The spatial distributions of pre-mRNAs suggest post-transcriptional splicing of specific introns within endogenous genes

Allison Coté[1], Chris Coté[1]*, Sareh Bayatpour[1]*, Heather L. Drexler[2], Katherine A. Alexander[3], Fei Chen[4], Asmamaw T. Wassie[5], Edward S. Boyden[5], Shelley Berger[3], L. Stirling Churchman[2], Arjun Raj[1#]

[1] Department of Bioengineering, University of Pennsylvania, Philadelphia, PA 19104
[2] Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, MA 02115
[3] Department of Cell and Developmental Biology, Penn Institute of Epigenetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104
[4] Broad Institute of MIT and Harvard, Cambridge, MA 02142
[5] Departments of Biological Engineering and Brain and Cognitive Sciences, Media Lab and McGovern Institute, Massachusetts Institute of Technology, Cambridge, MA 02139
* contributed equally # corresponding author

Abstract

Splicing is the molecular process by which introns are removed from pre-mRNA and exons are joined together to form the sequence of the mature mRNA. Measuring the timing of splicing relative to the transcription of nascent RNA has yielded conflicting interpretations. Biochemical fractionation suggests that RNA is spliced primarily during the process of transcription, but imaging of nascent RNA suggests that splicing happens after the process of transcription has been completed. We use single molecule RNA FISH together with expansion microscopy to measure the spatial distribution of nascent and partially spliced transcripts in mammalian cells, allowing us to infer the delay between when an intron is transcribed and when it is spliced out of a pre-mRNA. We show that 4 out of 4 genes we interrogated exhibit some post-transcriptional splicing, and that introns can be spliced in any order. We also show that completely synthesized RNA move slowly through a transcription site proximal zone while they undergo additional splicing and potentially other processing after transcription is completed. In addition, upon leaving this zone, some genes' transcripts localize to speckles during the process of splicing but some appear to traffic freely through the nucleus without localizing to any other nuclear compartment. Taken together, our observations suggest that the regulation of the timing and localization of splicing is specific to individual introns, as opposed to the previously surmised immediate excision of introns after transcription.

Introduction

The mRNA transcribed by most eukaryotic genes are spliced, a process in which the intronic RNA sequence is removed and the exonic RNA are joined together to form the ultimate mature mRNA sequence. A major question in the field is how tightly associated the process of transcription and splicing are, with some work suggesting that splicing occurs very shortly after the RNA polymerase transcribes a particular splice junction (Beyer et al., 1981; Carrillo Oesterreich et al., 2016), while other work suggests that many pre-mRNAs are fully transcribed before splicing occurs (Coulon et al., 2014; Drexler et al., 2019; Tsai et al., 1980). The relative spatial locations of nascent pre-mRNA, fully transcribed pre-mRNA, and mature mRNA species have the potential to directly reveal where—and consequently in what order—the processes of transcription and splicing occur. However, to date, the use of molecular imaging to systematically measure the locations of these partially-processed RNA intermediates has been limited in scope (interrogating single intron reporter genes or single introns within endogenous genes) (Coulon et al., 2014; Vargas et al., 2011; Waks et al., 2011). This limitation in scope is due largely to lack of probes designed to target specific splicing intermediates of endogenous genes.

In lieu of direct visualization, many have utilized biochemical fractionation to infer the location of various intermediates (Bhatt et al., 2012; Drexler et al., 2019; Mayer et al., 2015; Pandya-Jones and Black, 2009; Pandya-Jones et al., 2013; Tilgner et al., 2012; Wuarin and Schibler, 1994). Fractionation methods separate cellular RNA into different compartments, such as the putatively chromatin associated RNA, nucleoplasmic RNA, and cytoplasmic RNA, by centrifuging the cellular components in different lysis buffers and sedimentation gradients (Mayer and Churchman, 2017; Wuarin and Schibler, 1994). The implicit assumption made by such fractionation-based methods is that the RNA species in the “chromatin fraction” represent nascent pre-mRNA that are tethered to the gene body by the RNA polymerase II itself, and that once the pre-mRNA disengages with RNA polymerase II, it moves immediately and directly into the nucleoplasm. Under these assumptions, any splicing observed in the chromatin fraction would be assumed to be co-transcriptional. However, this assumption may not hold: it is possible that pre-mRNA remains in a chromatin associated compartment for some time after transcription completes, and thus splicing observed in the chromatin compartment may in fact still be post-transcriptional (Brody et al., 2011). Some groups have further purified nascent RNA via metabolic labeling or by using RNA Polymerase II antibodies, but these methods still have the potential to co-purify mature RNA (Nojima et al., 2015). Ultimately, such alternative explanations are difficult to eliminate without an independent and explicit verification of which RNA intermediates reside in particular compartments.

Advances in RNA imaging have enabled researchers to image RNA intermediates with single molecule resolution, both in fixed and living cells (Coulon et al., 2014; Levesque and Raj, 2013; Martin et al., 2013; Vargas et al., 2011; Waks et al., 2011; Zhang et al., 1994). Imaging using probes targeting both exonic and intronic regions of RNA has revealed bright nuclear foci that represent nascently transcribing RNA (Levesque and Raj, 2013; Vargas et al., 2011). The general lack of intronic signal away from these transcription sites has been taken as evidence

for co-transcriptional splicing, with notable cases of post-transcriptional splicing at speckles being observed in special cases (Vargas et al., 2011). However, owing to the diffraction limit for optical microscopy, it has been difficult to visualize RNA intermediates in the immediate vicinity of the gene undergoing transcription, thus making it difficult to observe whether RNA are still actively being transcribed during splicing or remain at the site of transcription for some time after transcription is complete. Since splicing would appear to occur in the “chromatin fraction” in both of these scenarios, it is possible that much of splicing actually occurs in this transcription proximal region after transcription is complete. Indeed, recent live-cell imaging methods showed that the splicing of a reporter gene is 85% post-transcriptional (Coulon et al., 2014), suggesting the latter, but as RNA from endogenous genes may be processed differently, the use of reporter genes leaves open the question of when endogenous genes undergo splicing relative to their transcription.

Here, we designed probes to comprehensively interrogate the spatial localization of several RNA intermediates using a combination of RNA FISH and expansion microscopy (Chen et al., 2015, 2016). We revealed the ordering of transcriptional and splicing processes with single molecule resolution. We found that the proportion of splicing that occurs post-transcriptionally varies from intron to intron within a single gene, but that all endogenous genes we tested displayed some degree of post-transcriptional splicing. We also employed expansion microscopy to demonstrate that newly synthesized RNA dwell and undergo continuous splicing near the site of transcription after transcription is complete. These RNA are untethered to the site of transcription and eventually diffuse into either the nucleoplasm or near nuclear speckles.

Results

At least one intron of each observed endogenous gene is spliced post-transcriptionally

To directly visualize the locations of spliced and unspliced RNA relative to the site of transcription, we used single molecule RNA FISH (Femino et al., 1998; Raj et al., 2008) to simultaneously fluorescently label the exons and several individual introns for a number of genes of interest: *CPS1*, *EEF2*, *TM4SF1*, and *FKBP5*. *CPS1* was chosen because it is a long and highly expressed gene, *EEF2* was chosen because it is a highly expressed housekeeping gene, *TM4SF1* was chosen because it is highly expressed and sequencing data suggested it was highly post-transcriptionally spliced (see methods for details), and *FKBP5* was chosen because it is inducible in A549 cells by application of dexamethasone.

By distinguishing the separate fluorescent signals from probes bound to exons and introns, we could visualize splicing intermediates (represented by colocalized intron and exon spots) relative to the site of transcription (represented by bright colocalized intron and exon spots, see Figure 1B) and fully spliced products (represented by exon spots alone). We were particularly interested in visualizing the location of splicing intermediates in order to distinguish between two possibilities: 1) pre-mRNA being spliced *at* the site of transcription and 2) pre-mRNA being spliced *away* from the site of transcription (Figure 1A; right; i and ii versus iii). Observing spliced RNA at the site of transcription could represent RNA that is spliced either co-transcriptionally (tracking along with or shortly behind the polymerase; Figure 1Ai) or proximally

post-transcriptionally (near the site of transcription but after the process of transcription has been completed) (Figure 1Aii). However, observing pre-mRNA with unspliced introns that were sufficiently far away from the site of transcription such that they could no longer be attached to the location of the gene itself would suggest that that intron would have to be spliced post-transcriptionally (distal post-transcriptional splicing, Figure 1Aiii).

To determine what fraction of splicing occurred far from the site of transcription, we first needed to classify each colocalized exon and intron spot as either a transcription site or a dispersed pre-mRNA. We computationally identified spots for both introns and exons of a particular gene, then each intron spot that was within 0.65μm of an exon spot was designated a colocalized exon and intron spot (0.65μm is the 75th percentile of mRNA lengths determined by measuring the distance from 5' to 3' signal; Figure 1B, see Supplementary Figure 1C for mRNA lengths), which we assumed represents at least one nascent pre-mRNA (potentially several nascent pre-mRNA when found at the site of transcription). Intron spots that did not colocalize with an exon spot were presumed to be degradation products or non-specific background and were discarded (these were generally <25% of intron spots; Table 1). We used an intensity threshold to categorize each colocalized exon and intron spot as either a transcription site or a dispersed pre-mRNA (Figure 1B, see Supplementary Figure 1A,B for the scheme for the classification of transcription sites).We also tried several other methods for choosing transcription sites, all of which showed qualitatively similar results (Supplementary Figure 1A). We then calculated the distance of each dispersed pre-mRNA from the nearest transcription site for all cells, yielding population-wide measurements of dispersal per intron (Figure 1D). We found that for the genes *CPS1*, *EEF2*, *TM4SF1*, and *FKBP5*, at least one intron displayed some pre-mRNA away from the site of transcription (3-5 introns tested per gene, Figure 1D, Figure 5C). The presence of these intermediates away from the site of transcription showed that for all the genes we tested, some introns were spliced post-transcriptionally and away from the gene body itself. We never observed introns in cytoplasmic mRNA, suggesting that all of these introns are eventually spliced out; i.e., the introns are not retained indefinitely. We assumed that this dispersion represents a difference in the relative splicing rates of different introns, reasoning that introns that are spliced more slowly relative to the speed of transcription (which might be variable from intron to intron as well) would be further dispersed because these introns would remain part of the pre-mRNA as the pre-mRNA travels away from the site of transcription. In summary, this dispersion suggests that at least one intron of each observed gene is spliced partly post-transcriptionally.

Different introns within the same gene are spliced largely independently of each other

The observation of individual pre-mRNA species away from the site of transcription with some but not all introns retained (Supplementary Figure 1E for number of introns retained in *FKBP5* as a function of distance from the transcription site) suggested that introns can be spliced independently of one another. We thus wanted to measure the extent to which introns were spliced independently by quantifying the number of intermediates that contained different introns in the same molecule. For each RNA containing a pair of introns A and B, we could place it in one of the following categories:

1) Both introns retained in the RNA (fully unspliced)
2) Intron A retained in the RNA, intron B spliced (partially spliced)
3) Intron B retained in the RNA, intron A spliced (partially spliced)
4) Both introns spliced (fully spliced)

Evidence for dependency would take the form of, say, intron B only being spliced if intron A is already spliced, whereas independence would be signified by intron B splicing regardless of the splicing status of intron A. In terms of the splicing intermediates listed above, splicing would be seen to be dependent if we always saw either both introns retained in the RNA or fully spliced (1, fully unspliced and 4, fully spliced), whereas independence would be signified by a mixture of partially spliced intermediates (2 and 3). One can capture dependencies by calculating the mutual information (measured in nats, a unit of mutual information based on natural logarithms) from this categorized frequency data. A mutual information of 0 would suggest independence of splicing, whereas a mutual information greater than 0 would suggest some degree of dependence.

To provide a realistic practical upper bound on mutual information for our measurements, we estimated, based on an 80% detection efficiency, the mutual information of two probes binding to the exact same target to be 0.137 nats (see methods for calculation).

We then analyzed probe pairs targeting different introns from the same gene for 4 different genes. We measured mutual information values ranging from 0.0003 to 0.1280 nats (Table 2). Many of these values were markedly lower than our rough upper bound of 0.137 nats calculated based on technical detection variability, suggesting that the splicing status of one intron often yields very little information about the splicing status of other introns from the same pre-mRNA. We classified introns as “effectively dependent” or “effectively independent” based on mutual information (using the Fisher 2x2 exact test to determine statistical significance). Intron pairs with high mutual information (> 0.025 nats) were classified as “effectively dependent”, while intron pairs with low mutual information (< 0.025 nats) were classified as “effectively independent”. The majority of introns pairs (11 out of 20 tested) were classified as “effectively independent” due to low mutual information. 0.025 nats appears to be the noise floor for mutual information in this system, given that many pairs of introns had mutual information values below this threshold with high P values for the Fisher 2x2 exact test, suggesting that our experimental system was unable to resolve more finely mutual information below these values.

Pairs of introns that had very little dispersal (<5% of intensity away from the transcription site) were classified as “ambiguous” (5 out of 20) because it is impossible to distinguish whether they are spliced together or separately at the transcription site due the inability of RNA FISH alone to distinguish single RNA molecules at the site of transcription.

The overall low mutual information and the classification of both neighboring introns (such as FKBP5 introns 8 and 9, or TM4SF1 introns 3 and 4) and introns that are genomically distant from one another (such as FKBP5 introns 1 and 8) (Table 2) as “independent” from one another suggests that genomic proximity of introns does not influence whether they are spliced independently or not.

We also asked whether the two ends of a single intron, separately labeled, would show dependence or independence in their removal. We labeled the 5’ and 3’ end of intron 1 from *FKBP5*, and measured a mutual information of 0.0551 nats (Table 2). This mutual information

was similar to our theoretical upper bound of 0.137, suggesting that these are indeed informative about each other intron's splicing status.

In summary, our dispersal data from multiple introns suggests that each intron has a particular splicing rate which governs the time it takes until that intron is spliced, apparently independently of neighboring introns.

Transcripts are untethered to the site of transcription and move slowly away from the site of transcription through a slow-moving transcription site proximal zone after transcription is completed

While conventional single molecule RNA FISH allowed us to determine what portion of splicing is happening post-transcriptionally and far from the site of transcription (Table 1, Supplementary Figure 1D), the resolution limits of conventional light microscopy made it impossible to distinguish whether transcripts are being spliced during the process of transcription or after the completion of transcription but before the pre-mRNA moves away from the site of transcription. Conventional light microscopy cannot easily distinguish these possibilities because all the RNA at or near the site of transcription are sufficiently close together that they are typically only visible as one large transcription site spot (see Figure 1B for example).

To distinguish splicing intermediates at the site of transcription, we used expansion microscopy to physically expand the transcription site (by around 3.3x, Supplementary Figure 2A) followed by staining by RNA FISH (Figure 2A) and imaging, thereby separating the single bright transcription site blob into visually distinct individual RNA intermediates at the site of transcription (Figure 2B) (Chen et al., 2015, 2016). We labeled the 5' and 3' regions, as well as one interior intron, of the gene *CPS1*, for which the unspliced transcript would be quite long (~87kb) but the spliced transcript is comparatively short (~5kb) (Figure 2A). This labeling scheme allowed us to measure the locations of 5' exons, a middle intron, and 3' exons of pre-mRNA or processed mRNA in the vicinity of the expanded transcription site.

Upon expanding, labeling and imaging, we observed that both the 5' and 3' probe signals formed small clouds where we had previously observed the tight transcriptional focus (Figure 2D). The presence of these clouds precluded any model of immediate, free movement of the transcript away from the RNA polymerase upon completion of transcription (Figure 2C). For instance, if transcripts freely diffused or were rapidly transported away after transcription, then we would have observed a gradient of decreasing pre-mRNA concentration with increasing distance away from the transcription site, which is incompatible with the sharp transition from the high concentration in the cloud region to the low concentration outside of the cloud (Figure 2C,D,E). Rather, the presence of clouds suggests that there is a distinct "proximal zone" surrounding the transcription site through which pre-mRNA move more slowly than in the more distant nucleoplasm. Furthermore, the presence of clouds for both the 5' and 3' ends of the pre-mRNAs suggest that no part of the pre-mRNA is tethered to the transcription site (e.g., 3' ends being tethered by remaining bound to RNA polymerase II or some other RNA processing factor), because otherwise we would expect the cloud around the tethered portion of the pre-mRNA to be considerably smaller than the other. Also, neither the 5' nor 3' signals

overlapped with actively elongating RNA Polymerase II immunofluorescence signals (see methods), further suggesting that pre-mRNA in the slow moving proximal zone were not being actively transcribed (Figure 2F). Based on the size of the clouds, we estimate the size of this zone to be around 0.3 microns in diameter around the transcription site (Figure 2G).

The existence of a slow-moving proximal zone surrounding the site of transcription suggested the possibility that splicing could indeed happen post-transcriptionally while pre-mRNA were moving through this zone. (Splicing that occurred while in this zone would appear to be co-transcriptional by conventional single molecule RNA FISH.) To test for the occurrence of splicing while pre-mRNA were in this zone, we simultaneously labeled an intron while labeling the 5' and 3' ends of the pre-mRNA. Similar to the signals from the 5' and 3' ends, the intronic signal also formed a cloud, showing that splicing has not yet been completed as the pre-mRNA move through this proximal zone (Figure 2H,I). Furthermore, the clouds of the 5' and 3' ends of the pre-mRNA are typically non-overlapping and are further apart than the mature mRNA we found in the cytoplasm, suggesting that these clouds do not represent fully mature, spliced mRNA (Figure 2J). Interestingly, the relative positions of the 5', 3', and middle intron clouds adopted a wide variety of conformations, suggesting that the linear genomic order of the pre-mRNA may not be strictly maintained during transit through the proximal zone.

Thus, in sum, expansion microscopy revealed that after the completion of transcription, pre-mRNA move slowly through a slow-moving proximal zone, during which splicing may be ongoing, and that without resolving pre-mRNA within this zone, it can be difficult to distinguish co-transcriptional splicing from post-transcriptional splicing.

Sequencing is consistent with RNA FISH dispersal as a metric to quantify distal post-transcriptional splicing

To corroborate that our RNA FISH dispersal metric can quantify the portion of splicing happening post-transcriptionally but away from the site of transcription (>1 mRNA length away; see methods and Supplementary Figure 1C), we turned to sequencing of nascent RNA to determine the splicing status of RNA that has recently been transcribed and compared that to our dispersal metrics. We reasoned that we could use the splicing status of nascent RNA as a proxy for the splicing status of pre-mRNA at the site of transcription, and then, by subtracting that from the splicing status of total isolated RNA, we could calculate the splicing status of pre-mRNA away from the site of transcription. (In principle, some amount of splicing that is post-transcriptional but still occurring near the site of transcription might not be captured by this metric.) We isolated nascent RNA by a combination of metabolic labeling and cellular fractionation to capture RNAs that are both newly synthesized and co-sediment with the chromatin fractions of cells (Drexler et al., 2019). We then sequenced this RNA and calculated a "splicing index" (see (Drexler et al., 2019)) for each intron, defined as the number of spliced reads for a given intron-exon junction divided by the number of unspliced reads for that same junction (Figure 3A). This splicing index represents how "co-transcriptionally spliced" a particular intron is: introns spliced during active transcription are likely to have more exon-exon junction reads in metabolically labelled and chromatin fractionated RNA, thus yielding a higher splicing index. Under these assumptions, a high splicing index suggests that those introns are spliced

faster than other introns, and therefore are likely to be spliced “more co-transcriptionally” than other introns. Conversely, introns with a low splicing index are thought to be spliced more slowly, and therefore may be post-transcriptionally spliced.

Measuring the splicing index for different introns within the same endogenous genes (Supplementary Figure 3A) revealed a variety of splicing indices, similar to the varying dispersal seen with RNA FISH. Directly comparing splicing index and mean dispersal of these introns showed an inverse relationship: as expected, introns with a high splicing index (more “co-transcriptionally” spliced) have low dispersal, and introns with a low splicing index (more “post-transcriptionally” spliced) have high dispersal (Figure 3B). However, there were also a number of introns for which both the splicing index and mean dispersal were low, potentially due to technical noise in either the sequencing or dispersal measurements, or from a high degree of “proximal post-transcriptional” splicing for those introns, which would result in low dispersal but also a low splicing index.

To look for evidence of “proximal post-transcriptional” splicing of this category of introns, we measured transcription site size, reasoning that large transcription sites may represent high numbers of pre-mRNA in the transcription proximal zone which we identified with expansion microscopy. We analyzed transcription site size (defined as the size of the mask of a binarized image of the transcription site; see code for calculation) for each intron for which we had RNA FISH data (Supplementary Figure 3B). Similar to mean dispersal, RNA FISH-based analysis also revealed an overall inverse relationship between transcription site size and splicing index; however, several introns still had both low splicing index (more “post-transcriptionally” spliced) and low transcription site size (Supplementary Figure 3B). Small transcription site size suggested that there were not a lot of pre-mRNA in the transcription proximal zone, suggesting high levels of “proximal post-transcriptional splicing”, as we might have anticipated for introns with both a low splicing index and a low dispersal. These introns may be spliced post-transcriptionally but without any drift of pre-mRNA away from the site of transcription. The anticorrelation of splicing index with both mean dispersal and transcription site size shows that sequencing of nascent RNA corroborates these two methods to assess the amount of post-transcriptional splicing occurring, but while there is an association, splicing index alone was insufficient to predict the dispersal pattern of an intron, pointing to the utility of molecular imaging as a complementary approach for measuring the properties of splicing at specific introns.

Unspliced pre-mRNA of some genes localize to a speckle-proximal compartment after exiting the slow moving zone proximal to the transcription site

We wondered where transcripts went after they were released from the slow-moving transcription-site-proximal zone that was revealed by expansion microscopy (see Figure 2C). We hypothesized that the transcripts could do one of three things: 1. freely diffuse away from the transcription-site-proximal zone through the nucleoplasm (nuclear dispersal), 2. be tethered to the transcription site proximal zone in some manner (tethering), 3. fill a compartment, potentially around or adjacent to the transcription-site-proximal zone or other nuclear bodies (compartmentalization) (Figure 4A). Owing to the fast degradation rate of introns, relatively few

unspliced RNA make it out of the transcription-proximal zone, which made it difficult to discriminate between these hypotheses. Thus, we inhibited splicing to generate more pre-mRNAs, making it easier to track their localization after leaving the transcription-site-proximal zone.

Upon splicing inhibition, we observed three distinct trafficking behaviors for pre-mRNA species: one in which there are increased numbers of dispersed pre-mRNA throughout the nucleus (nuclear dispersal; consistent with scenario 1), one in which the pre-mRNA are located in a large blob, likely around the transcription-site-proximal zone or another nuclear body (blobs; scenario 2 or 3), and one in which the pre-mRNA dispersal pattern looked identical with or without splicing inhibition (non-splicing inhibited) (Figure 4B,C). The compartmentalization pattern only appeared for three out of the 16 genes we tested *EEF2*, *GAPDH*, and *RPL13A*; Figure 4F,H), whereas we observed a nuclear dispersal phenotype for 7 genes and no change in dispersal for 6 genes.

For the genes whose pre-mRNA showed the blob behavior, one possibility was that some specific part of these pre-mRNA might be tethered to some location within the nucleus, whether that be the genomic locus, the transcription-site-proximal region, or from some other point within the blob, while the remaining portion of the pre-mRNA is able to freely diffuse around the tethered part. To test for tethering, we labeled the 5' and 3' ends of the pre-mRNA in different colors, reasoning that if either end of the pre-mRNA was tethered to a particular location, then the signal from that particular end would form a tighter spot in the nucleus while the other end would fill the compartment (Figure 4D, Supplementary Figure 4A). We found, however, that the 5' and 3' ends of the pre-mRNA both filled the entire blob, showing that neither end was specifically tethered to a particular point within the blob, suggesting instead that the pre-mRNA spread to fill the entire putative compartment (as in scenario 3 described above, compartmentalization) (Figure 4A,D).

We wondered if these compartmentalized pre-mRNA were located near any other nuclear structures. One candidate was nuclear speckles, which are compartments in the nucleus that contain concentrated splicing and transcription factors (Zhang et al., 1994). To determine whether the pre-mRNA compartments we observed colocalize with nuclear speckles, we performed RNA FISH simultaneously with immunofluorescence for SC35, a component of speckles, and saw that these compartmentalized pre-mRNA did indeed appear near nuclear speckles both before and after splicing inhibition (Figure 4G, Supplementary Figure 3C). We corroborated the spatial association we observed between speckles and compartmentalized pre-mRNA by analyzing previously published high throughput sequencing data that quantified the distance of transcripts from speckles and other nuclear compartments (Chen et al., 2018). From those data, we found that compartmentalized genes were all within the most speckle associated transcripts (top 10% of the data), while all other tested genes (both nuclear dispersal and non-responsive) exhibited a much broader range of distances to speckles (Figure 4H).

Speckles form a set of subcompartments within the nucleus. We thus wondered whether pre-mRNA from the genes exhibiting "compartmentalization" in their post-transcriptional trafficking would go to all of these speckle compartments, or rather just a subset. Moreover, if only a subset, do the pre-mRNA from different genes go to different subsets of speckles? Can pre-mRNA from the same gene but with different introns retained fill different subsets of

speckles? To test these possibilities, we performed RNA FISH on multiple introns within the same “compartmentalization” gene (EEF2) as well as introns from from several different “compartmentalization” genes simultaneously (EEF2, GAPDH, and RPL13A). We observed that multiple introns retained in pre-mRNA from the same gene colocalized to the same subset of speckles (Figure 4E), suggesting that all unspliced pre-mRNA from a particular gene localize to the same subset of speckles. We also observed that pre-mRNA from multiple “compartmentalization” genes (EEF2, GAPDH, and RPL13A) localize to a similar set of speckles after splicing inhibition (Figure 4F), although there are some differences. However, when observing the intron distributions of pre-mRNA from *EEF2*, *GAPDH*, and *RPL13A before* splicing inhibition, they do not appear to colocalize with one another or with the same speckles (Supplementary Figure 3C), suggesting perhaps that the speckles they colocalize with after splicing inhibition may be somehow targeted by these genes after splicing inhibition.

Our results show that pre-mRNA from specific genes remain in speckle-proximal compartments for some time before the process of splicing has been completed.

Dispersal is not an inherent trait of individual introns and can vary with transcription level

We wondered whether the degree of dispersal was an inherent property of each intron or whether the degree of dispersal could vary due to other factors such as the level of transcription. To test whether the level of transcription affected the degree of dispersal, we treated A549 cells with dexamethasone to induce transcription of the gene *FKBP5*, and then performed RNA FISH against introns 1, 8, and 9 at various time points in dexamethasone to measure the degree of dispersion (Figure 5A, B). We saw an increase in both exon and intron spot counts (Figure 5B) over time and a corresponding increase in the dispersal of some, but not all, introns (Figure 5C). The fact that intron dispersal increased with transcription level for at least some introns shows that dispersal is not an inherent property of each intron but can depend on other variables like the level of transcription (Figure 5C). Intron 9 did not exhibit an increase in dispersal even with long exposure to dexamethasone (8 hours) (Figure 5C).

We believe that the increased dispersal represented an escape of unprocessed pre-mRNA from the site of transcription. This increase would only happen in cases where splicing was slow enough that splicing of all introns was not completed before termination of transcription, therefore allowing time for these pre-mRNA to disperse away from the site of transcription. The lack of dispersal of some introns, even with increased transcription, suggested that these introns are spliced so quickly that pre-mRNA containing those introns have no time to disperse away from the site of transcription even in case of increased transcription.

We also wondered if local depletion of splicing factors could explain the increased dispersal exhibited with increased transcription level. A local depletion of splicing factors could occur as more pre-mRNA fill the local area and absorb the local pool of splicing factors (or particular splicing cofactors), resulting in more pre-mRNA escaping from the transcription proximal region before undergoing splicing. A signal for local depletion would be a change in the percentage of pre-mRNA detected far from the transcription site. For a given constant concentration of splicing factors (i.e., no local depletion), we would expect that the rate of splicing, defined as probability of a splicing event per unit time, would remain constant, and thus the percentage of pre-mRNA

remaining unspliced far from the transcription site should remain unchanged irrespective of overall transcription rate. However, a local depletion of splicing factors would manifest as a change in the percentage of pre-mRNA that are detected far from the transcription site because the local splicing rate would be slowed, allowing a higher percentage of pre-mRNA to leave the transcription site region.

We observed that two introns showed significant increases in the percent of intron signal that was far from the site of transcription over time in dexamethasone (from 0 to 22% for intron 1 and from 0 to 29% for intron 8; see Table 3 for all percentages), consistent with a local depletion of splicing factors that allowed both of these introns to escape their co-transcriptional regulation at low levels of transcription. These increases in the percent of distal post-transcriptional splicing show that even the same intron can exhibit a range of spatiotemporal patterns of splicing depending on expression levels and potentially on limiting splicing factors in the vicinity of the gene itself.

Discussion

Our results suggest the following model of splicing: splicing occurs continuously and independently between different introns while transcription is occurring and also while the pre-mRNA moves slowly through a region proximal to the transcription site, after which it moves either to the speckle compartment or freely in the nucleoplasm. This model stands in contrast to prevailing models of splicing, which suggest that splicing happens immediately (within 15-20 seconds or 45-100 nucleotides of transcription) after transcription of each intron is completed (Alpert et al., 2016; Carrillo Oesterreich et al., 2016; Eser et al., 2016; Huranová et al., 2010; Martin et al., 2013; Wallace and Beggs, 2017). This difference may arise from a variety of reasons, including but not limited to species-specific differences and the use of different assays to measure the timing of splicing (Alpert et al., 2016). Our model *is* in agreement with fractionation work by Pandaya-Jones *et al*. that suggests that specific transcripts are retained in the chromatin and their exons are not ligated until after transcription is complete (Pandya-Jones et al., 2013). It is also in agreement with new fractionation work by Drexler *et al.*, in which they find that indeed many genes are not yet fully spliced, even when transcription is completed (Drexler et al., 2019).

One of the original models for splicing is the “first come, first serve” model in which each intron is immediately spliced upon the completion of transcription, in a 5’ to 3’ order (Aebi and Weissman, 1987). Our results suggest that first come, first serve is not the case, based on seeing low splicing rates (or high dispersion) for even the 5’ most introns of some genes. This lack of first come, first serve splicing is confirmed by others in several different situations (Kessler et al., 1993; de la Mata et al., 2010; Yang et al., 2012). Our results also suggest that pre-mRNA are not tethered to the site of transcription while they move through a transcription-site-proximal compartment. This lack of tethering is in contrast to what is suggested by Dye et al., where the authors suggest that exons are tethered to polymerase II as splicing is occuring (Dye et al., 2006). The lack of colocalization of introns with polymerase II also suggests that splicing is not happening close to polymerase II. This conclusion stands in contrast to the data of Alexander et al., which suggested that almost all splicing occurs while

polymerase II is still paused proximal to the intron that was recently transcribed (Alexander et al. 2010). This discrepancy may be due to species-specific differences. In yeast, polymerase II has been found to pause at several sites involved in splicing, including the terminal exon, the 3' SS, and internal exons (for a review, see (Carrillo Oesterreich et al., 2011)). It is also possible that increased resolution made possible by expansion microscopy has allowed us to detect previously indistinguishable distances between introns and polymerase II.

Our expansion microscopy results suggest that the distance between the 5' end and the 3' end of RNA at the site of transcription is greater than the distance between the 5' end and 3' end of mRNA. This increased 5'-3' distance suggests that the transcripts at the site of transcription are either unpackaged, perhaps due to decreased RNA binding proteins occupancy, or are simply longer because there are likely more introns incorporated into the transcripts close to the site of transcription than away from it. If this 5' to 3' distance increase is due to increased incorporation of introns, this incorporation of introns even after transcription is completed would also challenge the idea that splicing is happening immediately post-transcription.

We see that the introns of most endogenously expressed genes are spliced in a variety of spatial distributions, with all the genes we probed demonstrating at least some pre-mRNA containing one intron with some splicing occurring distally from the transcription site. These dispersal patterns suggest that each intron has a distinct rate of splicing out of the pre-mRNA, and introns with dispersal are at least partially spliced post-transcriptionally. Taken together with our other data, this strongly opposes the "first come, first serve" model of splicing.

Our results also suggest that introns are spliced independently of one another, whether they are genomically proximal or distal to each other. This conclusion is in contrast to other work which suggests splicing of particular introns is controlled or gated by the splicing of other introns or exons within the same gene (Drexler et al., 2019; Kim et al., 2017). These discrepancies could be due to cell-type specific differences, or potentially specific regulation of the introns or genes looked at in this paper (perhaps due to the nature of their high expression). Also, the degree to which splicing of neighboring introns serves as a gate is modest, and thus perhaps not within our power to detect given the limited number of genes we interrogated.

Our splicing inhibition results show that the trafficking of transcripts after they escape the slow moving transcription proximal zone varies by gene. Our results were consistent with sequencing-based metrics, and those metrics are in turn largely consistent across multiple cell types, suggesting that the speckle-associative property is not subject to cell-type specific regulation (Chen et al., 2018). We also observed that pre-mRNA of compartmentalization genes seemed to localize exclusively to this speckle-proximal compartment, perhaps as a way to sequester the unspliced RNA. It is possible that these RNA, because of their abundance, sequence factors, or associated RNA binding proteins, are more likely to integrate into compartments, such as nuclear speckles (Chen et al., 2018).

We also see that increased transcription level correlates with intron dispersal, which suggests that the percentage of splicing that happens away from the transcription site is regulated by transcription level for at least some introns. This increased dispersal with transcription level may also explain why we observe post-transcriptional splicing of all genes we measured because all of them were highly expressed. Work by Ding and Elowitz suggests that,

for high expressing genes, splicing can act as an “economy of scale” filter, in which the expression of already highly expressed genes is amplified by increased splicing efficiency, potentially due to association with nuclear speckles (Ding and Elowitz, 2019). Our data rather suggests that as transcription level increases the splicing efficiency at the site of transcription either stays the same or decreases, resulting in an increased percentage of dispersed introns. This discrepancy could reflect cell type differences or specific regulation of the two inducible genes looked at compared to our constitutively expressed genes (Ding and Elowitz, 2019).

A major outstanding question is what sequence-based features determine the distinct behaviors of different introns. Large-scale imaging and potentially synthetic libraries of introns may be required to make such conclusions.

Methods

Cell culture, splicing inhibition, and FKBP5 induction

HeLa (kind gift of the lab of Dr. Phillip Sharp, MIT) and A549 (human lung carcinoma, A549, ATCC CCL-185) cells were cultured in DMEM (Gibco) supplemented with 50 U/mL penicillin, 50 µg/mL streptomycin, and 10% fetal bovine serum (FBS, Fisher). Splicing inhibition was accomplished by treating HeLa cells with 1uM Pladienolide B (Tocris Biosciences, 6070500U) for 4 hours, as described by (Nojima et al., 2015). HeLa cells were then fixed and used for RNA FISH as described below. FKBP5 was induced by treating A549 cells with 25nM dexamethasone (Sigma, D2915) for the specified lengths of time. A549 cells were then fixed and used for RNA FISH as described below.

RNA fluorescence in situ hybridization and expansion microscopy

Single-molecule RNA FISH was performed on samples as described previously(Raj et al., 2006). Cells were fixed in 4% formaldehyde and permeabilized with 70% ethanol before in situ hybridization was performed using the probes described in supplementary table 1. Samples were simultaneously co-stained with probes for the exon of gene of interest (labelled in cy3), two introns of the gene of interest (labelled in alexa594 or atto647N), and cyclin mRNA (labelled in either atto700 or atto647N) (Stellaris oligonucleotides, Biosearch Technologies). Samples were then washed twice with 2 X saline sodium citrate buffer (SSC) containing 10% formamide (Ambion), and then 2XSSC supplemented with DAPI (Molecular Probes D3571) to stain the cell nuclei. Cells were submerged in 2XSSC with DAPI for imaging. For combined expansion microscopy and RNA FISH, expansion microscopy was performed as described by Chen *et al.* (Chen et al., 2016). Cells were fixed and permeabilized as described above, before a poly-acrylamide gel was polymerized on top of the sample for 1 hour at 37°C. Cells were digested with proteinase K overnight at room temperature, then RNA FISH was performed as described above. Samples were then expanded in nuclease free water for 2 hours at room temperature, then expanded for an additional 2 hours at room temperature in nuclease free water with DAPI. Samples were submerged in nuclease free water with DAPI for imaging.

Imaging

Cells were imaged using a Leica DMI600B automated widefield fluorescence microscope equipped with a X100 Plan Apo objective, a Pixis 1024BR cooled CCD (charge-coupled device) camera, a Prior Lumen 220 light source, and filter sets specific for each fluorophore. Images in each fluorescence channel were taken as a series of optical z-sections (0.3 microns per section).

Immunofluorescence

Staining for SC35 and polymerase II were performed with antibodies against SC35 (abcam ab11826, 1:200) and phospho S2 polymerase II (Active Motif, 61083, 1:200), respectively. Briefly, staining was performed on cells fixed and permeabilized as described above for RNA FISH. Primary antibody hybridization was carried out in 1XPBS overnight at 4°C. Samples were

then washed with 1XPBS and incubated with secondary antibody for 1 hour in 1XPBS at room temperature. Samples were then fixed for an additional 10 minutes in formaldehyde, washed with 1XPBS, and RNA FISH was performed as described above.

RNA FISH quantification

RNA FISH was quantified as described previously (Raj et al., 2006). Briefly, cells were manually segmented, a gaussian filter was applied to all spots, signal was distinguished from noise through semi-automated thresholding, each called spot was further fit to a gaussian to get sub pixel resolution, and transcription sites were chosen based on a global brightness threshold (Supplementary Figure 1A,B). Data was processed to assess distances and graphed in R. Calculation of mRNA length was based on the 75th percentile of 5'-3' distances of labelled mRNA seen in Supplementary Figure 1C.

4sU labeled chromatin-associated RNA sequencing and splicing index analysis.

HeLa S3 cells (ATCC, CCL-2.2) were maintained in DMEM media containing 10% FBS, 100 U/ml penicillin, and 100 µg/ml streptomycin to 75% confluency. Cells were labeled in media containing 500 µM 4-thiouridine (4sU, Sigma, T4509) for 7.5 minutes. Plates were washed twice with 1X PBS and cells were lifted by scraping. Labeled cells were collected by centrifugation at 500 g for 2 minutes. To purify chromatin associated RNA, steps 8-21 were followed exactly as described in (Mayer and Churchman, 2016). In brief, nuclei were collected by lysing samples of 10M cells in 200 µl cytoplasmic lysis buffer (0.15% (vol/vol) NP-40 (Thermo Fisher Scientific, 28324), 10 mM Tris-HCl (pH 7.0), and 150 mM NaCl) for 2 min, layering over a 500 µl sucrose cushion (10 mM Tris-HCl (pH 7.0), 150 mM NaCl, 25% (wt/vol) sucrose), and centrifuging at 16,000 g for 10 minutes. The nuclei pellet was washed in 800 µl wash buffer (0.1% (vol/vol) Triton X-100, 1 mM EDTA, in 1X PBS) and collected by centrifuging at 1,150 g for 1 minute. Nuclei were resuspended in 200 µl glycerol buffer (20 mM Tris-HCl (pH 8.0), 75 mM NaCl, 0.5 mM EDTA, 50% (vol/vol) glycerol, 0.85 mM DTT), and mixed with 200 µl nuclei lysis buffer (1% (vol/vol) NP-40, 20 mM HEPES (pH 7.5), 300 mM NaCl, 1 M urea, 0.2 mM EDTA, 1 mM DTT) by pulse vortex and incubated on ice for 2 minutes. The chromatin pellet was collected by centrifugation at 18,500 g for 2 minutes and resuspended in 1X PBS. All steps were performed at 4°C and buffers were prepared with 25 µM α-amanitin (Sigma, A2263), 0.05U/µl SUPERase.In (ThermoFisher Scientific, AM2694), and protease inhibitor mix (Roche, 11873580001). Chromatin-associated RNA was extracted using Qiazol lysis reagent (Qiagen, 79306) following the manufacturer's instructions.

50 µg RNA per reaction was subjected to 4sU purification as described in (Dölken et al., 2008; Schwalb et al., 2016). In brief, labeled RNA (1 µg / 10 µl) was incubated with 10% biotinylation buffer (100mM Tris pH 7.5, 10mM EDTA) and 20% EZ-Link Biotin-HPDP (1 mg/mL resuspended in DMF, Thermo Fisher Scientific, 21341) for 1.5 hours 24°C in the dark and 800 rpm to mix. RNA was purified by shaking the sample with a 1:1 volume of chloroform/isoamylacohol (24:1), separating using a phase-lock tube at 16,000 g for 5 min, and performing isopropanol precipitation. Biotinylated RNA was separated using the µMACS streptavidin kit (Miltenyi Biotec, 130-074-101) by mixing with µMACS streptavidin beads at a 2:1 ratio by volume at 800 rpm and 24°C for 15 min. RNA-streptavidin beads mix was transferred to

the μMACS column and washed with wash buffer (100 mM Tris pH 7.5, 10 mM EDTA, 1 M NaCl, 0.1% Tween 20) at 65°C and room temperature 3 times each. Selected RNA was eluted off the magnet using the reducing agent, DTT (0.1M), and purified using the miRNeasy micro kit (Qiagen, 217084) with on-column DNase I treatment (Qiagen, 79254). For the poly(A) depleted sample, the RNA was first concentrated using the RNA Clean and Concentrator kit (ZymoResearch, R1015). 10 μL Oligo(dT) Dynabeads (ThermoFisher, 61002) were washed in 10 μL Binding Buffer (20mM Tris.HCl pH 7.5, 1M LiCl, and 2mM EDTA). The sample was mixed with 10 μl binding buffer, heated to 65°C for 2 min, moved to ice, and mixed with 1 μl SUPERase.In (ThermoFisher Scientific, AM2694). The sample and beads were mixed thoroughly and annealed by rotating continuously on a mixer for 5 minutes at room temperature. Poly(A) RNAs were collected on a magnet while depleted supernatant was removed and purified using the RNA Clean and Concentrator kit (ZymoResearch, R1015). Illumina sequencing libraries were prepared using the Ovation Universal RNA-seq System (NUGEN, 0343-32) with Universal Human rRNA strand selection reagent (NUGEN, S01859) following the manufacturer's instructions.

All samples were sequenced 2x80 on a NEXTseq 500 sequencer (Illumina, San Diego, CA, USA) in the Biopolymers Facility at Harvard Medical School. Paired-end reads were aligned to the ENSEMBLE GRCh38 (release-86) reference genome using STAR (v2.5.1a)(Dobin et al., 2013) with default parameters (except for readFilesCommand=cat, limitIObufferSize=200000000, limitBAMsortRAM=64000000000, outReadsUnmapped=Fastx, outSAMtype=BAM SortedByCoordinate, outSAMattributes=All, outFilterMultimapNmax=101, outSJfilterOverhangMin=3 1 1 1, outSJfilterDistToOtherSJmin=0 0 0 0, alignIntronMin=11, alignEndsType=EndToEnd). Splicing index calculations were determined by summing the number of spliced and unspliced read pairs that span exon junctions by at least 3 nucleotides and calculating the total spliced read pairs divided by the total unspliced read pairs for each gene; splicing index = 2 × spliced read pairs / (5'SS unspliced + 3'SS unspliced read pairs).

Calculation of MI

Mutual information was calculated using the "mutinformation" command from the "infotheo" R package. Maximum mutual information was calculated based off of an estimate for what 80% efficiency of probe binding would yield, which we approximated with 80 "unspliced" RNA, 20 of each "partially spliced" intermediate, and 50 "fully spliced" RNA. These numbers were based on the approximate number of exon only RNA found within the transcription site proximal region.

Data and Code Availability

All data and code to generate figures can be found at https://www.dropbox.com/sh/sqr9zk2icptemfa/AABHRHroWbcowgPr_4VS2x9ba?dl=0.

Tables

Gene	Intron #	# of Probes	% coloc with exon	mean dispersal (um)	% of intensity that is not local to txn site
CPS1	1/37	10	75.5814	0	0
CPS1	13/37	12	88.2716	0.067485645	1.337543
CPS1	21/37	24	74.71264	2.0966995	27.7981585
CPS1	34/37	16	88.88889	0.03229656	1.952689
EEF2	1/14	21	91.66667	0.93700431	15.69285
EEF2	2/14	21	91.73554	0.41633783	6.482196
EEF2	11/14	13	94.08867	1.67628235	24.386486
TM4SF1	1/4	17	98.59155	0	0
TM4SF1	3/4	24	94.78261	0.34731502	2.422979
TM4SF1	4/4	24	87.84648	3.00480873	47.094267
FKBP5**	1/10	32	77.52809	1.2286272	22.451107
FKBP5**	3/10	32	94.64286	0.4514456	3.692255
FKBP5**	5/10	32	73.4375	2.852407	6.836791
FKBP5**	8/10	32	81.41593	1.4870159	33.077603
FKBP5**	9/10	14	93.89671	0.5677152	3.679758
FKBP5**	10/10	11	71.60494	0	0

** reflects values after 8 hours in dexamethasone

Table 1: Per probe set dispersal and distal post-transcriptionality of splicing.

Gene	Intron 1	Intron 2	Classification	Fisher 2x2 Exact Test P-value	Mutual Information (nats)
CPS1	intron 1	intron 13	ambiguous	0.00014291	0.06061384
CPS1	intron 13	intron 21	independent	0.52542181	0.002127
CPS1	intron 21	intron 34	independent	0.042609	0.00871542
CPS1	intron 34	intron 1	ambiguous	0.02729512	0.01883053
CPS1	intron 13	intron 34	ambiguous	0.24468035	0.00440814
CPS1	intron 21	intron 1	independent	0.00199224	0.01614408
EEF2	intron 1	intron 11	independent	3.38E-05	0.01572976
EEF2	intron 2	intron 1	dependent	3.98E-07	0.0308474

EEF2	intron 11	intron 2	dependent	1.27E-06	0.03756791
TM4SF1	intron 1	intron 4	independent	0.43288281	0.00101288
TM4SF1	intron 3	intron 1	ambiguous	0.00568284	0.02056082
TM4SF1	intron 4	intron 3	independent	0.15205967	0.00181136
FKBP5**	intron 1.2	intron 1.1	dependent	6.97E-10	0.05519327
FKBP5**	intron 3	intron 10	ambiguous	0.07999788	0.0115184
FKBP5**	intron 5	intron 3	independent	0.01954313	0.01823341
FKBP5**	intron 8	intron 5	independent	0.23672561	0.00411029
FKBP5**	intron 8	intron 1.1	independent	0.3248287	0.00343541
FKBP5**	intron 8	intron 9	independent	0.6915974	0.0002986
FKBP5**	intron 9	intron 8	independent	0.69304665	0.00035716
FKBP5**	intron 10	intron 9	dependent	1.27E-05	0.12801797

** reflects values after 8 hours in dexamethasone

Table 2: Independence, mutual information, and classification of dependent, independent, or ambiguous for intron pairs.

Gene	**Intron**	**Time in Dexamethasone (min)**	**Percent of Intensity that is "Distally Post-transcriptional"**	**Replicate**
FKBP5	1/10	0	0	A
FKBP5	1/10	10	0	A
FKBP5	1/10	20	0	A
FKBP5	1/10	30	0	A
FKBP5	1/10	45	0	A
FKBP5	1/10	60	0.04347121	A
FKBP5	1/10	90	0.22236464	A
FKBP5	1/10	120	0.03700375	A
FKBP5	1/10	150	0.0771233	A
FKBP5	1/10	180	0.04102216	A
FKBP5	1/10	240	0.09652356	A
FKBP5	1/10	300	0.14042251	A
FKBP5	1/10	360	0.0638648	A
FKBP5	1/10	420	0.02351768	A
FKBP5	1/10	480	0.04788421	A
FKBP5	8/10	0	0	A

FKBP5	8/10	10	0	A
FKBP5	8/10	20	0	A
FKBP5	8/10	30	0	A
FKBP5	8/10	45	0	A
FKBP5	8/10	60	0	A
FKBP5	8/10	90	0.28510057	A
FKBP5	8/10	120	0.00410379	A
FKBP5	8/10	150	0.2430057	A
FKBP5	8/10	180	0.14650536	A
FKBP5	8/10	240	0.18797157	A
FKBP5	8/10	300	0.24780309	A
FKBP5	8/10	360	0.28302486	A
FKBP5	8/10	420	0.20445177	A
FKBP5	8/10	480	0.25101165	A
FKBP5	8/10	0	0	B
FKBP5	8/10	10	0	B
FKBP5	8/10	20	0	B
FKBP5	8/10	30	0	B
FKBP5	8/10	45	0	B
FKBP5	8/10	60	0.06022585	B
FKBP5	8/10	90	0.25481359	B
FKBP5	8/10	120	0.38613684	B
FKBP5	8/10	150	0.1799548	B
FKBP5	8/10	180	0.38312298	B
FKBP5	8/10	240	0.3589871	B
FKBP5	8/10	300	0.34120934	B
FKBP5	8/10	360	0.17685889	B
FKBP5	8/10	420	0.28945287	B
FKBP5	8/10	480	0.27975627	B
FKBP5	9/10	0	0	B
FKBP5	9/10	10	0	B
FKBP5	9/10	20	0	B
FKBP5	9/10	30	0	B
FKBP5	9/10	45	0	B

FKBP5	9/10	60	0	B
FKBP5	9/10	90	0	B
FKBP5	9/10	120	0	B
FKBP5	9/10	150	0	B
FKBP5	9/10	180	0	B
FKBP5	9/10	240	0	B
FKBP5	9/10	300	0	B
FKBP5	9/10	360	0	B
FKBP5	9/10	420	0	B
FKBP5	9/10	480	0	B

Table 3: Percent of intensity that is “distally post-transcriptional” for three FKBP5 introns over time in dexamethasone.

Acknowledgements

We would like to thank members of the Raj, Churchman, and Berger labs for critical reading of the manuscript. AR acknowledges R01 CA238237, NIH Director's Transformative Research Award R01 GM137425, R01 CA232256, NSF CAREER 1350601, P30 CA016520, SPORE P50 CA174523, NIH U01 CA227550, NIH 4DN U01 HL129998, NIH Center for Photogenomics (RM1 HG007743), and the Tara Miller Foundation. AJC was also supported by NIH training grant T32 GM- 07229. LSC acknowledges support from R21-HG009264 and R01-GM117333, and F31-GM122133 to HLD. SB acknowledges support from 3R01CA078831-20S1 and KA acknowledges support from 5F32CA221010-02.

Author Contributions

AJC and AR conceived of the research with preliminary input and method development by SB. AJC and CC designed, performed experiments and analysed the data with guidance from AR. HLD designed and performed nascent RNA sequencing experiments with guidance from LSC and analyzed the data in conjunction with AJC. AJC and KAA contributed to design of speckle experiments with guidance from SB. AJC wrote the paper with guidance from AR. FC, AW, and AJC conducted initial optimization of expansion microscopy experiments with guidance from EB.

References

Aebi, M., and Weissman, C. (1987). Precision and orderliness in splicing. Trends Genet. *3*, 102–107.

Alpert, T., Herzel, L., and Neugebauer, K.M. (2016). Perfect timing: splicing and transcription rates in living cells. Wiley Interdiscip. Rev. RNA.

Beyer, A.L., Bouton, A.H., and Miller, O.L., Jr (1981). Correlation of hnRNP structure and nascent transcript cleavage. Cell *26*, 155–165.

Bhatt, D.M., Pandya-Jones, A., Tong, A.-J., Barozzi, I., Lissner, M.M., Natoli, G., Black, D.L., and Smale, S.T. (2012). Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. Cell *150*, 279–290.

Brody, Y., Neufeld, N., Bieberstein, N., Causse, S.Z., Böhnlein, E.-M., Neugebauer, K.M., Darzacq, X., and Shav-Tal, Y. (2011). The in vivo kinetics of RNA polymerase II elongation during co-transcriptional splicing. PLoS Biol. *9*, e1000573.

Carrillo Oesterreich, F., Bieberstein, N., and Neugebauer, K.M. (2011). Pause locally, splice globally. Trends Cell Biol. *21*, 328–335.

Carrillo Oesterreich, F., Herzel, L., Straube, K., Hujer, K., Howard, J., and Neugebauer, K.M. (2016). Splicing of Nascent RNA Coincides with Intron Exit from RNA Polymerase II. Cell *165*, 372–381.

Chen, F., Tillberg, P.W., and Boyden, E.S. (2015). Optical imaging. Expansion microscopy. Science *347*, 543–548.

Chen, F., Wassie, A.T., Cote, A.J., Sinha, A., Alon, S., Asano, S., Daugharthy, E.R., Chang, J.-B., Marblestone, A., Church, G.M., et al. (2016). Nanoscale imaging of RNA with expansion microscopy. Nat. Methods *13*, 679–684.

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

Coulon, A., Ferguson, M.L., de Turris, V., Palangat, M., Chow, C.C., and Larson, D.R. (2014). Kinetic competition during the transcription cycle results in stochastic RNA processing. Elife *3*.

Ding, F., and Elowitz, M.B. (2019). Constitutive splicing and economies of scale in gene expression. Nat. Struct. Mol. Biol. *26*, 424–432.

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

Dölken, L., Ruzsics, Z., Rädle, B., Friedel, C.C., Zimmer, R., Mages, J., Hoffmann, R., Dickinson, P., Forster, T., Ghazal, P., et al. (2008). High-resolution gene expression profiling for

simultaneous kinetic parameter analysis of RNA synthesis and decay. RNA *14*, 1959–1972.

Drexler, H.L., Choquet, K., and Churchman, L.S. (2019). Splicing Kinetics and Coordination Revealed by Direct Nascent RNA Sequencing through Nanopores. Mol. Cell.

Dye, M.J., Gromak, N., and Proudfoot, N.J. (2006). Exon tethering in transcription by RNA polymerase II. Mol. Cell *21*, 849–859.

Eser, P., Wachutka, L., Maier, K.C., Demel, C., Boroni, M., Iyer, S., Cramer, P., and Gagneur, J. (2016). Determinants of RNA metabolism in the Schizosaccharomyces pombe genome. Mol. Syst. Biol. *12*, 857.

Femino, A.M., Fay, F.S., Fogarty, K., and Singer, R.H. (1998). Visualization of single RNA transcripts in situ. Science *280*, 585–590.

Huranová, M., Ivani, I., Benda, A., Poser, I., Brody, Y., Hof, M., Shav-Tal, Y., Neugebauer, K.M., and Stanek, D. (2010). The differential interaction of snRNPs with pre-mRNA reveals splicing kinetics in living cells. J. Cell Biol. *191*, 75–86.

Kessler, O., Jiang, Y., and Chasin, L.A. (1993). Order of intron removal during splicing of endogenous adenine phosphoribosyltransferase and dihydrofolate reductase pre-mRNA. Mol. Cell. Biol. *13*, 6211–6222.

Kim, S.W., Taggart, A.J., Heintzelman, C., Cygan, K.J., Hull, C.G., Wang, J., Shrestha, B., and Fairbrother, W.G. (2017). Widespread intra-dependencies in the removal of introns from human transcripts. Nucleic Acids Res. *45*, 9503–9513.

Levesque, M.J., and Raj, A. (2013). Single-chromosome transcriptional profiling reveals chromosomal gene expression regulation. Nat. Methods *10*, 246–248.

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

de la Mata, M., Lafaille, C., and Kornblihtt, A.R. (2010). First come, first served revisited: factors affecting the same alternative splicing event have different effects on the relative rates of intron removal. RNA *16*, 904–912.

Mayer, A., and Churchman, L.S. (2016). Genome-wide profiling of RNA polymerase transcription at nucleotide resolution in human cells with native elongating transcript sequencing. Nat. Protoc. *11*, 813–833.

Mayer, A., and Churchman, L.S. (2017). A Detailed Protocol for Subcellular RNA Sequencing (subRNA-seq). Curr. Protoc. Mol. Biol. *120*, 4.29.1–4.29.18.

Mayer, A., di Iulio, J., Maleri, S., Eser, U., Vierstra, J., Reynolds, A., Sandstrom, R., Stamatoyannopoulos, J.A., and Churchman, L.S. (2015). Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. Cell *161*, 541–554.

Nojima, T., Gomes, T., Grosso, A.R.F., Kimura, H., Dye, M.J., Dhir, S., Carmo-Fonseca, M., and Proudfoot, N.J. (2015). Mammalian NET-Seq Reveals Genome-wide Nascent Transcription

Coupled to RNA Processing. Cell *161*, 526–540.

Pandya-Jones, A., and Black, D.L. (2009). Co-transcriptional splicing of constitutive and alternative exons. RNA *15*, 1896–1908.

Pandya-Jones, A., Bhatt, D.M., Lin, C.-H., Tong, A.-J., Smale, S.T., and Black, D.L. (2013). Splicing kinetics and transcript release from the chromatin compartment limit the rate of Lipid A-induced gene expression. RNA *19*, 811–827.

Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., and Tyagi, S. (2006). Stochastic mRNA synthesis in mammalian cells. PLoS Biol. *4*, e309.

Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A., and Tyagi, S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. Nat. Methods *5*, 877–879.

Schwalb, B., Michel, M., Zacher, B., Frühauf, K., Demel, C., Tresch, A., Gagneur, J., and Cramer, P. (2016). TT-seq maps the human transient transcriptome. Science *352*, 1225–1228.

Tilgner, H., Knowles, D.G., Johnson, R., Davis, C.A., Chakrabortty, S., Djebali, S., Curado, J., Snyder, M., Gingeras, T.R., and Guigó, 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.

Tsai, M.J., Ting, A.C., Nordstrom, J.L., Zimmer, W., and O'Malley, B.W. (1980). Processing of high molecular weight ovalbumin and ovomucoid precursor RNAs to messenger RNA. Cell *22*, 219–230.

Vargas, D.Y., Shah, K., Batish, M., Levandoski, M., Sinha, S., Marras, S.A.E., Schedl, P., and Tyagi, S. (2011). Single-molecule imaging of transcriptionally coupled and uncoupled splicing. Cell *147*, 1054–1065.

Waks, Z., Klein, A.M., and Silver, P.A. (2011). Cell-to-cell variability of alternative RNA splicing. Mol. Syst. Biol. *7*, 506.

Wallace, E.W.J., and Beggs, J.D. (2017). Extremely fast and incredibly close: cotranscriptional splicing in budding yeast. RNA *23*, 601–610.

Wuarin, J., and Schibler, U. (1994). Physical isolation of nascent RNA chains transcribed by RNA polymerase II: evidence for cotranscriptional splicing. Mol. Cell. Biol. *14*, 7219–7225.

Yang, M., Wu, J., Wu, S.-H., Bi, A.-D., and Liao, D.J. (2012). Splicing of mouse p53 pre-mRNA does not always follow the "first come, first served" principle and may be influenced by cisplatin treatment and serum starvation. Mol. Biol. Rep. *39*, 9247–9256.

Zhang, G., Taneja, K.L., Singer, R.H., and Green, M.R. (1994). Localization of pre-mRNA splicing in mammalian nuclei. Nature *372*, 809–812.

Figure Captions

Figure 1: At least one intron from every tested endogenous gene is post-transcriptionally spliced.
A) Schematic depicting three categories of splicing. B) Schematic depicting RNA FISH method and translation of RNA FISH images into dispersal graphs. C) Example images of dispersed (intron 4) and not dispersed (introns 1 and 3) transcription sites, via RNA FISH. D) Graphs indicating dispersal distance or the distance of each detected pre-mRNA to its nearest transcription site. CPS1, EEF2, and TM4SF1 data is from HeLa cells, FKBP5 data is from A549 cells after 8 hours in dexamethasone. All scale bars represent 5um unless otherwise noted.

Figure 2: Transcripts dwell at the site of transcription after transcription is completed.
A) Gene and probe position diagram for CPS1 pre-mRNA and mRNA. B) example images of RNA FISH before and after expansion microscopy for individual mRNA and transcription sites (in HeLa cells). C) Example images of 5' and 3' CPS1 RNA FISH after expansion. D) Schematic depicting RNA FISH signal for diffusion or a slow moving proximal zone for completed (but not necessarily spliced) transcripts. E) Line scans of fluorescence intensity (arbitrary units) from transcription sites like in C. F) Example images of co-IF (polymerase II) and RNA FISH for CPS1 after expansion. Quantification of images like that in J, and others, representing distance from each RNA spot to the nearest polymerase II IF spot. G) Graph depicting area (in pixels) of polygons drawn around individual transcription sites after expansion. H) Schematic depicting RNA FISH signal that would result from either fast or slow splicing at the site of transcription. I) Example images of a middle intron (intron 13) of CPS1 RNA FISH after expansion. J) Distance from 5' to 3' of expanded and unexpanded mRNA and transcription sites, as detected by CPS1 RNA FISH. All scale bars represent 5 μm unless otherwise noted.

Figure 3: Sequencing corroborates RNA FISH dispersal results. A) Schematic of sequencing experiment design (in HeLa cells). B) Comparison of mean dispersal (as assessed by RNA FISH, see Figure 1) and mean splicing index (as assessed by sequencing).

Figure 4: Some transcripts localize to a speckle proximal compartment after splicing inhibition.
A) Schematic of possibilities for RNA movement post- transcription. B) Example images of introns of specified genes before and after pladienolide B treatment. Outline represents nucleus. C) Description of 3 response types to pladienolide B. D) EEF2 RNA FISH exon and intron images. Scale bar = 5um. E) EEF2 RNA FISH intron images. Scale bar = 5um. F) Specified RNA FISH intron images. Scale bar = 5um. G) Costaining of RNA FISH for specified intron and SC35 IF in the same cell. Scale bar = 5um. H) Quantification of speckle decile from previously published data, compared with PlaB response.

Figure 5: Intron dispersal varies with transcription level and is therefore not an inherent property of each intron. A) Gene and probe diagram for FKBP5 and schematic of dexamethasone

treatment schedule. B) Quantification of FKBP5 RNA FISH exon and intron spots over time of treatment in dexamethasone (in A549 cells). C) Dispersal graphs (as quantified from RNA FISH) of FKBP5 introns 1, 8, and 9 over time in dexamethasone (in A549 cells). D) Schematic of local depletion of splicing factors.

Supplementary Figure 1: Transcription site choice and defining post-transcriptionality. A) Histograms of intron intensities for TM4SF1 intron 1 and intron 4, before, after, and during defining a global thresholding cutoff and other transcription site selection methods, as well as dispersal graphs generated based on those transcription site selections. B) Example classification of pre-mRNA as either transcription sites or dispersed pre-mRNA. C) Distances between 5' and 3' ends of RNA as detected by RNA FISH for either the 3' or 5' end of the RNA of interest. D) Example classification of RNA as "distal" or "proximal" to the transcription site E) Histogram showing type of pre-mRNA as defined by RNA FISH for stated introns in *FKBP5*, as a function of distance to the nearest transcription site.

Supplementary Figure 2: Expansion microscopy yields a 3.3 fold linear expansion. A) Comparison of radii of nuclei (based on DAPI staining, and assuming spherical nuclei) before and after expansion.

Supplementary Figure 3: Splicing index varies on a per intron basis. A) Splicing index of each intron for which we obtained RNA FISH measurements. Error bars represent mean +/- sd. n = 2 B) Comparison of transcription site size (as assessed by RNA FISH) and splicing index, as assessed by sequencing.

Supplementary Figure 4: Compartmentalization genes before splicing inhibition.
A) Schematic of compartmentalization phenotype with and without tether. B) RNA FISH of RPL13A, EEF2, and GAPDH introns before pladienolide B treatment. Scale bar = 5 µm. C) Combined RNA FISH for the stated introns and IF for SC35, before pladienolide B treatment. Scale bar = 5 µm.

Figure 1

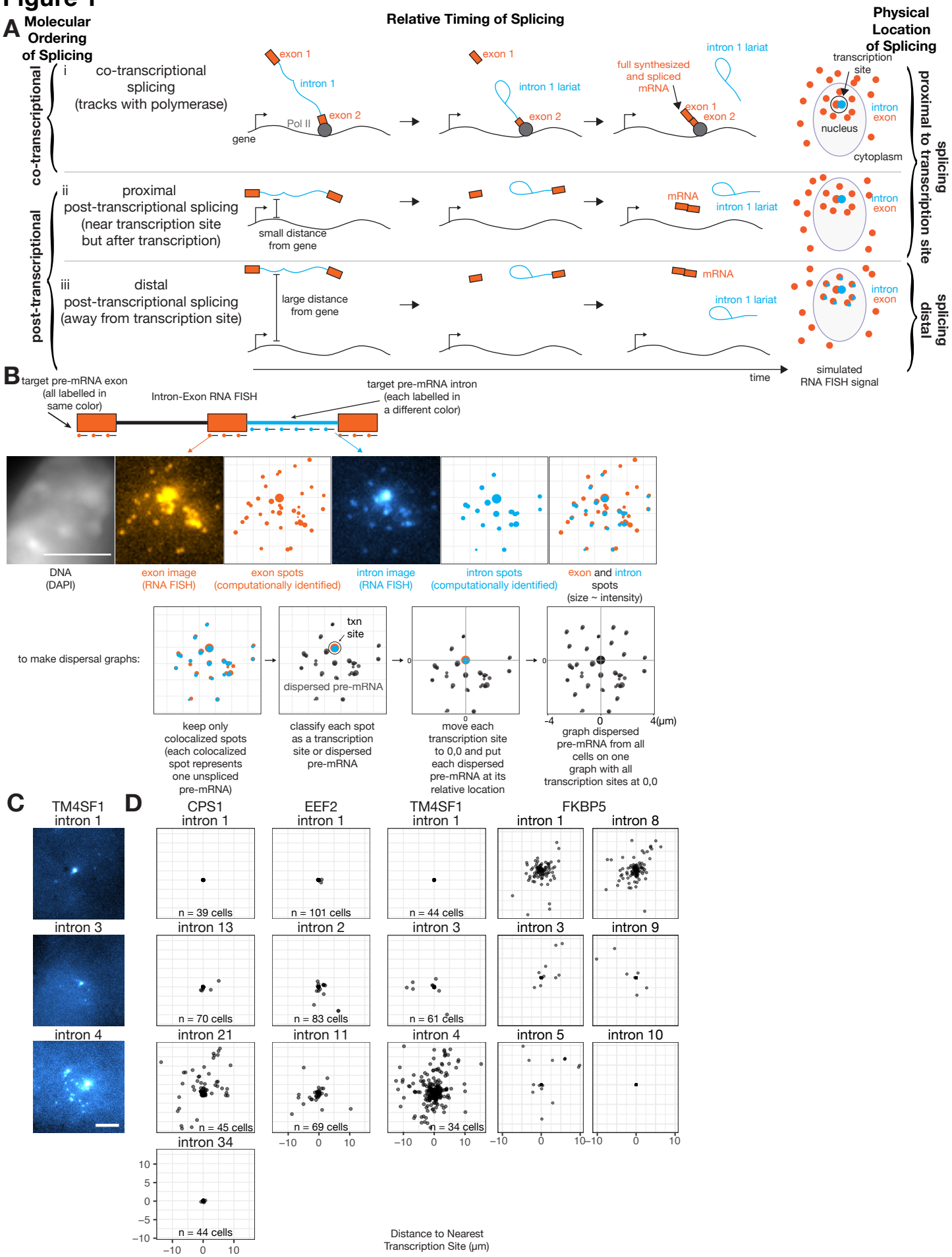

A
Molecular Ordering of Splicing
Relative Timing of Splicing
Physical Location of Splicing
co-transcriptional
post-transcriptional
i co-transcriptional splicing (tracks with polymerase)
ii proximal post-transcriptional splicing (near transcription site but after transcription)
iii distal post-transcriptional splicing (away from transcription site)
exon 1
intron 1
exon 2
Pol II
gene
intron 1 lariat
full synthesized and spliced mRNA
mRNA
small distance from gene
large distance from gene
time
transcription site
nucleus
cytoplasm
intron
exon
simulated RNA FISH signal
proximal to transcription site splicing
distal splicing
B
target pre-mRNA exon (all labelled in same color)
Intron-Exon RNA FISH
target pre-mRNA intron (each labelled in a different color)
DNA (DAPI)
exon image (RNA FISH)
exon spots (computationally identified)
intron image (RNA FISH)
intron spots (computationally identified)
exon and intron spots (size ~ intensity)
to make dispersal graphs:
keep only colocalized spots (each colocalized spot represents one unspliced pre-mRNA)
txn site
dispersed pre-mRNA
classify each spot as a transcription site or dispersed pre-mRNA
move each transcription site to 0,0 and put each dispersed pre-mRNA at its relative location
-4
0
4(μm)
graph dispersed pre-mRNA from all cells on one graph with all transcription sites at 0,0
C
TM4SF1 intron 1
intron 3
intron 4
D
CPS1 intron 1
EEF2 intron 1
TM4SF1 intron 1
FKBP5 intron 1
FKBP5 intron 8
intron 13
intron 2
intron 3
intron 9
intron 21
intron 11
intron 4
intron 5
intron 10
intron 34
n = 39 cells
n = 101 cells
n = 44 cells
n = 70 cells
n = 83 cells
n = 61 cells
n = 45 cells
n = 69 cells
n = 34 cells
n = 44 cells
Distance to Nearest Transcription Site (μm)

Figure 2

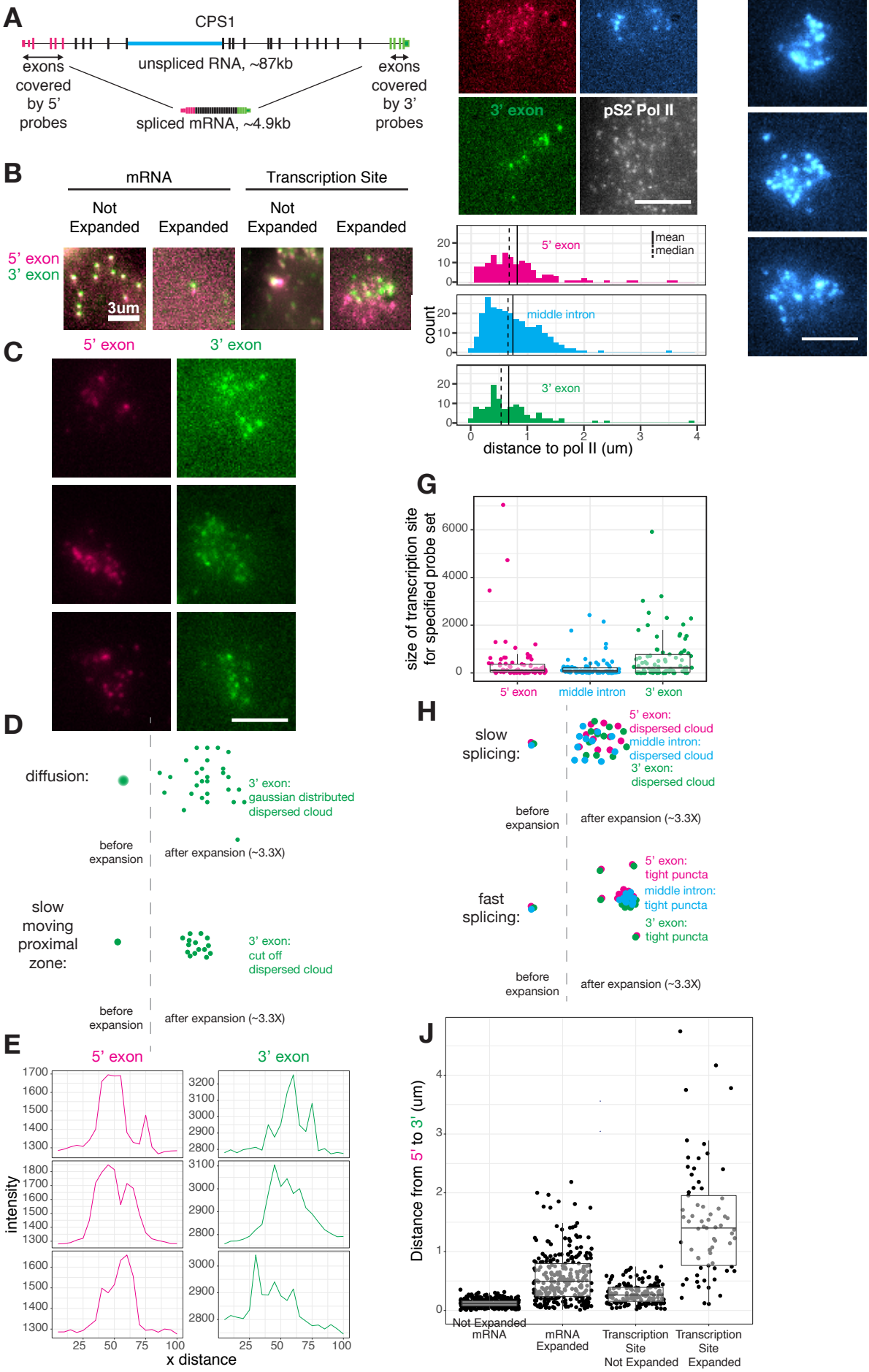
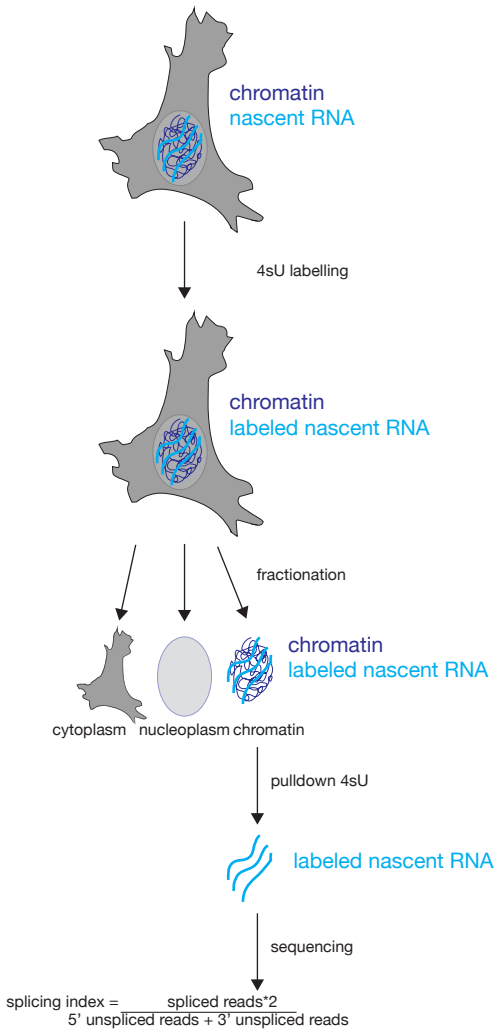

A
CPS1
exons covered by 5' probes
unspliced RNA, ~87kb
exons covered by 3' probes
spliced mRNA, ~4.9kb
B
mRNA
Transcription Site
Not Expanded
Expanded
Not Expanded
Expanded
5' exon
3' exon
3um
C
5' exon
3' exon
D
diffusion:
3' exon: gaussian distributed dispersed cloud
before expansion
after expansion (~3.3X)
slow moving proximal zone:
3' exon: cut off dispersed cloud
before expansion
after expansion (~3.3X)
E
5' exon
3' exon
intensity
x distance
F
5' exon
middle intron
3' exon
pS2 Pol II
count
mean
median
distance to pol II (um)
G
size of transcription site for specified probe set
5' exon
middle intron
3' exon
H
slow splicing:
5' exon: dispersed cloud
middle intron: dispersed cloud
3' exon: dispersed cloud
before expansion
after expansion (~3.3X)
fast splicing:
5' exon: tight puncta
middle intron: tight puncta
3' exon: tight puncta
before expansion
after expansion (~3.3X)
I
middle intron
J
Distance from 5' to 3' (um)
Not Expanded mRNA
mRNA Expanded
Transcription Site Not Expanded
Transcription Site Expanded

Figure 3

A


chromatin
nascent RNA
4sU labelling
chromatin
labeled nascent RNA
fractionation
chromatin
labeled nascent RNA
cytoplasm
nucleoplasm
chromatin
pulldown 4sU
labeled nascent RNA
sequencing
splicing index = spliced reads*2 / 5' unspliced reads + 3' unspliced reads

B

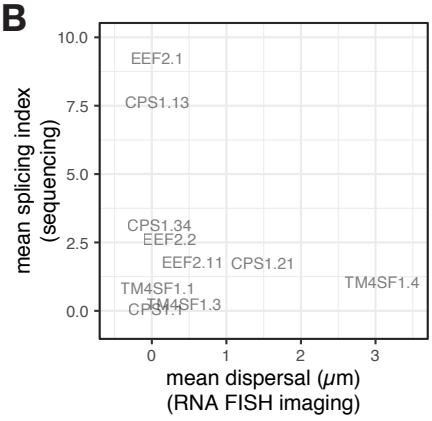
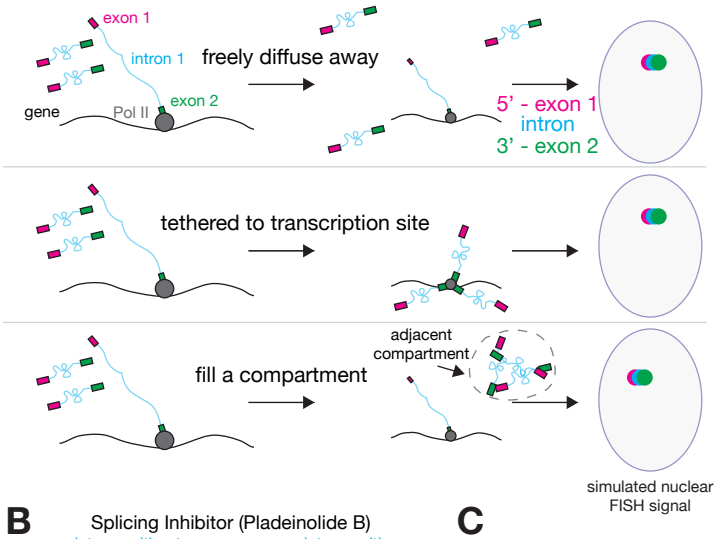
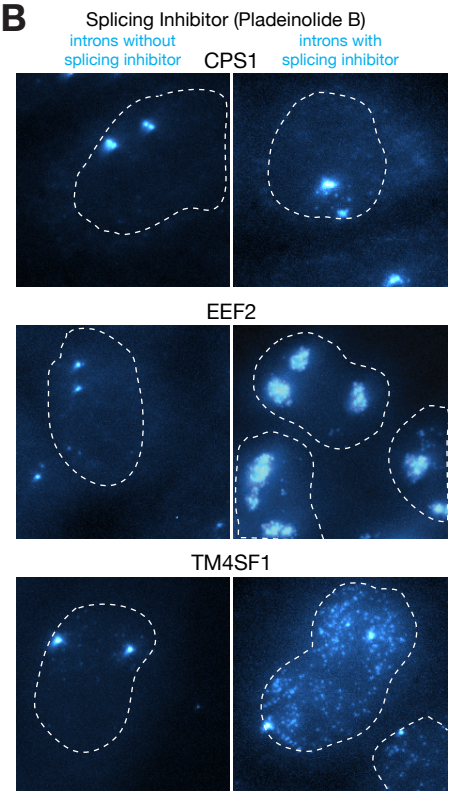

mean splicing index
(sequencing)
10.0
7.5
5.0
2.5
0.0
EEF2.1
CPS1.13
CPS1.34
EEF2.2
EEF2.11
CPS1.21
TM4SF1.4
TM4SF1.1
TM4SF1.3
CPS1.1
0
1
2
3
mean dispersal (μm)
(RNA FISH imaging)

Figure 4

A Potential Pre-mRNA Behaviors after Leaving Transcription Site



B



C

Splicing Inhibitor Response Categories:

Unaffected

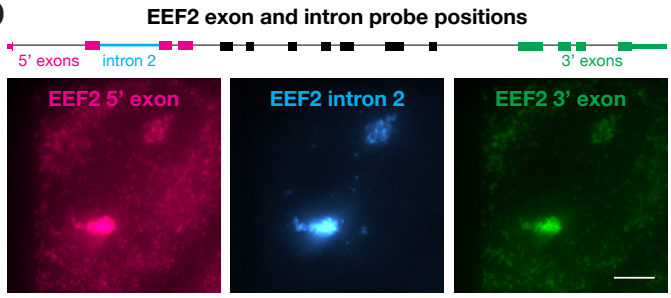
Compartmentalization

- increased pre-mRNA and dispersal but only within a compartment
- never fills nucleus

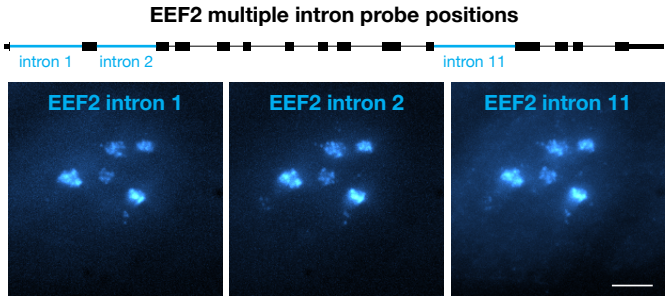
Nuclear Dispersal

- increased pre-mRNA and dispersal
- sometimes fills nucleus

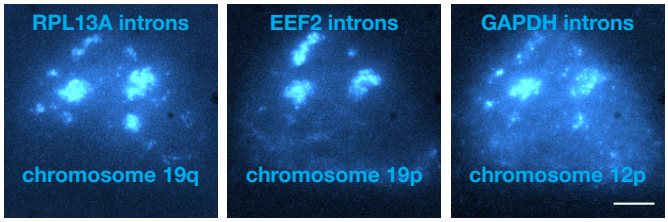
D



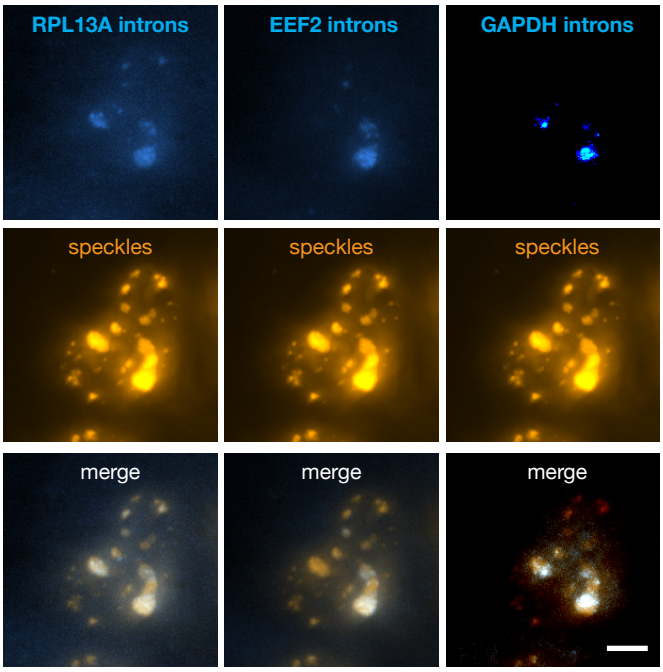
E



F



G



H

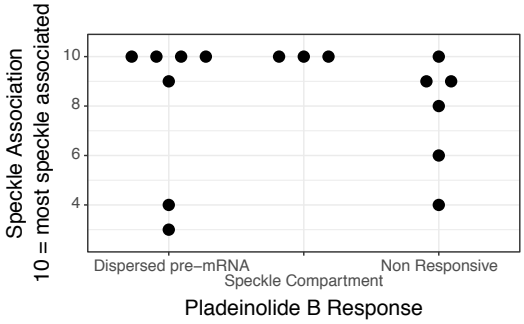
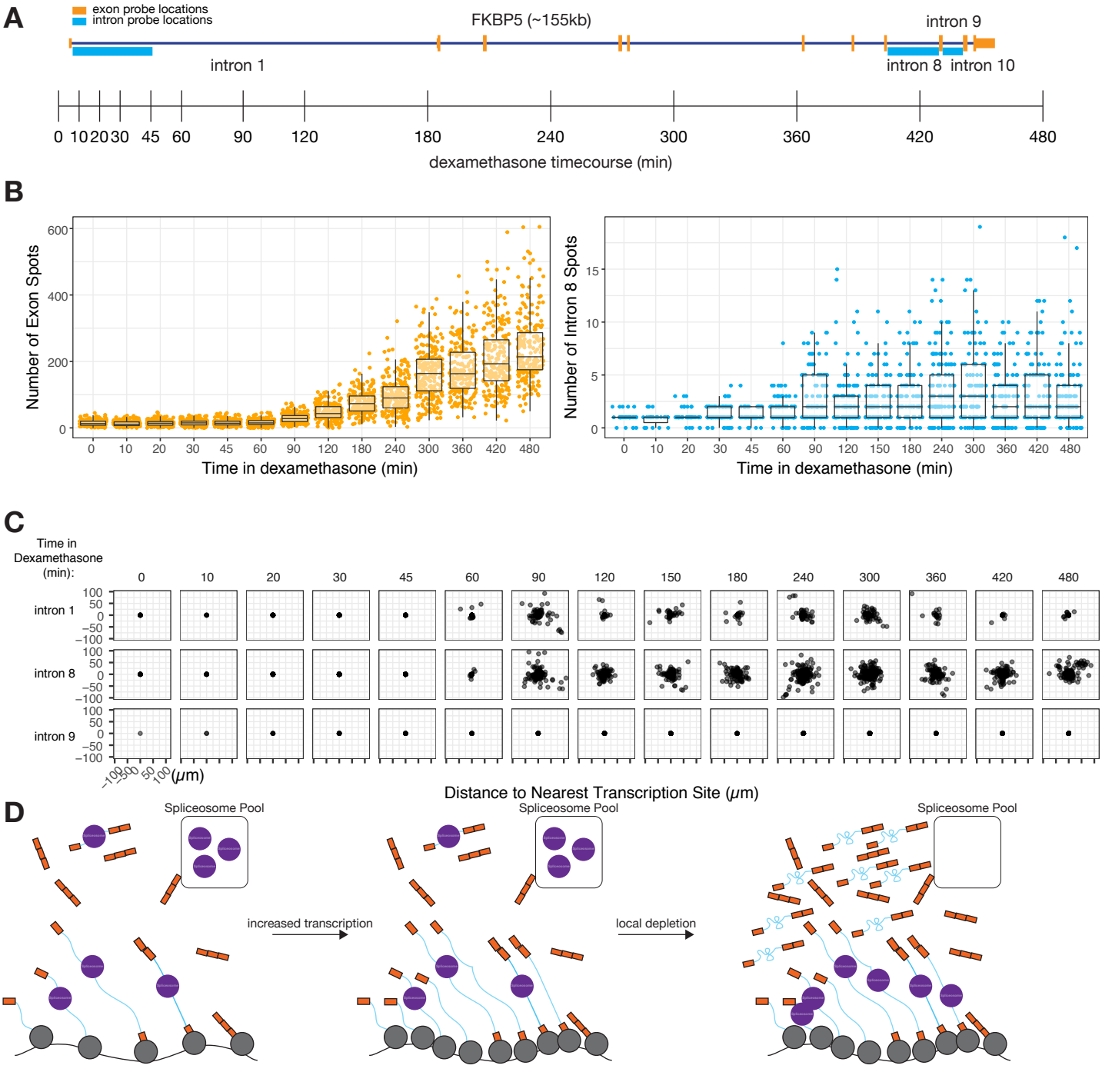


Figure 5


A
exon probe locations
intron probe locations
FKBP5 (~155kb)
intron 9
intron 1
intron 8
intron 10
0 10 20 30 45 60 90 120 180 240 300 360 420 480
dexamethasone timecourse (min)
B
Number of Exon Spots
Time in dexamethasone (min)
Number of Intron 8 Spots
Time in dexamethasone (min)
C
Time in Dexamethasone (min):
intron 1
intron 8
intron 9
Distance to Nearest Transcription Site (μm)
D
Spliceosome Pool
increased transcription
local depletion