_004097; https://github.com/topics/ucsf-chimera
cutadapt version 3.4	Marcel Martin	RRID:SCR_011841; https://doi.org/10.14806/ej.17.1.200
Juicer version 1.6	Durand et al., 2016	RRID:SCR_017226; https://github.com/aidenlab/juicer
BSGenome version 1.60.0	Hervé Pagès	https://github.com/Bioconductor/BSgenome/blob/master/DESCRIPTION

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
VAST-TOOLS version 2.5.1	Braunschweig et al., 2014 Irimia et al., 2014 Han et al., 2017 Tapial et al., 2017	https://github.com/vastgroup/vast-tools
ANAT version 2.0	Yosef et al., 2011 Almozlino et al., 2017	http://www.cs.tau.ac.il/%7Ebnet/ANAT
Cytoscape	Saito et al., 2012	RRID:SCR_003032; https://github.com/cytoscape
DAVID version 6.8	Huang et al., 2009a, 2009b	RRID:SCR_001881; david.ncifcrf.gov
ImageJ	Schneider, 2012	RRID:SCR_003070; https://imagej.nih.gov/ij/
Imaris version 9.5.1	Bitplane, Inc.	RRID:SCR_007370; https://imaris.oxinst.com/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Gil Ast.

Materials availability

All unique/stable materials generated for or used within this study are available upon request to the Lead Contact.

Data and code availability

High throughput sequencing data generated in this study by ChIP-seq of lamin A/C and processed data files were deposited at GEO under accession number GEO: GSE174396. RT-PCR original gel images have been deposited to Mendeley Data: <https://doi.org/10.17632/mtsbg6brmw.1>.

All bioinformatics pipelines used to analyze the data are described in the relevant STAR Methods sections. This paper does not report original code.

Any additional information required to reanalyze the data reported in this paper is available from the Lead Contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell culture

Flp-In-HEK293 (Invitrogen) and HEK293 (ATCC) cells were cultured in complete Dulbecco's Modified Eagle's Medium (DMEM) (Biological Industries), K562 (ATCC) and GM12878 cells (Coriell Institute) were maintained in RPMI 1640 Medium (Biological Industries). Medium was supplemented with 10% fetal bovine serum (Biological Industries), 2cmg/mL L-alanyl-L-glutamine (Biological Industries), 100 U/mL penicillin and 0.1cmg/mL streptomycin (Biological Industries). Cells were grown at 37°C in a humidified atmosphere with 5% CO₂.

METHOD DETAILS

Chromatin immunoprecipitation (ChIP)

Approximately 1 × 10⁷ GM12878 or K562 cells per sample were harvested, resuspended at a concentration of 1 × 10⁶ per ml medium, and crosslinked in 1% formaldehyde for 10 min at 37°C with mixing. The crosslinking reaction was quenched by the addition glycine at a final concentration of 125 mM and incubation for 5 min at room temperature. Cells were washed twice with PBS, centrifuged, and pellets were frozen at −80°C.

Pellets were thawed on ice for at least 15 min, resuspended in ice-cold Hi-C Lysis Buffer (0.2% NP-40, 10 mM NaCl, 10 mM Tris-HCl, pH 8, complete protease inhibitor cocktail (Roche)), and incubated on ice for at least 15 min. Nuclei were centrifuged, the supernatant was discarded, and nuclei were washed with ice-cold Hi-C Lysis Buffer and centrifuged. Nuclei pellets were re-suspended in RIPA FA buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM EDTA, 0.1% SDS, and complete protease inhibitor cocktail (Roche)) and incubated for 10 min. Samples were sonicated with a Vibra-Cell VCX600 (Sonics & Materials) for 75 min total (2.2 s ON, 9.9 s OFF, 40% amplitude) to obtain DNA fragments with lengths averaging 150–350 nt (maximum 500 nt). After centrifugation at 20,000 g for 8 min, the supernatant was diluted 1:10 with Dilution Buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl, pH 8.1, 167 mM NaCl, and complete protease inhibitor cocktail (Roche)) and aliquoted.

Chromatin from an equivalent of 10 × 10⁶ cells was used per immunoprecipitation reaction. For lamin A/C immunoprecipitation, 104 μl of protein-A and 104 μl of protein-G Dynabeads (Invitrogen) were washed twice with RIPA buffer (0.1% deoxycholate, 0.1% SDS, 1% Triton X-100, 10 mM Tris-HCl, pH 8.1, 1 mM EDTA, 140 mM NaCl), resuspended in ChIP Blocking Buffer (PBS, 0.5% TWEEN,

0.5% BSA), and incubated with 50 μg of anti-Lamin A/C antibody (Santa-Cruz Biotechnology, sc-7292) for at least 2 h at 4°C with rotation. Sonicated chromatin was added to the conjugated beads, and samples were incubated for 16 h at 4°C. The beads were washed six times with RIPA buffer, twice with RIPA-high salt buffer (0.1% deoxycholate, 0.1% SDS, 1% Triton X-100, 10 mM Tris-HCl, pH 8.1, 1 mM EDTA, 360 mM NaCl), twice with LiCl wash buffer (250 mM LiCl, 0.5% NP-40 (Sigma IGEPA), 0.5% deoxycholate, 1 mM EDTA, 10 mM Tris-HCl, pH 8.1) and twice with TE buffer (10 mM Tris-HCl, pH 8.1, 1 mM EDTA). DNA was eluted from the beads with Elution buffer (0.5% SDS, 300 mM NaCl, 5 mM EDTA, 10 mM Tris-HCl, pH 8.1) in a 30-min incubation in a thermo-shaker at 65°C.

From this stage on, input tubes were processed similarly to elution tubes. To the input and elution samples was added 1 μl of 10 mg/mL RNase A (Sigma), and samples were incubated for 30 min at 37°C. Next, 1.5 μl Proteinase K (NEB) was added, and samples were incubated for 16 h at 65°C. DNA was purified using phenol:chloroform:isoamyl extraction. DNA was eluted in 30 μl Elution buffer (10 mM Tris-HCl pH 8.5). DNA concentration was measured using a Qubit assay (ThermoFisher).

3D Genomic models

GPSeq (Girelli et al., 2020), TSA-seq (Chen et al., 2018), Dip-C (Tan et al., 2018), Hi-C and lamin A/C binding peaks datasets are listed in Table S2. Briefly, GPSeq relies on gradual restriction digestion of chromatin from the nuclear lamina toward the nuclear center, and TSA-seq relies on an immunochemical method to map chromatin segments relative to nuclear speckles. Lastly, Dip-C relies on transposon-based whole-genome amplification, and the spatial distances deduced are not relative to nuclear structures. The Chrom3D models were created in the following steps: for our lamin A/C ChIP-seq data in K562 and GM12878 cells: lamin A/C broad enrichment domain peaks were identified using EDD (version 1.1.19) (Lund et al., 2014) with an 8 kb bin size and a gap penalty of 3. The resolutions of the Hi-C matrices were 5 kb for GM12878 cell data and 10 kb for K562 and HeLa cell data. The K562 TAD list was devoid of chromosome 9 due to a chromosomal aberration, thus the Chrom3D model of K562 lacks this chromosome. Chrom3D running parameters were set according to the protocol reported in (Paulsen et al., 2017) (<https://github.com/Chrom3D>) as follow: *Chrom3D -y 0.15 -r 5.0 -n 2000000 -nucleus*. Mean GC content of TADs was extracted using bwtool (version 1.0, *bwtool summary*). GC content was normalized using HiCNorm (Hu et al., 2012) for GM12878 matrices at 50kb resolution (Rao et al., 2014). TADs were identified using TopDom (version 0.0.2) (Shin et al., 2016), and were divided into five equal groups according to their mean GC content (with R *bin()* function). Subcompartment annotations (A1, A2, B1, B2, B3, and B4) were obtained from Rao et al. (2014) (GSE63525) and added to the TAD list of the GM12878 cells using bedtools (*bedtools intersect*). An in-house Perl script was used to add colors to the GC content groups and the subcompartment annotations. Visualization of the Chrom3D spatial models was done using UCSF Chimera (version 1.13.1) (Pettersen et al., 2004).

Euclidian distances from nuclear center were calculated by transforming the Cartesian spatial coordinates (x, y, z) of the Chrom3D and Dip-C models into distances from the center (0, 0, 0) using the formula: $d = \sqrt{x^2 + y^2 + z^2}$: Radial scopes from nuclear center (or speckles, in the case of TSA-seq) were defined by dividing the Euclidian distances into five groups of intervals with R *bin()* (*method = "content"*) function.

To examine the gene architecture in each spatial region, human Ensembl gene annotations (hg19 assembly) were downloaded from the UCSC table browser (genome.ucsc.edu) and the coordinates of exons and introns were extracted. Duplicated sequences as well as the first and last exons were removed. Annotations of subcompartments and radial scopes per spatial modeling method were added using bedtools (*bedtools intersect*, (Quinlan and Hall, 2010)). GC percent data in the human genome (hg19 assembly) was downloaded from the UCSC website (<https://genome.ucsc.edu/>). Exon-intron GC content profiles across exon-intron structures were extracted using bwtool (*bwtool aggregate 75:150*, (Pohl and Beato, 2014)).

LMNB1 KO versus WT 3D Genome models and RNA-seq analysis

Raw sequences were downloaded from the Sequence Read Archive (SRA). Reads were trimmed using cutadapt (version 3.4, <https://doi.org/10.14806/ej.17.1.200>), with *-q 30,30*. Subsequently, we used the Juicer pipeline (Juicer version 1.6; BWA 0.7.17; java version 1.8.0_191, (Durand et al., 2016), with default parameters, restriction enzyme Mbol and hg19 assembly files) to process the Hi-C reads and generate the Hi-C contact maps (in .hic format) after merging the biological replicates (using mega.sh, Juicer pipeline). Contact domains were called using Arrowhead (Juicer pipeline, Juicer version 1.6, *juicer_tools* version 1.22.01, using default parameters). Contact maps were extracted at 50 kb (intrachromosomal) and 1Mb (interchromosomal) (*juicer_tools.jar dump command + in-house bash and R scripts* to get the files into the genome-wide contact matrix sparse three-column format required for preprocessing using the Chrom3D preprocessing scripts). The genome-wide contact maps were then preprocessed and the gtrack input files were generated for Chrom3D using the scripts provided by Paulsen et al., 2017, (Chrom3D, version: 1.0.2). Lamin A ChIP-Seq was integrated while processing the contact maps for Chrom3D to generate the input gtrack file (files provided in BED format via GEO, see also Table S2). Chrom3D was run using different seed values to generate 1000 models (thereby ensuring that the models were generated starting from different random chromosomal conformations) using the following parameters: *Chrom3D -y 0.15 -r 5.0 -n 2000000 -l 10000 -seed i-nucleus*. The output file contains all relevant information for the models including the x, y, z -coordinates per TAD (represented via beads) for each model. The Euclidean distances (d) were calculated using the Cartesian coordinates (x, y, z) from the center (0, 0, 0) using the formula $d = \sqrt{x^2 + y^2 + z^2}$ for each bin per model, and then averaged over all 1000 models. The mean GC content of TADs was obtained using the BSgenome package in R (BSgenome_1.60.0, BSgenome.Hsapiens.UCSC.hg19 version 1.4.3, (DOI:10.18129/B9.bioc.BSgenome) Bioconductor (Huber et al., 2015) and required packages, in-house R script).

We obtained RNA-seq datasets of LMNB1 KO and WT in MDA-MB-231 cells from the SRA (see [Table S2](#)). Reads were processed using cutadapt (version 3.4, <https://doi.org/10.14806/ej.17.1.200>), with $-m20$ and $-q 30,30$. AS and gene expression analyses were performed using the Vertebrate Alternative Splicing and Transcription Tools (VAST-TOOLS) v.2.5.1 ([Braunschweig et al., 2014](#); [Han et al., 2017](#); [Irimia et al., 2014](#); [Tapial et al., 2017](#)). Reads were aligned to the hg19 genome assembly (VASTDB: [vastdb.hsa.23.06.20](#)) using vast-tools' *align* command with default parameters and the $-expr$ option. Biological replicates were merged using vast-tools' *merge* command with $-expr$. PSI values of ES and IR events, as well as cRPKM values of genes were taken from the vast-tools' *combine* output tables. Constitutive AS events which correspond to those with $PSI < 5$ (for IR) or $PSI > 95$ (for ES) were excluded, as well as cryptic events, which correspond to those with $PSI > 95$ (for IR) or $PSI < 5$ (for ES) in all samples. In addition, lowly expressed genes with cRPKM values of less than 2 were excluded in each sample. To determine the shift in the spatial location of TADs following LMNB1 KO, annotations of TADs from LMNB1 KO and WT samples with their Euclidean distances were intersected using bedtools (version 2.29.1) *intersect* command with $-wa -wb -f 0.6$, and delta distances (WT - KO) were calculated. Overlapped TADs were divided to five groups according to scopes of Euclidean distances from the nuclear center of the WT sample (with R *cut* function) and the mean delta distance per scope was calculated. Two-tailed paired t tests were performed comparing the mean Euclidean distances of the two samples per scope. We then assigned the delta distances to the PSI and the cRPKM tables using bedtools (version 2.29.1) *intersect* command with $-wa -wb -f 1.0$. We divided the ES and IR events to smaller groups based on intervals of delta distance following KO and performed Pearson correlation tests between the mean delta distance and the mean delta PSI of the groups. The same process was done for the gene expression analysis.

Exon skipping and intron retention spatial distribution analysis

A catalog of human AS events (hg19) was obtained from the VastDB ([Tapial et al., 2017](#)) (<https://vastdb.org.eu>). In addition, annotations of single-nucleotide polymorphisms that cause donor disruption (5' SS mutations) were obtained from Shiraiishi et al., 2018 ([Shiraiishi et al., 2018](#)). Subcompartment annotations as well as radial scopes per spatial modeling method were added using bedtools (*bedtools intersect*). Fold enrichment of exon skipping and intron retention events in each radial scope and subcompartment were calculated as ratios of observed counts to expected counts. The expected counts were the number of ES or IR events in each spatial region based on its proportion in the dataset. Hypergeometric tests were applied (with R *dhyper()* function).

To examine whether the nuclear spatial distributions of ES and IR events are determined by exon-intron architecture rather than spatial location, we calculated the observed to expected ratios of ES and IR events that manifest the leveled and the differential architectures in different nuclear spatial regions. The expected counts were based on the proportions of the two architectures per scope. The leveled and the differential architectures ($n = 9828$ and 10168 , respectively) were defined by the exon-intron GC content differential levels of alternative exons (obtained from VastDB ([Tapial et al., 2017](#)) and their flanking introns, as described previously ([Gelfman et al., 2013](#))).

RNA FISH

Stellaris® FISH Probes were designed against the nascent RNA of the endogenous genes *NDUFS4*, *DDX41*, *ANKIB1* and *INTS1* in GM12878 cells, and against the nascent RNA of the integrated minigenes *FRG1* and *CORO1B* and of their plasmid UTRs, and against the *Zecin* cDNA in HEK293 cells, using the Stellaris® RNA FISH Probe Designer (version 4.2) ([Table S3](#)). It is important to note that we previously demonstrated that the *FRG1* minigene is co-transcriptionally spliced in HEK293 cells ([Leader et al., 2021](#)), thus the site of transcription and splicing co-localize. Probes were labeled with Quasar-570 (Biosearch Technologies). RNA FISH experiments with Stellaris® probes were performed following the manufacturer's instructions for adherent cells (<http://www.biosearchtech.com/support/resources/stellaris-protocols>). Cells were seeded on 10-mm coverslips and grown for 48 h followed by fixation with 3.7% formaldehyde solution (Sigma-Aldrich) in PBS for 10 min. After washing twice with PBS, cells were permeabilized using 70% ethanol for 1 h at 4°C. Cells were then incubated twice for 5 min in SSC (10% formamide in 2 × saline sodium citrate). Samples were incubated cell-side down with 125 nM RNA probe in 100 μl hybridization buffer (10% formamide, 2 × SSC, and 10% dextran sulfate) at 37°C sealed with parafilm in a humidified dark chamber overnight. Cells were transferred face up to a fresh well and washed twice in SSC at 37°C with 30-min incubations. After washing with PBS, nuclei were counterstained with Hoechst (Sigma-Aldrich) in PBS for 10 min. Coverslips were then mounted onto a microscope slide with p-phenylenediamine mounting medium. Finally, coverslips were sealed with nail polish. Images were acquired using the Cell^R system on an Olympus IX81 fully motorized inverted microscope (60 × PlanApo objective, 1.42 NA) fitted with an Orca-AG CCD camera (Hamamatsu) driven by the Cell^R acquisition software. The z-slice images were post-processed, and sites were localized and their distance from the nuclear envelope was measured in μm using Imaris (version 9.5.1) software (Bitplane, Inc.) available at <https://imaris.oxinst.com/>. The cell nuclei volumes of the HEK293 with integrated minigenes were measured by Imaris and used for normalization as the HEK293 cells manifest high variation in nuclear size. Hence, for accuracy the measured distance of each site of a minigene (μm) was divided by the volume (μm³) of their own nucleus. Distance is presented in (1/μm²).

Minigene cloning

Human genomic fragments were amplified using primers designed based on NCBI reference sequences. Two differential GC content minigenes were constructed based on exon 4 through exon 6 of *FRG1* (RefSeq: NM_004477.3) and exon 2 through exon 4 of *SR140* (RefSeq: NM_001320219.1). Eleven minigenes with leveled GC content were constructed: exon 31 through exon 33 of *PLXNB1* (RefSeq: NM_001130082.2), exon 7 through exon 9 of *PNPLA2* (RefSeq: NM_020376.3), exon 4 through exon 6 of *CORO1B* (RefSeq:

NM_001018070.2), exon 22 through exon 24 of *DHX37* (RefSeq: NM_032656), exon 14 through exon 16 of *MTA1* (RefSeq: NM_004689), exon 8 through exon 10 of *FES* (RefSeq: NM_002005), exon 33 through exon 35 of *FASN* (RefSeq: NM_004104), exon 24 through exon 26 of *PNPLA5* (RefSeq: NM_006702), exon 5 through exon 7 of *SUSD2* (RefSeq: NM_019601.3), exon 16 through exon 18 of *PLXNB2* (RefSeq: NM_012401.3), and exon 111 through exon 113 of *OBSCN* (RefSeq: NM_001271223.2). The minigenes *FRG1* and *SR140* were amplified and cloned into KpnI and BamHI restriction sites, minigenes *PLXNB1*, *PNPLA2*, *CORO1B*, *MTA1*, *FES*, *SUSD2*, *PLXNB2*, and *OBSCN* were cloned into EcoRI and BamHI sites, minigenes *DHX37* and *FASN* were cloned into EcoRI and KpnI sites, and minigene *PNPLA5* was cloned into EcoRI and *XhoI* sites. The PCR products were digested with the restriction enzymes and ligated into the vector pEGFP-C3. Primers for cloning are listed in Table S3. Four minigenes were re-cloned into pcDNA5/FRT/TO (Invitrogen) for stable transfection. *FRG1* and *SR140* were cut with KpnI and BamHI and *CORO1B* and *FES* were cut with HindIII and BamHI.

Site-directed mutagenesis

To inactivate the 5'SS, we mutated the first nucleotide of the intron (position +1), and to generate a weak 5'SS, we used a different combination of mutations at positions -2, or +3 to +6 of the 5'SS of the middle exon positions -1 and +1 are the last and first nucleotides of the exon and intron 5'SS junction. Mutations at the 5'SS of the second intron of the minigenes were selected based on the consensus 5'SS sequence (Ast, 2004) using Analyzer Splice Tool (Carmel et al., 2004). For the weak mutation, *FRG1* was mutated at positions +3G to A and +6T to A. *SR140* was mutated at positions +3G and +5G were mutated to A. *FES* was mutated at position -1G to C. *CORO1B* was mutated at position -2A to C. For the inactive mutation all minigenes were mutated at the same position +1G to A. Specific overlapping oligonucleotide primers containing the splice-site mutations were used in PCR reactions using KAPA HiFi Hot-Start ReadyMix PCR Kit (Kapa Biosystems). Primer sequences are listed in Table S3. All plasmids were confirmed by sequencing.

Manipulation of minigene introns and exons

To manipulate intron or exon lengths of minigenes cloned in pcDNA5/FRT/TO, PCR was performed using KAPA HiFi HotStart Ready-Mix PCR Kit (Kapa Biosystems) using primers containing 5' phosphate modification or restriction sites (primers are listed in Table S3). Changes in the intron/exon lengths were made in the middle of the unit and as far from the splice sites as possible. To avoid disruption of functional and predicted binding sites of splicing regulatory proteins by utilizing published methods (Lee and Ule, 2018; Schwartz et al., 2009a; Van Nostrand et al., 2016; Wheeler et al., 2018). *FRG1* introns 1 and 2 were shortened to 210 nt and 235 nt, respectively. *SR140* introns 1 and 2 were shortened to 222 nt and 197 nt, respectively. Lengthening the second intron of *FES* and *CORO1B* minigenes to 680 nt and 788 nt, respectively, was performed by inserting a 566-nt fragment with 77% GC content taken from an intergenic region with high GC content, *FES*'s intron was lengthened with a second sequence of 615-nt with 76% GC content. *FES* and *CORO1B* lengthened introns were shortened to several lengths from the middle of the inserted segment by PCR using 5' phosphorylated primers. Exons were lengthened by inserting a segment of 357 nt taken from intragenic region following shortening of the relevant exons to the desired length from the middle of this segment by PCR using 5' phosphorylated primers. For the second lengthening of the exon with a different DNA sequence, the same 357-nt segment was inserted again in the inverted direction. To generate differential GC content in the *FES* and *CORO1B* minigenes, we replaced 100 nt downstream of the 5'SS (from position +7 to +107 in intron 2) with a segment taken from a similar location of gene with differential architecture. For the second insertion of the differential segment, the same segment was inserted in the inverted direction to *CORO1B* minigene.

Generation of peripheral FRT-site using CRISPR-Cas9

In order to add a second sgRNA expression cassette to pSpCas9(BB)-2A-GFP (PX458; Addgene #48138) vector by standard molecular cloning techniques, the plasmid was digested using KpnI restriction enzyme (New England Biolabs). For the sgRNA expression cassette (insert), PCR reaction was performed on pX552 to amplify its sgRNA expression cassette sequence only, using primers with 5' tails carrying restriction digestion sites for KpnI sgRNA was amplified for cloning using KAPA HiFi HotStart ReadyMix (KAPA Biosystems), primers for amplification use are found in Table S3. The linearized plasmid and sgRNA expression cassette were ligated using T4 DNA ligase (New England Biolabs) according to manufacturer's instructions, positive colonies were selected and sequenced. The final plasmid containing both sgRNA was termed pX458-GFP-Cas9-2sgRNA.

For sgRNA design and cloning, sgRNA specific to the desired peripheral genomic location in GRCh37/hg19;chr1:172,497,478-172,497,500 with minimum off targets was chosen based on online web tools (Bae et al., 2014; Haeussler et al., 2016; Labun et al., 2019). In addition, a non-human genome targeting sgRNA was also designed; for this purpose, the sgRNA was designed against the mCherry gene. Each sgRNA was designed together with its complement sequence and nucleotides were added to the 5' end; for peripheral genomic targeted sgRNA (peripheral-sgRNA), ACC or CAA were added to the sgRNA or complement sequence respectively, to complement the SapI restriction site. For mCherry sgRNA, CACC or AAAC were added to the sgRNA or complement sequence respectively, to complement the BbsI restriction site.

Each sgRNA, oligonucleotide and complementary oligonucleotide were phosphorylated at the 5' end using T4 Polynucleotide Kinase (New England Biolabs) according to manufacturer's protocol and annealed together (thermo-cycling program: 30 min at 37°C PNK reaction, 20 min at 65°C heat inactivation, 5 min at 95°C denaturation and ramp to 4°C at 0.5°C per sec to anneal). For each ligation reaction, the double-stranded sgRNAs were diluted 1:25. First, the mCherry sgRNA insert was ligated with linearized pX458-GFP-Cas9-2sgRNA (BbsI digestion) using T4 DNA ligase (New England Biolabs) according to manufacturer's instructions

(50 ng vector with 1 μ L diluted sgRNA in a 20 μ L reaction volume). A 2 μ L aliquot of the ligation reaction was used to transform competent 10XL Gold *E. coli* after heat shock (ampicillin selection was used). Positive colonies were selected and verified by sequencing. Next, the same process was repeated for the peripheral-sgRNA and a pX458-GFP-Cas9-mCherry_sgRNA positive plasmid. pX458-GFP-Cas9 or pX458-GFP-Cas9-mCherry_sgRNA were linearized with BbsI or SapI (New England Biolabs), respectively, according to manufacturer's instructions. The final plasmid containing both sgRNAs was termed pX458-2gRNA_mCherry-A2.

To construct the donor vector, an insert was cloned into the plasmid pFRT/LacZeo (containing the FRT site) upstream to the CMV promoter. pFRT/LacZeo was double digested with KpnI and XhoI (New England Biolabs) according to manufacturer's instructions. The insert is comprised of several sequences: 40bp downstream and 40bp upstream to the genomic cleavage site at the target sgRNA which act as homology arms for the microhomology-mediated end-joining-dependent integration of the donor plasmid. In addition, between the two arms there is the mCherry sgRNA sequence with a PAM sequence where Cas9 will cleave the plasmid *in vivo*. At the ends are the sticky ends of the restriction enzymes (KpnI and XhoI) used in the cloning. For sequences see Table S3. The insert oligonucleotide and a complementary oligonucleotide were phosphorylated and annealed as the sgRNAs to get double-stranded insert. The insert was ligated with linearized pFRT/LacZeo using T4 DNA ligase (New England Biolabs) according to manufacturer's instructions. Positive colonies were selected and verified by sequencing. The final plasmid was termed pFRT/LacZeo-Periphery.

Generation of peripheral FRT-site stable expression in HEK293 cell line

Flp-In-293 cell line (Invitrogen) contains an FRT recombination site for stable integration which is found in GRCh37/hg19;chr12:1,332,498-1,332,531 (Lin et al., 2014) in HEK293 cells. According to our Chrom3D analysis this site is found closer to the nuclear center. To generate the peripheral FRT recombination site we co-transfected HEK293 cells with 4 μ g pX458-2gRNA_mCherry-A2 and 2 μ g donor plasmid (pFRT/LacZeo-Periphery). After 48 h, cells are re-seeded in 15cm plate and maintained for 2-3 weeks in complete DMEM supplemented with 400 μ g/mL ZeocinTM (InvivoGen) to select for stable colonies. The medium was replaced every 3-4 days until resistant foci were observed. Several foci (colonies) were transferred to 30-mm plates, each colony representing a monoclonal stable cell line. Genomic DNA was extracted using Tissue/Blood DNA Mini Kit (GeneAid Biotech Ltd) from the monoclonal stable cell-lines and the new FRT recombination site was verified by PCR and Sanger sequencing.

Transfection of Flp-In-293 HEK cells

Flp-In-293 HEK cells (250,000 cells) were first seeded on 6-well plates. For transient transfection: after 24 h cells were transfected with 1 μ g plasmid using 3 μ L TransIT[®]-LT1 Transfection Reagent (MirusBio), according to the manufacturer's protocol, and incubation was continued for 48 h. For stable transfection in central and peripheral integration FRT-sites: after 24 h of seeding cells were co-transfected with 500 ng Flp-In expression plasmid (pcDNA5/FRT/TO) containing cloned minigene, and 4.5 μ g of Recombinase Expression Vector (pOG44) at a ratio of 1:9. After 48 h, cells were harvested and moved to 15-cm culture dishes. Stable transfectants were selected using hygromycin B. Two weeks later genomic DNA was extracted using Tissue/Blood DNA Mini Kit (GeneAid Biotech Ltd) from monoclonal stable cell-lines and verified by PCR and Sanger sequencing.

RNA isolation RT-PCR amplification and RT-qPCR

Cells were harvested 48 h after transient transfection. For stable transfection, positive cell colonies were harvested from 6-well plates after reaching 80%–90% confluence. Total RNA extraction was performed using TRI reagent (Sigma-Aldrich). cDNA synthesis was performed with RT-FLEX (Quanta) according to the manufacturer's instructions. To specifically amplify the spliced cDNA products derived from the expressed minigenes (and not the endogenous genes), we used a set of primers that are pcDNA5/FRT/TO specific and recognize the plasmid sequence located on either side of the inserted minigene (primers are listed in Table S3). Amplification was performed using Red Load Taq Master (Larova GmbH). The products were separated in 1.5% agarose gel; each PCR product band was cut from gel, cleaned using Gel/PCR DNA Fragments Kit (GeneAid Biotech Ltd) and sequenced. RT-qPCR was performed using KAPA SYBR FAST Universal qPCR kit (KAPA Biosystems) according to the manufacturer's instructions. Alternative splicing and expression Ct levels were calculated according to (Harvey and Cheng, 2016). Primers used span the exon-exon junctions, to calculate the inclusion levels of the middle exon, and other primers that span the intron-exon junctions, to calculate IR. For the expression levels primers span constitutive exons for endogenous genes, or plasmidic sequences for integrated minigenes. Primers are listed in Table S3.

Exon dataset construction for the eCLIP-seq analysis

Human Ensembl gene annotations (hg19) were downloaded from the UCSC table browser (<https://genome.ucsc.edu/cgi-bin/hgTables>). Annotations of unplaced scaffolds, mitochondrial chromosomes, and haplotypes on the MHC region of chromosome 6 were removed. In addition, transcripts with less than four exons were removed, leaving only those with more than one internal exon. Using in-house Perl scripts, flanking intron coordinates and lengths were extracted, and first and last exons as well as duplicates were removed. To examine the GC content of the exon-intron structures, we used GC percent data in the human genome (hg19 assembly) from the UCSC genome browser (<https://genome.ucsc.edu/>). The bwtool (Pohl and Beato, 2014) (*bwtool matrix 100:100:100*) was used to extract the GC content of the exon-intron structures. The exon-intron GC content differential level was calculated as $\%GC_{exon} - \%GC_{downstream\ intron} (\leq 100bp)$. Subcompartment annotations, the Euclidian distances from nuclear center in

K562 cells, and TSA-seq percentiles defining a nuclear lamina-to-speckle axis (Chen et al., 2018) were added using *bedtools* (*bedtools intersect*). Transcript quantification data (hg19 assembly) derived from poly A plus RNA-seq of K562 cells was downloaded from the ENCODE web portal (<https://www.encodeproject.org/>, accession ID: ENCFF322HPV) and added to the exon list using an in-house R script. We included only transcripts expressed in K562 cells (FPKM > 0) and excluded introns shorter than 30 nt. The final dataset comprises 122,608 internal exons.

eCLIP-seq analysis

eCLIP-seq datasets for 55 experimentally validated splicing factors in K562 cells were obtained from the ENCODE web portal (<https://www.encodeproject.org/>, (Consortium, 2012)) as BED files of binding peaks (hg19 assembly). The binding of each SF was examined in a window of 75 nt taken from both sides of each exon and the adjacent 100 nt of its flanking introns. In cases where introns were shorter than 100 nt, we used their full length instead (in house Perl script). We used *bedtools* (*bedtools intersect* with *-c* and *-s* specifications) to overlap the extended exon coordinates with the peak coordinates for each eCLIP assay (Table S4). The relative distances from nuclear center of each set of exons bound by each SF were determined using the following formula:

$$\text{Percentage change}(d_S, d_T) = \frac{d_S - d_T}{d_T} \times 100$$

Where *d* stands for mean Euclidian distance (μm), and *S* and *T* represent the set of bound exons and total exons in the dataset ($n = 122,608$), respectively. We next applied two-tailed Wilcoxon rank sum tests (with R *wilcox.test()* function) to compare the Euclidian distances of each set of exons with the Euclidian distances of total exons. *P*-values were adjusted using the Benjamini-Hochberg False Discovery Rate (FDR) correction (with R *p.adjust()* function). SFs that bind sets of exons significantly close to or remote from the nuclear center ($FDR < 0.05$) were labeled as cSFs ($n = 27$) or pSFs ($n = 15$), respectively. The rest of the SFs ($n = 13$) were labeled as Unbiased.

The enrichment factor of each set of exons per: i) concentric radial scope from nuclear center, ii) nuclear subcompartment, and iii) TSA-seq percentile, was calculated as the ratio between the observed and the expected counts. The expected number of bound exons per spatial region was calculated based on its proportion in the dataset. Chi-Square goodness of fit tests (with R *chisq.test()* function) were used to compare the observed distribution with the expected probability distributions. *P*-values were adjusted using the FDR correction (with R *p.adjust()* function). We applied the two-tailed Wilcoxon rank sum test (with R *wilcox.test()* function) to compare the exon/intron lengths and mean GC percent values between each set of bound exons and the values obtained from total exons in the dataset. *P*-values were adjusted using the FDR correction (with R *p.adjust()* function).

A Pearson correlation test was used (with R *cor.test()* function) to examine the association between the mean exon-intron GC content (%) differential levels and the mean Euclidian distance (μm) from nuclear center of the sets of exons that are bound by the SFs.

To examine whether the binding of the SFs is directed by the GC content architecture rather than the spatial location of the transcribed pre-mRNA, we computed the enrichment of exon-intron structures with the leveled versus the differential architectures among those that are bound by cSFs and pSFs (10 most proximal and distal to/from the nuclear center, respectively) and the unbiased group. The leveled and the differential architectures ($n = 24373$ and 24405 , respectively) were defined according to exon-intron GC content differential level as described previously (Gelfman et al., 2013). The enrichments were calculated as the observed/expected ratio. Expected counts of each architecture were based on the proportions of the two architectures in each scope. Hypergeometric tests were applied (with R *dhyper()* function).

Central and peripheral subnetworks construction

The human protein-coding gene list was obtained from BioMart (Ensemble's data-mining tool). Active genes (cRPKM > 0) in K562 cells were filtered using gene expression data from the VastDB (Tapial et al., 2017) (<https://vastdb.org.eu/wiki/Downloads>). We used ANAT 2.0 (Almozilino et al., 2017; Yosef et al., 2011), a Cytoscape (Saito et al., 2012) plug-in for network analysis (<http://www.cs.tau.ac.il/~bnet/ANAT>) with the "general" non-anchored setting and default parameters to construct the most probable subnetworks created by the SFs that were classified to the central and the peripheral groups. The human protein-protein interaction network provided by ANAT was reduced to the active protein-coding genes in K562 cells ($n = 16,291$) for a total of $\sim 415\text{k}$ interactions. FTO was excluded from the central network, as it was isolated from the rest. The resulting central subnetwork consists of 32 proteins, with the peripheral network including 19 proteins.

Subnetworks; separation analysis

We used the human protein-protein interactions available through ANAT to test the degree of separation between the central and the peripheral subnetworks. HNRNPK was excluded from this analysis because it exists in both subnetworks, resulting in a unified splicing-related network of 376 interactions among 49 proteins. To evaluate the degree of separation between the two subnetworks, we counted the number of interactions between them (165) and compared this count to those obtained in 1000 randomized networks on the 49 splicing-related proteins that preserve node degrees. The number of separating edges was smaller in only 14 of the randomized cases, indicating that the subnetworks are well separated ($p < 0.015$).

Gene ontology (GO) enrichment analysis

SFs from the peripheral and the central subnetworks were used for the GO enrichment analysis for cellular component terms (*GO-TERM_CC_DIRECT*), compared to the *Homo sapiens* background gene list. SFs originally classified as central and pSFs that were added to the opposite subnetwork were excluded. Analyses were performed using DAVID version 6.8 (Huang et al., 2009a, 2009b) (<https://david.ncifcrf.gov>). Fisher's exact test was used to measure the gene-enrichment in annotation terms. GO terms with *FDR* < 0.05 are presented.

Evolutionary history and gene family size analysis

For ancestry-calling, consensus ages were identified for each splicing factor using publicly available ortholog databases collected from (Litman and Stein, 2019). Following phylostrata categories from Liebeskind et al. (Liebeskind et al., 2016), clades were collapsed into five different groups: Last eukaryotic common ancestor group (which encompassed age categories from eubacteria, bacteria, and all their descendants to unikonta), Unikonta (which comprises age categories from opisthokonta to choanozoa), Eumetazoa (which comprises age categories from eumetazoa to bilateria), Chordata (which comprises age categories from chordata to vertebrata) and Vertebrata. A consensus age was then calculated using the modal value of the age estimates in all ortholog databases. For paralog-calling, gene orthology clusters were used. Briefly, each splicing factor was manually searched in the orthology clusters and paralogs in the clusters were counted. A curated ohnolog list from Touceda-Suárez et al. (Touceda-Suárez et al., 2020) was employed for ohnolog calling.

QUANTIFICATION AND STATISTICAL ANALYSIS

Details regarding the statistical tests, sample sizes, and biological replicates are indicated in the figure legends. More information on how the analyses were performed can be found in the Methods section. The statistical significance threshold was set to 0.05.

Supplemental information

**Gene architecture directs splicing outcome
in separate nuclear spatial regions**

**Luna Tammer, Ofir Hameiri, Ifat Keydar, Vanessa Rachel Roy, Asaf Ashkenazy-
Titelman, Noélia Custódio, Itay Sason, Ronna Shayevitch, Victoria Rodríguez-Vaello, José
Rino, Galit Lev Maor, Yodfat Leader, Doha Khair, Erez Lieberman Aiden, Ran
Elkon, Manuel Irimia, Roded Sharan, Yaron Shav-Tal, Maria Carmo-Fonseca, and Gil Ast**

Supplemental Information

Gene architecture directs splicing outcome in separate nuclear spatial regions

Luna Tammer, Ofir Hameiri, Ifat Keydar, Vanessa Rachel Roy, Asaf Ashkenazy-Titelman, Noélia Custódio, Itay Sason, Ronna Shayevitch, Victoria Rodríguez Vaello, José Rino, Galit Lev Maor, Yodfat Leader, Doha Khair, Erez Lieberman Aiden, Ran Elkon, Manuel Irimia, Roded Sharan, Yaron Shav-Tal, Maria Carmo-Fonseca, and Gil Ast

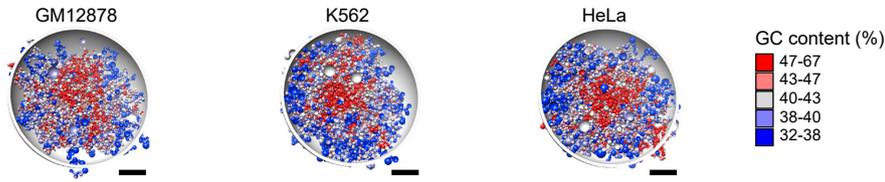
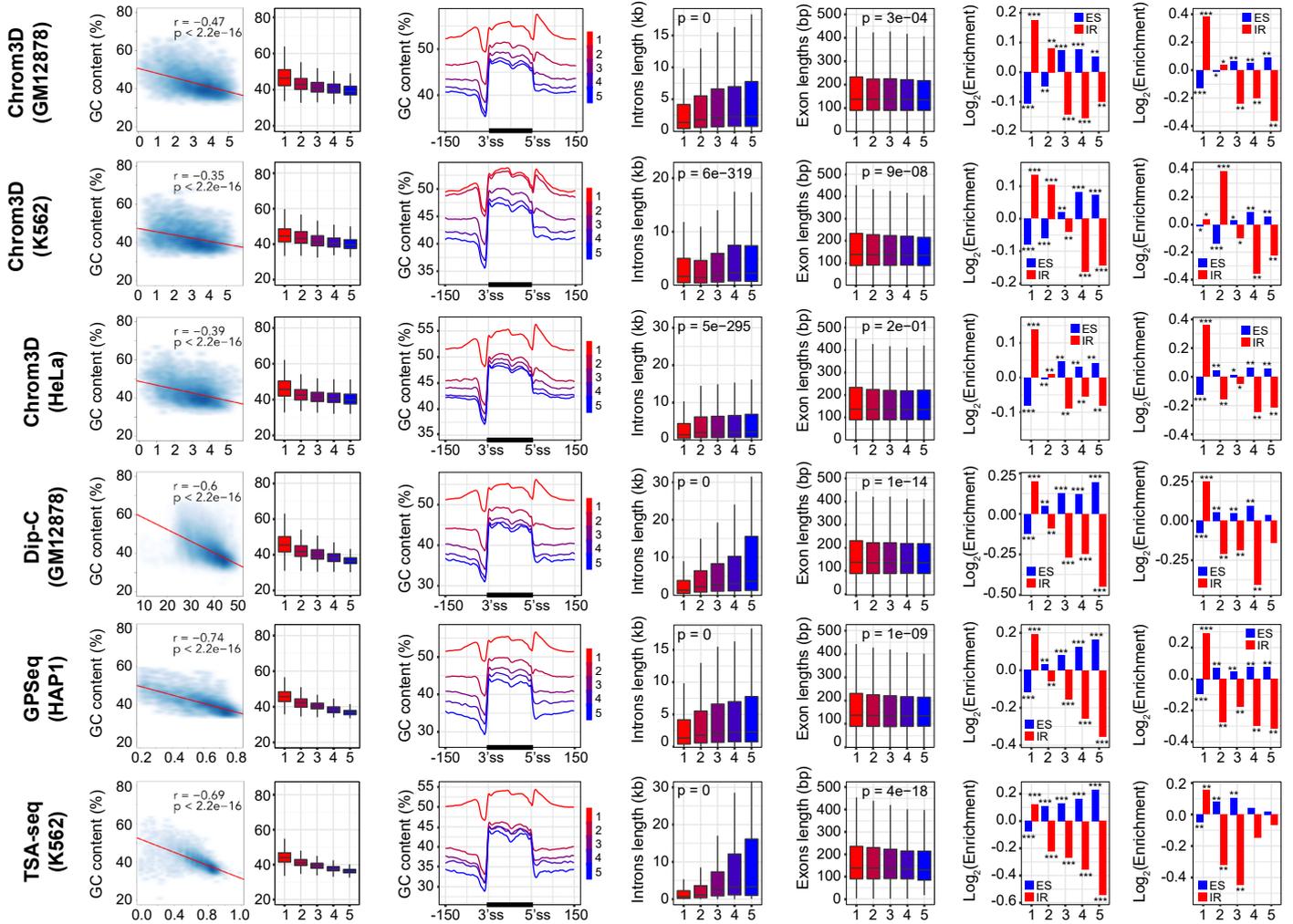
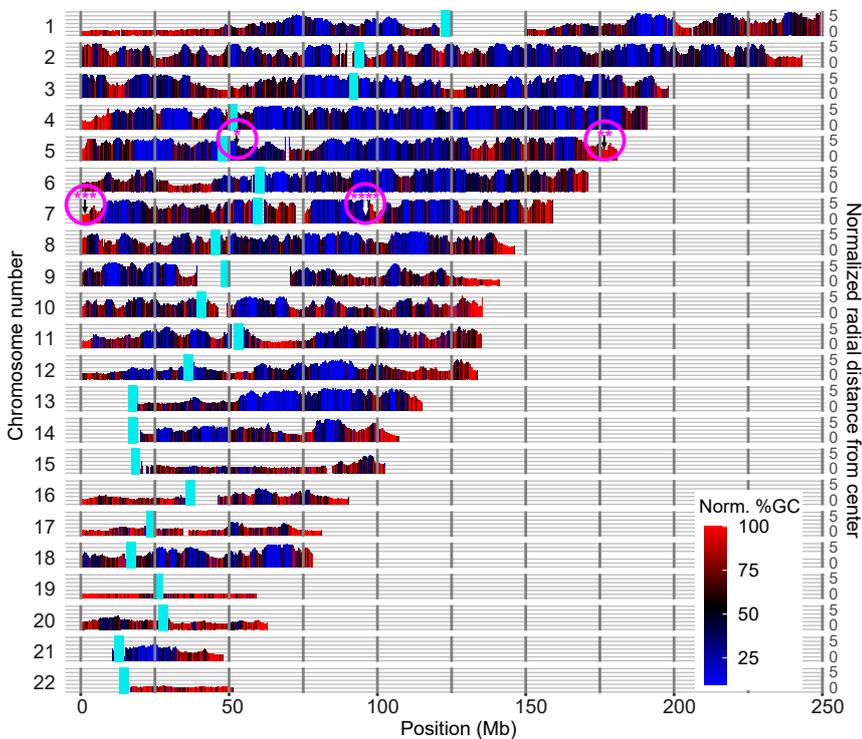
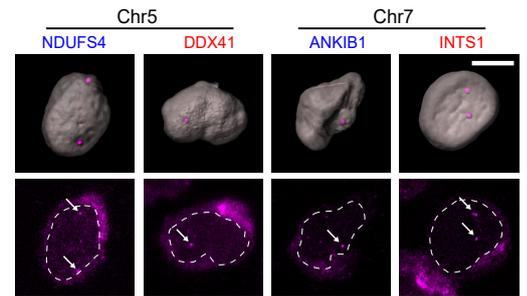
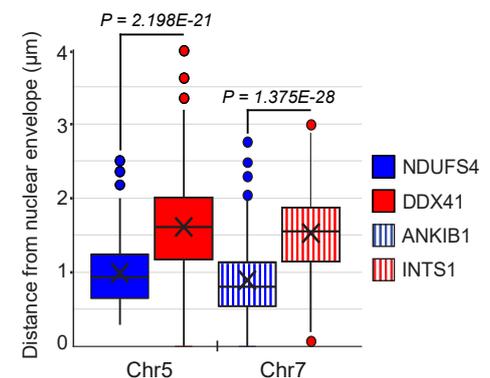
A**B****H****I****J**

Figure S1. The spatial distribution of GC content in interphase nuclei. Related to Figure 1.

(A) Nuclear Chrom3D models depicting the mean normalized GC content (%) of TADs (shown as spheres) in interphase nuclei. Sphere size corresponds to the amount of DNA. Black bars represent a scale of 2 μm . (B) Left: Pearson correlation tests between the GC content (%) of whole genome chromatin segments and the distance from the nuclear center (Chrom3D, GPSeq, Dip-C) or from nuclear speckles (TSA-seq) (see also Table S1); Right: Box plots depicting the GC content (%) distribution in each radial scope (1-to-5) from the nuclear center (Chrom3D, GPSeq, Dip-C) or from nuclear speckles (TSA-seq). (B) to (G) Features analyzed per concentric radial scope (x-axis; 1-to-5 representing center-to-periphery for Chrom3D, Dip-C, and GPSeq, or lowest-to-highest proximity to the speckles for TSA-seq): (C) Mean GC content (%) in a sliding window across exons (represented by 75 nt on each end) and the adjacent 150 nt of the flanking introns, excluding splice site regions (20 nt from the 3' end of the upstream introns, the first 6 nt at the 5' end of the downstream intron, and the first 2 and the last 3 nt of the exonic sequences). (D) Intron lengths distribution. (E) Exon lengths distribution. (D) and (E) P-value correspond to one-way ANOVA tests. (F) Fold enrichment of ES ($n = 33883$) and IR ($n = 18496$) events obtained from VastDB. (G) Fold enrichment of cancer-related 5'SS mutations that lead to ES ($n = 4290$) and IR ($n = 1225$) events (Jayasinghe et al., 2018). (F) and (G) P-values correspond Hypergeometric tests: $*(p \leq 0.05)$, $** (p \leq 0.01)$, $*** (p \leq 1e-10)$. (H) Bar plots depicting the nuclear radial distance from the center and mean GC content (%) along the indicated chromosomes (1Mb resolution). Values of the radial distances from nuclear center are the means of 48 single cells (GM12878) and were determined by the Dip-C modeling method. Centromeres are marked by light blue rectangles. Asterisks mark the positions of genes manifesting the leveled and the differential architectures as follows: $* NDUFS4$, $** DDX41$ (chromosome 5) and $*** INTS1$, $**** ANKIB1$ (chromosome 7). (I) Representative RNA FISH images of fixed GM12878 cells hybridized with probes that specifically detect genes with differential or leveled architectures. The predicted radial distances of the selected genes were determined using Dip-C data (GM12878, Figure S1H). The selected genes are as follow: $NDUFS4$, $DDX41$ (chromosome 5) and $ANKIB1$, $INTS1$ (chromosome 7). Genes with the differential and the leveled architectures are colored in blue and red, respectively. (J) Box plots depicting the distances of the indicated genes from the nuclear envelope ($n > 160$ nuclei). Mean distance per gene is marked by X. P-values correspond to two-tailed t-tests.

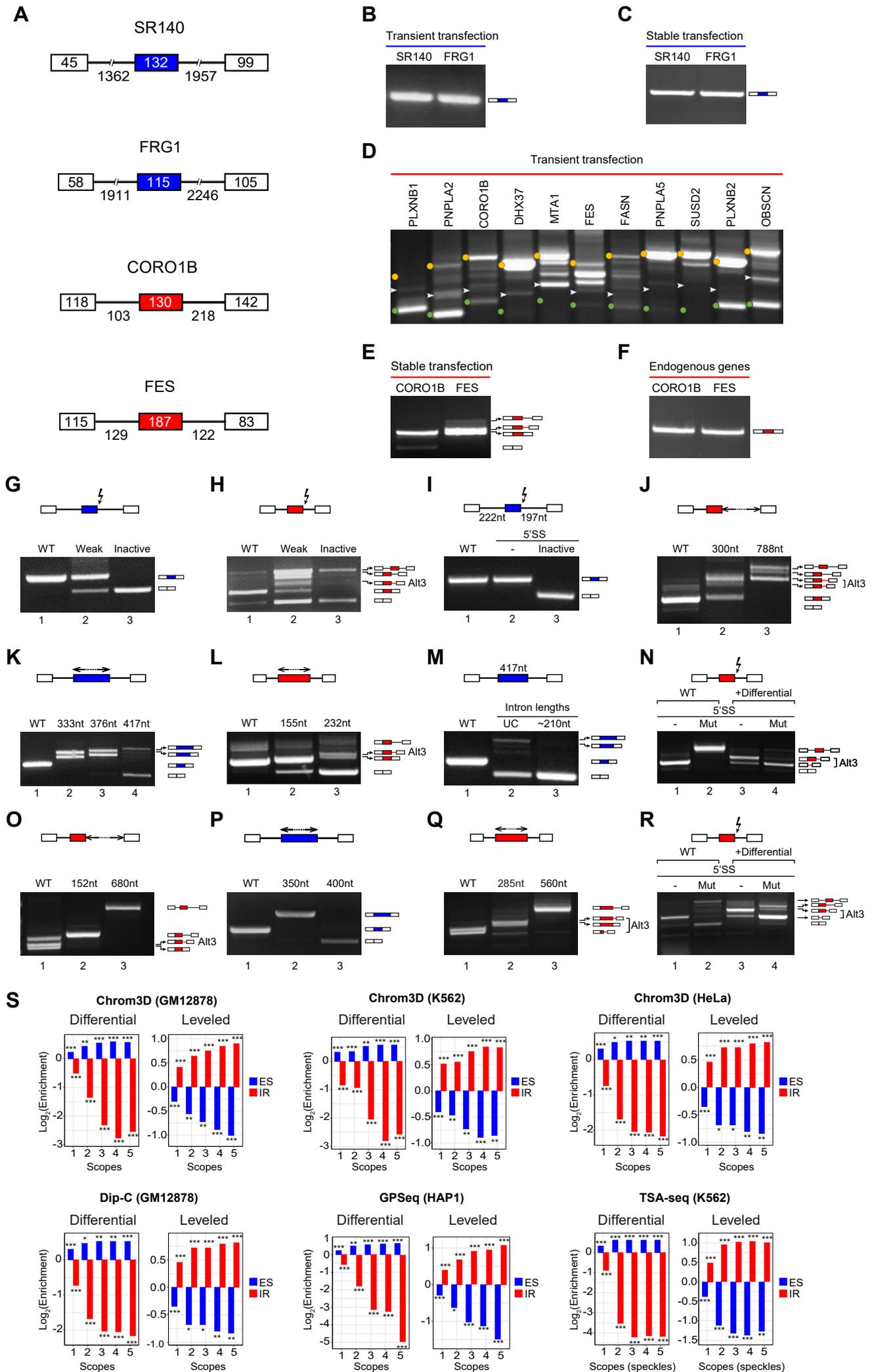


Figure S2. Nuclear spatial organization of exon-intron gene architectures and alternative splicing outcomes. Related to Figure 2.

(A) Minigenes composed of three exons and two introns cloned from human genomic regions with differential architecture (*FRG1* and *SR140*, colored in blue) and leveled architecture (*CORO1B* and *FES*, colored in red). Boxes indicate exons, lines represent introns; lengths of exons and introns of the wild-type (WT) minigenes are as indicated in nucleotides. (B) to (R) each minigene was introduced to 293HEK cells, total RNA was extracted, desired regions were amplified by RT-PCR, and products were separated on 1.5% agarose gels. The mRNA products are shown schematically on the right side of each gel image. Boxes define exons, black lines define retained introns and red lines represent alternative 3'SS. Splicing patterns of *SR140* and *FRG1* minigenes with the differential architecture when transiently transfected (B) or stably transfected (C) into the central integration FRT-site. (D) Splicing patterns of minigenes containing sequences from eleven human genes with the leveled architecture were transiently transfected. Arrowheads indicate the exon inclusion isoforms. Yellow dots indicate the unspliced isoforms and green dots indicate the location of ES isoform. (E) Stably transfected *CORO1B* and *FES* minigenes with the leveled architecture. (F) Splicing of the endogenous *CORO1B* and *FES* genes (the same three exons). Panels (G) to (N) are similar analyses as in panels (B) to (I) in Figure 2, but with the second set of minigenes (*SR140* represents the differential, and *CORO1B* the leveled minigene) that were stably integrated into the central integration FRT-site. (G) and (H) The splicing patterns of the differential (G) and the leveled (H) minigenes with the WT 5'SS or the mutated 5'SS sequence of the middle exon (weak or inactive 5'SS as indicated), see Methods for the exact mutations performed. (I) The splicing patterns of the differential minigenes before (*lane 1*) and after shortening the upstream intron to 222 nt and the downstream intron to 197 nt (*lanes 2 and 3*), and after inserting a 5'SS inactive mutation to the minigene with the shortened introns (*lane 3*). Intron length in the second lane was unchanged (indicated as UC). (J) The splicing patterns of the leveled minigenes before and after lengthening the second intron as indicated. (K) and (L) The splicing patterns of differential (K) and leveled (L) minigenes before and after lengthening of the middle exon as indicated. (M) The splicing patterns of the differential minigenes before (*lane 1*) and after lengthening the middle exon to 417 nt (*lanes 2 and 3*) and after shortening both introns to the same indicated lengths in panel I (*lane 3*). (N) The splicing patterns of the leveled minigenes before (WT: *lanes 1 and 2*) and after replacement of a DNA segment with lower GC content located downstream of the 5'SS of the second intron (+differential: *lanes 3 and 4*), and with or

without a 5'SS mutation (minus (-) represents the WT 5'SS and Mut represents a weak 5'SS mutation). **(O)** This panel is identical to Figure 2F, except that a different DNA sequence for exon lengthening was used. **(P)** This panel is identical to Figure 2E, except that a different DNA sequence was used for intron lengthening. **(Q)** This panel is identical to Figure 2G, except that a different DNA sequence for exon lengthening was used. **(R)** This panel is identical to Figure 2N, except that a different DNA sequence was inserted in the downstream intron. **(S)** We defined the differential and leveled architectures for exons with the highest and lowest GC content compared to their flanking introns, respectively (Gelfman et al., 2013). Enrichment patterns of ES and IR events (VastDB) among exon-intron structure with the leveled or differential architectures per concentric radial scope from the nuclear center (Chrom3D, TSA-Seq, Dip-C) or nuclear speckles (TSA-Seq). P-values correspond Hypergeometric tests: $*(p \leq 0.05)$, $** (p \leq 0.01)$, $*** (p \leq 1e-10)$.

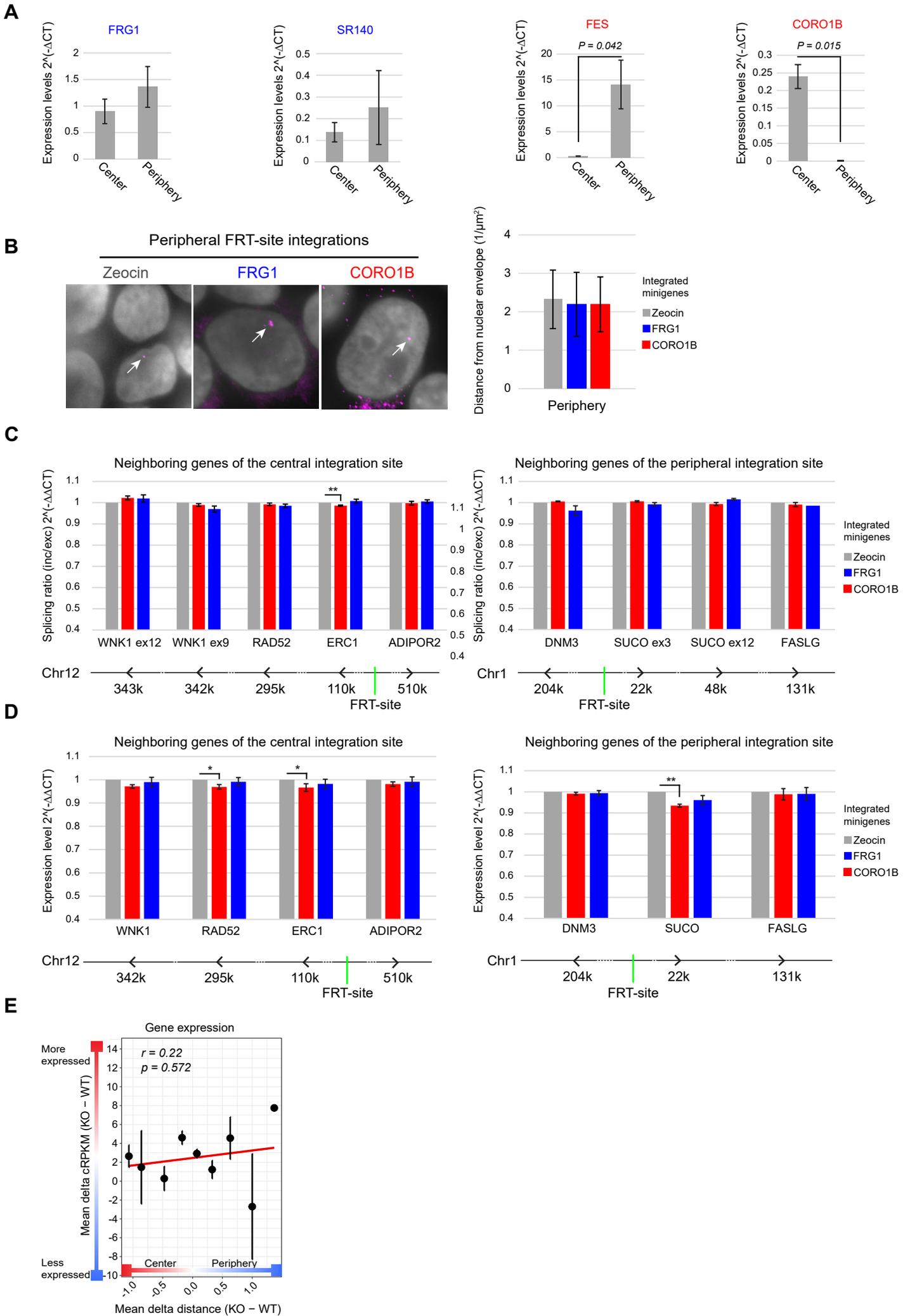


Figure S3. The effect of nuclear spatial organization on alternative splicing and gene expression. Related to Figure 3.

(A) Bar plots showing the gene expression levels of the minigenes with (left) the differential architecture (*FRG1*, *SRI40*) and (right) the leveled architecture (*FES*, *CORO1B*). All four minigenes were inserted into the integration sites found in the center and the periphery of the nucleus. Gene expression level is normalized to the housekeeping gene *ACTB*. (B) Left: Representative RNA FISH images of fixed HEK293 cells hybridized with probes that detect the *Zeocin* transcribed gene (control conditions before minigene integration) and the integrated *FRG1* and *CORO1B* minigenes (representing the differential and the leveled architectures, respectively) that were stably inserted into the peripheral integration site (FRT-site). White bar represents a scale of 5 μm . Right: Bar plots depicting the mean distance of the indicated genes from the nuclear envelope after insertion into the peripheral integration site. The distance was normalized to the appropriate nuclear volume (see Methods, distance given as $1/\mu\text{m}^2$), nuclei analyzed: *Zeocin* n = 44, *FRG1* n = 54, and *CORO1B* n = 44. (C) Bar plots demonstrating splicing ratio (exon inclusion/exon exclusion) of alternatively spliced exons located in endogenous genes that are proximal to the central integration site (*left panel*) or proximal to the peripheral integration site (*right panel*), upon integration of the differential (*FRG1*) or leveled (*CORO1B*) minigenes, and with respect to the splicing ratio prior to the integration of the minigenes (referred to as *Zeocin*). The distance and name of upstream and downstream genes from the FRT-site are shown below the bar plots. (D) As panel (C) but for the gene expression normalized to the expression of housekeeping gene *ACTB*. (A) to (D) P-values correspond to two-tailed t-tests: *($p \leq 0.05$), **($p \leq 0.01$). (E) Error plot with mean delta cRPKM values of genes (n = 7032) per mean delta distance after LMNB1 KO in MDA-MB-231 cells. Pearson correlation test was applied. Regression line is shown in red.

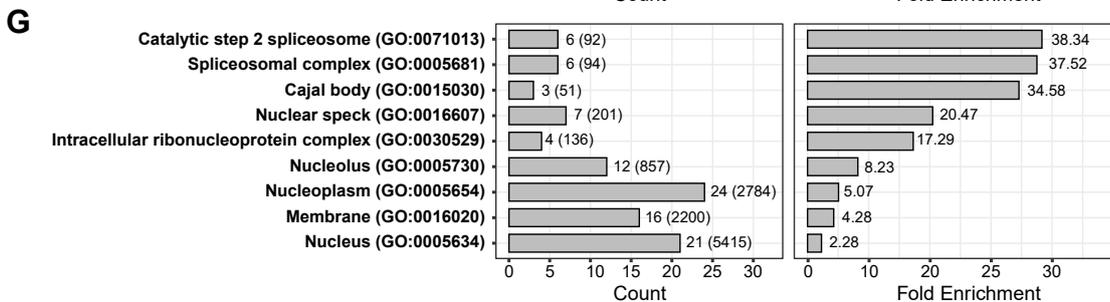
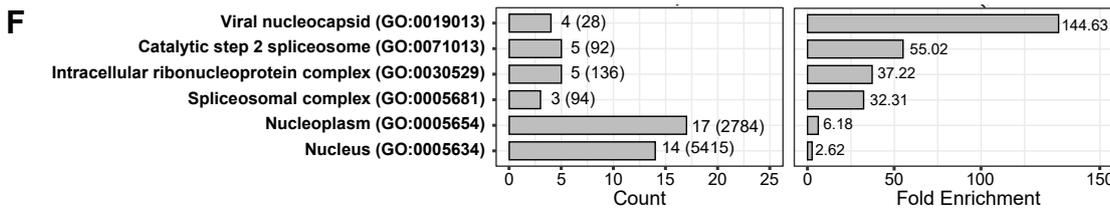
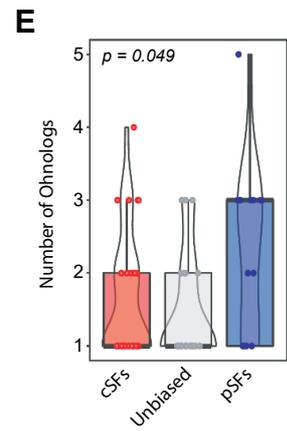
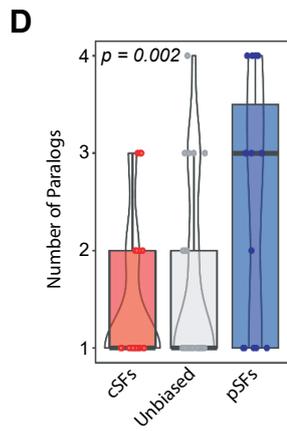
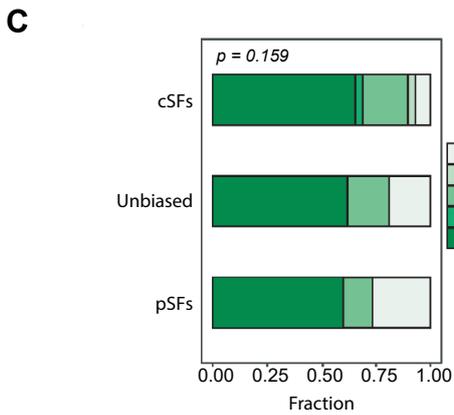
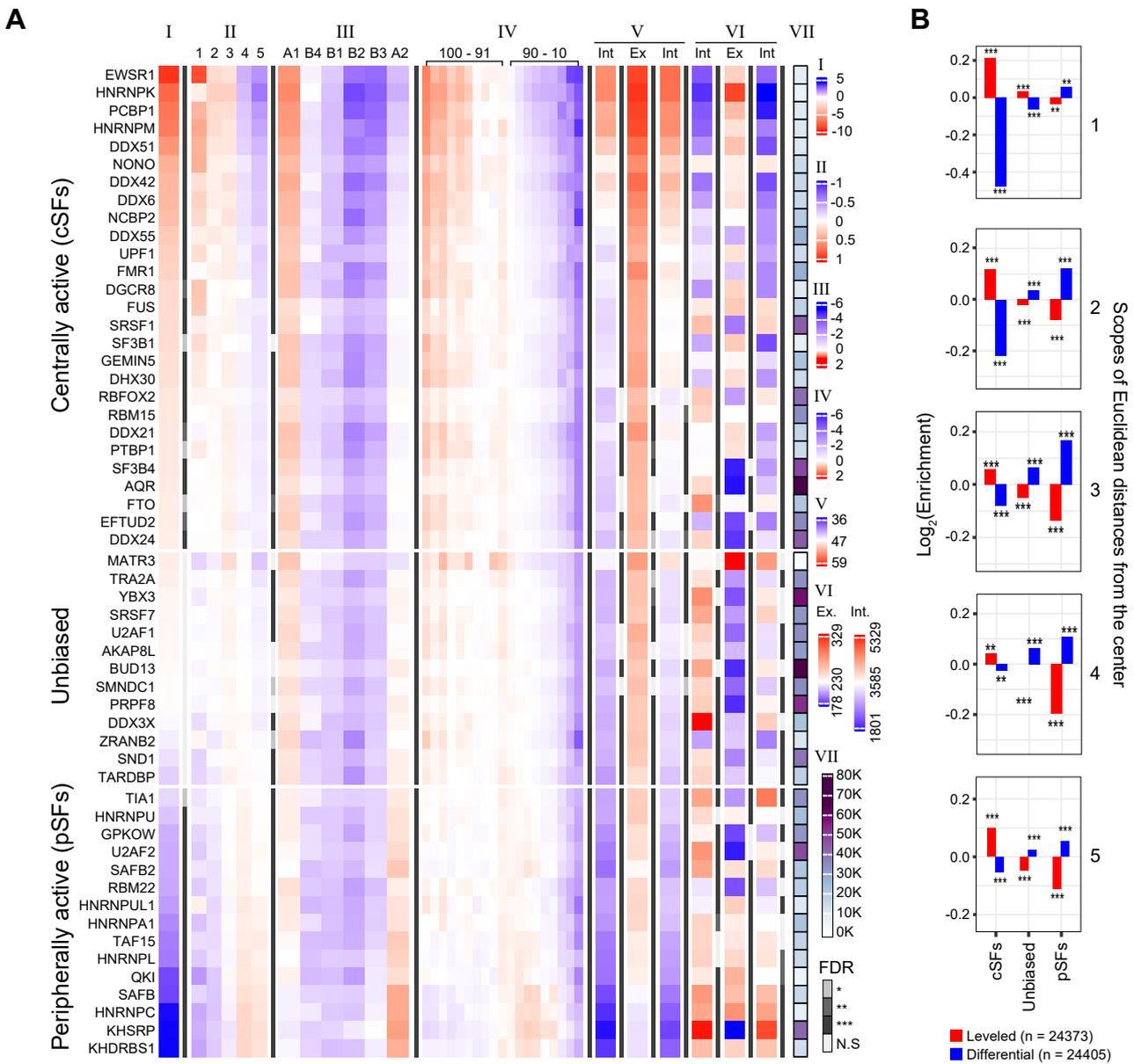


Figure S4. Characteristics of the pre-mRNA targets of the centrally and peripherally active SFs. Related to Figure 4.

(A) Heatmaps I-to-VII depicting the results for analyzed features of exon-intron structures (a window of 75 nt taken from both sides of the exons and the adjacent 100 nt of the flanking introns) containing cross-linking sites to each SF in K562 cells. (I) Mean relative distances (see Methods) from the nuclear center. Two-tailed Wilcoxon rank sum tests were applied. SFs with a significant distance were classified as cSFs and pSFs. (II) Fold enrichment in each concentric radial scope from the nuclear center. (III) As in (II) for subcompartments. (IV) As in (II) for TSA percentile (10-to-100 indicate lowest-to-highest proximity to the speckles). (II) to (IV) Chi-square goodness of fit tests followed by FDR corrections were applied. (V) Mean GC content (%) of exons (middle) and their upstream (left) and downstream (right) introns. Two-tailed Wilcoxon rank sum tests were applied (compared to the mean GC content (%) of the combined SF bound structures ($n = 122,608$)). (VI) Mean length of exons (middle) and their upstream (left) and downstream (right) introns. Two-tailed Wilcoxon rank sum tests were applied (compared to the mean length of the combined SF bound exons/introns). (VII) Total bound structures per SF. (I) to (VI). Significance levels are adjacent to each heatmap on the right. $*(FDR \leq 0.05)$, $** (FDR \leq 0.01)$, $*** (FDR \leq 1e-10)$. (B) Enrichment patterns of exon-intron structures that manifest the leveled vs. the differential architectures in transcripts bound by pSFs and cSFs (10 most proximal and distal to/from the nuclear center, respectively (see Figure 4A)) and the Unbiased group. Hypergeometric tests were applied. $*(p \leq 0.05)$, $** (p \leq 0.01)$, $*** (p \leq 1e-10)$. (C) to (E) Evolutionary history and gene family size of the cSFs and the pSFs. (C) Stacked plot showing ancestry-calling of cSFs, unbiased and pSFs. P -value corresponds to a two-tailed Fisher's Exact test comparing the vertebrate vs. the other categories together. (D) and (E) Bar and violin plots showing the number of paralogs (D) or ohnologs (E) of the cSFs, pSFs, and the unbiased SFs. P -values correspond to Fisher's Exact tests. (F) and (G) Gene ontology analyses of the central and peripheral subnetworks. Enrichment for cellular component terms in *Homo sapiens* using the DAVID classification system (Huang da et al., 2009b). Significantly associated terms ($FDR < 0.05$, Fisher's exact test) for genes included in the peripheral (F) and the central (G) subnetworks are plotted with the number of genes in each term (left panels, total number of reference genes are shown in parentheses) along with the fold enrichment in each term (right panels).

Table S1. Pearson correlation tests between the GC content (%) of chromatin segments and the distance from nuclear center (Chrom3D, GPSeq, Dip-C) or from nuclear speckles (TSA-seq). ‘N’ indicates the number of analyzed segments per test. Related to Figure 1 and Figure S1B.

Method	Cell type	Segment	N	Alternative	Pearson's r	P value
Chrom3D	GM12878	Whole genome	9089	2-tailed	-0.47	0.00E+00
Chrom3D	K562	Whole genome	6355	2-tailed	-0.35	1.45E-184
Chrom3D	HeLa	Whole genome	6939	2-tailed	-0.39	2.22E-256
Dip-C	GM12878	Whole genome	140678	2-tailed	-0.6	0.00E+00
GPSeq	HAP1	Whole genome	2714	2-tailed	-0.74	0.00E+00
TSA-seq	K562	Whole genome	9518	2-tailed	-0.69	0.00E+00
Chrom3D	GM12878	Genes	33707	2-tailed	-0.4	0.00E+00
Chrom3D	K562	Genes	32360	2-tailed	-0.31	0.00E+00
Chrom3D	HeLa	Genes	33707	2-tailed	-0.32	0.00E+00
Dip-C	GM12878	Genes	33249	2-tailed	-0.52	0.00E+00
GPSeq	HAP1	Genes	31354	2-tailed	-0.54	0.00E+00
TSA-seq	K562	Genes	33707	2-tailed	-0.65	0.00E+00
Chrom3D	GM12878	Introns	332089	2-tailed	-0.37	0.00E+00
Chrom3D	K562	Introns	319432	2-tailed	-0.29	0.00E+00
Chrom3D	HeLa	Introns	332089	2-tailed	-0.3	0.00E+00
Dip-C	GM12878	Introns	329491	2-tailed	-0.49	0.00E+00
GPSeq	HAP1	Introns	310509	2-tailed	-0.51	0.00E+00
TSA-seq	K562	Introns	332089	2-tailed	-0.61	0.00E+00
Chrom3D	GM12878	Exons	372821	2-tailed	-0.27	0.00E+00
Chrom3D	K562	Exons	359028	2-tailed	-0.22	0.00E+00
Chrom3D	HeLa	Exons	372821	2-tailed	-0.22	0.00E+00
Dip-C	GM12878	Exons	370160	2-tailed	-0.36	0.00E+00
GPSeq	HAP1	Exons	348860	2-tailed	-0.38	0.00E+00
TSA-seq	K562	Exons	372821	2-tailed	-0.45	0.00E+00