Published in final edited form as:

Wiley Interdiscip Rev RNA. 2017 March ; 8(2): . doi:10.1002/wrna.1401.

Perfect timing: splicing and transcription rates in living cells

Tara Alpert, Lydia Herzel, and Karla M. Neugebauer*

Department of Molecular Biophysics and Biochemistry, Yale University, New Haven CT

Abstract

An important step towards understanding gene regulation is the elucidation of the time necessary for the completion of individual steps. Measurement of reaction rates can reveal potential nodes for regulation. For example, measurements of *in vivo* transcription elongation rates reveal regulation by DNA sequence, gene architecture, and chromatin. Pre-mRNA splicing is regulated by transcription elongation rates and vice versa, yet the rates of RNA processing reactions remain largely elusive. Since the 1980s, numerous model systems and approaches have been used to determine the precise timing of splicing *in vivo*. Because splicing can be co-transcriptional, the position of Pol II when splicing is detected has been used as a proxy for time by some investigators. In addition to these “distance-based” measurements, “time-based” measurements have been possible through live cell imaging, metabolic labeling of RNA, and gene induction. Yet splicing rates can be convolved by the time it takes for transcription, spliceosome assembly and spliceosome disassembly. The variety of assays and systems used has, perhaps not surprisingly, led to reports of widely differing splicing rates *in vivo*. Recently, single molecule RNA-seq has indicated that splicing occurs more quickly than previously deduced. Here we comprehensively review these findings and discuss evidence that splicing and transcription rates are closely coordinated, facilitating the efficiency of gene expression. On the other hand, introduction of splicing delays through as yet unknown mechanisms provide opportunity for regulation. More work is needed to understand how cells optimize the rates of gene expression for a range of biological conditions.

Graphical Abstract

*corresponding author: karla.neugebauer@yale.edu.

The authors have no conflict of interest.

Introduction

Gene expression begins with the synthesis of RNA from the DNA template by RNA polymerase. In eukaryotes, this nascent RNA must be further processed, transported within the cell, translated (in the case of mRNA), and ultimately degraded. The steady-state level of any RNA in the cell reflects the balance of its synthesis and degradation; yet, many reactions take place between these two end points. Each step of mRNA processing – 5' end capping, splicing, editing, 3' end cleavage and polyadenylation, nucleotide modification, and mRNP maturation – also takes time. If all of these reactions were to take place post-transcriptionally, one could add the time it takes to transcribe a gene to the time it takes to perform RNA processing and determine the total time for mRNA synthesis. However, because many of these reactions at least begin during transcription (1-3), transcription and RNA processing rates are convoluted, raising the possibility that gene expression is on the whole more efficient. For example, during the reductive cleavage divisions of early embryos, cell cycle lengths are as short as 15 minutes; yet, early gene expression takes place in this time interval, which appears to limit the length of genes that can be transcribed (4). Transcription and RNA processing rates in early embryos are currently unknown but are required to fully understand the time constraints placed on gene expression in rapidly dividing cells and possible regulatory mechanisms that contribute. In contrast, post-mitotic cells would appear to lack such time constraints; however, other constraints may arise due to other regulatory processes, such as cell signaling or responses to environmental cues. These scenarios illustrate the necessity to measure rates *in vivo*.

The early observation that spliced nascent transcripts are attached to the DNA axis in chromosome spreads (Fig 1A) stimulated the field to consider the possibility that the physical proximity of RNA polymerase II (Pol II), chromatin, the splicing machinery and the nascent RNP provide opportunities for regulation. Several RNA processing events, including 5' end capping and 3' end cleavage and polyadenylation, are physically coupled to transcription through factor binding to Pol II (3). For example, capping enzymes bind the

Pol II C-terminal domain (CTD) phosphorylated on the serine 5 residue of the CTD repeat (5). The central role of splicing – the removal of introns and ligation of exons – in gene expression has motivated many groups to ask when exactly splicing occurs during this window of transcription. Understanding the potential coupling between splicing and transcription could reveal significant insights into the relationship between variation in Pol II elongation rates over introns and exons, pause sites, chromatin and alternative splicing. For example, local changes in elongation rates affect the gene position of splicing and/or alternative splicing in a variety of systems (2, 3, 6-10). In humans, transcription can take anywhere from a few seconds up to several hours for a single gene, seemingly an eternity for a chemical reaction. In spite of this, evidence from fern spores, platelets and induced macrophage provide examples of biological scenarios in which splicing delays lead to stable transcripts with retained introns that are removed at a later time depending on signaling (11-13). Clearly, co-transcriptionality is not required for splicing; yet, most splicing – from yeast to man – is co-transcriptional (1, 3, 14-22).

Splicing is accomplished by a macromolecular complex, called the spliceosome, which must assemble in a step-wise fashion *de novo* on every intron (23). In instances of co-transcriptional splicing, spliceosome assembly and catalysis clearly take place before transcription termination (Fig 1B). Note that even in instances of post-transcriptional catalysis, spliceosome assembly may well proceed co-transcriptionally (16). In yeast, the U1 snRNP, which recognizes the 5' splice site (5'SS), accumulates on nascent RNA and is detectable by chromatin immunoprecipitation (ChIP) over introns, consistent with its recruitment to 5'SSs as soon as they are transcribed (24, 25). The accumulation of U2 and U5 snRNPs as well as the nineteen complex (NTC, present in the active spliceosome) downstream of 3' splice sites (3'SSs) is accompanied by the loss of U1 snRNP, consistent with the step-wise maturation and function of the spliceosome (26-28). Intriguingly, histone post-translational modifications, such as acetylation by GCN5, modify co-transcriptional spliceosome assembly patterns (29-32), indicating that regulation can be mediated by chromatin (2, 3).

Measurement of co-transcriptional spliceosome assembly by ChIP speaks to the co-transcriptionality of splice site recognition by snRNP and non-snRNP splicing factors but does not reveal splicing rates. The abundance of any of these required factors along the length of a gene does not reveal the frequency or efficiency of catalysis and could also reflect their retention during product RNA release from spliceosomes and spliceosome disassembly. Thus, the field has explored many methods that aim to determine splicing kinetics *in vivo*; these can be divided into two categories: distance-based techniques, which measure the gene location of Pol II when splicing has occurred, and time-based techniques, which measure the time it takes for splicing to complete (Fig 2). Integration of spatial and temporal information will ultimately advance our understanding of cross-regulation between transcription and splicing and provide insight into how the timing of these processes contributes to gene expression in a variety of biological contexts.

Distance-based measurements of *in vivo* splicing kinetics

Historically, co-transcriptional splicing was discovered and measured by analyzing electron micrographs of chromatin spreads (Fig 1A). Osheim et al observed two particular species of RNP particles on nascent transcripts, a 25 nm particle and a 40 nm particle that appeared to be a combination of two 25 nm particles (33). These particles appeared at predictable sites along nascent RNA transcripts of early embryo genes from *Drosophila melanogaster* and were predicted to represent subunits of the spliceosome based on several observations including RNA looping between the two particles and eventual disappearance of the particle after loop removal. By measuring the DNA distance in micrometers (μm) from known locations on the gene and assuming a maximum chromatin compaction of 4.8 kb/ μm , this study measured one of the first instances of *in vivo* splicing occurrence 4.5 kb after the 3' splice site (Fig 2) and began a decades-long investigation (34). Chromatin spreads may not provide the most accurate analysis of splicing kinetics due to the disruption of chromatin with harsh conditions. This extensive disruption could break interactions between Pol II, DNA and nascent RNA and lead to an underrepresentation of co-transcriptional splicing. For instance, a simple decrease in pH from 9.0 to 7.5 during chromatin preparation gave different results (34). Nevertheless, many of the observations made using chromosome spreading methods are consistent with findings in a different model system, the Balbiani ring of *Chironomus tentans*, where it has been possible to dissect out chromosomal regions and infer the progress of splicing by RT-PCR (35). Thus, we garnered much of our early knowledge about co-transcriptional splicing from such model systems. Although direct visualization of single RNA molecules in chromatin spreads is something modern techniques have lacked, recent nascent RNA sequencing strategies (see below) have enabled their visualization by other means.

When it first became clear that co-transcriptional splicing was a likely scenario, the most urgent question being posed was precisely when does splicing occur during the process of transcription? ChIP is a versatile tool for acquiring information about co-transcriptional spliceosome assembly (see above). Rosbash and colleagues utilized MS2 stem loops to assay splicing status along the length of HZ18 reporter genes, which harbor MS2 RNA stem loops in either the intron or expressed as two halves of the MS2 RNA stem loop in its exons, such that the loop forms after splicing (8). MS2-coat protein binds the stem loop and serves as a target in ChIP experiments that aimed to detect the Pol II position along the gene the moment splicing occurred. MS2 ChIP signal 1.5 kb downstream of the intron suggested that splicing takes place long after intron synthesis (8). Moreover, these studies employed an intronic ribozyme sequence to measure splicing in a more direct way, in which RT-qPCR was used to determine the relative fraction of RNAs that had been self-cleaved by the ribozyme as opposed to those species that co-transcriptionally spliced the ribozyme before cleavage. These assays identified the gene location of splicing between ~500 bp and ~1.5 kb downstream of the 3' SS. Taking an average measured transcription elongation rate of 1.5kb/min in budding yeast into account (36), these authors reckoned that splicing occurs within 20-60 seconds *in vivo*. Because the majority of genes in budding yeast have terminal exons that are shorter than 1.5kb (median 434bp), Rosbash and colleagues concluded that most splicing must be post-transcriptional in yeast (27). A direct test of co-transcriptional

splicing frequency in yeast refuted this conclusion by showing that the majority of introns are removed in nascent RNA (37); these authors identified terminal exon pausing (TEP) by Pol II and suggested that TEP might compensate for short terminal exons by providing more time for splicing. Meanwhile additional transcriptional pausing events have been linked to splicing (38-40), further obscuring the time axis.

Time-based measurement of *in vivo* splicing kinetics

The development of alternative approaches that measure splicing in terms of time compensates for the limitations of ChIP-based approaches, which are unavoidably indirect, and imaging of nascent RNA through methods such as chromatin spreads, which rely on specialized model systems and are difficult to adapt to global determinations (Fig 2). These methods include live cell imaging to quantify intron lifetimes and snRNP dynamics and metabolic labeling to quantify intron lifetimes and the emergence of spliced RNAs. Several studies have set out to determine when splicing occurs by quantitatively measuring the amount of spliced and unspliced RNA transcripts using RT-qPCR or RNA-seq. Some time-based studies have implemented a system of inducible transcription that allows tracking of (pre-)mRNA intermediates from a given gene over a time course. To date, the highest time resolution achieved with RT-qPCR analysis is 30 seconds, during which an integrated, tetracycline-inducible reporter gene in budding yeast yielded observed spliced transcripts 60 seconds after pre-mRNA transcription was induced (41).

In mammalian cells, where genes and introns are very large, longer time points are needed to sample the transcription and splicing of induced transcripts. The stimulation of mouse and human macrophage is a classic experimental system for activating a program of gene expression. Making use of this paradigm, Bhatt et al characterized the differential appearance of unspliced and spliced transcripts in chromatin, nucleoplasm and cytoplasm by RNA-seq over a time course of 15, 30, 60, and 120 minutes after Lipid-A stimulation (11, 42). These widely spaced time points inform on cellular dynamics and regulation but are not designed to provide precise measurement of splicing rates. Notably, the 11 most quickly transcribed genes displayed rapid 3' end cleavage and polyadenylation but delays in splicing of at least some introns (peak at 30 min) relative to transcription (peak at 15 min). In a related approach, the Regev lab used metabolic labeling of LPS-stimulated mouse dendritic cells and RNA-seq (43), yielding a range of pre-mRNA half-lives from 1 to 30 minutes, with a median intron half-life of 14 minutes (Fig 3A and see below). In contrast to the above results, quickly spliced genes were generally longer transcripts with many introns and typically expressed quickly in response to a stimulus, while slowly spliced genes were shorter and constantly expressed at high levels (43, 44). Although the differences between the findings from these two labs are unclear, these observations show that slow rates of splicing can affect transcript release and eventual translation.

Analysis of nascent RNA at different time points has enabled investigators to address specific hypotheses. For example, the Padgett group set out to measure transcription and splicing kinetics for very long genes and introns to determine whether special regulation accounts for the expression of such genes (45). To achieve this, they developed a system to observe splicing of endogenous human genes *in vivo*. Using treatment and then wash out of

a reversible inhibitor of Pol II elongation, 5,6-dichloro-1- β -D-ribofuranosylbenzimidazole (DRB), the authors collected time points of newly synthesized transcripts that were then quantified using RT-qPCR. They found that despite the incredible length of introns assayed (>100 kb) all introns were spliced in the first or second time point (5 and 10 min), which thereby delimits a maximum time window for splicing in human Tet21 cells. A related study conducted in human mammary epithelial MCF10A cells stimulated with EGF measured pre-mRNA half-life of 2-3 minutes by RT-qPCR (46), placing splicing within the same time window in another cell type. Interestingly, the Padgett study also quantitated Pol II elongation rates of over these long introns at 3.8 kb/min, a relatively high value that agrees with the range of transcription rates measured more recently in human cells (0.5-4 kb/min) and the report of faster elongation rates along introns (47-49). The Padgett study was therefore critical in confirming that splicing is relatively quick and often cotranscriptional, even in endogenous human genes with extremely long introns.

Studies utilizing live cell imaging for the analysis of splicing kinetics rely primarily on reporter genes in human cells (Fig 3B). Results from these experiments vary widely, even when using a similar reporter element. For example, two studies used stably integrated β -globin reporter genes with MS2 or PP7 stem loops inserted into intronic or exonic sequences to track pre-mRNA (50, 51). The first study reported splicing in HEK293 cells within 20 and 30 seconds for the first and second introns, respectively. The second study observed splicing 267 seconds after transcription of the 3' SS of the terminal intron in U2OS cells. The different analyses carried out in these two studies may contribute to the deduced splicing rates. Coulon et al. applies an autocorrelation function to the fluorescent fluctuations to determine how the fluorescence of the intronic/exonic signal is correlated to itself after a certain time delay. Martin et al. identifies short bursts of transcription and quantifies the number of introns in each burst and the lifetime in seconds of the burst of fluorescence. Median lifetimes for each intron is considered the time window in which splicing takes place.

Another possible source of variation among live cell imaging studies is the choice of cell line. Perhaps U2OS and HEK293 cells are characterized by different splicing rates. Interestingly, Schmidt et al. used a very similar approach in U2OS cells, by measuring the fluorescent half-life of the MINX reporter intron labeled with MS2-GFP (52). This study reports an intronic half-life of 105 seconds (1.75 min), in between the two values discussed above, arguing against cell-specificity. On the other hand, Schmidt et al. incorporated a bGH polyadenylation/cleavage signal that is known to be slowly processed. This signal was used to ensure that the construct would splice co-transcriptionally; however, slower 3' end processing could potentially feed back to splicing and affect their measurement.

The attraction of live-cell imaging of fluorescent reporters, such as those described in the above experiments, is that these measurements in real time provide higher time resolution than metabolic labeling or time points taken for RT-qPCR after gene induction. If splicing is or can be a very fast reaction, 5- and 10-minute time points are not sufficient to deepen our understanding of splicing kinetics. Additionally, single cell information can provide insight into cell-to-cell variation and compensate for the lack of synchronization in a large cell population. However, chromosomally integrated reporter genes will never report the

diversity of kinetics available in endogenous genomes. Splicing rates likely differ among genes as well as introns and exons within each gene; thereby, reporter genes miss out on this source of information. Differences between cell lines and reporters (see above) may also be attributable to unknown regulatory features of the insertion site; indeed, one of the β -globin reporters used above was less efficiently spliced than the wild type transcript (50), indicating that even subtle modifications to this gene had a noticeable effect on splicing. To understand splicing kinetics in complex cellular systems, it is prudent to look at a sampling of endogenous genes.

While most groups measure intron half-lives with fluorescent probes, one study took a global approach by measuring the residency times of fluorescently-labeled spliceosomal snRNPs on pre-mRNA transcripts using FRAP and FCS in HeLa cells (53). While U1 and U4 snRNPs that transiently associate with the assembling spliceosome display shorter residency times, subunits of the active spliceosome – U2 and U5 snRNPs – reside on pre-mRNA for 15-30 seconds. This indicates that the average splicing duration at steady state in HeLa cells lies within a 30 second window. These results are in good agreement with Martin et al., however much shorter than reports by Coulon et al. and Schmidt et al. While the range of splicing rates derived from fluorescent assays seems quite variable, it is clear that the majority of splicing in mammalian cells is faster than estimates from induced systems, such as stimulated macrophage. Finally it is important to note that measurements of intron half-lives, as well as snRNP dynamics will encompass remaining transcription of intron and exon elements, spliceosome assembly, splicing, spliceosome disassembly and/or intron release, intron debranching and degradation. Thus, the overall range of times observed – 0.5-3 minutes – could reflect differences in the rates of any one or more of these processes.

Complementary to fluorescence measurements, metabolic labeling affords time-resolved datasets (Fig 3C). In addition, finding a way to purify only the youngest cellular transcripts greatly increases the depth of any analysis to follow. Typically, cells are fed with 4-thiouracil (4tU) or 4-thiouridine (4sU), which can be incorporated into newly synthesized RNA (54), for different time intervals. Thiol-specific biotinylation and subsequent purification on streptavidin-coated magnetic beads of metabolically labeled RNA follows total RNA purification from cells. Transcript levels are determined by RNA-seq, microarray analysis or RT-qPCR. This approach has been used to quantify transcription and degradation dynamics by labeling newly transcribed RNA for minutes to hours (55, 56). Technical advances in sample preparation and sequencing allow for shorter labeling times from 1.5 to 5 minutes and thus allow us to zoom in on transcriptome-wide RNA processing dynamics in a relatively unperturbed way (17, 43, 44, 57, 58).

The amount of metabolically labeled RNA depends on cellular uptake of modified bases, labeling time, incorporation rate, transcription rate, processing rate and degradation rate. The first three are pre-defined or can be determined in initial experiments by measuring the relative proportion of labeled versus unlabeled RNA (43). The latter can be inferred by applying mathematical models to the data (43, 44, 58) or through analyzing the resulting fraction of pre-mRNA and mRNA at different experimental time points (17, 57). A steady state total RNA-seq dataset serves as reference dataset in both approaches.

All five above mentioned studies detect substantial amounts of pre-mRNA splicing at the very first labeling time point: 1.5 minutes in *S. cerevisiae* (57), 2 minutes in *S. pombe* with a median intron splicing time of 37 seconds (58), 5 minutes in human B cells (17), and 10 minutes in LPS-stimulated mouse dendritic cells with a median intron splicing time of 14 minutes (44). These genome-scale datasets identify broad ranges in splicing duration from 1-30 minutes (43), 1-223 minutes (44), and 4 seconds-6 minutes (58). Direct comparison among these values is difficult, given differences in experimental systems and analysis methods. It is also important to note that these time windows reflect the convolution of transcription, processing and degradation, as is the case for live-cell imaging approaches addressed above. Nevertheless, the global nature of metabolic labeling has enabled the discovery of gene architecture and sequence motif correlations linked to synthesis, degradation and splicing kinetics (17, 43, 57, 58), which drive hypotheses for future investigation.

Revisiting the distance-based approach through single molecule RNA-seq

The most recently published methods – Single Molecule Intron Tracking (SMIT) and long read sequencing of nascent RNA – use 3' end linker ligation and single molecule RNA-seq to identify the position of Pol II along the length of genes at the moment introns are removed (59). SMIT uses paired-end sequencing to associate splicing status with Pol II position; when compiled, the data indicate the percentage of molecules spliced at a given Pol II position with approximately 300 reads per position (Fig 4A). Direct, long read sequencing of nascent transcripts with Pol II positions marked by their 3' ends (Fig 4B) provide images reminiscent of chromatin spreads (see Fig 1A). SMIT analysis of 87 endogenous genes and long read sequencing of nascent RNA from *S. cerevisiae* and *S. pombe* reveal exon-exon ligation detectable when Pol II is 26 and 36 nt downstream of the 3' SS, respectively. These findings indicate that the active spliceosome is physically very close to Pol II, harkening back to tomography images of nascent RNPs in the Balbiani ring system (Fig 4C). In these images, the U2 snRNP and Pol II are shown to be in the same electron dense particle adjacent to the DNA axis (60). Because 15 nt of nascent RNA is embedded in the elongating polymerase (5), it seems likely that the active spliceosome is positioned at the exit channel of Pol II similar to the 5' end capping enzymes (Fig 4D). These findings revitalize models that consider direct interactions between Pol II and spliceosome components (3). The mammalian U2AF and FUS proteins, which interact directly with Pol II and components of the splicing machinery, are examples of factors that bridge the two machineries (61, 62). Given the observed association of particular Pol II CTD phosphorylation states with spliceosome assembly and splicing (63, 64), it is intriguing to consider the regulation of such interactions during the elongation process.

SMIT precisely measures the occurrence of the second step of splicing – exon-exon ligation – in distance. However, evidence that Pol II did not pause at or near 3'SSs of the 87 genes analyzed permits an estimation of splicing rate in time; SMIT curves reached saturation at a median Pol II position of 129 nt downstream of 3'SSs, indicating that splicing completes within ~5 seconds if transcription elongation proceeds at 1.5kb/min (36, 59). This fast rate of splicing in budding and fission yeasts can be considered one component of the whole splicing process, including spliceosome disassembly and intron degradation, as measured by

others (see above, Fig 2). The impressive speed of the splicing reaction as measured by SMIT begs the question of how alternative splice sites could be chosen if exon ligation is already complete a mere 129 nt after the 3'SS. Given the extensive network of splicing regulation in metazoans, a myriad of elements could play a role in delaying splicing until alternative sites and regulatory sequences have been transcribed. Moreover, trans-acting splicing factors, histone post-translational modifications, and nucleosome positioning may alter transcription and splicing kinetics (16, 29, 65-69). The observed fast rates of splicing in *S. pombe* support the idea that splicing is inherently quick, regardless of evolutionary distance. More experiments are needed to understand the kinetics of splicing in higher metazoans.

Pausing has been reported in yeast and human cells under various conditions and may impact the time estimate derived from Pol II position (38-40, 63, 64, 70). Changes in splicing kinetics measured with distance-based methods would be expected to reflect such changes in transcription elongation. Indeed, a mutation in Pol II that causes faster transcription elongation, leads to lengthening of the distance measurement by SMIT and an increase in levels of unspliced transcripts (59, 71). Thus, faster Pol II elongation rates result in greater physical separation between the spliceosome and Pol II, while slower elongation would be expected to bring the two machines closer together, consistent with the observation that slow Pol II mutants correlate with increased levels of spliced transcripts (71). Together with recent findings that mammalian transcription and splicing rates *in vivo* are optimized (72), the data suggest that the two rates have co-evolved. The matching of the reaction rates of splicing and transcription elongation rates observed in budding and fission yeast suggest perfect timing in the coordination of these two macromolecular machines.

Outlook

With the diverse panel of methods for investigating transcription and splicing rates now available, we can expect future work to illuminate the mechanisms responsible for achieving perfect timing in cells. As *in vivo* rates of transcription and splicing come online, it will become increasingly possible to use modeling to predict gene expression outcomes and/or to extract rate parameters from steady state RNA-seq data. A current black box is the transition from co-transcriptional to post-transcriptional splicing, in which it can be anticipated that transcripts that are incompletely spliced during transcription will continue to be spliced after poly-A cleavage. As metabolically labeled RNA corresponds to young but not necessarily nascent RNA, metabolic labeling in combination with cellular fractionation offers an opportunity to address co- and post-transcriptional splicing within the same cellular system. Additional features such as polyA cleavage rates and RNA stability could be addressed with this approach. In order to distinguish between co- and post-transcriptional processes even shorter labeling times, differential analysis of polyA⁺ and polyA⁻ nascent RNA and technical improvements, such as enhancing 4sU-/4tU-labeled RNA recovery with more efficient biotinylation methods (73), can be imagined. Computational approaches (44, 58, 74-76) will be increasingly important for data interpretation and to create testable hypotheses that will ultimately provide new insights into modes of regulation in multiple species and a variety of biological contexts.

Acknowledgments

This work was supported by the National Institutes of Health (CMB TG T32GM007223 to Tara Alpert and NIH R01 GM112766 from the NIGMS to Karla Neugebauer).

References

1. Brugiolo M, Herzel L, Neugebauer KM. Counting on co-transcriptional splicing. *F1000Prime Rep.* 2013; 5:9. [PubMed: 23638305]
2. Naftelberg S, Schor IE, Ast G, Kornbliht AR. Regulation of alternative splicing through coupling with transcription and chromatin structure. *Annu Rev Biochem.* 2015; 84:165–98. [PubMed: 26034889]
3. Saldi T, Cortazar MA, Sheridan RM, Bentley DL. Coupling of RNA Polymerase II Transcription Elongation with Pre-mRNA Splicing. *Journal of molecular biology.* 2016; 428(12):2623–35. [PubMed: 27107644]
4. Heyn P, Kalinka AT, Tomancak P, Neugebauer KM. Introns and gene expression: cellular constraints, transcriptional regulation, and evolutionary consequences. *Bioessays.* 2015; 37(2):148–54. [PubMed: 25400101]
5. Martinez-Rucobo FW, Kohler R, van de Waterbeemd M, Heck AJR, Hemann M, Herzog F, et al. Molecular Basis of Transcription-Coupled Pre-mRNA Capping. *Molecular cell.* 2015; 58(6):1079–89. [PubMed: 25959396]
6. Howe KJ, Kane CM, Ares M Jr. Perturbation of transcription elongation influences the fidelity of internal exon inclusion in *Saccharomyces cerevisiae*. *RNA.* 2003; 9(8):993–1006. [PubMed: 12869710]
7. Roberts GC, Gooding C, Mak HY, Proudfoot NJ, Smith CW. Co-transcriptional commitment to alternative splice site selection. *Nucleic Acids Res.* 1998; 26(24):5568–72. [PubMed: 9837984]
8. Lacadie SA, Tardiff DF, Kadener S, Rosbash M. In vivo commitment to yeast cotranscriptional splicing is sensitive to transcription elongation mutants. *Genes Dev.* 2006; 20(15):2055–66. [PubMed: 16882983]
9. de la Mata M, Alonso CR, Kadener S, Fededa JP, Blaustein M, Pelisch F, et al. A slow RNA polymerase II affects alternative splicing in vivo. *Mol Cell.* 2003; 12(2):525–32. [PubMed: 14536091]
10. Dujardin G, Lafaille C, de la Mata M, Marasco LE, Munoz MJ, Le Jossic-Corcus C, et al. How slow RNA polymerase II elongation favors alternative exon skipping. *Mol Cell.* 2014; 54(4):683–90. [PubMed: 24793692]
11. Bhatt DM, Pandya-Jones A, Tong AJ, Barozzi I, Lissner MM, Natoli G, et al. Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell.* 2012; 150(2):279–90. [PubMed: 22817891]
12. Boothby TC, Zipper RS, van der Weele CM, Wolniak SM. Removal of retained introns regulates translation in the rapidly developing gametophyte of *Marsilea vestita*. *Dev Cell.* 2013; 24(5):517–29. [PubMed: 23434411]
13. Denis MM, Tolley ND, Bunting M, Schwartz H, Jiang H, Lindemann S, et al. Escaping the nuclear confines: signal-dependent pre-mRNA splicing in anucleate platelets. *Cell.* 2005; 122(3):379–91. [PubMed: 16096058]
14. Herzel L, Neugebauer KM. Quantification of co-transcriptional splicing from RNA-Seq data. *Methods.* 2015; 85:36–43. [PubMed: 25929182]
15. Tilgner H, Knowles DG, Johnson R, Davis CA, Chakraborty S, Djebali S, et al. Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res.* 2012; 22(9):1616–25. [PubMed: 22955974]
16. Vargas DY, Shah K, Batish M, Levandoski M, Sinha S, Marras SA, et al. Single-molecule imaging of transcriptionally coupled and uncoupled splicing. *Cell.* 2011; 147(5):1054–65. [PubMed: 22118462]

17. Windhager L, Bonfert T, Burger K, Ruzsics Z, Krebs S, Kaufmann S, et al. Ultrashort and progressive 4sU-tagging reveals key characteristics of RNA processing at nucleotide resolution. *Genome Res.* 2012; 22(10):2031–42. [PubMed: 22539649]
18. Ameer A, Zaghlool A, Halvardson J, Wetterbom A, Gyllenstein U, Cavelier L, et al. Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. *Nature structural & molecular biology.* 2011; 18(12):1435–40.
19. Khodor YL, Rodriguez J, Abruzzi KC, Tang CH, Marr MT 2nd, Rosbash M. Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in *Drosophila*. *Genes Dev.* 2011; 25(23):2502–12. [PubMed: 22156210]
20. Khodor YL, Menet JS, Tolan M, Rosbash M. Cotranscriptional splicing efficiency differs dramatically between *Drosophila* and mouse. *RNA.* 2012; 18(12):2174–86. [PubMed: 23097425]
21. Girard C, Will CL, Peng J, Makarov EM, Kastner B, Lemm I, et al. Post-transcriptional spliceosomes are retained in nuclear speckles until splicing completion. *Nature communications.* 2012; 3:994.
22. Carrillo Oesterreich F, Preibisch S, Neugebauer KM. Global analysis of nascent RNA reveals transcriptional pausing in terminal exons. *Mol Cell.* 2010; 40(4):571–81. [PubMed: 21095587]
23. Wahl MC, Will CL, Luhrmann R. The spliceosome: design principles of a dynamic RNP machine. *Cell.* 2009; 136(4):701–18. [PubMed: 19239890]
24. Kotovic KM, Lockshon D, Boric L, Neugebauer KM. Cotranscriptional recruitment of the U1 snRNP to intron-containing genes in yeast. *Molecular and Cellular Biology.* 2003; 23(16):5768–79. [PubMed: 12897147]
25. Gornemann J, Barrandon C, Hujer K, Rutz B, Rigaut G, Kotovic KM, et al. Cotranscriptional spliceosome assembly and splicing are independent of the Prp40p WW domain. *RNA.* 2011; 17(12):2119–29. [PubMed: 22020974]
26. Gornemann J, Kotovic KM, Hujer K, Neugebauer KM. Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. *Molecular cell.* 2005
27. Tardiff DF, Lacadie SA, Rosbash M. A genome-wide analysis indicates that yeast premRNA splicing is predominantly posttranscriptional. *Molecular cell.* 2006; 24(6):917–29. [PubMed: 17189193]
28. Lacadie SA, Rosbash M. Cotranscriptional spliceosome assembly dynamics and the role of U1 snRNA:5'ss base pairing in yeast. *Mol Cell.* 2005; 19(1):65–75. [PubMed: 15989965]
29. Gunderson FQ, Johnson TL. Acetylation by the transcriptional coactivator Gcn5 plays a novel role in co-transcriptional spliceosome assembly. *PLoS genetics.* 2009; 5(10):e1000682. [PubMed: 19834536]
30. Gunderson FQ, Merkhofer EC, Johnson TL. Dynamic histone acetylation is critical for cotranscriptional spliceosome assembly and spliceosomal rearrangements. *Proc Natl Acad Sci U S A.* 2011; 108(5):2004–9. [PubMed: 21245291]
31. Herissant L, Moehle EA, Bertaccini D, Van Dorsselaer A, Schaeffer-Reiss C, Guthrie C, et al. H2B ubiquitylation modulates spliceosome assembly and function in budding yeast. *Biol Cell.* 2014; 106(4):126–38. [PubMed: 24476359]
32. Patrick KL, Ryan CJ, Xu J, Lipp JJ, Nissen KE, Roguev A, et al. Genetic interaction mapping reveals a role for the SWI/SNF nucleosome remodeler in spliceosome activation in fission yeast. *PLoS Genet.* 2015; 11(3):e1005074. [PubMed: 25825871]
33. Samson D, Dittmore QM, Beyer CW Jr. Intravascular use of isobutyl 2-cyanoacrylate: Part 1 Treatment of intracranial arteriovenous malformations. *Neurosurgery.* 1981; 8(1):43–51. [PubMed: 7207772]
34. Beyer AL, Osheim YN. Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev.* 1988; 2(6):754–65. [PubMed: 3138163]
35. Bjork P, Wieslander L. The Balbiani Ring Story: Synthesis, Assembly, Processing, and Transport of Specific Messenger RNA-Protein Complexes. *Annu Rev Biochem.* 2015; 84:65–92. [PubMed: 26034888]
36. Mason PB, Struhl K. Distinction and relationship between elongation rate and processivity of RNA polymerase II in vivo. *Mol Cell.* 2005; 17(6):831–40. [PubMed: 15780939]

37. Oesterreich FC, Preibisch S, Neugebauer KM. Global analysis of nascent RNA reveals transcriptional pausing in terminal exons. *Molecular Cell*. 2010; 40(4):571–81. [PubMed: 21095587]
38. Alexander RD, Innocente SA, Barrass JD, Beggs JD. Splicing-dependent RNA polymerase pausing in yeast. *Molecular Cell*. 2010; 40(4):582–93. [PubMed: 21095588]
39. Chathoth KT, Barrass JD, Webb S, Beggs JD. A splicing-dependent transcriptional checkpoint associated with prespliceosome formation. *Molecular Cell*. 2014; 53(5):779–90. [PubMed: 24560925]
40. Kwak H, Fuda NJ, Core LJ, Lis JT. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science*. 2013; 339(6122):950–3. [PubMed: 23430654]
41. Alexander RD, Barrass JD, Dichtl B, Kos M, Obtulowicz T, Robert M-C, et al. RiboSys, a high-resolution, quantitative approach to measure the in vivo kinetics of pre-mRNA splicing and 3'-end processing in *Saccharomyces cerevisiae*. *RNA*. 2010; 16(12):2570–80. [PubMed: 20974745]
42. Pandya-Jones A, Bhatt DM, Lin C-H, Tong A-J, Smale ST, Black DL. Splicing kinetics and transcript release from the chromatin compartment limit the rate of Lipid A-induced gene expression. *RNA*. 2013; 19(6):811–27. [PubMed: 23616639]
43. Rabani M, Levin JZ, Fan L, Adiconis X, Raychowdhury R, Garber M, et al. Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nature Biotechnology*. 2011; 29(5):436–42.
44. Rabani M, Raychowdhury R, Jovanovic M, Rooney M, Stumpo DJ, Pauli A, et al. High-resolution sequencing and modeling identifies distinct dynamic RNA regulatory strategies. *Cell*. 2014; 159(7):1698–710. [PubMed: 25497548]
45. Singh J, Padgett RA. Rates of in situ transcription and splicing in large human genes. *Nature Structural & Molecular Biology*. 2009; 16(11):1128–33.
46. Zeisel A, Kostler WJ, Molotski N, Tsai JM, Krauthgamer R, Jacob-Hirsch J, et al. Coupled pre-mRNA and mRNA dynamics unveil operational strategies underlying transcriptional responses to stimuli. *Molecular systems biology*. 2011; 7:529. [PubMed: 21915116]
47. Jonkers I, Kwak H, Lis JT. Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *Elife*. 2014; 3:e02407. [PubMed: 24843027]
48. Jonkers I, Lis JT. Getting up to speed with transcription elongation by RNA polymerase II. *Nat Rev Mol Cell Biol*. 2015; 16(3):167–77. [PubMed: 25693130]
49. Veloso A, Kirkconnell KS, Magnuson B, Biewen B, Paulsen MT, Wilson TE, et al. Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. *Genome Res*. 2014; 24(6):896–905. [PubMed: 24714810]
50. Martin RM, Rino J, Carvalho C, Kirchhausen T, Carmo-Fonseca M. Live-cell visualization of pre-mRNA splicing with single-molecule sensitivity. *Cell reports*. 2013; 4(6):1144–55. [PubMed: 24035393]
51. Coulon A, Ferguson ML, de Turris V, Palangat M, Chow CC, Larson DR. Kinetic competition during the transcription cycle results in stochastic RNA processing. *eLife*. 2014;3.
52. Schmidt U, Basyuk E, Robert M-C, Yoshida M, Villemin J-P, Auboeuf D, et al. Real-time imaging of cotranscriptional splicing reveals a kinetic model that reduces noise: implications for alternative splicing regulation. *The Journal of Cell Biology*. 2011; 193(5):819–29. [PubMed: 21624952]
53. Huranova M, Ivani I, Benda A, Poser I, Brody Y, Hof M, et al. The differential interaction of snRNPs with pre-mRNA reveals splicing kinetics in living cells. *J Cell Biol*. 2010; 191(1):75–86. [PubMed: 20921136]
54. Melvin WT, Milne HB, Slater AA, Allen HJ, Keir HM. Incorporation of 6-thioguanosine and 4-thiouridine into RNA. Application to isolation of newly synthesised RNA by affinity chromatography. *European journal of biochemistry / FEBS*. 1978; 92(2):373–9.
55. Dolken L, Ruzsics Z, Radle B, Friedel CC, Zimmer R, Mages J, et al. High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA*. 2008; 14(9):1959–72. [PubMed: 18658122]
56. Miller C, Schwalb B, Maier K, Schulz D, Dumcke S, Zacher B, et al. Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. *Molecular systems biology*. 2011; 7:458. [PubMed: 21206491]

57. Barrass JD, Reid JEA, Huang Y, Hector RD, Sanguinetti G, Beggs JD, et al. Transcriptome-wide RNA processing kinetics revealed using extremely short 4tU labeling. *Genome biology*. 2015; 16:282. [PubMed: 26679539]
58. Eser P, Wachutka L, Maier KC, Demel C, Boroni M, Iyer S, et al. Determinants of RNA metabolism in the *Schizosaccharomyces pombe* genome. *Molecular systems biology*. 2016; 12(2): 857. [PubMed: 26883383]
59. Carrillo Oesterreich F, Herzel L, Straube K, Hujer K, Howard J, Neugebauer KM. Splicing of Nascent RNA Coincides with Intron Exit from RNA Polymerase II. *Cell*. 2016; 165(2):372–81. [PubMed: 27020755]
60. Wetterberg I, Zhao J, Masich S, Wieslander L, Skoglund U. In situ transcription and splicing in the Balbiani ring 3 gene. *The EMBO Journal*. 2001; 20(10):2564–74. [PubMed: 11350946]
61. Yu Y, Reed R. FUS functions in coupling transcription to splicing by mediating an interaction between RNAP II and U1 snRNP. *Proceedings of the National Academy of Sciences*. 2015; 112(28):8608–13.
62. Ujvari A, Luse DS. Newly Initiated RNA encounters a factor involved in splicing immediately upon emerging from within RNA polymerase II. *The Journal of biological chemistry*. 2004; 279(48):49773–9. [PubMed: 15377657]
63. Harlen KM, Trotta KL, Smith EE, Mosaheb MM, Fuchs SM, Churchman LS. Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue. *Cell Rep*. 2016; 15(10):2147–58. [PubMed: 27239037]
64. Nojima T, Gomes T, Grosso AR, Kimura H, Dye MJ, Dhir S, et al. Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell*. 2015; 161(3):526–40. [PubMed: 25910207]
65. Schor IE, Rascovan N, Pelisch F, Allo M, Kornblihtt AR. Neuronal cell depolarization induces intragenic chromatin modifications affecting NCAM alternative splicing. *Proc Natl Acad Sci U S A*. 2009; 106(11):4325–30. [PubMed: 19251664]
66. Munoz MJ, Perez Santangelo MS, Paronetto MP, de la Mata M, Pelisch F, Boireau S, et al. DNA damage regulates alternative splicing through inhibition of RNA polymerase II elongation. *Cell*. 2009; 137(4):708–20. [PubMed: 19450518]
67. Fu XD, Ares M Jr. Context-dependent control of alternative splicing by RNA-binding proteins. *Nature reviews Genetics*. 2014; 15(10):689–701.
68. Moehle EA, Braberg H, Krogan NJ, Guthrie C. Adventures in time and space: splicing efficiency and RNA polymerase II elongation rate. *RNA biology*. 2014; 11(4):313–9. [PubMed: 24717535]
69. Berget SM. Exon recognition in vertebrate splicing. *The Journal of biological chemistry*. 1995; 270(6):2411–4. [PubMed: 7852296]
70. Mayer A, di Iulio J, Maleri S, Eser U, Vierstra J, Reynolds A, et al. Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. *Cell*. 2015; 161(3):541–54. [PubMed: 25910208]
71. Braberg H, Jin H, Moehle EA, Chan YA, Wang S, Shales M, et al. From structure to systems: high-resolution, quantitative genetic analysis of RNA polymerase II. *Cell*. 2013; 154(4):775–88. [PubMed: 23932120]
72. Fong N, Kim H, Zhou Y, Ji X, Qiu J, Saldi T, et al. Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. *Genes Dev*. 2014; 28(23):2663–76. [PubMed: 25452276]
73. Duffy EE, Rutenberg-Schoenberg M, Stark CD, Kitchen RR, Gerstein MB, Simon MD. Tracking Distinct RNA Populations Using Efficient and Reversible Covalent Chemistry. *Mol Cell*. 2015; 59(5):858–66. [PubMed: 26340425]
74. Davis-Turak JC, Allison K, Shokhirev MN, Ponomarenko P, Tsimring LS, Glass CK, et al. Considering the kinetics of mRNA synthesis in the analysis of the genome and epigenome reveals determinants of co-transcriptional splicing. *Nucleic Acids Res*. 2015; 43(2):699–707. [PubMed: 25541195]
75. Honkela A, Peltonen J, Topa H, Charapitsa I, Matarese F, Grote K, et al. Genome-wide modeling of transcription kinetics reveals patterns of RNA production delays. *Proc Natl Acad Sci U S A*. 2015; 112(42):13115–20. [PubMed: 26438844]

76. Aitken S, Alexander RD, Beggs JD. Modelling reveals kinetic advantages of cotranscriptional splicing. *PLoS Comput Biol*. 2011; 7(10):e1002215. [PubMed: 22022255]
77. Osheim YN, Miller OL, Beyer AL. RNP particles at splice junction sequences on *Drosophila* chorion transcripts. *Cell*. 1985; 43(1):143–51. [PubMed: 3935315]
78. Kwak H, Lis JT. Control of transcriptional elongation. *Annual review of genetics*. 2013; 47:483–508.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1. Observation of co-transcriptionally spliced transcripts leads to distance-based view of splicing kinetics *in vivo*

A) Electron micrograph of chromatin spreads from *Drosophila melanogaster* (left panel) visualize electron dense spliceosomes as they assemble near the 5' ends of nascent transcripts; shortening of transcript 10 is indicative of intron removal (77). Camera lucida drawing of the chromatin spread is shown in the right panel. B) Diagram of a simple gene with a single intron undergoing transcription by several active Pol II molecules. The 5' methyl cap (black ball) is added shortly after transcription of the 5' end of the RNA transcript. Spliceosomal components (red balls) bind the 5'SS and 3'SS of the pre-mRNA co-transcriptionally.

Figure 2. Comparison of distance-based and time-based measurements of splicing *in vivo*
 Many studies, indicated by flags on the gene ruler (upper panel) or time line (lower panel), have addressed when or where splicing occurs in multiple species and obtained the indicated results. This summary relates the measured half-maximum splicing values (or similar value if half-max not provided). Translation of distance to time is herein assumed at 3 kb/min (78). See text for details.

Figure 3. Time-based experiments vary widely in methodology and results

A) The median number of exon-intron junctions are processed in 14 min according to RNA-seq reads of metabolically labeled mouse RNA (44). B) Use of fluorescent reporter genes permits imaging of introns and quantitation of their half-lives (50). C). Fraction spliced values from a metabolic labeling experiment are plotted at different time points for three pairs of gene paralogs in yeast. Paralogs have identical exonic sequences and different intronic sequences. Splicing values appear highly similar between paralogs, yet different between genes (57).

Figure 4. Recent distance-based techniques predict spliceosome adjacent to polymerase during splicing

A) Representative SMIT trace from budding yeast shows splicing reaches half-maximum levels at ~62 bp past the 3' SS (59). B) Long read sequencing of budding yeast genes enables splicing analysis of single molecules (59). C) 3D reconstruction of nascent transcription and splicing complex (NTS) bound to Balbiani ring 3 locus on chromatin (dashed line) from *Chironomus tentans* with antibody gold particles against U2 snRNP (left) and Pol II CTD (right) (60). D) Crystal structures are docked into electron microscopy density show that capping enzymes bind RNA exit tunnel of RNA Polymerase II to modify the 5' end of RNA immediately (5).