pre-mRNA spatial distributions suggest that splicing can occur post-transcriptionally

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. While the biochemical steps of splicing have been largely worked out, the ordering and spatial distribution of these steps, especially in relation to the transcriptional process itself, remain controversial. Here, 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 order in which an intron is transcribed and spliced out of the pre-mRNA. We show that 4 out of 4 genes we interrogated exhibit clear post-transcriptional splicing of at least one intron, and that introns can be spliced in any order. Expansion microscopy further revealed the presence of a transcription site proximal zone in which the motion of RNA is slower than in the nucleoplasm. Full length pre-mRNA undergo continuous splicing as they move through this zone after transcription. Upon leaving this zone, some genes' transcripts localize to speckles during the process of splicing while others appear to traffic freely through the nucleus without localizing to any other nuclear compartment. Our results suggest a model in which transcription and splicing are largely independent, and the existence of a slow moving zone around the transcription site allows for the unification of co-transcriptional and post-transcriptional models of splicing.

Introduction

The mRNA transcribed by many eukaryotic genes are spliced, a process in which the intronic RNA are removed and the exonic RNA are joined together to form the ultimate mature mRNA. 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 (*1*, *2*), while other work suggests that many pre-mRNAs are fully transcribed before splicing occurs (*3–5*). 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) (*5–7*).

In lieu of direct visualization, many studies have used biochemical fractionation to infer the location of various intermediates (*4*, *8–13*). 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 (*9*, *14*). 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 (*15*). 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 (*16*). 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 (*5–7*, *17–19*). Imaging using probes targeting both exonic and intronic regions of RNA has revealed bright nuclear foci that represent nascently transcribing RNA (*7*, *19*). 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 (*7*). 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 (*5*),

suggesting the latter possibility, 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 (*20*, *21*). 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 at least 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. These results suggest a model for splicing dynamics that unifies existing data.

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 (*22*, *23*) 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) an intron from a pre-mRNA being spliced out *at* the site of transcription and 2) an intron from a pre-mRNA being spliced out *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 out 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 1E 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). We chose this colocalization threshold such that at least 78% (depending on the gene, CPS1 = 78%, EEF2 = 85%, TM4SF1 = 87%) of exons colocalized with other exons of the same gene in the cytoplasm; however, varying this threshold from 0.13µm (one pixel) to 2.6µm (20 pixels) did not significantly change dispersal patterns (Supplementary Figure 1C).

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), showing that the detection of dispersal is not dependent on this choice of method. Our chosen strategy likely grouped together the signal from multiple pre-mRNA (introns and exons) at the transcription site, rendering them indistinguishable at this level of spatial resolution; we separately analyzed these transcription sites further later (see Figure 2). 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 was present in dispersed pre-mRNA in both transformed cell lines (HeLa and A549) and non-transformed cells (CRL-2097) (3-5 introns tested per gene, Figure 1D), suggesting that pre-mRNA dispersal is a generic feature of transcription and splicing that is not dependent on the particular cell line used. However, we found that the dispersal patterns for any one particular intron were not necessarily consistent between HeLa and CRL-2097 cells. For example, *TM4SF1* intron 4 dispersed strongly in HeLa cells but did not disperse at all in CRL-2097 cells, even though it was expressed at similar levels. However, *EEF2* introns 2 and 11 dispersed to a similar extent in both HeLa cells and CRL-2097 cells, suggesting that some introns may potentially disperse regardless of the cellular context. 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 observed a large range in the number of pre-mRNA observed per cell (from 0 to ~20), which also varied from intron to intron (Supplementary Figure 1B).

It is possible that pre-mRNA away from the transcription site are ultimately degraded without ever being spliced and thus represent an alternative "dead end" fate for the pre-mRNA rather than an intermediate on the path to a mature mRNA. This alternative possibility is difficult to eliminate with data relying on a snapshot in time, and hence studies using sequencing or fixed

cell imaging (such as ours) suffer from this limitation. The further development of live cell imaging capable of distinguishing putative splicing intermediates away from the transcription site may help resolve these issues. However, we do note that pre-mRNA contain fewer introns as they increase in distance from the transcription site (Supplementary Figure 1G), which is at least consistent with our assumption that these introns are spliced as they travel away from the transcription site. Furthermore, pulse-chase experiments have shown that the "yield" of splicing events is high for mRNA, suggesting that most pre-mRNA do not go down a dead-end pathway (*24, 25*). Given this assumption, the dispersion we observe 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 1G 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 of 0.055 was surprisingly low compared to our theoretical upper bound of 0.137, although the value was still high enough to suggest that these two are indeed informative about each other intron's splicing status. One explanation for the low mutual information could be recursive splicing of the very long intron 1 of FKBP5, in which 3' splice sites within the intron are spliced to additional intermediate 5' splice sites within the intron to splice an intron in a piecemeal manner, as opposed to the entire intron being spliced at once (*26*). Recursive splicing would yield a lower mutual information because there would be several 3' spots observed separately from 5' spots (which had already been spliced and degraded). Other explanations may include unknown technical errors.

We also calculated odds ratio, which describes the association between two events (odds ratios above 1 signify correlation, odds ratios below 1 signify mutual exclusion, and equal to 1 signify independence). For most intron pairs, the odds ratios observed were well below 1, suggesting that these splicing events are likely in competition with one another (the occurence of one intron being spliced makes it less likely that the other intron is spliced). There were a few exceptions, primarily in cases where the pair analyzed consisted of a highly dispersed intron and a less dispersed intron (TM4SF1 intron 1 vs. intron 4, CPS1 intron 13 vs. intron 21). In these cases, the odds ratio was typically slightly above 1, suggesting either independence (the splicing status of one has no effect on the other) or the possibility of mild cooperativity or processivity between the splicing of the two introns.

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 4.6x, Supplementary Figure 2B) 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) (*20, 21*). 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. Based on previous work (*20*), we expected expansion of transcription sites to be isotropic. To confirm isotropic expansion in our system, we imaged the same cell before and after expansion with and without a perturbation that changed the localization of introns (pladeinolide B treatment, which inhibits splicing and causes the introns of EEF2 to form a well-defined blob, see Figure 4) and indeed observed that both the transcription site and the blob of EEF2 introns expanded isotropically (Supplementary Figure 2B, displaying the same cells before and after expansion).

Upon expanding, labeling, and imaging, we observed that the 5', 3', and middle intron probe signals formed small clouds where we had previously observed the tight transcriptional focus (Figure 2C). 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 away right after their transcription, then a mathematical model predicts an exponential decrease in pre-mRNA concentration with increasing distance away from the transcription site (*27*) (see methods for details). We fit such a model to pre-mRNA that were between 3 and 10 microns away from the site of transcription, in which range the diffusion model fit well. From these data, we were able to estimate what the concentration should be at or near the transcription site given the free diffusion model (essentially, the y-intercept of the concentration curve). We found that the predicted concentration (18 molecules per cubic micron) was far lower than the actual concentration (235 molecules per cubic micron) in the vicinity of the transcription site measured by expansion microscopy (Fig. 2D). This high concentration in the "cloud" is not consistent with free diffusion subsequent to the completion of transcription, but is compatible with the existence of a distinct

"proximal zone" surrounding the transcription site through which pre-mRNA move more slowly than in the more distant nucleoplasm.

We wondered whether the pre-mRNA present in this zone were partially spliced intermediates or full pre-mRNA, the former signifying co-transcriptional splicing and the latter signifying predominantly post-transcriptional splicing. If splicing occurred concurrently or shortly after transcription, then we would expect to see a potential cloud of 5' ends but a tight spot of 3' ends corresponding to the point of transcriptional termination (Fig. 2H). However, if splicing occurred some time after the entire full length pre-mRNA was transcribed, one would expect to see separate clouds for both the 5' and 3' ends, representing pre-mRNAs that have completed transcription and are slowly diffusing through the proximal zone. We observed clouds for both the 5' and 3' ends of the pre-mRNA in the proximal zone. Based on the size of the clouds, we estimated the diameter of this zone to be around 0.3 microns around the transcription site (Figure 2G). We concluded that pre-mRNA are not spliced immediately upon transcription, but rather that a large proportion of the splicing occurs post-transcriptionally while the pre-mRNA were moving through the proximal zone. (Splicing that occurred while in this zone would appear to be co-transcriptional by conventional single molecule RNA FISH.)

To further 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 (Figure 2C). There does appear to be some polarity to the conformations of RNA at the transcription site, with each cloud of 5' exon, intron, or 3' exon spots only partially overlapping the other clouds for the same transcription site. With some regularity, it appears that the intron cloud is more likely to overlap the 3' exon cloud than the 5' exon cloud, but the overlap of the 5' exon and 3' exon clouds is more variable from site to site. The polarity we observe may represent the physical configuration of the gene itself.

Also, neither the 5' exon nor 3' exon signals overlapped with actively elongating RNA Polymerase II immunofluorescence signals (see methods), potentially suggesting the relatively few mRNA spots that are quite close to RNA polymerase II signal are undergoing active transcription while the majority of pre-mRNA in the slow moving proximal zone were not being actively transcribed (Figure 2F). It is important to note, however, that we do not label the entirety of the pre-mRNA molecule, nor can we be sure that we can detect single molecules of RNA polymerase II, thus it is difficult to eliminate the possibility that some other part of the pre-mRNA is close to RNA polymerase II or that a pre-mRNA is localized to an undetected single RNA polymerase II molecule.

Another interesting observation was that the size of these 5', intron, and 3' clouds, as well as the number of spots in each cloud, varied considerably from cell to cell (Figure 2C,G). In some cases, we observed more 5' spots than 3' spots, which may be due to the interception of an

ongoing transcriptional burst (where RNA polymerase II has already synthesized the 5' ends of the RNA but has not yet synthesized the 3' ends). We also observed some cases where there were more 3' spots than 5' spots and indeed many cases where the 3' clouds are more dispersed than the 5' clouds, which may suggest that the spots of 5' probe may represent a pileup of transcripts at the 5' end, such that we only see one bright spot where we should see several. When the dyes of the two probes are swapped, the 3' clouds are still consistently larger than the 5' clouds, suggesting that it's unlikely to be a dye-bias effect (Supplementary Figure 2C).

Thus, 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 (*4*). We then sequenced this RNA and calculated a "splicing index" (see (*4*)) 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 low splicing index (more "post-transcriptionally" spliced) had high dispersal (Figure 3B). 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.

We also calculated a "percentage spliced" metric (the spliced reads over the total number of reads), which we expected to be anticorrelated with mean dispersal (high dispersal = low percentage spliced, low dispersal = high percentage spliced), however, there was not a clear inverse correlation between these two metrics (Figure 3C). Rather, introns that are intermediately spliced based on sequencing reads were those with the highest dispersal.

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 lack of dispersal of some introns, relatively few unspliced RNA were detectable outside of the transcription-proximal zone, making it difficult to discriminate between these hypotheses in those cases. 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 (*17*). 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 (Supplementary Figure 4C) and after (Figure 4G) splicing inhibition. We corroborated the spatial association we observed between speckles and compartmentalized pre-mRNA by analyzing previously published high throughput sequencing data (TSA Seq 2.0) that quantified the distance of genes from speckles and other nuclear compartments (*28*, *29*)(24). We found that even in the absence of pladeinolide B, the genes that localized to compartments post-splicing inhibitor treatment were indeed the closest to speckles. Furthermore the distance from speckles (without splicing inhibitor treatment) was anticorrelated with the signal from TSA (Tyramide Signal Amplification)-Seq 2.0, which measures the distance from all genes to various physical anchors in the nucleus (in this case, speckles). Those data showed that 3 compartmentalized genes were all within the most speckle associated transcripts while all other tested genes (both nuclear dispersal and non-responsive; 11 genes) exhibited a much broader range of distances to speckles (Figure 4H). This anticorrelation is to be expected, because the genes closest to speckles will receive the most reads in TSA-Seq and should have the smallest distance to speckles as measured by IF-FISH.

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 4C), 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 some number of prior studies of the timing 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 (*2*, *18*, *24*, *30–33*). This discrepancy 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 (a thorough review is available in reference (*31*)). 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 (*11*). 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 (*4*). Our model is also in agreement with several other studies which suggest that there can be significant delays between RNA transcription and splicing (*25*, *34–39*). Additionally, several studies have been performed specifically on the products of the Balbiani rings in polytene chromosomes in the salivary gland cells of *Chironomus tentans* and have seen that while the 5' intron is spliced cotranscriptionally, the 3' intron is most often (~90-95%) spliced post-transcriptionally (*33*, *40–42*), suggesting that post-transcriptional splicing of specific introns could be a general biological phenomenon.

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 (*43*). 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 (*44–46*). 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 (*47*). 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 (*48*). 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 (*49*)). 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 note that it is difficult to know which particular 5' and 3' spots correspond to the same pre-mRNA molecule near the site of transcription; further expansion might help make those connections and refine our measurement of distances between 5' and 3' ends of transcripts.

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 (*4*, *50*). 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 (*28*). 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 (*28*). The association between specific genes and speckles has been observed before in several studies by Jeanne Lawrence (*51–53*). Our compartmentalization genes localized close to speckles, and it may be that they have similar speckle localization to *COL1A1* as observed by Shopland *et al.* (*51*) (an example of a Type I gene as per Lawrence's nomenclature). Our observations are also consistent with those of Girard *et al.* in which certain genes are retained post-transcriptionally at speckles as splicing occurs and then are released and immediately exported from the nucleus (*54*). Wang et al. also observed speckle localization of RNA upon microinjection into nuclei (*55*). Our data further support the conclusion that speckle association can be gene specific and can help retain pre-mRNA in the gene's vicinity until post-transcriptional splicing is completed.

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 (*56*). 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 (*56*). These data are consistent with the findings of Hochberg-Laufer et al., who also observed that transcripts can accumulate near the site of transcription, and that the rate of leaving this area depended on the local availability of splicing factors (*57*).

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 (*16*). 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(*58*). 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.* (*20*). 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, NOTE: it has recently been described that this antibody may instead target SON) 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 (1:200) 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 (*58*). 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 (*59*). 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 (*60, 61*). 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)(62) 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 Independence Statistics

For all independence calculations, it was first determined what proportion of RNA fell into each of the following categories:
1. Partially spliced (exon spot appearing with only one of two tested introns)
2. Fully spliced
3. Fully unspliced

Then those numbers were fed into a variety of statistical methodologies to determine essentially whether the splicing status of one intron can tell anything about the splicing status of another intron measured simultaneously.

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. This yielded our theoretical upper bound of 0.137 nats. Higher mutual information would suggest that the knowledge of the splicing status of one intron yields some information of the

splicing status of the other intron (complete correspondence yields a mutual information of 0.137 based on 80% probe detection efficiency). Low mutual information would suggest the opposite, namely that knowledge of the splicing status of one intron says little about the splicing status of the second intron.

Fisher exact test was calculated using the "fisher.test" command from the "stats" R package. Odds ratio was calculated by multiplying the number of partially spliced intermediates and dividing by the number of fully spliced RNA times the number of unspliced RNA.

Comparison to point source diffusion with degradation model

In order to evaluate whether diffusion of pre-mRNA from the site of transcription was compatible with our observations of clouds at the site of transcription revealed by expansion microscopy, we used a model of pre-mRNA diffusion that assumed a source at the site of transcription and a constant rate of degradation. The equation describing this model is given by (*27*):

$$\frac{\partial C}{\partial t} = D\Delta C - \gamma C + R_0 \delta(r),$$

Where C is the concentration as a function of the radial distance and time, D is the diffusion constant, gamma is the degradation rate, R_0 is the rate of production, and \delta (r) is the delta function. The steady state solution in 3D is given by:

$$C(r) = C_0 \exp(\frac{-r}{\lambda}),$$

where C_0 is the concentration at the site of transcription, and lambda is sqrt(D/gamma). We fit this formula to the radial distribution of pre-mRNA density from the region from 3 to 10 microns away from the transcription site, in which region we assume that the pre-mRNA are freely diffusing in the nucleoplasm. Note that our estimate for lambda in this region was roughly 3.34 microns. As a check for numerical consistency, we used an estimate of nucleoplasmic diffusion of 0.034 microns^2/sec (*7*) to obtain a degradation rate gamma of approximately 0.183 per minute, which yields an average lifetime of around 5.46 minutes for the pre-mRNA. This value is in line with what is generally accepted in the field and thus provided validation for the accuracy of our model and fit procedure. In order to see what concentration the diffusion and degradation model would predict at the transcription site itself, we extrapolated to the y-axis (radial distance = 0), yielding an estimated concentration of 18 molecules per cubic micron. We then measured the actual transcription site density at the site of transcription by counting the number of molecules in the vicinity using expansion microscopy and dividing by an estimated transcription site volume, yielding 235 molecules per cubic micron. (The volume was estimated by taking the area of the transcription site spots in 2D, estimating the radius of a putatively circular transcription site, and then using that radius to compute the volume of a 3D sphere.)

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)	Odds Ratio
CPS1	intron 1	intron 13	dependent	0.00014291	0.06061384	0.16641566
CPS1	intron 13	intron 21	independent	0.52542181	0.002127	1.97202797
CPS1	intron 21	intron 34	ambiguous	0.042609	0.00871542	0.47826087
CPS1	intron 34	intron 1	ambiguous	0.02729512	0.01883053	0.35416667
CPS1	intron 13	intron 34	independent	0.24468035	0.00440814	0.58811881
CPS1	intron 21	intron 1	ambiguous	0.00199224	0.01614408	0.37878788

EEF2	intron 1	intron 11	ambiguous	3.38E-05	0.01572976	0.40192029
EEF2	intron 2	intron 1	dependent	3.98E-07	0.0308474	0.28132992
EEF2	intron 11	intron 2	dependent	1.27E-06	0.03756791	0.26775956
TM4SF1	intron 1	intron 4	independent	0.43288281	0.00101288	1.4454776
TM4SF1	intron 3	intron 1	ambiguous	0.00568284	0.02056082	0.34687111
TM4SF1	intron 4	intron 3	independent	0.15205967	0.00181136	1.38855091
FKBP5	intron 1.2	intron 1.1	dependent	6.97E-10	0.05519327	0.25227021
FKBP5	intron 3	intron 10	independent	0.07999788	0.0115184	0.47619048
FKBP5	intron 5	intron 3	ambiguous	0.01954313	0.01823341	0.44331797
FKBP5	intron 8	intron 5	independent	0.23672561	0.00411029	0.55555556
FKBP5	intron 8	intron 1.1	independent	0.3248287	0.00343541	1.44230769
FKBP5	intron 8	intron 9	independent	0.6915974	0.0002986	1.19822347
FKBP5	intron 9	intron 8	independent	0.69304665	0.00035716	0.80300752
FKBP5	intron 10	intron 9	dependent	1.27E-05	0.12801797	0.1025641

** reflects values after 8 hours in dexamethasone

Table 2: Independence, mutual information, odds ratio, 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

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

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

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

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

4. H. L. Drexler, K. Choquet, L. S. Churchman, Splicing Kinetics and Coordination Revealed by Direct Nascent RNA Sequencing through Nanopores. *Mol. Cell* (2019), doi:10.1016/j.molcel.2019.11.017.

5. A. Coulon, M. L. Ferguson, V. de Turris, M. Palangat, C. C. Chow, D. R. Larson, Kinetic competition during the transcription cycle results in stochastic RNA processing. *Elife*. **3** (2014), doi:10.7554/eLife.03939.

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

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

8. H. Tilgner, D. G. Knowles, R. Johnson, C. A. Davis, S. Chakrabortty, S. Djebali, J. Curado, M. Snyder, T. R. Gingeras, R. Guigó, 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 (2012).

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

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

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

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

13. A. Mayer, J. di Iulio, S. Maleri, U. Eser, J. Vierstra, A. Reynolds, R. Sandstrom, J. A. Stamatoyannopoulos, L. S. Churchman, Native elongating transcript sequencing reveals

human transcriptional activity at nucleotide resolution. *Cell*. **161**, 541–554 (2015).

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

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

16. T. Nojima, T. Gomes, A. R. F. Grosso, H. Kimura, M. J. Dye, S. Dhir, M. Carmo-Fonseca, N. J. Proudfoot, Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell*. **161**, 526–540 (2015).

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

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

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

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

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

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

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

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

25. L. Wachutka, L. Caizzi, J. Gagneur, P. Cramer, Global donor and acceptor splicing site kinetics in human cells. *Elife*. **8** (2019), doi:10.7554/eLife.45056.

26. Y. Wan, D. G. Anastasakis, J. Rodriguez, M. Palangat, P. Gudla, G. Zaki, M. Tandon, G. Pegoraro, C. C. Chow, M. Hafner, D. R. Larson, Dynamic Imaging of Nascent RNA Reveals General Principles of Transcription Dynamics And Stochastic Splice Site Selection. *papers.ssrn.com › sol3 › paperspapers.ssrn.com › sol3 › papers* (2019), , doi:10.2139/ssrn.3467157.

27. T. Maire, H. Youk, Molecular-Level Tuning of Cellular Autonomy Controls the Collective

Behaviors of Cell Populations. *Cell Syst*. **1**, 349–360 (2015).

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

29. L. Zhang, Y. Zhang, Y. Chen, O. Gholamalamdari, J. Ma, A. S. Belmont, TSA-Seq 2.0 reveals both conserved and variable chromosomal distances to nuclear speckles (2019), p. 824433.

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

31. T. Alpert, L. Herzel, K. M. Neugebauer, Perfect timing: splicing and transcription rates in living cells. *Wiley Interdiscip. Rev. RNA* (2016), doi:10.1002/wrna.1401.

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

33. G. Baurén, L. Wieslander, Splicing of Balbiani ring 1 gene pre-mRNA occurs simultaneously with transcription. *Cell*. **76**, 183–192 (1994).

34. A. Honkela, J. Peltonen, H. Topa, I. Charapitsa, F. Matarese, K. Grote, H. G. Stunnenberg, G. Reid, N. D. Lawrence, M. Rattray, Genome-wide modeling of transcription kinetics reveals patterns of RNA production delays. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 13115–13120 (2015).

35. S. Hao, D. Baltimore, RNA splicing regulates the temporal order of TNF-induced gene expression. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 11934–11939 (2013).

36. J. Singh, R. A. Padgett, Rates of in situ transcription and splicing in large human genes. *Nat. Struct. Mol. Biol.* **16**, 1128–1133 (2009).

37. M. Rabani, R. Raychowdhury, M. Jovanovic, M. Rooney, D. J. Stumpo, A. Pauli, N. Hacohen, A. F. Schier, P. J. Blackshear, N. Friedman, I. Amit, A. Regev, High-resolution sequencing and modeling identifies distinct dynamic RNA regulatory strategies. *Cell*. **159**, 1698–1710 (2014).

38. L. Windhager, T. Bonfert, K. Burger, Z. Ruzsics, S. Krebs, S. Kaufmann, G. Malterer, A. L'Hernault, M. Schilhabel, S. Schreiber, P. Rosenstiel, R. Zimmer, D. Eick, C. C. Friedel, L. Dölken, Ultrashort and progressive 4sU-tagging reveals key characteristics of RNA processing at nucleotide resolution. *Genome Res.* **22**, 2031–2042 (2012).

39. J. E. Darnell Jr, Reflections on the history of pre-mRNA processing and highlights of current knowledge: a unified picture. *RNA*. **19**, 443–460 (2013).

40. B. Daneholt, Transcription in polytene chromosomes. *Cell*. **4**, 1–9 (1975).

41. B. Daneholt, Assembly and transport of a premessenger RNP particle. *Proc. Natl. Acad.*

Sci. U. S. A. **98**, 7012–7017 (2001).

42. I. Wetterberg, G. Baurén, L. Wieslander, The intranuclear site of excision of each intron in Balbiani ring 3 pre-mRNA is influenced by the time remaining to transcription termination and different excision efficiencies for the various introns. *RNA*. **2**, 641–651 (1996).

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

44. M. de la Mata, C. Lafaille, A. R. Kornblihtt, 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 (2010).

45. M. Yang, J. Wu, S.-H. Wu, A.-D. Bi, D. J. Liao, 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 (2012).

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

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

48. R. D. Alexander, S. A. Innocente, J. D. Barrass, J. D. Beggs, Splicing-dependent RNA polymerase pausing in yeast. *Mol. Cell*. **40**, 582–593 (2010).

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

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

51. L. S. Shopland, C. V. Johnson, J. B. Lawrence, Evidence that all SC-35 domains contain mRNAs and that transcripts can be structurally constrained within these domains. *J. Struct. Biol.* **140**, 131–139 (2002).

52. C. Johnson, D. Primorac, M. McKinstry, J. McNeil, D. Rowe, J. B. Lawrence, Tracking COL1A1 RNA in osteogenesis imperfecta. splice-defective transcripts initiate transport from the gene but are retained within the SC35 domain. *J. Cell Biol.* **150**, 417–432 (2000).

53. K. P. Smith, P. T. Moen, K. L. Wydner, J. R. Coleman, J. B. Lawrence, Processing of endogenous pre-mRNAs in association with SC-35 domains is gene specific. *J. Cell Biol.* **144**, 617–629 (1999).

54. C. Girard, C. L. Will, J. Peng, E. M. Makarov, B. Kastner, I. Lemm, H. Urlaub, K. Hartmuth, R. Lührmann, Post-transcriptional spliceosomes are retained in nuclear speckles until splicing completion. *Nat. Commun.* **3**, 994 (2012).

55. J. Wang, L. G. Cao, Y. L. Wang, T. Pederson, Localization of pre-messenger RNA at

discrete nuclear sites. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 7391–7395 (1991).

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

57. H. Hochberg-Laufer, N. Neufeld, Y. Brody, S. Nadav-Eliyahu, R. Ben-Yishay, Y. Shav-Tal, Availability of splicing factors in the nucleoplasm can regulate the release of mRNA from the gene after transcription. *PLoS Genet.* **15**, e1008459 (2019).

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

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

60. L. Dölken, Z. Ruzsics, B. Rädle, C. C. Friedel, R. Zimmer, J. Mages, R. Hoffmann, P. Dickinson, T. Forster, P. Ghazal, U. H. Koszinowski, High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA*. **14**, 1959–1972 (2008).

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

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

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. Computed spots are calculated by fitting a gaussian to the distribution of intensity signals and pinpointing the center of that gaussian at sub-pixel resolution. Gene depicted is *FKBP5*. C. Example images of dispersed (intron 4) and not dispersed (introns 1 and 3) transcription sites, via RNA FISH. Each image is a max merge of 30 optical Z sections, 0.3µm step distance. D. Graphs indicating dispersal distance or the distance of each detected pre-mRNA to its nearest transcription site. First graphs of CPS1, EEF2, and TM4SF1 data are from HeLa cells, FKBP5 data is from A549 cells after 8 hours in dexamethasone, second graphs of EEF2 and TM4SF1 data are from CRL-2097. 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 for CPS1 before and after expansion microscopy for individual mRNA and transcription sites (in HeLa cells). C. Example images of 5', 3', and middle intron 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. Distance from 5' to 3' of expanded and unexpanded mRNA and transcription sites, as detected by CPS1 RNA FISH. This distance was calculated using nearest neighbors with replacement. All scale bars represent 5 µm unless otherwise noted. All expanded images and calculations are shown in expanded space scale (reduce by 4.65X to yield original space scale).

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). FKBP5 is not present on this graph due to low read counts in HeLa cells (FISH was performed for FKBP5 in A549 cells). C. Comparison of mean dispersal (as assessed by RNA FISH, see figure 1) and mean spliced reads over total (as assessed by sequencing). FKBP5 is not in this graph due to low read counts in HeLa cells.

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 (treatment status as indicated on figure). Outline represents nucleus. C. Description of 3 response types to pladienolide B. D. EEF2 RNA FISH exon and intron images with pladeinolide B treatment. Scale bar = 5um. E. EEF2 RNA FISH intron only images with pladeinolide B treatment. Scale bar = 5um. F. RPL13A, GAPDH, and EEF2 RNA FISH intron images with pladeinolide B treatment. Scale bar = 5um. G. Costaining of RNA FISH for specified intron and SC35 IF in the same cell treated with pladeinolide B. Scale bar = 5um. H. Quantification of speckle decile from previously published data (TSA Seq 2.0), compared with distance from nearest speckle calculated based on RNA FISH of specified genes without pladeinolide B treatment.

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.
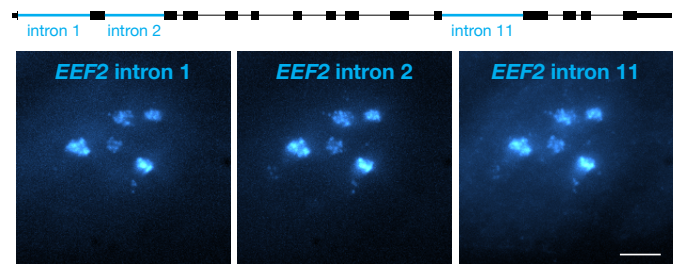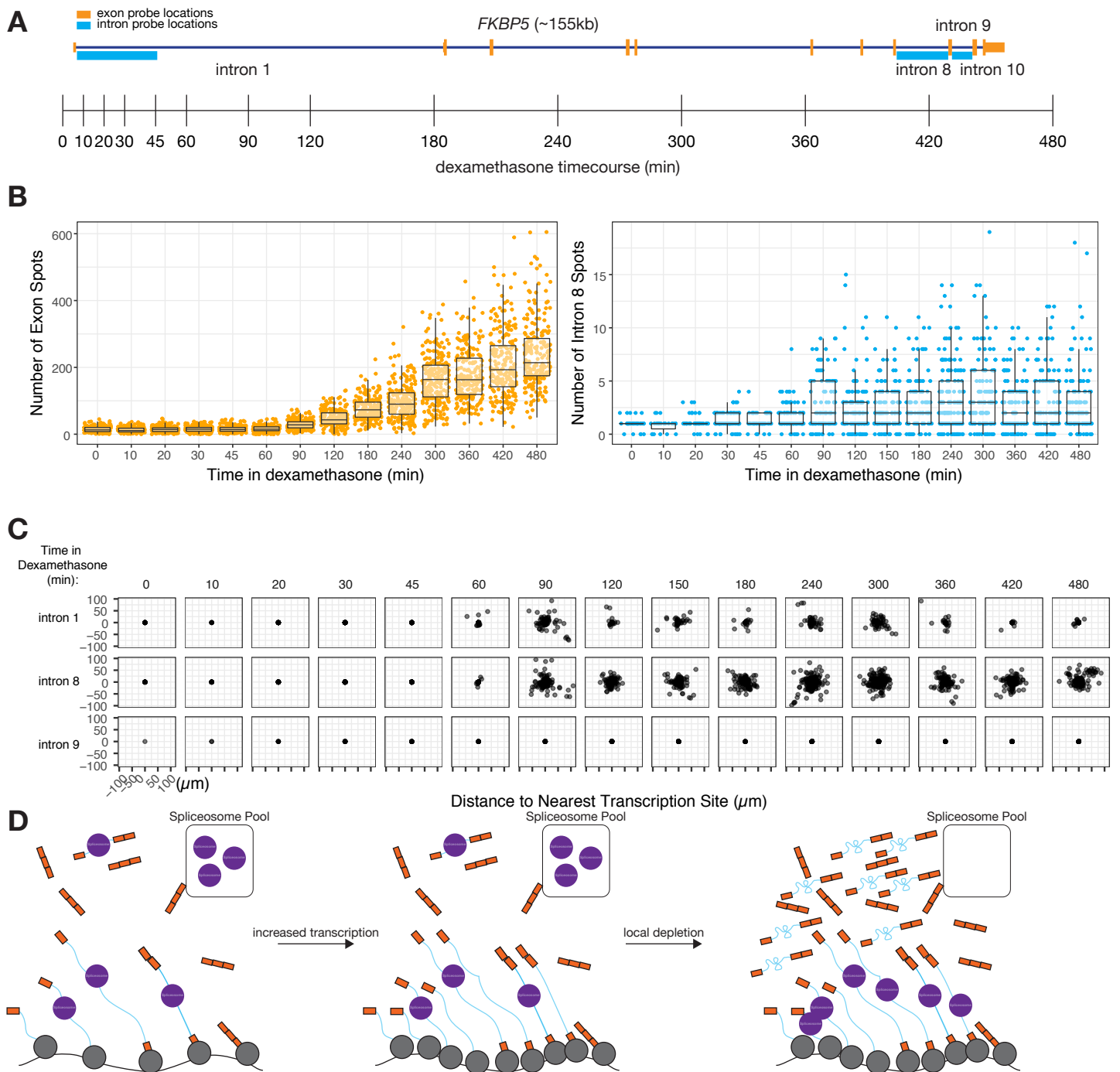
Supplementary Figure 1: Transcription site choice and defining post-transcriptionality.
A. Histograms of intron intensities for TM4SF1 intron 1 and intron 4, before, after, and during defining a global thresholding cutoff and other transcription site selection methods, as well as dispersal graphs generated based on those transcription site selections.
B. Number of dispersed pre-mRNA per cell.
C. Parameter sweep of colocalization threshold and associated dispersal graphs.
D. Example classification of pre-mRNA as either transcription sites or dispersed pre-mRNA.
E. 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.
F. Example classification of RNA as "distal" or "proximal" to the transcription site
G. 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 4.65 fold linear expansion and expands isotropically. A. Compari- son of radii of nuclei (based on DAPI staining, and assuming spherical nuclei) before and after expansion. B. Images of the same cell before and after expansion, with or without pladeinolide B treatment (as noted). Scale bars = 5µm. C. Transcription site area (microns squared) for 5' and 3' probes, with and without dye swap.
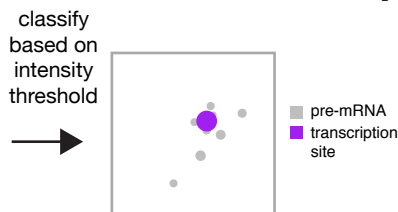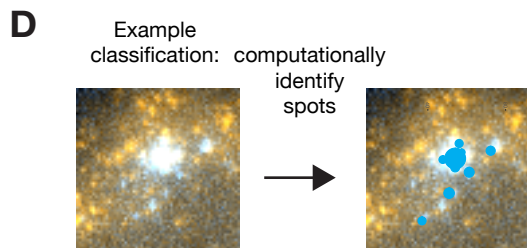
Supplementary Figure 3: Splicing index varies on a per intron basis. A. Splicing index of each intron for which we obtained RNA FISH measurements. Error bars represent mean +/- sd. n = 2
B. Comparison of transcription site size (as assessed by RNA FISH) and splicing index, as assessed by sequencing.

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

Figure 2

Figure 3

A

splicing index = $\dfrac{\text{spliced reads}*2}{\text{5' unspliced reads + 3' unspliced reads}}$

spliced reads over total = $\dfrac{\text{spliced reads}}{\text{5' unspliced reads + 3' unspliced reads + spliced reads}}$

B

C

Figure 4

A Potential Pre-mRNA Behaviors after Leaving Transcription Site
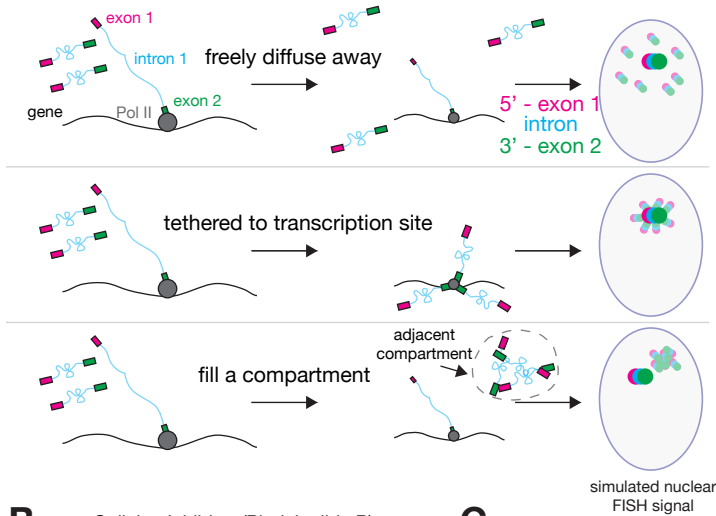
B Splicing Inhibitor (Pladeinolide B)
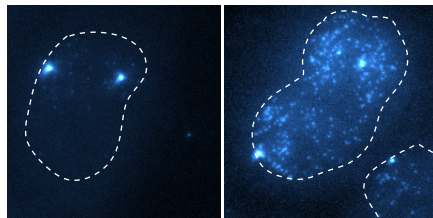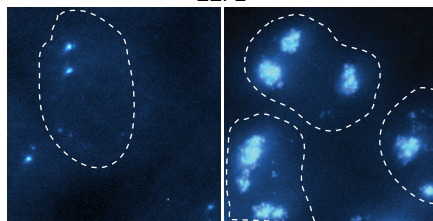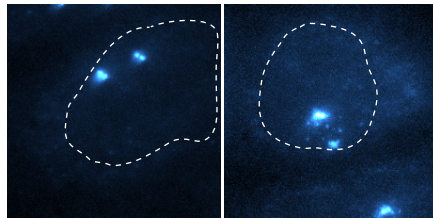
C Splicing Inhibitor Response Categories:
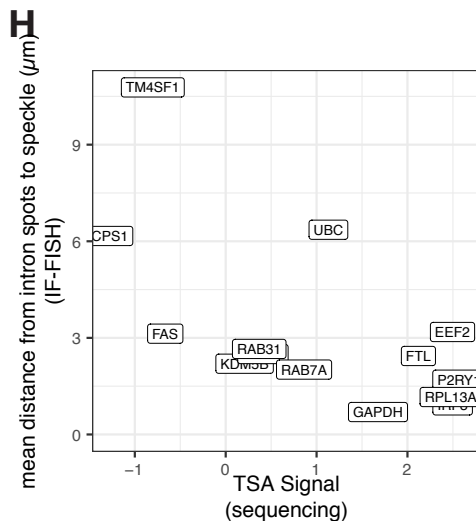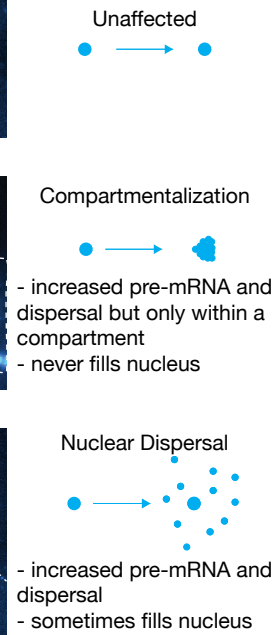
D *EEF2* exon and intron probe positions
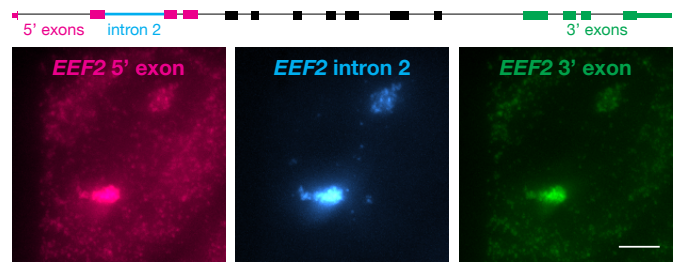
E *EEF2* multiple intron probe positions

F

G

H
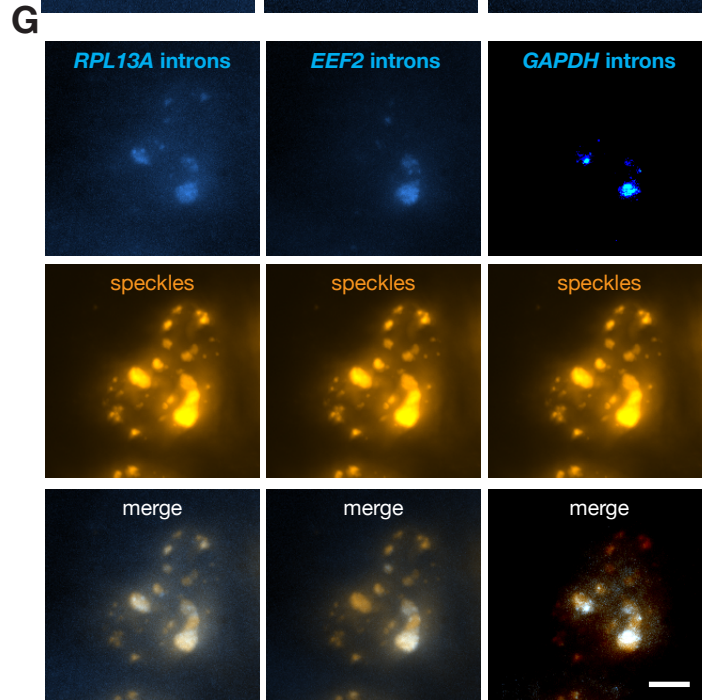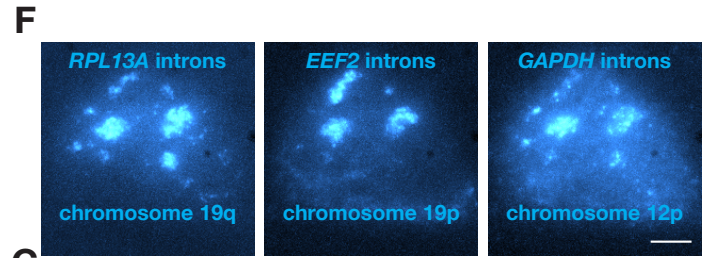
Figure 5

Supplementary Figure 1

A

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. Number of dispersed pre-mRNA per cell.
C. Parameter sweep of colocalization threshold and associated dispersal graphs.
D. Example classification of pre-mRNA as either transcription sites or dispersed pre-mRNA.
E. 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.
F. Example classification of RNA as "distal" or "proximal" to the transcription site
G. 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.
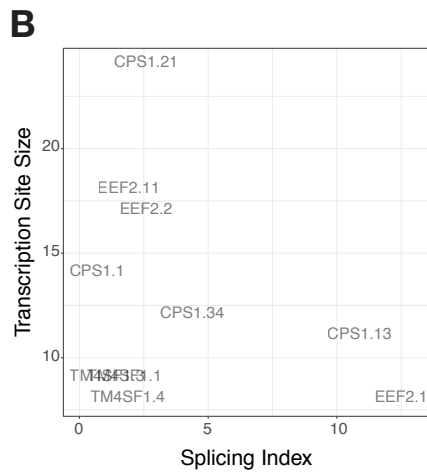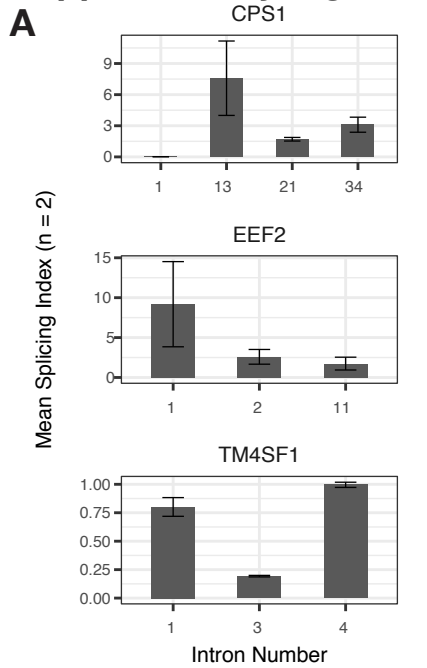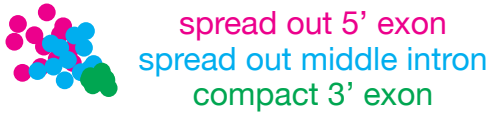
Supplementary Figure 1

Supplementary Figure 2
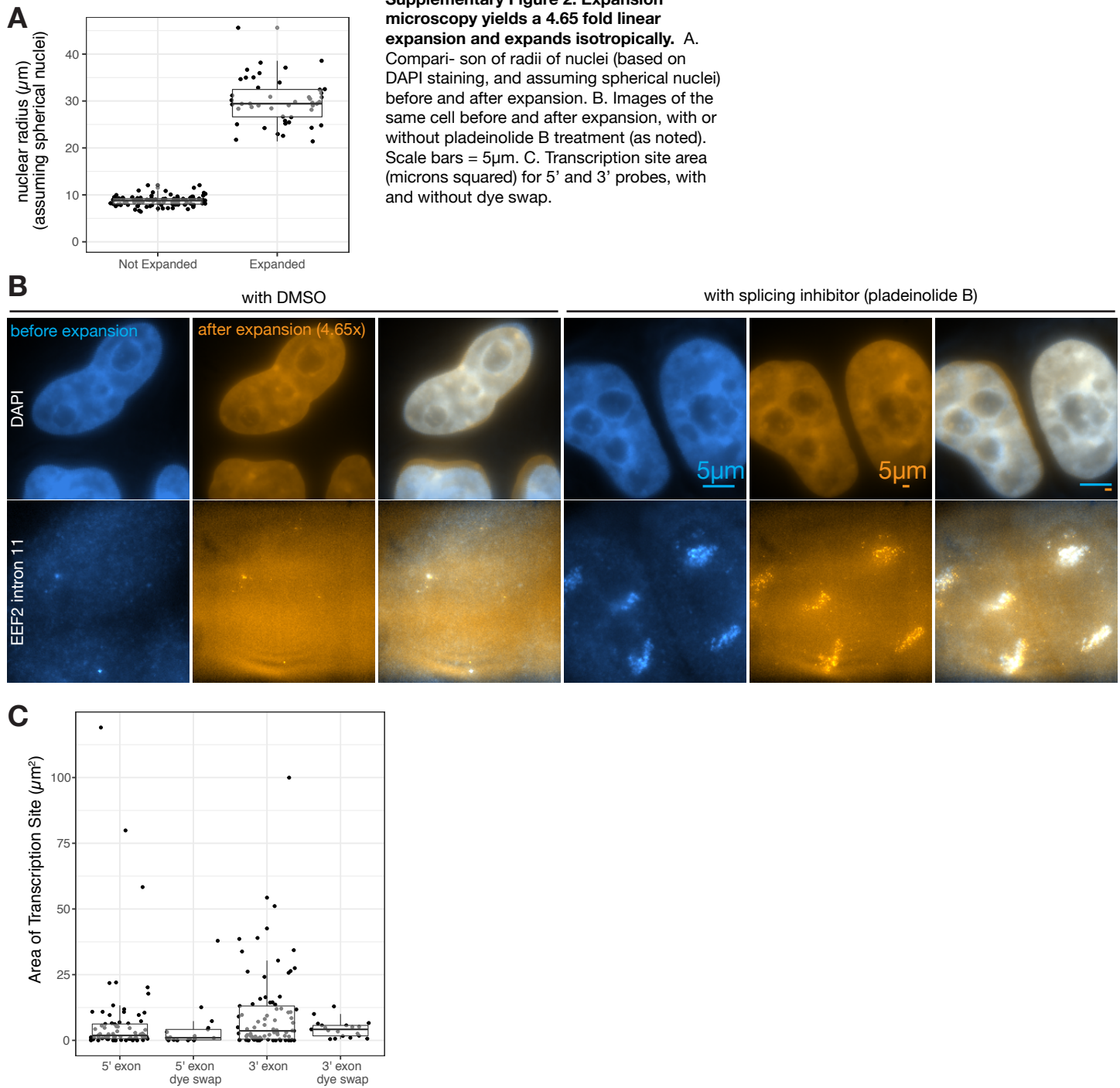
A

Supplementary Figure 2: Expansion microscopy yields a 4.65 fold linear expansion and expands isotropically. A. Compari- son of radii of nuclei (based on DAPI staining, and assuming spherical nuclei) before and after expansion. B. Images of the same cell before and after expansion, with or without pladeinolide B treatment (as noted). Scale bars = 5μm. C. Transcription site area (microns squared) for 5' and 3' probes, with and without dye swap.

B

C
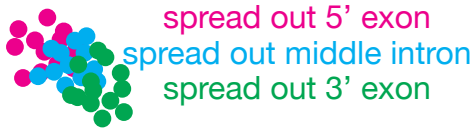
Supplementary Figure 3

A

B

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.
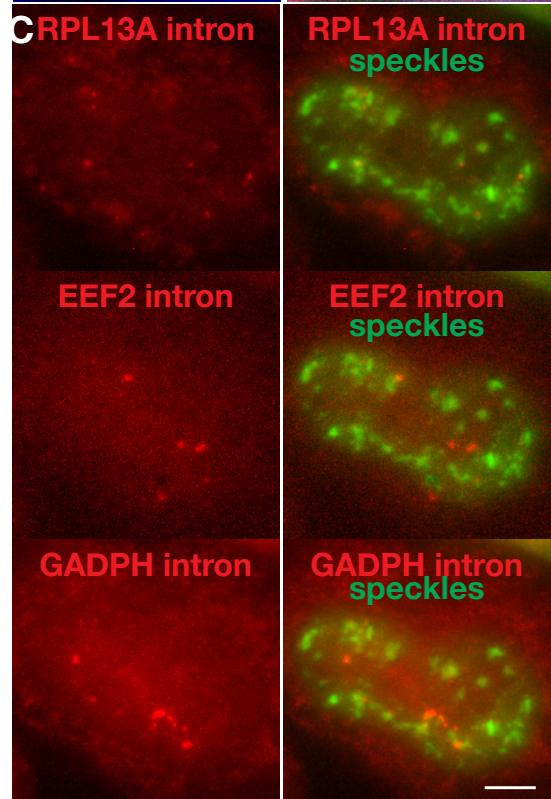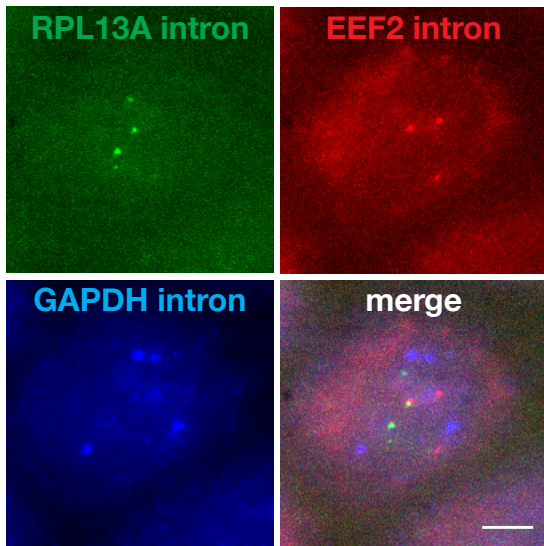
Supplementary Figure 4

A Compartment with 3' tether:

spread out 5' exon
spread out middle intron
compact 3' exon

Compartment without tether:

spread out 5' exon
spread out middle intron
spread out 3' exon

B

C

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.