

Cotranscriptional splicing efficiencies differ within genes and between cell types

KARAN BEDI,^{1,2,3} BRIAN MAGNUSON,^{2,3} ISHWARYA VENKATA NARAYANAN,^{1,3}
MICHELLE T. PAULSEN,^{1,3} THOMAS E. WILSON,⁴ and MATS LJUNGMAN^{1,3}

¹Department of Radiation Oncology, University of Michigan, Ann Arbor, Michigan 48109, USA

²Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, Michigan 48109, USA

³Rogel Cancer Center and Center for RNA Biomedicine, University of Michigan, Ann Arbor, Michigan 48109, USA

⁴Department of Pathology and Department of Human Genetics, University of Michigan Medical School, University of Michigan, Ann Arbor, Michigan 48109, USA

ABSTRACT

Pre-mRNA splicing is carried out by the spliceosome and involves splice site recognition, removal of introns, and ligation of exons. Components of the spliceosome have been shown to interact with the elongating RNA polymerase II (RNAPII), which is thought to allow splicing to occur concurrently with transcription. However, little is known about the regulation and efficiency of cotranscriptional splicing in human cells. In this study, we used Bru-seq and BruChase-seq to determine the cotranscriptional splicing efficiencies of 17,000 introns expressed across six human cell lines. We found that less than half of all introns across these six cell lines were cotranscriptionally spliced. Splicing efficiencies for individual introns showed variations across cell lines, suggesting that splicing may be regulated in a cell type-specific manner. Moreover, the splicing efficiency of introns varied within genes. The efficiency of cotranscriptional splicing did not correlate with gene length, intron position, splice site strengths, or the intron/neighbor exons GC content. However, we identified binding signals from multiple RNA-binding proteins (RBPs) that correlated with splicing efficiency, including core spliceosomal machinery components—such as SF3B4, U2AF1, and U2AF2 showing higher binding signals in poorly spliced introns. In addition, multiple RBPs, such as BUD13, PUM1, and SND1, showed preferential binding in exons that flank introns with high splicing efficiencies. The nascent RNA splicing patterns presented here across multiple cell types add to our understanding of the complexity in RNA splicing, wherein RNA-binding proteins may play important roles in determining splicing outcomes in a cell type- and intron-specific manner.

Keywords: splicing; RNA-binding proteins; cotranscriptional; spliceosome; cell lines

INTRODUCTION

Primary transcripts generated by RNA polymerase II (RNAPII) undergo several cotranscriptional processing steps, including capping, splicing, and 3' end processing (Maniatis and Reed 2002). To carry out splicing, the spliceosome must selectively recognize bona fide splice sites among a sea of pseudosplice sites, and for splicing to occur cotranscriptionally, this must be done rapidly given that the average rate of RNA synthesis is estimated to be ~2–3 kb/min in cells (Singh and Padgett 2009; Fuchs et al. 2014; Jonkers et al. 2014; Veloso et al. 2014). While the timing and location at which splicing occurs is still under debate, it has been reported that splicing and polyadenylation take place while the RNA is still anchored to the chromatin, suggesting that these events occur

cotranscriptionally (Beyer and Osheim 1988; Bauren and Wieslander 1994; Ameer et al. 2011; Bhatt et al. 2012; Djebali et al. 2012; Tilgner et al. 2012; Darnell 2013; Naftelberg et al. 2015; Oesterreich et al. 2016; Reimer et al. 2021; Sousa-Luis et al. 2021).

Further support for the existence of cotranscriptional splicing arose from the postulation of three models of splicing: the recruitment, kinetic and the goldilocks models. According to the recruitment model, splicing factors, along with factors responsible for 5' end capping and 3' end processing, are recruited to the elongating RNAPII via the carboxy-terminal domain (CTD) of the polymerase (Bentley 2014; Naftelberg et al. 2015). In the kinetic model

Corresponding author: ljungman@umich.edu

Article is online at <http://www.majournal.org/cgi/doi/10.1261/rna.078662.120>.

© 2021 Bedi et al. This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see <http://majournal.cshlp.org/site/misc/terms.xhtml>). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

of splicing, the rate of RNAPII elongation influences alternative splicing, where a reduced elongation rate allows more time for recognition of splice junction and recruitment of regulatory splicing factors (de la Mata et al. 2003; Schor et al. 2009; Braberg et al. 2013; Darnell 2013; Fong et al. 2014; Naftelberg et al. 2015). The goldilocks model suggests an optimal elongation rate is required for kinetic coupling of transcription and splicing, wherein elongation rates could affect the roles of parameters such as RNA folding and RNA-binding proteins in determining the splicing outcome (Fong et al. 2014; Saldi et al. 2021). Conversely, splicing has been suggested to regulate the rate of transcription elongation via an “elongation checkpoint” that is activated if splicing is compromised and prevents the release of the transcripts from chromatin (Martins et al. 2011; Chathoth et al. 2014). Splicing efficiency may also be regulated indirectly via genomic features that slow the rate of RNAPII translocation. The strong phasing of nucleosomes over exons has been shown to slow the elongation rate of RNAPII as it traverses exons, thereby increasing the chances of productive splicing events (Schwartz et al. 2009; Tilgner et al. 2009; Nojima et al. 2015). This slowing effect of phased nucleosomes has also been observed in genome-wide studies, where the density of exons in the path of RNAPII correlates with a slower elongation rate (Jonkers et al. 2014; Veloso et al. 2014).

As a consequence of cotranscriptional splicing, introns are expected to be rapidly and sequentially removed from the nascent RNA in the direction of transcription elongation. However, some splicing events have been suggested to occur after transcripts have been polyadenylated, and introns are not always spliced in the order they were synthesized, suggesting that some splicing may occur post-transcriptionally (Pandya-Jones and Black 2009; Jia et al. 2020). This is supported by findings that many introns are post-transcriptionally spliced in yeast, fruit fly, mouse, and plants (Tardiff et al. 2006; Khodor et al. 2012; Jia et al. 2020). Some reports suggest that nascent RNA is held attached to the chromatin and not polyadenylated until fully spliced (Brody et al. 2011; Chathoth et al. 2014). In contrast, other studies have found evidence of intron retention in fractions of total RNA that are enriched for polyadenylated RNA [poly(A)⁺] (Ameur et al. 2011; Braunschweig et al. 2014; Boutz et al. 2015). Splicing efficiency and retention of introns is also influenced by the splicing machinery used, as introns recognized in a U12-dependent mechanism and spliced by the minor spliceosome are retained to a higher degree than introns served by the canonical spliceosomes (Singh and Padgett 2009; Ameur et al. 2011). Transcripts with retained introns appear to be restricted to the nucleus where they either undergo post-transcriptional splicing followed by exit to the cytoplasm or are targeted for degradation by the nuclear exosome (Khodor et al. 2012; Windhager et al.

2012; Boutz et al. 2015; Pan et al. 2015; Jia et al. 2020). Recent studies using long read, direct RNA sequencing show that splicing efficiency is variable, with some transcripts being fully spliced, while others are only partially spliced or completely unspliced (Oesterreich et al. 2016; Herzel et al. 2018; Drexler et al. 2020). These findings suggest that cotranscriptional splicing is inefficient and unspliced transcripts are either purged by quality control mechanisms or spliced post-transcriptionally.

In this study, we used nascent RNA Bru-seq and BruChase-seq to assess splicing efficiencies of about 17,000 introns expressed across six human cell lines, including one noncancer (HF1) and five cancer cell lines (HCT116, HeLa, K562, MiaPaCa, and Panc1). Furthermore, we correlated the splicing efficiencies with both *cis*- and *trans*-acting genomic features. The data revealed that less than half of the introns interrogated were spliced within 30 min of synthesis, and the splicing efficiencies for individual introns differed across cell lines and within genes. Furthermore, the first introns of genes were the least efficiently spliced across all cell lines. Importantly, specific RNA-binding proteins (RBPs) showed binding patterns associated with the efficiency of splicing, suggesting that they may regulate gene expression at the level of splicing.

RESULTS

Splicing estimates vary across RNA sequencing assays

In this study, we used nascent RNA Bru-seq to interrogate cotranscriptional splicing efficiencies across six human cell lines (Fig. 1A; Paulsen et al. 2013, 2014). In order to accurately assess cotranscriptional splicing efficiencies, which we here define as splicing events recorded after the 30-min bromouridine labeling period, we first needed to define the splice junctions which only appear in the dominant isoform (isoform with the highest FPKM) of each gene across the cell lines. We used BruChase-seq (Fig. 1A) to obtain records of fully mature RNA (i.e., 6 h old RNA) which allowed us to define cell line-specific annotations to accurately define exon–intron boundaries for use in splicing index calculations across the cell lines. Transcript-level quantifications were performed with Cufflinks (Trapnell et al. 2010; Roberts et al. 2011) for K562 cells using BruChase-seq data, and we used ENCODE whole cell poly(A)⁺, and cytoplasmic poly(A)⁺ data for isoform conformations. The dominant isoforms were concordant in 99.57% (10,163/10,206) of the expressed genes across the three assays (Supplemental Fig. S1A). An intron-centric splicing index (SI) (Herzel et al. 2018) was then calculated as described in Figure 1B. SI values range between 0 and 1, with 1 being fully spliced and 0 being unspliced. We calculated SI values of K562 cells from nine different assays, namely Bru-seq, BruChase-seq, TT-seq (Schwalb et al.

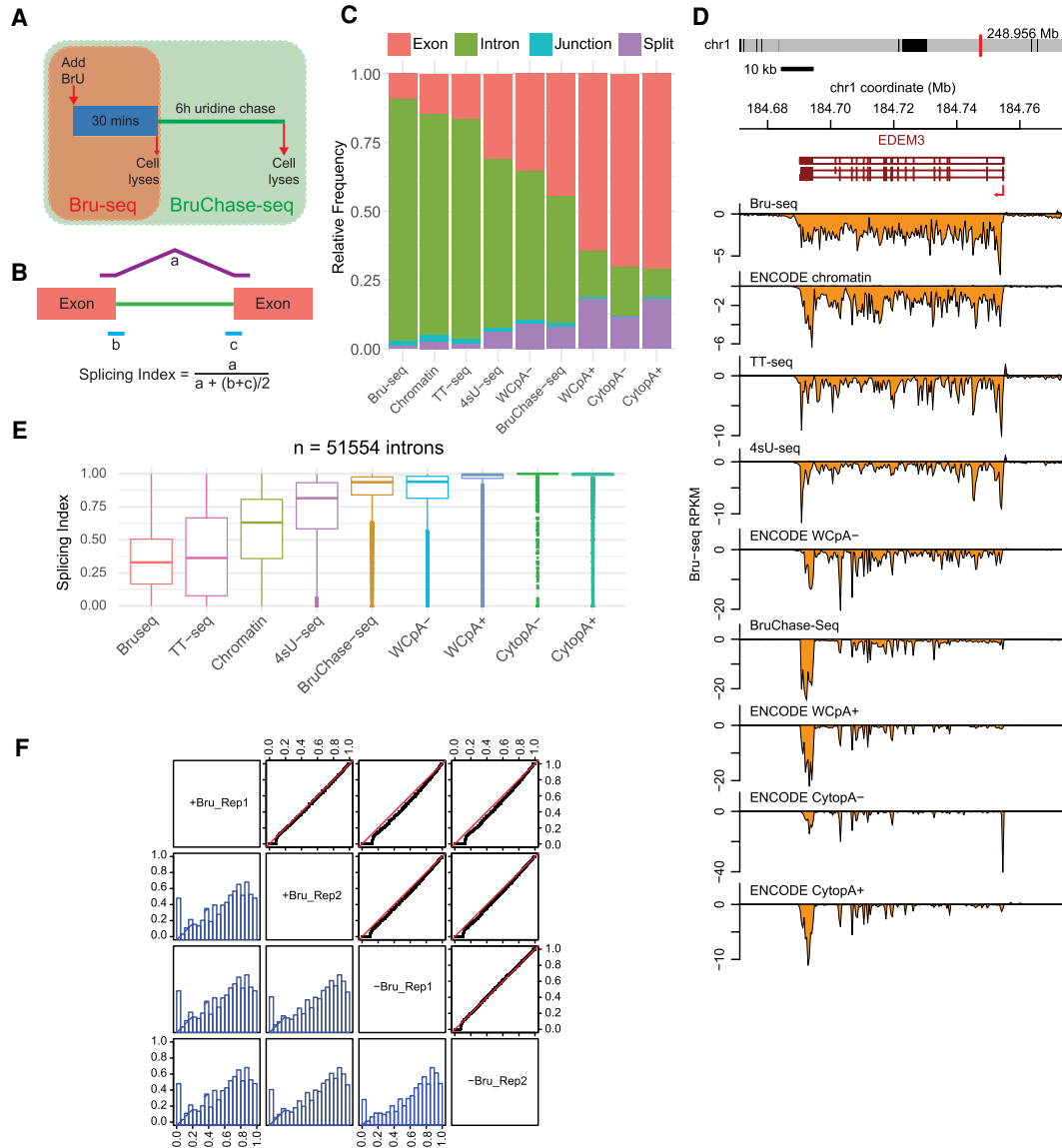


FIGURE 1. Assessment of RNA splicing. (A) Schematic of Bru-seq and BruChase-seq techniques. Bru-seq involves labeling cells with bromouridine (BrU) for 30 min followed by cell lysis and sequencing, while BruChase-seq involves a 30 min BrU labeling period followed by BrU removal and chase in uridine for 6 h. (B) Definition of splicing index, calculated as a ratio of the split reads (exon–exon reads, *a*) to the sum of split (*a*) and half of junction reads (exon–intron junction reads [*b*] and intron–exon junction reads [*c*]). (C) Genome-wide relative distribution of RNA sequencing reads from exons, introns, split (exon–exon), and junction (both 5' exon–intron and 3' intron–exon) for different RNA isolation procedures from K562 cells. (WC) Whole cell, (pA) poly(A), (Cyto) cytoplasmic, (TT-seq) transient transcriptome sequencing. (D) RNA-seq read coverage profile of the *EDEM3* gene from the nine different RNA isolation procedures. The gene is present on the negative strand and transcribes from right to left (shown with red arrow). (E) Splicing index (SI) distributions for 51,554 introns in K562 cell line across nine assays (of the type all-introns-included). (F) qqplot of SIs for two K562 replicates comparing SI in the presence (+Bru) or absence (–Bru) of bromouridine. The sample names are along the diagonal, the qqplots are presented in the panels above the diagonal, with the red line representing intercept = 0 and slope = 1, while the distributions are presented in panels below the diagonal.

2016), 4sU-seq (Schwalb et al. 2016), and five ENCODE RNA-seq data sets representing various cell fractions {chromatin, cytoplasmic, and whole cell poly(A)[–] [poly(A) depleted] and poly(A)⁺ (polyadenylated)}. Nascent RNA assays, such as Bru-seq, ENCODE chromatin RNA-seq, and TT-seq had more than 80% of reads covering intronic regions within gene bodies and between 1.8%–2.4% of reads cov-

ering exon–exon junctions, pointing toward the nascent nature of these data sets. In contrast, ENCODE derived samples of cytoplasmic poly(A)⁺ (CytopA⁺) and poly(A)[–] (CytopA[–]) had more than 70% of reads covering exonic regions (Fig. 1C). Highly expressed genes, such as *EDEM3* (Fig. 1D) and *ZBTB37* (Supplemental Fig. S1B) showcase the distinct read coverages of the different assays.

The dominant isoform analysis from the K562 BruChase-seq data resulted in one dominant isoform for 20,665 transcripts, containing 165,944 introns. Of these introns, 82,050 (49.4%) met the junction reads sum (jrsum, see Materials and Methods) threshold of eight in the K562 Bru-seq sample (Supplemental Fig. S1D). A subset of these introns where only those that were found in genes in which all introns of that gene met the jrsum threshold of eight ($n = 5913$ genes) were selected, amounting to 63% (51,554/82,050) of our initial/parent intron pool. These introns had varied SI distributions across our analyzed assays, with medians ranging from 0.32 (for Bru-seq) to 1 (for CytopA+) (Fig. 1E). The BruChase-seq median was close to 1, further establishing that RNA is almost fully spliced 6 h after synthesis. Principal component analysis (PCA) on SI values resulted in our BruChase-seq and ENCODE poly(A) fractions clustering tightly, while Bru-seq, chromatin and TT-seq clustered away from the techniques that enrich for more mature RNAs (Supplemental Fig. S1C,E). To ensure BrU does not interfere with splicing, we isolated the chromatin fraction of K562 cells that were either Bru-labelled or not and sequenced these in duplicate. SI qqplots (quantile-quantile plot) between the labeled and unlabeled cells did not show any difference, thus ensuring splicing was not affected by the presence of BrU (Fig. 1F). Our findings confirm that cotranscriptional splicing efficiencies in K562 are low using our Bru-seq assay (Paulsen et al. 2013, 2014), and in a similar nascent RNA sequencing assay, TT-seq.

Splicing measurements vary across multiple cell types

In an effort to further understand cotranscriptional splicing, we explored this process across multiple human cell lines. We calculated SI values across Bru-seq data sets from six human cell lines, including five cancerous (HCT116, HeLa, K562, MiaPaCa, and Panc1) and one noncancerous fibroblast line (HF1). The jrsum thresholds calculated for these cell lines ranged between 4–8 (Supplemental Fig. S2A). These thresholds were decided by determining the minimum jrsum at which the cumulative median SI of introns (with lower jrsums) was a nonzero value (Supplemental Fig. S2B). This resulted in >82,000 introns (of type any-introns-included, i.e., any intron that met the jrsum threshold was included in the analysis) and >46,000 introns (of type all-introns-included, i.e., all introns of a gene that met the jrsum threshold were included in the analysis) for each of the six cell lines (Supplemental Fig. S2C; Supplemental Table S2). The medians of the SI value distributions for ~17,000 introns (belonging to 2003 genes) that were common among all six cell lines varied between 0.31 and 0.48, with HeLa cells showing the highest median SI value (Fig. 2A; Supplemental Fig. S2D). Less than 3% of these common introns ($n = 410$) were not spliced (SI = 0) in all cell lines. The medians remained almost identical when

the same jrsum threshold (=10) was used across all cell lines ($n = 10,247$ introns, Supplemental Fig. S2E). The RPKMs of these 2003 genes were very similar among five of the six cell lines, with HF1 showing slightly lower median RPKM (Supplemental Fig. S2F,G) and weaker correlations to other cell lines (Supplemental Fig. S2H). Gene RPKM showed no correlation with their SI values (Supplemental Fig. S2H). Interestingly, SI values of different introns within the same gene were also variable, as can be seen for the ~6 kb *ARCN1* gene (Fig. 2B). Introns 2 and 3 had higher SI values (ranging between 0.63–0.91) compared to other introns in *ARCN1* (Fig. 2B, dot plot). Interestingly, these two introns had higher SI values in all six cell lines, suggesting common splicing regulatory mechanisms between cell types. Consistently, the RPKMs for these two introns were lower than for the other introns both in the Bru-seq and BruChase-seq data (Fig. 2B heatmap, top and middle panel). Intron 5, on the other hand, was poorly spliced. *TPR*, a ~63 kbp gene, showed considerable variations in splicing across its 50 introns and across the different cell lines (Supplemental Fig. S2I). These findings show that splicing efficiencies for individual introns can vary within genes and across cell lines.

Across all cell lines, first introns were less efficiently spliced than internal and last introns, which showed similar SI's independent of their position (Supplemental Fig. S2J). To further elaborate on the differential splicing efficiencies across cell lines, we performed hierarchical clustering of introns of the top 10% highest expressing protein-coding genes (527 introns) (Fig. 2C). These genes were highly enriched for functions such as RNA-binding and translation initiation, and predominantly associated with ribosomal subunits (Supplemental Fig. S2K). Clusters A and B showed medium to high splicing across all cell lines, with cluster A consisting of introns that were highly spliced in all six cell lines. Cluster C and D consist of introns that had low to medium SI values, while clusters E–H showed a mixed pattern (Fig. 2C). Overall, SI values of the ~17,000 introns across the six cell lines correlated well (with a minimum $r > 0.73$), with cell type differences accounting for lower than perfect correlations. SI values did not correlate with gene lengths, while there was a slight negative correlation to intron length (Fig. 2D; Supplemental Fig. S2L). Our results show low cotranscriptional splicing efficiencies across the six different human cell lines tested.

Role of cis-acting elements in splicing efficiency

Splicing efficiencies and alternative patterns can be regulated by *cis*-acting elements within introns and exons in pre-mRNAs presumably through the binding of specific proteins (Sheth et al. 2006; Juan et al. 2014). The essential *cis*-acting elements include the 5' splice site, 3' splice site, as well as the branchpoint sequences that are recognized by *trans*-acting factors. In order to interrogate whether

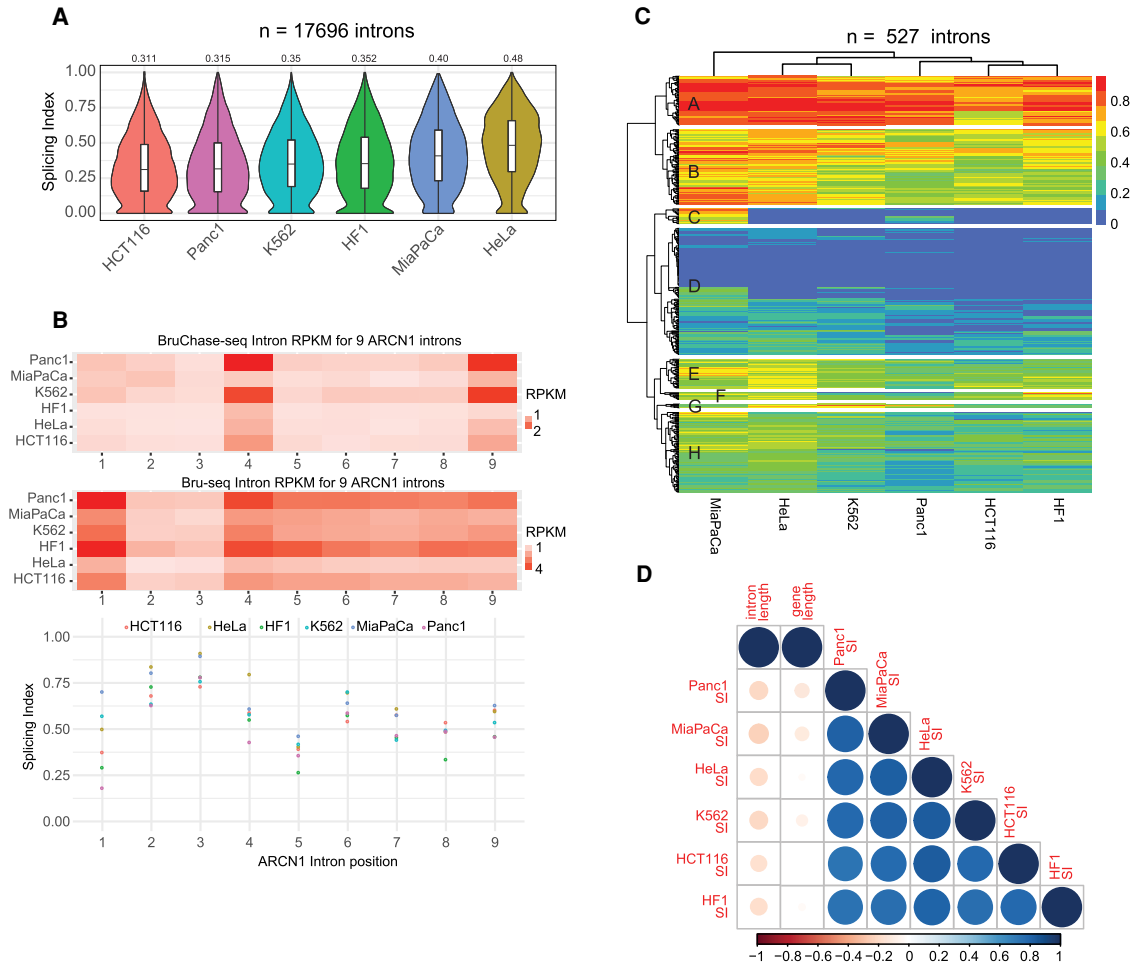


FIGURE 2. Cotranscriptional splicing across multiple cell types is inefficient. (A) Splicing index (SI) distributions for 17,696 introns common across six cell lines. The numbers at the top of violin plots represent the median splicing index in each cell line. (B) SI across nine introns of the *ARC1* gene in six cell lines. Bottom panel plots the SI (y-axis, range between 0 and 1) versus intron position in the gene (x-axis), with each colored dot representing the SI of the intron in the respective cell line (represented by one of six colors). The top and middle panels are heatmaps of intronic RPKMs for each of the nine introns assessed by BruChase-seq (upper panel) and Bru-seq (lower panel) assays. (C) Heatmap showing the SIs of introns ($n = 527$) of the top 10% highest expressed genes common across the six cell lines, split into eight clusters A–H (using cutree option in pheatmap R package) (D) Correlation plot depicting the correlations between the SIs of the six cell lines to each other, as well as to gene length and intron length.

the variability in SI values obtained from the Bru-seq data could be explained by *cis*-acting elements, we calculated the mean SI values for each intron common across six cell lines and split the introns into four quartiles based on their mean SI value. We obtained the range of SIs for each quartile after excluding single-intron transcripts, as well as first introns of remaining transcripts ($n = 15,671$ introns, Fig. 3A). When comparing the introns across the quartiles we found that intron length progressively decreased from the first to last quartile, with introns in the first quartile (q1) being roughly twice as long (median length = 2118 bp) compared with introns in the last quartile (q4, median length = 1077 bp) (Fig. 3B; Supplemental Fig. S3A). Furthermore, the introns were distributed evenly across the four quartiles when considering their start distances

to the transcription start site (TSS) (Fig. 3C). Regarding AT/GC content of the introns and their respective upstream and downstream exons, we found no differences across the four quartiles (Fig. 3D; Supplemental Fig. S3B [1–3]). Poorly spliced introns (q1) had a higher intron RPKM both in our Bru-seq and BruChase-seq samples compared with well spliced introns across all six cell lines (Supplemental Fig. S3C–E). Splice site consensus sequence scores, as calculated by the MaxEntScore, polypyrimidine tract lengths and distance of branchpoint sequence from the 3' splice site did not show any differences between the quartiles (Supplemental Fig. S3F–I). Our results are in line with previous findings (Drexler et al. 2020) that these *cis*-acting elements do not influence splicing efficiencies in human cells.

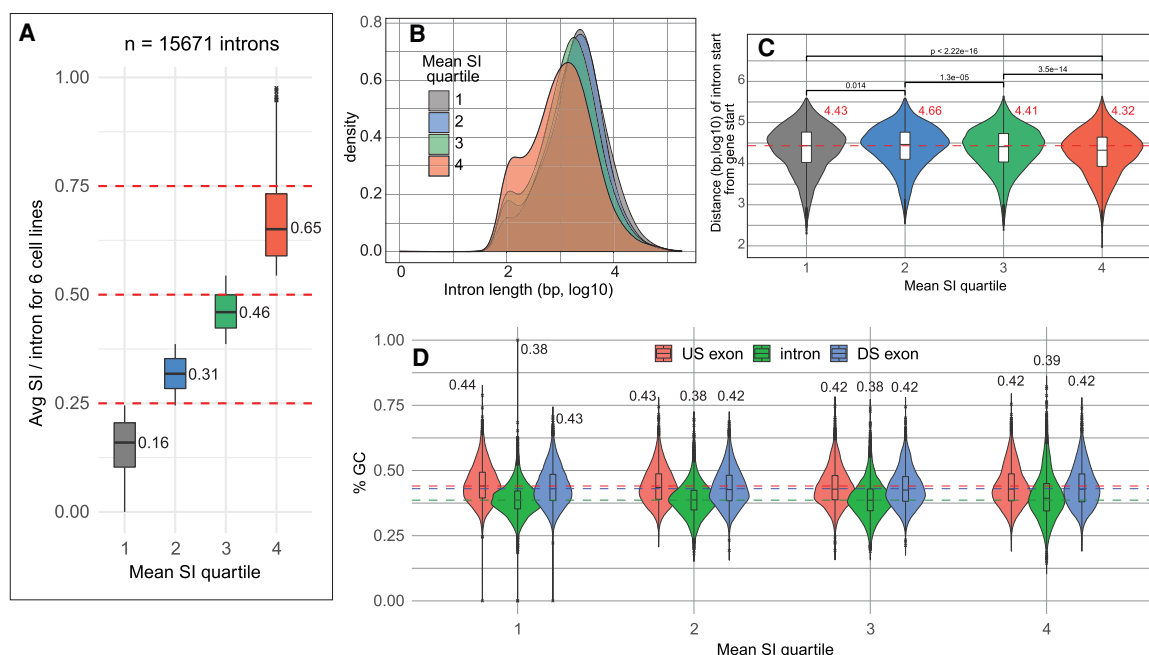


FIGURE 3. Correlations of *cis*-acting features with cotranscriptional splicing efficiency across six human cell lines. (A) The distribution of mean values of SI for 15,671 introns (common between six cell lines, with first introns and single-intron genes removed) when split into four quartiles based on the mean SI value (mean SI quartile) are shown. The numbers next to the boxplots are the median values. (B) Density plot showing the intron lengths for the four quartiles. (C) Violin and boxplots showing distribution of distance between intron start and TSS across the four quartiles. The red dotted line crosses at the median of quartile 1 (gray violin plot). The numbers (in red) next to the violin plots are the median values. *P*-values of pairwise mean comparisons are shown above the plots. (D) Distribution of GC content of introns and their flanking exons (upstream and downstream) across four quartiles. The red, green, and blue dotted lines cross the medians of GC contents of upstream exon, intron and downstream exon, respectively, for quartile 1.

Role of genomic organization and *trans*-acting features in regulating splicing

To explore whether chromatin and *trans*-acting features are associated with splicing regulation, we downloaded functional genomics data from the ENCODE portal for four cell lines: HCT116, HeLa, K562, and Panc1. For each cell line, introns (where all introns of a transcript met the respective cell line's *jrsum* threshold) were split into four quartiles based on their respective SIs (Supplemental Fig. S4A). DNase-seq, histone ChIP-seq and transcription factor ChIP-seq signals for all four cell lines were profiled at both the 5' and 3' splice sites of introns ≥ 500 bp and 100 bp upstream and downstream from the splice sites and did not show any distinct signal density patterns across the exon–intron junctions for the four quartiles (Supplemental Fig. S4B–D). We next interrogated whether *trans*-acting RNA-binding proteins showed differential associations with intron or exon regions across the quartiles using ENCODE eCLIP data on RNA-binding proteins (RBPs) for the K562 cells (Van Nostrand et al. 2020). Importantly, these data showed enrichment of specific RBPs in both exons and introns of differentially spliced introns. Five RBPs (KHSRP, PRPF8, SF3B4, U2AF1, and U2AF2) showed a greater binding signal in introns that were poorly spliced (Fig. 4). Both KHSRP and PRPF8 showed peak enrichment

just downstream from the 5' splice site, while the other three RBPs were more active on the 3' end of the introns. Twelve RBPs in K562 showed a higher binding signal in exons flanking introns with higher splicing efficiency (q4) compared with medium and poorly spliced introns (q1–q3) (Supplemental Fig. S4E). All 17 RBPs showed statistically significant differences in average binding signals between introns with the highest and lowest splicing efficiencies (Supplemental Fig. S4F,G). The remaining RBPs did not show any discernable differences in the binding signals across the quartiles (Supplemental Fig. S4H; Supplemental Table S3). We next ran the same analysis using the nascent RNA TT-seq data for K562 cells and found very similar results (Supplemental Fig. S4I). The five RBPs enriched in the intronic regions are known to play crucial roles in pre-mRNA splicing (Brody and Abelson 1985; Lustig et al. 1986; Lin et al. 1987; Zamore et al. 1992; Min et al. 1997; Igel et al. 1998). Our findings show a correlation between multiple RNA-binding proteins and splicing outcome.

DISCUSSION

Splicing is a critical part of post-transcriptional processing of primary transcripts and plays an important role in the

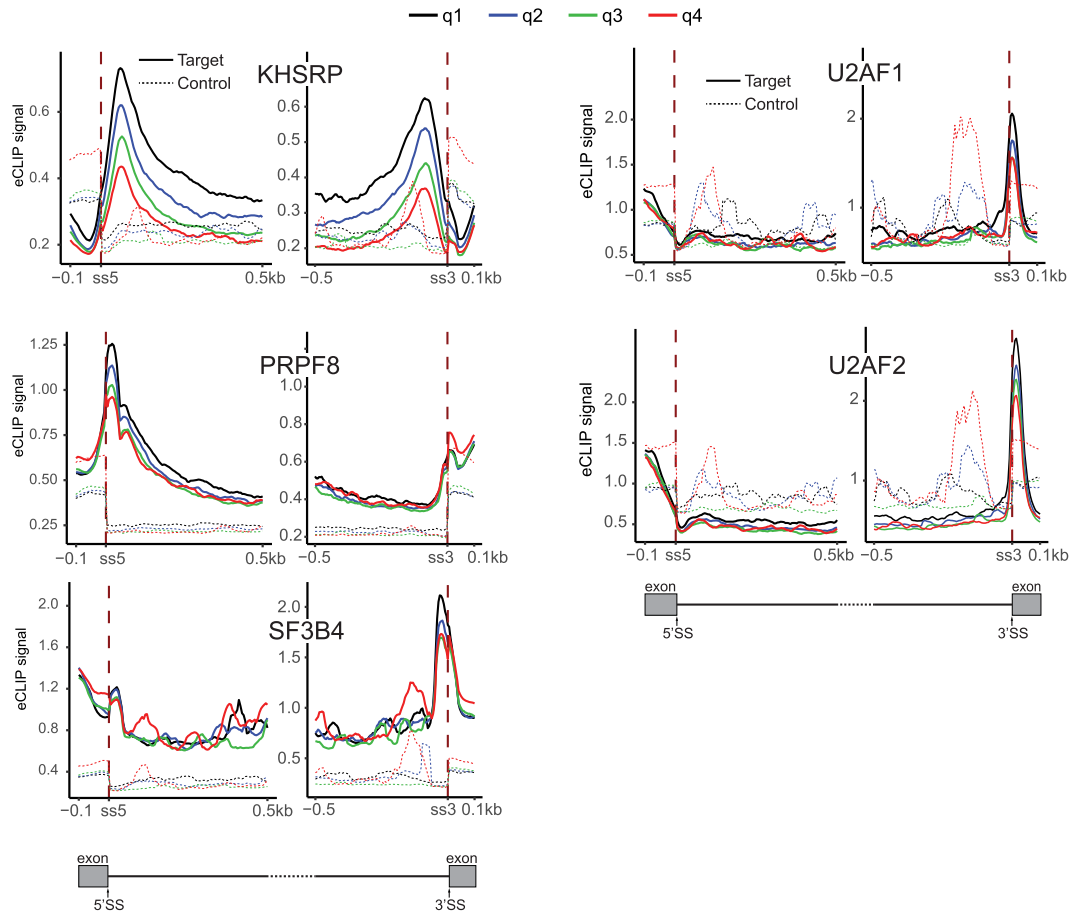


FIGURE 4. Correlations between RNA-binding proteins and splicing efficiency. Lines indicate target (solid) and control (dotted) signal tracks across splice junctions of introns in four SI quartiles for five eCLIP RBPs- KHSRP, PRPF8, SF3B4, U2AF1, and U2AF2. The regions of 100 bp upstream and 500 bp downstream from the 5' splice site, and 500 bp upstream and 100 bp downstream from the 3' splice site are profiled for introns ≥ 500 bp in length in K562.

regulation of gene expression. Splicing of nascent RNA has been analyzed in fission yeast using chromatin-bound RNA followed by long-read sequencing as well as direct sequencing of chromatin-bound RNA from *Drosophila*, human and plant cells (Jia et al. 2020). In this study, we used a combination of Bru-seq and BruChase-seq to measure genome-wide cotranscriptional intron splicing efficiencies in nascent RNA across six human cell lines. Our results indicate that cotranscriptional splicing is a rather inefficient process where more than half of the 17,000 introns analyzed across the cell lines were “missed” and remained in the transcripts following a 30-min labeling time.

The mammalian genome contains more than 200,000 introns of varying sizes (20 bp–1.24 Mb) and these introns are thought to be spliced mainly cotranscriptionally (Singh and Padgett 2009; Tilgner et al. 2012). To accomplish this, the splicing machinery must identify splice sites and promote splicing of the nascent RNA polymers as they are being synthesized at a rate of ~ 2 –3 kb/min (Veloso et al. 2014). With our sequencing depths ranging from ~ 195

to ~ 360 million reads and using jrsum thresholds, we captured the splicing patterns for about half of all annotated introns. An almost complete overlap between the dominant isoforms of genes selected based on our K562 analysis (Supplemental Fig. S1A) gives us confidence in the transcript annotation that we selected for all our downstream analyses. Enrichment of exonic and exon–exon split reads, with a simultaneous reduction of intronic as well as exon–intron junction reads was observed after the 6 h chase period in the BruChase-seq samples. The median SI value obtained from RNA collected from chromatin was higher than for Bru-seq, likely owing to the higher average age of the RNA collected from chromatin. Surprisingly, SI values calculated from 4sU-seq data were considerably higher than for Bru-seq despite the fact that these cells were only labeled for 5 min compared to the 30 min for Bru-seq (Fig. 1E). It is possible that the lower ratio of labeled to unlabeled RNA obtained with shorter labeling times results in a larger proportion of unlabeled, fully spliced RNA being captured as background thereby

driving up the SI values for the 4sU samples. Alternatively, the lower SI values in the Bru-seq analyses compared to 4sU-seq may stem from a more stringent purification procedure involving immunocapturing using magnetic beads. Cotranscriptional splicing efficiencies showed an overall strong correlation across our six cell lines ($0.73 \leq r \leq 0.84$). Analysis of splicing efficiency showed that most human introns are not spliced cotranscriptionally (medians 0.31–0.48), which is lower than similar studies (based on different assays though) performed in *S. pombe* (median 0.59) (Herzel et al. 2018), *S. cerevisiae* (median 0.74) (Brugiolo et al. 2013) but closer to estimates obtained for mouse (median 0.45) (Khodor et al. 2012). We found a slight negative correlation between intron length and cotranscriptional splicing efficiency (Figs. 2D, 3B; Supplemental Fig. S3A) in all six cell lines, which is in accord with previous nascent RNA based studies reported for *Drosophila*, mouse and humans (Khodor et al. 2011, 2012; Tilgner et al. 2012), but unlike findings in *S. pombe* (Herzel et al. 2018). Lower SI's of first introns in all our cell lines (also reported in *S. pombe* [Herzel et al. 2018]) suggests that splicing of first intron does not preclude the splicing of downstream introns. Indeed, higher overall splicing efficiencies of introns 2 and 3 in *ARCN1* (Fig. 2B) and introns 4 and 26 in *TPR* (Supplemental Fig. S2I) support the notion that the first intron can persist while downstream introns are spliced out. We observed that the splice site strength, as calculated by MaxEnt score, did not correlate with SI. There are multiple factors other than splice sites that are also involved in the splicing process. Factors such as *cis*-elements (such as exonic splicing enhancers [ESEs] or intronic splicing enhancers [ISEs]), secondary proteins that are part of the spliceosome complex or other independent RBPs could play a more important role than splice site strengths in determining splicing outcomes. Indeed, weaker splice sites have been predicted to contain more abundant auxiliary *cis*-elements, such as ESEs and ISEs (Faustino and Cooper 2003), perhaps allowing for even more intricate regulation of alternative splicing. As a result, splicing outcome does not only depend on the nucleotide variation of these consensus sequences. It is also possible that our splicing measurements represent initial splicing outcomes, with the possibility that poorly spliced introns (quartile 1) may eventually get spliced out.

Our findings that specific RBPs bind to introns and flanking exons of differentially spliced introns presents interesting avenues of further investigations. Drexler et al. (2020) also recently reported that the total occupancy of the aggregated signal from all RBPs differed significantly between spliced and unspliced introns. Specifically, they observed 27 RBPs that were significantly more likely to be bound to unspliced introns than spliced introns within pairs, both in terms of the number of introns bound and RBP occupancy levels. Our study provides further resolu-

tion of these findings by identifying five RBPs with occupancy levels enriched at either the 5'- or 3'-splice sites (or both) in introns with low splicing efficiency, and 12 RBPs enriched in the exons flanking introns with high splicing efficiency, with clear separation of enrichments according to the SI value quartiles. Of these 17 RBPs, seven play an active role in pre-mRNA splicing, five of them being members of the core spliceosomal machinery (Brody and Abelson 1985; Lustig et al. 1986; Lin et al. 1987; Zamore et al. 1992; Min et al. 1997; Igel et al. 1998; Zhang et al. 2018). The remaining RBPs have various assigned functions, ranging from regulation of transcription and translation, and mRNA export (Li and Chang 1996; Zhao et al. 2000; Didiot et al. 2008; Gupta et al. 2008; Dong et al. 2011; Filipovska et al. 2011; Bot et al. 2017; Elbarbary et al. 2017). An enriched RBP binding in introns that are poorly spliced could mean longer residual times spent in preparing these introns for splicing. Alternatively, they rely on other regulatory factors, like the RBPs enriched in neighboring exons to complete the splicing process. SR proteins that function as general activators of exon definition by binding to exonic splicing enhancer elements (ESEs) facilitate recruitment and/or stabilization of components of the core spliceosome (Blencowe 2000; Graveley 2000; Lam and Hertel 2002; Black 2003), and thereby affect splicing outcomes. We observe enrichment of one such component of spliceosome P complex, BUD13, over exons flanking introns with high splicing efficiency (Supplemental Fig. S4E). Whether the enriched RBPs function independently, or in coordination with other proteins and factors remains to be explored. Our observation that the *cis*-acting elements do not show any differences across our splicing index quartiles lends support to the hypothesis that secondary factors may have a bigger role to play in determining splicing outcomes.

Conclusions

Our results show that cotranscriptional splicing only removes about half of the introns present in newly synthesized RNA. This surprising finding held true across six different human cell lines that we analyzed. Furthermore, we found that splicing efficiencies differed for introns within the same gene and for introns across cell lines. Our study also revealed a specific set of RNA-binding proteins whose RNA-binding patterns correlated with splicing efficiencies suggesting that these proteins may play regulatory roles in gene and cell type-specific splicing. While over half of the introns across the six cell lines remained unspliced following the 30-min bromouridine labeling period, they were found to be almost completely removed following a 6-h chase period. This would suggest that transcripts that are not cotranscriptionally spliced are either spliced posttranscriptionally or purged from the transcriptome by a quality control mechanism involving degradation by the RNA

exosome. A scenario where the fate of a transcript is dependent on whether it is fully spliced or not puts splicing into a central regulatory role of gene expression. Here, specific RNA-binding proteins are likely to influence the efficiency of the splicing machinery and hence, fine-tune which transcripts will go on to translation and which ones will fall short.

MATERIALS AND METHODS

Cell lines and cell culture

A list of cell lines used in this study is provided in [Supplemental Table S1](#). HCT116 were grown in RPMI (RPMI 1640, 10% FBS, 100 U/mL penicillin and 100 U/mL streptomycin). HeLa, MiaPaCa, and Panc1 were grown in DMEM (DMEM, 10% FBS, 100 U/mL penicillin and 100 U/mL streptomycin). MEM growth media (Minimal Essential Medium, 10% FBS, 1 × MEM amino acids, 1 × nonessential amino acids, 2 mM L-glutamine, 1 × antibiotic-antimycotic, 1 × MEM vitamin mixture and 0.15% [w/v] sodium bicarbonate) was used for HF1. K562 were grown in IMDM with 10% FBS and antibiotics.

Bru-seq nascent RNA sequencing and read mapping

For Bru-seq, cells were incubated in media containing bromouridine (Bru) (Aldrich) at a final concentration of 2 mM for 30 min at 37°C to label nascent RNA. Following labeling, cells were lysed directly in TRIzol followed by isolation of total RNA and immunocapturing of Bru-labeled RNA using anti-BrdU antibodies. Bru-labeled RNA was incubated with superscript II first-strand buffer containing MgCl₂ at 85°C for 10 min to achieve fragmentation, followed by preparation of strand-specific DNA libraries with the Illumina TruSeq Kit (Illumina) and deep sequencing using the Illumina sequencing platform, all as previously described (Paulsen et al. 2013, 2014). Sequenced reads were strand-specific, single-ended, and of read lengths from 40–65 bp. Reads were pre-mapped to the ribosomal RNA (rRNA) repeating unit (GenBank U13369.1) and the mitochondrial and EBV genomes (from the hg38 analysis set) using Bowtie2 (2.3.3) (Langmead and Salzberg 2012). Unaligned reads were subsequently mapped to human genome build hg38/GRCh38 using STAR (v 2.5.3a) (Dobin et al. 2013) and a STAR index created from GENCODE annotation version 27 (ftp://ftp.ebi.ac.uk/pub/databases/genCODE/GenCODE_human/release_27/genCODE.v27.basic.annotation.gtf.gz) (Frankish et al. 2019). This strategy allowed mapping of reads throughout the genome, including any spliced reads found in purified nascent RNA. For all cell lines where multiple replicates were available, the mapped reads were merged for deeper coverage. Read statistics for each cell line and library are provided in [Supplemental Table S1](#).

Exon–intron definition

Cufflinks (<http://cole-trapnell-lab.github.io/cufflinks/>) (Trapnell et al. 2010; Roberts et al. 2011) was used to estimate the abundance of transcripts (Ensembl) for all six BruChase-seq cell lines.

After filtering for all transcripts with FPKM ≥ 1, for every expressed gene in each cell line, the transcript with the highest FPKM (hereby named the “dominant isoform”) was selected, and the exon and intron boundaries of this dominant isoform were used to define splice junctions. The K562 ENCODE whole-cell poly(A)⁺ and cytoplasmic poly(A)⁺ fractions were also subjected to cufflinks and the dominant isoform determined in a similar manner in both fractions.

Splicing index

An intron-centric splicing index (SI) was calculated for all the introns of the dominant isoform in all six cell lines and both Bru-seq and BruChase-seq assays. The SI was defined as:

$$\text{Splicing Index (SI)} = \frac{a}{a + (b + c)/2},$$

where

a = split reads (i.e., reads from exon to exon)

b = 5′ Exon–Intron junction reads (with a minimum 3 bp overlap on either side of the junction)

c = 3′ Intron–Exon junction reads (with a minimum 3 bp overlap on either side of the junction)

jrsum = Junction reads sum (a + b + c).

A minimum jrsum threshold was calculated per cell line at which the cumulative median SI of introns (with lower jrsums) was a non-zero value (see [Supplemental Fig. S2B](#)), instead of choosing an arbitrary number as the jrsum threshold.

cis-elements and functional genomics data influencing splicing

Splice site consensus sequence scores were extracted using MaxEnt with default parameters (Yeo and Burge 2004). Polypyrimidine tract lengths and branchpoint distances from 3′ splice site were calculated using the SVM-BPfinder software (http://regulatorygenomics.upf.edu/Software/SVM_BP/) (Corvelo et al. 2010). DNase-seq, TF-ChIP-seq, Histone ChIP-seq, and eCLIP data for the various cell lines were obtained from the ENCODE (<https://www.encodeproject.org/>). The plots utilizing this functional genomics data were created using deepTools (<https://deeptools.readthedocs.io/en/develop/index.html>) (Ramirez et al. 2016). The RBP binding signal plots generated using deepTools were first visually inspected to determine which RBP's showed clear binding signal separation between the quartiles, followed by statistical analysis in the specified regions.

ENCODE K562 RNA-seq, 4sU-seq, and TT-seq data sets

The download links and SRAs for the ENCODE K562 RNA-seq data, 4sU-seq, and TT-seq public data sets are provided in [Supplemental File 1](#).

DATA DEPOSITION

The data sets generated and/or analyzed during the current study are available in the GEO repository GSE160759.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

We would like to thank personnel at University of Michigan Advanced Genomics Core for professional technical assistance and Fan Meng and Manhong Dai for administration and maintenance of the University of Michigan Molecular and Behavioral Neuroscience Institute (MBNI) computing cluster. We also acknowledge former and current members of the Ljungman laboratory for their input and contributions to this work. This work was supported by the National Cancer Institute (NCI-R01 CA213214) and the National Human Genome Research Institute (NHGRI-UM1 HG009382).

Author contributions: K.B. and M.L. designed the study; M.T.P. performed Bru-seq experiments; K.B. and B.M. created the data analysis pipeline; K.B., B.M., and I.V.N. analyzed the data; K.B. and M.L. prepared the figures and wrote the manuscript.

Received December 24, 2020; accepted May 5, 2021.

REFERENCES

- Ameur A, Zaghlool A, Halvardson J, Wetterbom A, Gyllenstein U, Cavelier L, Feuk L. 2011. Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. *Nat Struct Mol Biol* **18**: 1435–1440. doi:10.1038/nsmb.2143
- Bauren G, Wieslander L. 1994. Splicing of Balbiani ring 1 gene pre-mRNA occurs simultaneously with transcription. *Cell* **76**: 183–192. doi:10.1016/0092-8674(94)90182-1
- Bentley DL. 2014. Coupling mRNA processing with transcription in time and space. *Nat Rev Genet* **15**: 163–175. doi:10.1038/nrg3662
- Beyer AL, Osheim YN. 1988. Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev* **2**: 754–765. doi:10.1101/gad.2.6.754
- Bhatt DM, Pandya-Jones A, Tong AJ, Barozzi I, Lissner MM, Natoli G, Black DL, Smale ST. 2012. Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell* **150**: 279–290. doi:10.1016/j.cell.2012.05.043
- Black DL. 2003. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* **72**: 291–336. doi:10.1146/annurev.biochem.72.121801.161720
- Blencowe BJ. 2000. Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochem Sci* **25**: 106–110. doi:10.1016/S0968-0004(00)01549-8
- Bot C, Pfeiffer A, Giordano F, Manjeera DE, Dantuma NP, Strom L. 2017. Independent mechanisms recruit the cohesin loader protein NIPBL to sites of DNA damage. *J Cell Sci* **130**: 1134–1146. doi:10.1242/jcs.197236
- Boutz PL, Bhutkar A, Sharp PA. 2015. Detained introns are a novel, widespread class of post-transcriptionally spliced introns. *Genes Dev* **29**: 63–80. doi:10.1101/gad.247361.114
- Braberg H, Jin H, Moehle EA, Chan YA, Wang S, Shales M, Benschop JJ, Morris JH, Qiu C, Hu F, et al. 2013. From structure to systems: high-resolution, quantitative genetic analysis of RNA polymerase II. *Cell* **154**: 775–788. doi:10.1016/j.cell.2013.07.033
- Braunschweig U, Barbosa-Morais NL, Pan Q, Nachman EN, Alipanahi B, Gonatopoulos-Pourmatzis T, Frey B, Irimia M, Blencowe BJ. 2014. Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res* **24**: 1774–1786. doi:10.1101/gr.177790.114
- Brody E, Abelson J. 1985. The “spliceosome”: yeast pre-messenger RNA associates with a 40S complex in a splicing-dependent reaction. *Science* **228**: 963–967. doi:10.1126/science.3890181
- Brody Y, Neufeld N, Bieberstein N, Causse SZ, Bohnlein EM, Neugebauer KM, Darzacq X, Shav-Tal Y. 2011. The in vivo kinetics of RNA polymerase II elongation during co-transcriptional splicing. *PLoS Biol* **9**: e1000573. doi:10.1371/journal.pbio.1000573
- Brugiolio M, Herzl L, Neugebauer KM. 2013. Counting on co-transcriptional splicing. *F1000Prime Rep* **5**: 9. doi:10.12703/P5-9
- Chathoth KT, Barrass JD, Webb S, Beggs JD. 2014. A splicing-dependent transcriptional checkpoint associated with prespliceosome formation. *Mol Cell* **53**: 779–790. doi:10.1016/j.molcel.2014.01.017
- Corvelo A, Hallegger M, Smith CW, Eyras E. 2010. Genome-wide association between branch point properties and alternative splicing. *PLoS Comput Biol* **6**: e1001016. doi:10.1371/journal.pcbi.1001016
- Darnell JE Jr. 2013. Reflections on the history of pre-mRNA processing and highlights of current knowledge: a unified picture. *RNA* **19**: 443–460. doi:10.1261/ma.038596.113
- de la Mata M, Alonso CR, Kadener S, Fededa JP, Blaustein M, Pelisch F, Cramer P, Bentley D, Kornblihtt AR. 2003. A slow RNA polymerase II affects alternative splicing in vivo. *Mol Cell* **12**: 525–532. doi:10.1016/j.molcel.2003.08.001
- Didiot MC, Tian Z, Schaeffer C, Subramanian M, Mandel JL, Moine H. 2008. The G-quartet containing FMRP binding site in FMR1 mRNA is a potent exonic splicing enhancer. *Nucleic Acids Res* **36**: 4902–4912. doi:10.1093/nar/gkn472
- Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, et al. 2012. Landscape of transcription in human cells. *Nature* **489**: 101–108. doi:10.1038/nature11233
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21. doi:10.1093/bioinformatics/bts635
- Dong S, Wang Y, Cassidy-Amstutz C, Lu G, Bigler R, Jezyk MR, Li C, Hall TM, Wang Z. 2011. Specific and modular binding code for cytosine recognition in Pumilio/FBF (PUF) RNA-binding domains. *J Biol Chem* **286**: 26732–26742. doi:10.1074/jbc.M111.244889
- Drexler HL, Choquet K, Churchman LS. 2020. Splicing kinetics and coordination revealed by direct nascent RNA sequencing through nanopores. *Mol Cell* **77**: 985–998.e988. doi:10.1016/j.molcel.2019.11.017
- Elbarbary RA, Miyoshi K, Myers JR, Du P, Ashton JM, Tian B, Maquat LE. 2017. Tudor-SN-mediated endonucleolytic decay of human cell microRNAs promotes G1/S phase transition. *Science* **356**: 859–862. doi:10.1126/science.aai9372
- Faustino NA, Cooper TA. 2003. Pre-mRNA splicing and human disease. *Genes Dev* **17**: 419–437. doi:10.1101/gad.1048803
- Filipovska A, Razif MF, Nygard KK, Rackham O. 2011. A universal code for RNA recognition by PUF proteins. *Nat Chem Biol* **7**: 425–427. doi:10.1038/nchembio.577
- Fong N, Kim H, Zhou Y, Ji X, Qiu J, Saldi T, Diener K, Jones K, Fu XD, Bentley DL. 2014. Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. *Genes Dev* **28**: 2663–2676. doi:10.1101/gad.252106.114
- Frankish A, Diekhans M, Ferreira AM, Johnson R, Jungreis I, Loveland J, Mudge JM, Sisu C, Wright J, Armstrong J, et al. 2019. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res* **47**: D766–D773. doi:10.1093/nar/gky955

- Fuchs G, Voichek Y, Benjamin S, Gilad S, Amit I, Oren M. 2014. 4sUDRB-seq: measuring genomewide transcriptional elongation rates and initiation frequencies within cells. *Genome Biol* **15**: R69. doi:10.1186/gb-2014-15-5-r69
- Graveley BR. 2000. Sorting out the complexity of SR protein functions. *RNA* **6**: 1197–1211. doi:10.1017/S1355838200000960
- Gupta YK, Nair DT, Wharton RP, Aggarwal AK. 2008. Structures of human Pumilio with noncognate RNAs reveal molecular mechanisms for binding promiscuity. *Structure* **16**: 549–557. doi:10.1016/j.str.2008.01.006
- Herzel L, Straube K, Neugebauer KM. 2018. Long-read sequencing of nascent RNA reveals coupling among RNA processing events. *Genome Res* **28**: 1008–1019. doi:10.1101/gr.232025.117
- Igel H, Wells S, Perriman R, Ares M Jr. 1998. Conservation of structure and subunit interactions in yeast homologues of splicing factor 3b (SF3b) subunits. *RNA* **4**: 1–10.
- Jia J, Long Y, Zhang H, Li Z, Liu Z, Zhao Y, Lu D, Jin X, Deng X, Xia R, et al. 2020. Post-transcriptional splicing of nascent RNA contributes to widespread intron retention in plants. *Nat Plants* **6**: 780–788. doi:10.1038/s41477-020-0688-1
- Jonkers I, Kwak H, Lis JT. 2014. Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *Elife* **3**: e02407. doi:10.7554/eLife.02407
- Juan WC, Roca X, Ong ST. 2014. Identification of *cis*-acting elements and splicing factors involved in the regulation of BIM Pre-mRNA splicing. *PLoS ONE* **9**: e95210. doi:10.1371/journal.pone.0095210
- Khodor YL, Rodriguez J, Abruzzi KC, Tang CH, Marr MT II, Rosbash M. 2011. Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in *Drosophila*. *Genes Dev* **25**: 2502–2512. doi:10.1101/gad.178962.111
- Khodor YL, Menet JS, Tolan M, Rosbash M. 2012. Cotranscriptional splicing efficiency differs dramatically between *Drosophila* and mouse. *RNA* **18**: 2174–2186. doi:10.1261/ma.034090.112
- Lam BJ, Hertel KJ. 2002. A general role for splicing enhancers in exon definition. *RNA* **8**: 1233–1241. doi:10.1017/S1355838202028030
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357–359. doi:10.1038/nmeth.1923
- Li X, Chang YH. 1996. Evidence that the human homologue of a rat initiation factor-2 associated protein (p67) is a methionine aminopeptidase. *Biochem Biophys Res Commun* **227**: 152–159. doi:10.1006/bbrc.1996.1482
- Lin RJ, Lustig AJ, Abelson J. 1987. Splicing of yeast nuclear pre-mRNA in vitro requires a functional 40S spliceosome and several extrinsic factors. *Genes Dev* **1**: 7–18. doi:10.1101/gad.1.1.7
- Lustig AJ, Lin RJ, Abelson J. 1986. The yeast RNA gene products are essential for mRNA splicing in vitro. *Cell* **47**: 953–963. doi:10.1016/0092-8674(86)90810-X
- Maniatis T, Reed R. 2002. An extensive network of coupling among gene expression machines. *Nature* **416**: 499–506. doi:10.1038/416499a
- Martins SB, Rino J, Carvalho T, Carvalho C, Yoshida M, Klose JM, de Almeida SF, Carmo-Fonseca M. 2011. Spliceosome assembly is coupled to RNA polymerase II dynamics at the 3' end of human genes. *Nat Struct Mol Biol* **18**: 1115–1123. doi:10.1038/nsmb.2124
- Min H, Turck CW, Nikolic JM, Black DL. 1997. A new regulatory protein, KSRP, mediates exon inclusion through an intronic splicing enhancer. *Genes Dev* **11**: 1023–1036. doi:10.1101/gad.11.8.1023
- Naftelberg S, Schor IE, Ast G, Kornblihtt AR. 2015. Regulation of alternative splicing through coupling with transcription and chromatin structure. *Annu Rev Biochem* **84**: 165–198. doi:10.1146/annurev-biochem-060614-034242
- Nojima T, Gomes T, Grosso ARF, Kimura H, Dye MJ, Dhir S, Carmo-Fonseca M, Proudfoot NJ. 2015. Mammalian NET-seq reveals genome-wide nascent transcription coupled to RNA processing. *Cell* **161**: 526–540. doi:10.1016/j.cell.2015.03.027
- Oesterreich FC, Herzel L, Straube K, Hujer K, Howard J, Neugebauer KM. 2016. Splicing of nascent RNA coincides with intron exit from RNA polymerase II. *Cell* **165**: 372–381. doi:10.1016/j.cell.2016.02.045
- Pan K, Lee JT, Huang Z, Wong CM. 2015. Coupling and coordination in gene expression processes with pre-mRNA splicing. *Int J Mol Sci* **16**: 5682–5696. doi:10.3390/ijms16035682
- Pandya-Jones A, Black DL. 2009. Co-transcriptional splicing of constitutive and alternative exons. *RNA* **15**: 1896–1908. doi:10.1261/rna.1714509
- Paulsen MT, Veloso A, Prasad J, Bedi K, Ljungman EA, Tsan YC, Chang CW, Tarrier B, Washburn JG, Lyons R, et al. 2013. Coordinated regulation of synthesis and stability of RNA during the acute TNF-induced proinflammatory response. *Proc Natl Acad Sci* **110**: 2240–2245. doi:10.1073/pnas.1219192110
- Paulsen MT, Veloso A, Prasad J, Bedi K, Ljungman EA, Magnuson B, Wilson TE, Ljungman M. 2014. Use of Bru-Seq and BruChase-Seq for genome-wide assessment of the synthesis and stability of RNA. *Methods* **67**: 45–54. doi:10.1016/j.ymeth.2013.08.015
- Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dundar F, Manke T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**: W160–W165. doi:10.1093/nar/gkw257
- Reimer KA, Mimoso CA, Adelman K, Neugebauer KM. 2021. Co-transcriptional splicing regulates 3' end cleavage during mammalian erythropoiesis. *Mol Cell* **81**: 998–1012.e1017. doi:10.1016/j.molcel.2020.12.018
- Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L. 2011. Improving RNA-seq expression estimates by correcting for fragment bias. *Genome Biol* **12**: R22. doi:10.1186/gb-2011-12-3-r22
- Saldi T, Riemondy K, Erickson B, Bentley DL. 2021. Alternative RNA structures formed during transcription depend on elongation rate and modify RNA processing. *Mol Cell* **81**: 1789–1801.e5. doi:10.1016/j.molcel.2021.01.040
- Schor IE, Rascovan N, Pelisch F, Allo M, Kornblihtt AR. 2009. Neuronal cell depolarization induces intragenic chromatin modifications affecting NCAM alternative splicing. *Proc Natl Acad Sci* **106**: 4325–4330. doi:10.1073/pnas.0810666106
- Schwalb B, Michel M, Zacher B, Fruhauf K, Demel C, Tresch A, Gagneur J, Cramer P. 2016. TT-seq maps the human transient transcriptome. *Science* **352**: 1225–1228. doi:10.1126/science.1249841
- Schwartz S, Meshorer E, Ast G. 2009. Chromatin organization marks exon-intron structure. *Nat Struct Mol Biol* **16**: 990–995. doi:10.1038/nsmb.1659
- Sheth N, Roca X, Hastings ML, Roeder T, Krainer AR, Sachidanandam R. 2006. Comprehensive splice-site analysis using comparative genomics. *Nucleic Acids Res* **34**: 3955–3967. doi:10.1093/nar/gkl556
- Singh J, Padgett RA. 2009. Rates of in situ transcription and splicing in large human genes. *Nat Struct Mol Biol* **16**: 1128–1133. doi:10.1038/nsmb.1666
- Sousa-Luis R, Dujardin G, Zukher I, Kimura H, Weldon C, Carmo-Fonseca M, Proudfoot NJ, Nojima T. 2021. POINT technology illuminates the processing of polymerase-associated intact nascent transcripts. *Mol Cell* **81**: 1935–1960.e6. doi:10.1016/j.molcel.2021.02.034
- Tardiff DF, Lacadie SA, Rosbash M. 2006. A genome-wide analysis indicates that yeast pre-mRNA splicing is predominantly post-transcriptional. *Mol Cell* **24**: 917–929. doi:10.1016/j.molcel.2006.12.002

- Tilgner H, Nikolaou C, Althammer S, Sammeth M, Beato M, Valcarcel J, Guigo R. 2009. Nucleosome positioning as a determinant of exon recognition. *Nat Struct Mol Biol* **16**: 996–1001. doi:10.1038/nsmb.1658
- Tilgner H, Knowles DG, Johnson R, Davis CA, Chakraborty S, Djebali S, Curado J, Snyder M, Gingeras TR, Guigo R. 2012. Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res* **22**: 1616–1625. doi:10.1101/gr.134445.111
- Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**: 511–515. doi:10.1038/nbt.1621
- Van Nostrand EL, Pratt GA, Yee BA, Wheeler EC, Blue SM, Mueller J, Park SS, Garcia KE, Gelboin-Burkhart C, Nguyen TB, et al. 2020. Principles of RNA processing from analysis of enhanced CLIP maps for 150 RNA binding proteins. *Genome Biol* **21**: 90. doi:10.1186/s13059-020-01982-9
- Veloso A, Kirkconnell KS, Magnuson B, Biewen B, Paulsen MT, Wilson TE, Ljungman M. 2014. Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. *Genome Res* **24**: 896–905. doi:10.1101/gr.171405.113
- Windhager L, Bonfert T, Burger K, Ruzsics Z, Krebs S, Kaufmann S, Malterer G, L'Hernault A, Schilhabel M, Schreiber S, et al. 2012. Ultrashort and progressive 4sU-tagging reveals key characteristics of RNA processing at nucleotide resolution. *Genome Res* **22**: 2031–2042. doi:10.1101/gr.131847.111
- Yeo G, Burge CB. 2004. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* **11**: 377–394. doi:10.1089/1066527041410418
- Zamore PD, Patton JG, Green MR. 1992. Cloning and domain structure of the mammalian splicing factor U2AF. *Nature* **355**: 609–614. doi:10.1038/355609a0
- Zhang X, Yan C, Zhan X, Li L, Lei J, Shi Y. 2018. Structure of the human activated spliceosome in three conformational states. *Cell Res* **28**: 307–322. doi:10.1038/cr.2018.14
- Zhao Y, Yu L, Fu Q, Chen W, Jiang J, Gao J, Zhao S. 2000. Cloning and characterization of human DDX24 and mouse Ddx24, two novel putative DEAD-Box proteins, and mapping DDX24 to human chromosome 14q32. *Genomics* **67**: 351–355. doi:10.1006/geno.2000.6255