Combined Manuscript File Click here to view linked References

Post-transcriptional splicing can occur in a slow-moving zone around the gene

Allison Coté[1]*, Aoife O'Farrell[1]*, Ian Dardani[1], Margaret Dunagin[1], Chris Coté[1], Yihan Wan[2], Sareh Bayatpour[1], Heather L. Drexler[3], Katherine A. Alexander[4] , Fei Chen[5] , Asmamaw T. Wassie[6], Rohan Patel[1], Kenneth Pham[6], Edward S. Boyden[7], Shelley Berger[4], Jennifer Phillips-Cremins[1], L. Stirling Churchman[3], Arjun Raj[1]#

1 Department of Bioengineering, University of Pennsylvania, Philadelphia, PA 19104
2 Westlake University, School of Life Sciences, Hangzhou, Zhejiang, China
3 Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, MA 02115
4 Department of Cell and Developmental Biology, Penn Institute of Epigenetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104
5 Broad Institute of MIT and Harvard, Cambridge, MA 02142
6 Department of Cell and Molecular Biology, University of Pennsylvania, Philadelphia, PA 19104
7 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 stepwise molecular process by which introns are removed from pre-mRNA and exons are joined together to form mature mRNA sequences. The ordering and spatial distribution of these steps remain controversial, with opposing models suggesting splicing occurs either during or after transcription. We used single molecule RNA FISH, expansion microscopy, and live cell imaging to reveal the spatiotemporal distribution of nascent transcripts in mammalian cells. At super-resolution levels, we found that pre-mRNA formed clouds around the transcription site. These clouds indicate the existence of a transcription site proximal zone through which RNA move more slowly than in the nucleoplasm. Full length pre-mRNA undergo continuous splicing as they move through this zone following transcription, suggesting a model in which splicing can occur post-transcriptionally but still within the proximity of the transcription site, thus seeming co-transcriptional by most assays. These results may unify conflicting reports of co-transcriptional versus post-transcriptional 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 (1-4), while other work suggests that many pre-mRNAs are fully transcribed before splicing occurs (5-9). 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) (7, 10-11).

In lieu of direct visualization, many studies have used biochemical fractionation to infer the location of various intermediates (6, 12-17). Fractionation methods separate cellular RNA into different compartments, such as the putatively chromatin associated RNA, nucleoplasmic RNA, and cytoplasmic RNA (13, 18). 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 (19). 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 (20). 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 (7, 10-11, 21-23). Imaging using probes targeting both exonic and intronic regions of RNA has revealed bright nuclear foci that represent nascently transcribing RNA (11, 23). 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 (11). 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 (“proximal post-transcriptional splicing”). Indeed, recent live-cell imaging methods showed that the splicing of a reporter gene is 85% post-transcriptional (7), 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 (24-25). 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 and live cell imaging to demonstrate that newly synthesized RNA dwell and undergo continuous splicing near the site of transcription after transcription is complete within a proximal slow-moving zone. 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

The extent to which splicing is coupled to transcription, both in space (distance from transcription site) and time (time since transcription has begun) has been difficult to measure. To address this question, we directly visualized the locations of spliced and unspliced RNA relative to the site of transcription in situ using single molecule RNA FISH (26-27), 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) and fully spliced products (represented by exon spots alone).

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) 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 <u>at</u> the site of transcription and 2) an intron from a pre-mRNA being spliced out <u>away from</u> 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 must 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-C, 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, suggesting splicing patterns may be influenced by cellular context. The presence of 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., transcripts with retained introns cannot be exported from the nucleus). 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).

As a corroboration of RNA FISH dispersal as a metric to quantify distal post-transcriptional splicing, we turned to nascent RNA sequencing, using 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 (6) (Supplementary Figure 2). We found a modest association between splicing index and both mean dispersal and transcription site size, suggesting that sequencing of nascent RNA was at least consistent with RNA FISH assessments of distal post-transcriptional splicing; however, more work is required to fully establish a quantitative correspondence between the metrics.

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 tracking individual mRNA throughout their lifetime 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 (28-29). 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

In these experiments, we observed individual pre-mRNA with some but not all introns retained, and wondered whether these introns could be spliced in any order. Thus, we quantified the number of transcript intermediates that contained different introns in the same RNA molecule. For each RNA containing a pair of introns A and B, there are two alternative routes to generate a fully spliced mature mRNA: either intron A is spliced before intron B, or intron B is spliced before intron A. Were the splicing of these introns ordered, for instance if B always gets spliced after A, then we would expect the vast majority of partially spliced transcripts to contain intron B only, and that very few transcripts would contain intron A without intron B. We thus identified and counted partially

spliced transcripts which contain only one intron without the other (Table 2). To estimate the confidence in our measurements, we labeled the fourth intron of *TM4SF1* with two colors; in this case, any transcripts detected as containing one color without the other must be due to technical reasons, thereby setting a “noise floor” (see methods for details). The dye used to label each intron also did not affect these measurements (*FKBP5* introns 8 and 9 dye-swap, rows 13 and 14 of Table 2). It was also difficult for us to discriminate between models if both introns are spliced so rapidly that intermediates are not observed.

We observed a variety of patterns of splicing intermediates, some showing strong ordering (94% of intermediate transcripts of *FKBP5* contain intron 8 without intron 5) and some showing a lack of ordering (50.3% of intermediate transcripts of *EEF2* contain intron 2 without intron 1). Lack of ordering between introns did not depend on distance between them, with no ordering observed for *EEF2* introns 1 and 2 as well as the distant introns *CSP1* introns 1 and 34.

Overall, our data support a model in which splicing of different introns within the same primary transcript can occur in any order.

Transcripts are untethered to the transcription site and move through a slow-moving transcription site proximal zone where splicing occurs 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 (co-transcriptional) or after the completion of transcription but before the pre-mRNA moves away from the site of transcription (proximal post transcriptional). 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). It is therefore possible that lower mobility of transcripts near the transcription site may lead to the classification of proximal post transcriptional splicing as co-transcriptional when using low-resolution imaging methods.

We thus used expansion microscopy to physically expand the transcription site (by around 4.6x, Supplementary Figure 3) 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) (24-25). 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 (24), 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 3B, displaying the same cells before and after expansion). To further ensure that expansion microscopy did not alter the morphology of the transcription site, we confirmed that the transcription sites in expanded samples were the same size as those imaged using stochastic optical reconstruction microscopy (STORM) (30), a superresolution imaging technique that does not rely on expansion (Supplementary Figure 4A-B).

Lower mobility of transcripts near the site of transcription may lead to a buildup of multiple transcripts in the vicinity of the transcription site. Upon expanding, labeling, and imaging, we indeed observed that the 5', 3', and middle intron probe signals formed small clouds where we had previously observed the tight transcriptional focus (Figure 2C, quantified in Supplementary Figure 4C). Of note, the expanded transcription site contains multiple 3' and 5' spots, while expanded cytoplasmic mRNA contain one 3' and 5' spot, indicating the presence of multiple transcripts at the transcription site as opposed to a single transcript broken into multiple parts as a result of expansion (Figure 2B). The variable brightness between spots in expanded versus unexpanded samples may have been due to variable lengths of transcribing RNA or RNA fragmentation during expansion, as each RNA is likely linked to the gel in multiple locations.

The presence of these clouds precluded any model of immediate, free movement of the transcript away from the RNA polymerase upon completion of transcription. If transcripts freely diffused away right after their transcription, then a mathematical model would predict an exponential decrease in pre-mRNA concentration with increasing distance away from the transcription site (31) (see methods for details, Figure 2D scenario II). 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 (Figure 2E). This high concentration in the "cloud" is not consistent with free diffusion subsequent to the completion of transcription (Figure 2D scenario II), 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 (Figure 2D scenario I).

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 (Figure 2D scenario III). 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 (Figure 2D scenario II). We observed similarly-sized clouds for both the 5' and 3' ends of the pre-mRNA in the proximal zone. Since RNA polymerase II can transcribe beyond the poly-A site (32-33), it may be possible that observed clouds are a result of transcripts tethered beyond the 3' poly-A site of the gene body. To eliminate this possibility, we performed RNA FISH for a sequence 2kb beyond the 3' poly-A site of *CSP1*. We found that the 3' and post 3' transcription sites were also approximately the same size, suggesting these transcripts are not tethered to chromatin as splicing is occurring (Supplementary Figure 4D-F).

Based on the size of these clouds, we estimated the diameter of this cylindrical slow-moving zone around the gene body to be around 0.3 microns (Figure 2F, see methods for calculations). This width does not depend on the length of an individual transcript or gene body, but rather extends radially outward from the gene body.

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 2G, H). 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 2I).

We next wanted to confirm the presence of this slow-moving zone in a live cell as transcription is actively taking place, as this method does not require manipulation through fixation or expansion. We obtained data from cells in which the 3' UTR of the gene *TFF1* was labeled with MS2-GFP (34), allowing us to observe the geometry of transcription sites in living cells. If transcripts immediately and rapidly diffused away from the transcription site after synthesis at the same diffusion rate as in the nucleoplasm, then the 3' end of the transcripts would form a tight, diffraction limited spot at the site of transcription, much like the 3' ends of individual transcripts we detected by mRNA FISH (Figure 2B). (Note that the limit of detection of MS2-GFP was such that only transcription sites, presumably with multiple transcripts on them, would be detectable; hence, individual mRNA were not detectable.) If, on the other hand, there were a gene-proximal zone in which diffusion was slower, one would expect to see primary transcripts at a higher concentration within the zone (see Figure 2E). This higher concentration would translate into a cloud of primary transcripts visible as a blob that was larger than the diffraction limit, reflecting the size of the putative slow moving zone. Thus, the presence of transcription site spots larger than the diffraction limit would be an indicator of the presence of a cloud of primary transcripts and hence a slow-moving zone around the site of transcription.

It is challenging to rigorously discriminate between a diffraction limited spot and a blob with larger spatial extent based solely on the area of the fluorescent region because the thresholds chosen in determining area can be somewhat arbitrary, and the underlying data is often noisy. Hence, we instead used eccentricity (a measure of deviation from a perfect circle, ranging from 0 to 1) as a means to distinguish diffraction limited spots from blobs, reasoning that blobs larger than the diffraction limit could potentially be asymmetric, thus exhibiting increased eccentricity as compared to pure diffraction limited spots. Visually, several images showed evidence for such asymmetry (Figure 3A-C). We quantified the asymmetry by measuring the eccentricity for 205 transcription site images from 36 total temporal transcription site tracks (Figure 3D). We found that the majority of images (141 of 205) showed evidence of substantial eccentricity (defined as eccentricity > 0.6) (Figure 3C). As these sites are unambiguously larger than a diffraction limited spot, we concluded that RNA were likely retained at the site for some time following transcription, corroborating the existence of the slow-moving zone.

It should be noted that not all transcription site images appeared eccentric. Possible reasons are that the transcription burst duration was short enough that there was not time for a buildup of pre-mRNA in the slow moving zone, or that the blob happened to grow in a symmetric fashion, or that the three-dimensional orientation of the asymmetry was such that it was not detectable in the single x-y plane of the microscope.

Similarly, in our expansion microscopy data, 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). The length of each intron, as well as the strength of its 3' and 5' splice site, did not affect its dispersal from the transcription site, suggesting that transcript size and splice site sequence are not major contributors to the dynamics of the slow moving zone (Supplemental Figure 5, Supplementary Table 2). 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 are not being actively transcribed (Figure 2G). 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 5', intron, and 3' clouds, the number of spots in each cloud, and the area of live cell transcription sites varied considerably from cell to cell (Figure 2C, 2H, 3A-B). 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 3C).

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. This finding potentially unites existing conflicting reports of co-transcriptional versus post-transcriptional splicing, as splicing within the slow moving zone would appear co-transcriptional when evaluated by conventional light microscopy but in reality occurs after transcription has been completed.

Localization of unspliced pre-mRNA to speckle-proximal compartments is gene specific

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: (Figure 4A)

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).

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. 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*), whereas we observed a nuclear dispersal phenotype for 7 genes and no change in dispersal for 6 genes. To test for potential tethering of transcripts to some location in the nucleus, 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 6A). We found, however, that the 5'

and 3' ends of the pre-mRNA both filled the entire blob, suggesting that the pre-mRNA spread to fill the entire putative compartment (as in scenario 3 described above, compartmentalization) (Figure 4A, 4D).

We wondered if these compartmentalized pre-mRNA were located near nuclear speckles, which are compartments in the nucleus that contain concentrated splicing and transcription factors (21). To test this hypothesis, 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 6C) and after (Figure 4E) 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 (35-36). 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 4F). 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 gene-specific subset. 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 4G), 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 4H), 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 6C), suggesting perhaps that splicing inhibition impacts these responses. Overall, our results suggest that mRNA trafficking upon leaving the slow-moving zone is gene specific.

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

We wondered whether the degree of dispersal (and thus, the degree of post-transcriptional splicing) was an inherent property of each intron or whether 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 represents 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, suggests 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 hypothesized that the increased dispersal of some introns during periods of increased transcription may be due to a local depletion of splicing factors. 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. If the concentration of splicing factors remained constant (i.e., no local depletion), we would expect the rate of splicing to remain constant, and the percentage of unspliced pre-mRNA far from the transcription site to be unchanged, regardless of transcription rate. Instead, we observed a significant increase in the proportion of intron signal for two introns over time (from 0 to 22% for intron 1 and from 0 to 29% for intron 8; Figure 5D), suggesting a possible local depletion of splicing factors at the transcription site. (Of note, we may potentially miss post-transcriptional splicing occurring after a burst has concluded.) 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

By using expansion microscopy combined with single molecule RNA FISH, we were able to obtain direct measurements of the spatial distribution of splicing intermediates in the vicinity of the transcription site. We found evidence for extensive post-transcriptional splicing that occurs throughout that zone. Our results support a model that unifies many of the observations made of both co-transcriptional and post-transcriptional splicing. We argue that splicing can occur continuously and post-transcriptionally while pre-mRNA move through this slow moving zone. Fractionation approaches may have interpreted such splicing intermediates as arising co-transcriptionally owing to proximity to the transcription site. We also found introns that seemed to be spliced out so close to the site of transcription that we were unable to distinguish whether they were spliced co- or post-transcriptionally. As methods are developed with ever increasing spatio-temporal resolution, it may soon be possible to discriminate splicing dynamics for these introns as well.

Our results suggest that pre-mRNA are not tethered to the site of transcription while they move through this transcription-site-proximal compartment, 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 (37). 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 (50). 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 (39)). 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 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, 4, 22, 28, 40-43). 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 (41)). Of note, we see that increased transcription level correlates with intron dispersal, suggesting that the percentage of splicing occurring away from the transcription site is regulated by transcription level for at least some introns. This may explain why we observe post-transcriptional splicing of all genes we measured, as all were highly expressed. Our model is in agreement with fractionation work by Pandaya-Jones et al. and by Drexler et al. that suggest that specific transcripts are retained in the chromatin and that splicing is not completed until transcription has finished (6, 15). Our model is also in agreement with several other studies which suggest that there can be significant delays between RNA transcription and splicing (29, 44-49).

What might the slow moving zone represent, physically? A natural candidate is the chromatin in the vicinity of the gene itself. Live imaging studies have shown that mRNA exhibit slow diffusion in chromatin-dense regions (50). Furthermore, recent work tracking cells over time by sequencing shows that mRNA spends relatively more time on chromatin than in the nucleoplasm (51), further lending credence to this hypothesis.

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 due to progressive splicing over time. If this 5' to 3' distance increase were due to increased retention of introns, that would also support the post-transcriptional splicing model. 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 or live tracking of individual mRNA might help make those connections and refine our measurement of distances between 5' and 3' ends of transcripts.

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 (52). 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 (53-55). 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 (6, 9, 56).

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 (35). The association between specific genes and speckles has been observed before in several studies by Jeanne Lawrence (57-59). 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 (60). Wang et al. also observed speckle localization of RNA upon microinjection into nuclei (61). 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.

Study Limitations

With the notable exception of our live cell data in Figure 3, the experiments in this study utilize fixed cells and capture transcription at a single moment in time. As a result, many of the conclusions we draw can be strengthened by future work that evaluates active transcription through live cell imaging or other methods that do not involve fixation. Our investigation of intron localization in Figure 4 relies upon the inhibition of splicing, which may impact several aspects of cellular physiology. This paper also does not address potential relationships between splicing and other co-transcriptional events, such as 5' capping or 3' end formation. Future work may unravel dependencies between these events and the splicing of introns, and a potential role for the transcription-site-proximal slow moving zone in other aspects of mRNA processing. Lastly, a major outstanding question is what sequence-based features determine the distinct behaviors of different introns. Large-scale imaging and synthetic libraries of introns may be required to make such conclusions.

Acknowledgements

We would like to thank members of the Raj, Churchman, Phillips-Cremins, and Berger labs for critical reading of the manuscript. We thank Hyun Youk for discussions about radial distributions. AR and JPC acknowledge support from NIH 4DN U01 HL129998, NIH 4DN U01DK127405 and NSF EFMA19334000. AR additionally acknowledges R01 CA238237, NIH Director's Transformative Research Award R01 GM137425, R01 CA232256, NSF CAREER 1350601, P30 CA016520, SPORE P50 CA174523, NIH U01 CA227550, NIH Center for Photogenomics (RM1 HG007743), and the Tara Miller Foundation. AJC was also supported by NIH training grant T32 GM- 07229. AO acknowledges support from the NSF-GRFP. LSC acknowledges support from R21-HG009264 and R01-GM117333, and F31-GM122133 to HLD. SB acknowledges support from 3R01CA078831-20S1 and KA acknowledges support from 5F32CA221010-02.

Author Contributions

AJC and AR conceived of the research with preliminary input and method development by SB. AJC and CC designed and performed experiments with guidance from AR. AJC, AO, and CC analyzed 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 the design of speckle experiments with guidance from SB. AJC and AO wrote the paper with guidance from AR. FC, AW, and AJC conducted initial optimization of expansion microscopy experiments with guidance from EB. YW conducted the live cell imaging experiments. ID, MD, RP, KP, and JPC conducted STORM imaging.

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 simulated RNA FISH signal for three potential scenarios: a slow moving zone around the site of transcription (1), free diffusion away from the transcription site (2), or co-transcriptional splicing of tethered transcripts.
E. Schematic depicting RNA FISH signal for diffusion or a slow moving proximal zone for completed (but not necessarily spliced) transcripts
F. Line scans of fluorescence intensity (arbitrary units) from transcription sites like in C.
G. 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.
H. Graph depicting area (in pixels) of polygons drawn around individual transcription sites after expansion.
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: Live cell imaging corroborates a slow moving zone around the site of transcription
A-B. Representative images from two cells of a transcriptional burst of TFF1 labeled with MS2-GFP. Each image is taken one second apart. Shown are the raw, computationally identified transcription site images, as well as the binarized image used to calculate eccentricity.
C. Plot of eccentricity for each transcription site image, alongside representative images that show both eccentric and non-eccentric sites.
D. Graph of eccentricity measured in individual transcription sites tracked over time

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. Costaining of RNA FISH for specified intron and SC35 IF in the same cell treated with pladeinolide B. Scale bar = 5um.
F. 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.
G. EEF2 RNA FISH intron only images with pladeinolide B treatment. Scale bar = 5um.
H. RPL13A, GAPDH, and EEF2 RNA FISH intron images with pladeinolide B treatment. Scale bar = 5um.

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. Graph showing the percentage of intron intensity that is distally post-transcriptional for 3 *FKBP5* introns over time in dexamethasone

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: 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. Splicing index of each intron for which we obtained RNA FISH measurements. Error bars represent mean +/- sd. n = 2
D. Comparison of transcription site size (as assessed by RNA FISH) and splicing index, as assessed by sequencing.

Supplementary Figure 3: Expansion microscopy yields a 4.65 fold linear expansion and expands isotropically.
A. Comparison 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 4: Storm Validation and Post 3' RNA FISH
A. Sample STORM image to demonstrate identification of transcription sites

B. Comparison of *CSP1* transcription site size in cells imaged through expansion microscopy and STORM. Distances are in unexpanded space
C. Quantification of transcription site clouds shown in Figure 2C
D. Gene and probe position diagram for CPS1 pre-mRNA to test for a potential tether beyond the 3' end of the gene
E. Representative images of 3' and post-3' RNA FISH signal at transcription sites (in unexpanded cells)
F. Comparison of 3' and post-3' RNA FISH signal at transcription sites

Supplementary Figure 5: Intron characteristics do not influence dispersal
A. Comparison of intron size versus mean dispersal by RNA FISH
B-C. Comparison of 3' splice site (X3SS, B) and 5' splice site (X5SS, C) scores versus mean dispersal by RNA FISH.

Supplementary Figure 6: 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.

STAR Methods:

Key Resources Table:

Reagent or Resource	**Source**	**Identifier**
Antibodies		
SC35	Abcam	ab11826
Phospho S2 polymerase II	Active Motif	61083
Cell Lines and Cell Culture		
HeLa	Lab of Dr. Phillip Sharp (MIT)	N/A
HeLa-S3	ATCC	CCL-2.2
A549	ATCC	CCL-185
CRL-2097	ATCC	CCD-1079Sk
DMEM	Gibco	11995040
Penicillin / Streptomycin	Invitrogen	15140122
Fetal Bovine Serum (FBS)	Fisher	16000044
Pladienolide B	Tocris Biosciences	6070500U
Dexamethasone	Sigma	D2915
4sU (4-thiouridine)	Sigma	T4509
Oligonucleotides		
Single molecule RNA FISH probes targeting CSP1, FKBP5, TM4SF1, and EEF2	See Table S1	N/A
Software and Algorithms		
rajlabimagetools	Raj et al, 2008	https://github.com/arjunrajlaboratory/rajlabimagetools
STAR (v2.5.1a)	Dobin et al, 2013	N/A
All data and code used to produce figures in this paper	This paper	https://www.dropbox.com/sh/el816d3aqdgrh70/AAD0_uTOqjSpZTi1md0cKRKta?dl=0

Resource Availability:

Lead Contact: Further information and requests should be directed to and will be fulfilled by the lead contact, Arjun Raj (arjunrajlab@gmail.com)

Materials Availability: This study did not generate new materials.

Method Details:

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, STORM, 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. Chromatic aberration was evaluated using multi-color fluorescent beads and was far smaller than any biologically relevant distances we measured (see Supplementary Figure 1E).

Stochastic optical reconstruction microscopy (STORM) was performed on the 5' exons and middle intron of CSP1 in HeLa cells. Cells were imaged in Vutara d-STORM imaging buffer (20mM cysteamine (MEA) + 1% 2-mercaptoethanol + 1x Gloxy (glucose oxidase + catalase dissolved in 50mM Tris-HCl + 10mM NaCl) in buffer B (50mM Tris-HCl + 10mM HCL + 10% glucose). Samples were imaged on a Vutara VXL microscope and exported points were reconstructed in Matlab. Transcription sites were identified by alignment with widefield images, and area was calculated using the polyarea() function in R.

For combined expansion microscopy and RNA FISH, expansion microscopy was performed as described by Chen et al. (20). Briefly, Acryloyl-X, SE (6-((acryloyl)amino)hexanoic acid, succinimidyl ester, here abbreviated AcX; Thermo-Fisher) was resuspended in anhydrous DMSO at a concentration of 10 mg/mL, aliquoted and stored frozen in a desiccated environment. Label-IT ® Amine Modifying Reagent (Mirus Bio, LLC) was resuspended in the provided Mirus Reconstitution Solution at 1mg/ml and stored frozen in a desiccated environment. To prepare LabelX, 10 µL of AcX (10 mg/mL) was reacted with 100 µL of Label-IT ® Amine Modifying Reagent (1 mg/mL) overnight at room temperature with shaking. LabelX was subsequently stored frozen (−20 °C) in a desiccated environment until use.

Fixed cells were washed twice with 1× PBS and incubated with LabelX diluted to 0.002 mg/mL in MOPS buffer (20 mM MOPS pH 7.7) at 37 °C for 6 hours followed by two washes with 1× PBS.

Monomer solution (1x PBS, 2 M NaCl, 8.625% (w/w) sodium acrylate, 2.5% (w/w) acrylamide, 0.15% (w/w) N,N'-methylenebisacrylamide) was mixed, frozen in aliquots, and thawed before use. Prior to embedding, monomer solution was cooled to 4°C to prevent premature gelation. Concentrated stocks (10% w/w) of ammonium persulfate (APS) initiator and tetramethylethylenediamine (TEMED) accelerator were added to the monomer solution up to 0.2% (w/w) each. 100uL of gel solution specimens were added to each well of a Lab Tek 8 chambered coverslip and transferred to a humidified 37° C incubator for two hours.

Proteinase K (New England Biolabs) was diluted 1:100 to 8 units/mL in digestion buffer (50 mM Tris (pH 8), 1 mM EDTA, 0.5% Triton X-100, 0.8 M guanidine HCl) and applied directly to gels in at least ten times volume excess. The gels were then incubated in digestion buffer for at least 12 hours. Gels were then incubated with wash buffer (10% formamide, 2× SSC) for 2 hours at room temperature and hybridized with RNA FISH probes in hybridization buffer (10% formamide, 10% dextran sulfate, 2× SSC) overnight at 37 °C. Following hybridization, samples were washed twice with wash buffer, 30 minutes per wash, and washed 4 times with water, 1 hr per wash, for expansion. Samples were imaged in water with 0.1ug/mL DAPI.

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).

Intron splice site strength (Supplementary Figure 5, Supplementary Table 2) was calculated using MaxEnt (62).

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 4ºC. 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.

Live Cell Imaging

The live cell imaging was conducted using human bronchial epithelial cells (single cell clones containing 24 X MS2 stem-loops in RAB7A first intron). To capture the diffusing RNA with MS2 labeling, live cell imaging was performed on a Zeiss LSM780 laser scanning confocal microscope using 37C incubation and 5% C02. Imaging was performed using 488 nm excitation, pinhole size of 2 airy units. The pixel size was 0.066 microns. One z plane was imaged every 1 sec for 60 frames. Transcription site intensity tracks were analyzed using KNIME pipeline

https://github.com/CBIIT/Larson-Lab-CCR-NCI/tree/master/Wan_GeneTrap_2019/KNIME_Workflows. MSD analysis was performed using MATLAB.

Transcription site intensity tracks were then analyzed in Matlab. Briefly, we identified the brightest pixel in each track, and called its location as the transcription site. We then cropped a 30x30 pixel region around this site, and used imresize() to increase resolution without interpolation. We then set a threshold of (((max pixel intensity - median pixel intensity) / 2) + median pixel intensity) to create a binarized, Gaussian filtered image of the transcription site, for which we calculated the eccentricity using regionprops(). Images without clear, identifiable transcription sites were discarded.

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 Intron Splicing Order

For all independence calculations, it was first determined what proportion of RNA fell into each of the following categories:

1. Fully unspliced (exon spot with both introns)
2. Partially spliced with intron A removed
3. Partially spliced with intron B removed
4. Fully spliced (exon spot without either intron)

Spots within 0.65 microns of each other were considered to be colocalized spots (as described in Figure 1). Colocalized spots that contained exon, intron A, and intron B were designated as a "3 color spot" (category 1). Spots that contained only exon signal were designated as fully spliced (category 4). Colocalized spots with only one intron signal were designated by the intron that was removed (for example, if intron B is present, then intron A was removed, and the transcript is in category 2). Categories 2 and 3 represent mutually exclusive pathways to generate the final product of a mature transcript.

Once each transcript was categorized, we focused our attention on partially spliced transcripts, and determined the proportion of these transcripts in either category 2 or category 3. Should the splicing rate of different introns be remarkably different from each other, or should intron splicing be dependent on the splicing of other introns, we would expect an overrepresentation of one splicing pathway.

We labeled the 4th intron of TM4SF1 with two colored probes, and determined the number of intermediate transcripts identified using this method, allowing us to measure the noise inherent in this analysis method. With 100% detection efficiency, we would expect this set of two probes labelling the same intron to be spliced out of pre-mRNA simultaneously. To calculate a "noise floor" of intermediate transcripts, we took the number of intermediate transcripts divided by two (as both category 2 and category 3 in this case measure the same type of error), and divided by the total number of transcripts captured in this experiment. Values below this noise floor are likely below our ability to faithfully detect differences in intron splicing behavior. Rows marked with ** in Table 2 show intron pairs that fall below this threshold, potentially due to the low number of intermediate transcripts captured for fast-splicing introns.

Calculation of slow-moving zone dimension

To find the diameter of each (roughly circular) cloud, we used the equation $2 \cdot \sqrt{(median\ area) / \pi)}$, then converted these values from expanded space into unexpanded space by dividing by 4.65. Because the 5' end, 3' end, and intron clouds were of a similar size, we averaged the values from each cloud to give an approximate slow moving zone diameter of 0.36 microns. Given the uncertainties inherent to these calculations, these estimates are meant to provide an order of magnitude for the size of the zone, rather than precise measurements.

Below are the full calculations:

<u>3' cloud:</u>

Median cloud area = 3.596 microns^2

Cloud Diameter = $2 \cdot \sqrt{(3.596) / \pi)}$ = 2.14 microns

In unexpanded space, cloud diameter = **0.46** microns

<u>5' cloud:</u>

Median cloud area = 1.862 microns^2

Cloud Diameter = $2 \cdot \sqrt{(1.862) / \pi)}$ = 1.54 microns

In unexpanded space, cloud diameter = **0.33** microns

<u>Intron cloud:</u>

Median cloud area = 1.474 microns^2

Cloud Diameter = $2 \cdot \sqrt{(1.474) / \pi)}$ = 1.37 microns

In unexpanded space, cloud diameter = **0.29** microns

Approximate size of the slow moving zone: **0.36** microns

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):

$\partial C\ \partial t = D\Delta C - \gamma C + R\ 0\ \delta(r)$,

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:

$C(r) = C\ ,\ 0\ \exp(\ -r\ \lambda\ )$,

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.

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. Y. L. Khodor, J. Rodriguez, K. C. Abruzzi, C.-H. A. Tang, M. T. Marr 2nd, M. Rosbash, Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. *Genes Dev.* **25**, 2502–2512 (2011).

4. Reimer, K. A., Mimoso, C. A., Adelman, K. & Neugebauer, K. M. Co-transcriptional splicing regulates 3' end cleavage during mammalian erythropoiesis. *Mol. Cell* **81**, 998–1012.e7 (2021)

5. 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).

6. 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.

7. 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.

8. Y. L. Khodor, J. S. Menet, M. Tolan, M. Rosbash, Cotranscriptional splicing efficiency differs dramatically between Drosophila and mouse. *RNA*. **18**, 2174–2186 (2012).

9. Choquet, K. *et al.* Pre-mRNA splicing order is predetermined and maintains splicing fidelity across multi-intronic transcripts. *bioRxiv* 2022.08.12.503515 (2022) doi:10.1101/2022.08.12.503515

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

11. 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).

12. 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).

13. 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).

14. 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).

15. 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).

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

17. 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).

18. 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).

19. 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).

20. 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).

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

22. 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).

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

24. 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).

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

26. 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).

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

28. 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)

29. 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

30. Rust, M. J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods* **3**, 793–795 (2006)

31. T. Maire, H. Youk, Molecular-Level Tuning of Cellular Autonomy Controls the Collective Behaviors of Cell Populations. *Cell Syst*. **1**, 349–360 (2015).

32. Anamika, K., Gyenis, À., Poidevin, L., Poch, O. & Tora, L. RNA polymerase II pausing downstream of core histone genes is different from genes producing polyadenylated transcripts. *PLoS One* **7**, e38769 (2012)

33. Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845–1848 (2008)

34. Wan, Y. *et al.* Dynamic imaging of nascent RNA reveals general principles of transcription dynamics and stochastic splice site selection. *Cell* **184**, 2878–2895.e20 (2021)

35. 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).

36. 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.

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

38. 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).

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

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

41. 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.

42. 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).

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

44. 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).

45. 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).

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

47. 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).

48. 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).

49. 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).

50. Vargas, D. Y., Raj, A., Marras, S. A. E., Kramer, F. R. & Tyagi, S. Mechanism of mRNA transport in the nucleus. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 17008–17013 (2005)

51. Smalec, B. M. *et al.* Genome-wide quantification of RNA flow across subcellular compartments reveals determinants of the mammalian transcript life cycle. *bioRxiv* 2022.08.21.504696 (2022) doi:10.1101/2022.08.21.504696

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

53. 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).

54. 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).

55. 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).

56. 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).

57. 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).

58. 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).

59. 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).

60. 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).

61. 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).

62. Yeo, G. & Burge, C. B. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.* **11**, 377–394 (2004)

Figure 1

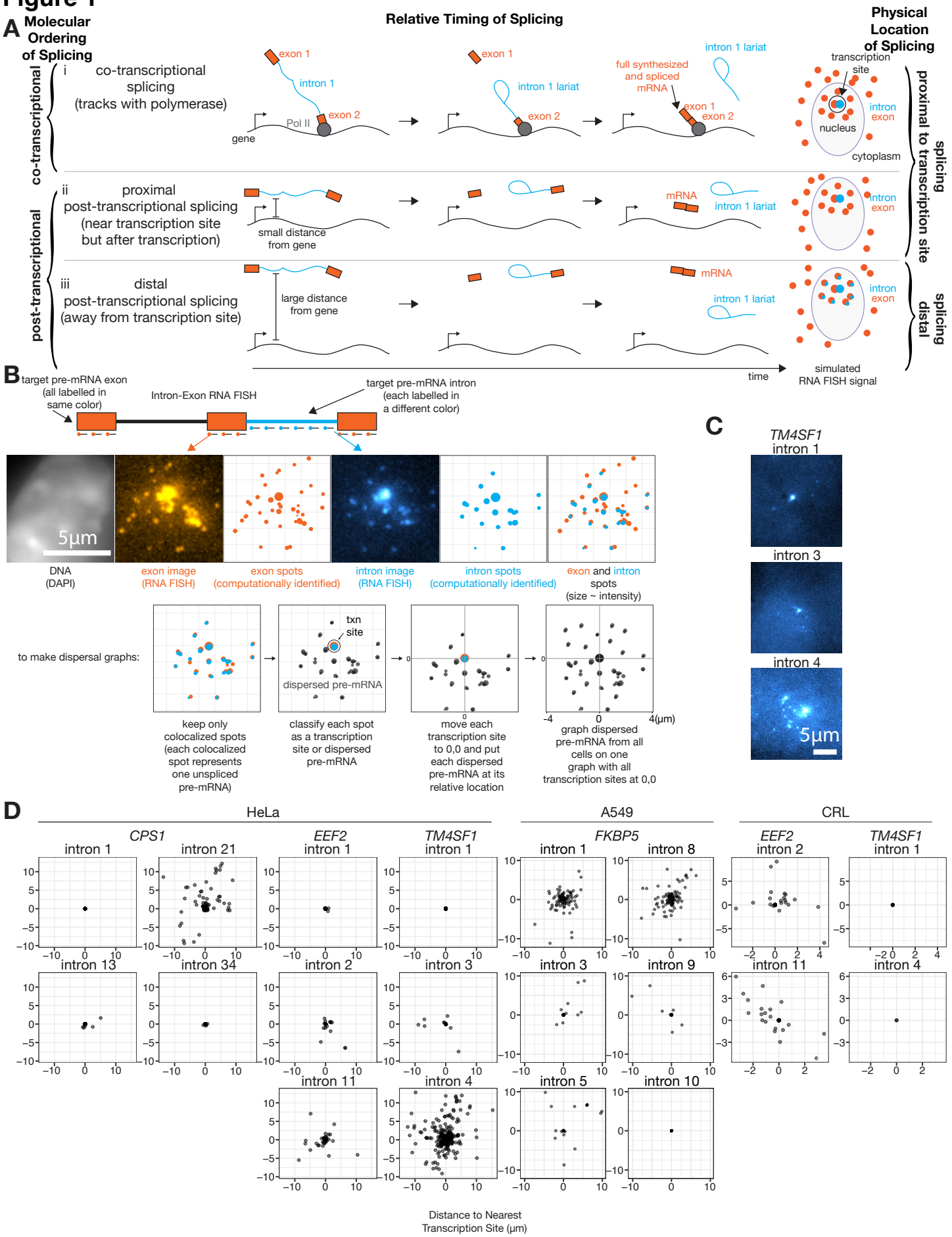

A
Molecular Ordering of Splicing
Relative Timing of Splicing
Physical Location of Splicing
co-transcriptional
post-transcriptional
i co-transcriptional splicing (tracks with polymerase)
ii proximal post-transcriptional splicing (near transcription site but after transcription)
iii distal post-transcriptional splicing (away from transcription site)
exon 1
intron 1
exon 2
Pol II
gene
intron 1 lariat
full synthesized and spliced mRNA
mRNA
small distance from gene
large distance from gene
time
transcription site
nucleus
cytoplasm
intron
exon
proximal to transcription site splicing
distal splicing
simulated RNA FISH signal
B
target pre-mRNA exon (all labelled in same color)
Intron-Exon RNA FISH
target pre-mRNA intron (each labelled in a different color)
5µm
DNA (DAPI)
exon image (RNA FISH)
exon spots (computationally identified)
intron image (RNA FISH)
intron spots (computationally identified)
exon and intron spots (size ~ intensity)
to make dispersal graphs:
keep only colocalized spots (each colocalized spot represents one unspliced pre-mRNA)
txn site
dispersed pre-mRNA
classify each spot as a transcription site or dispersed pre-mRNA
move each transcription site to 0,0 and put each dispersed pre-mRNA at its relative location
graph dispersed pre-mRNA from all cells on one graph with all transcription sites at 0,0
-4
0
4(µm)
C
TM4SF1
intron 1
intron 3
intron 4
5µm
D
HeLa
A549
CRL
CPS1
EEF2
TM4SF1
FKBP5
intron 1
intron 21
intron 13
intron 34
intron 2
intron 3
intron 9
intron 11
intron 4
intron 5
intron 10
Distance to Nearest Transcription Site (µm)

Figure 2

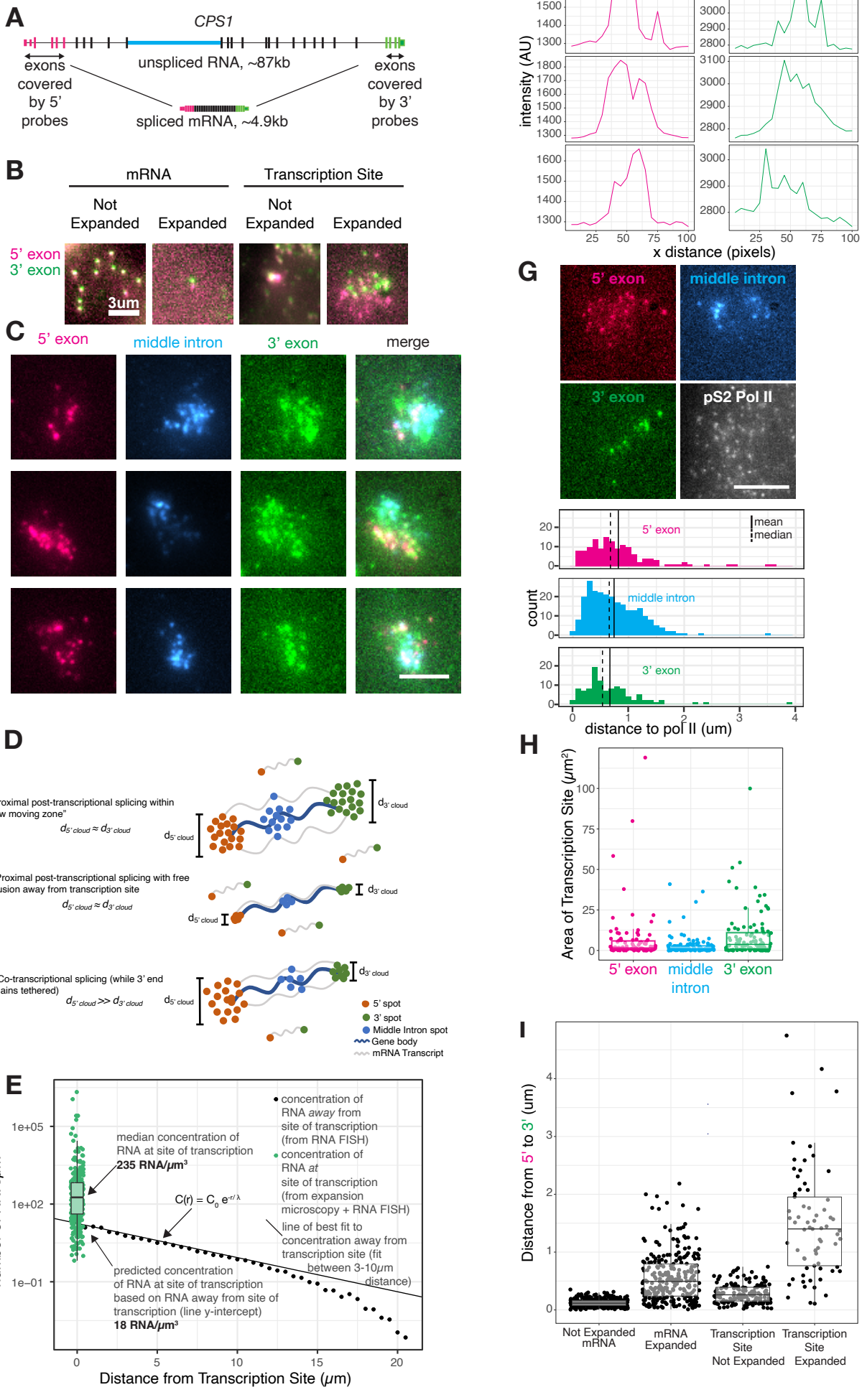

A
CPS1
exons covered by 5' probes
unspliced RNA, ~87kb
exons covered by 3' probes
spliced mRNA, ~4.9kb
B
mRNA
Transcription Site
Not Expanded
Expanded
Not Expanded
Expanded
5' exon
3' exon
3um
C
5' exon
middle intron
3' exon
merge
D
I. Proximal post-transcriptional splicing within "slow moving zone"
$d_{5' cloud} \approx d_{3' cloud}$
II. Proximal post-transcriptional splicing with free diffusion away from transcription site
$d_{5' cloud} \approx d_{3' cloud}$
III. Co-transcriptional splicing (while 3' end remains tethered)
$d_{5' cloud} >> d_{3' cloud}$
$d_{5' cloud}$
$d_{3' cloud}$
5' spot
3' spot
Middle Intron spot
Gene body
mRNA Transcript
E
number of RNA/μm³
Distance from Transcription Site (μm)
median concentration of RNA at site of transcription
235 RNA/μm³
C(r) = C₀ e^{-r/λ}
predicted concentration of RNA at site of transcription based on RNA away from site of transcription (line y-intercept)
18 RNA/μm³
concentration of RNA away from site of transcription (from RNA FISH)
concentration of RNA at site of transcription (from expansion microscopy + RNA FISH)
line of best fit to concentration away from transcription site (fit between 3-10μm distance)
F
5' exon
3' exon
intensity (AU)
x distance (pixels)
G
5' exon
middle intron
3' exon
pS2 Pol II
mean
median
count
distance to pol II (um)
H
Area of Transcription Site (μm²)
5' exon
middle intron
3' exon
I
Distance from 5' to 3' (um)
Not Expanded mRNA
mRNA Expanded
Transcription Site Not Expanded
Transcription Site Expanded

Figure 3

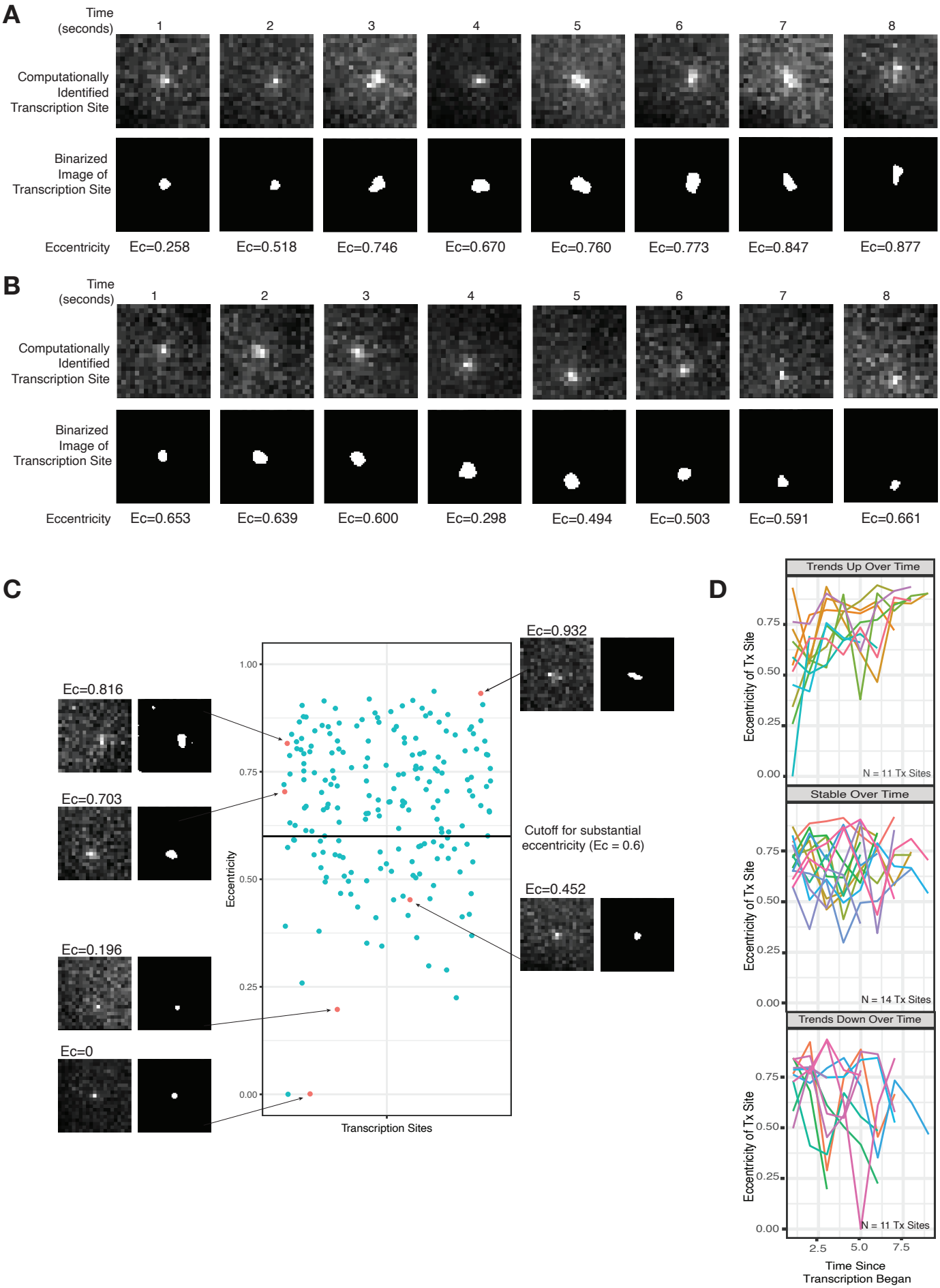
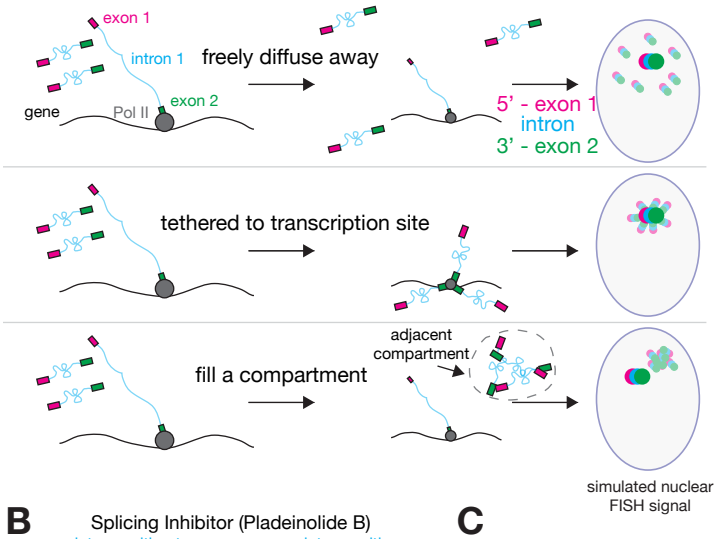
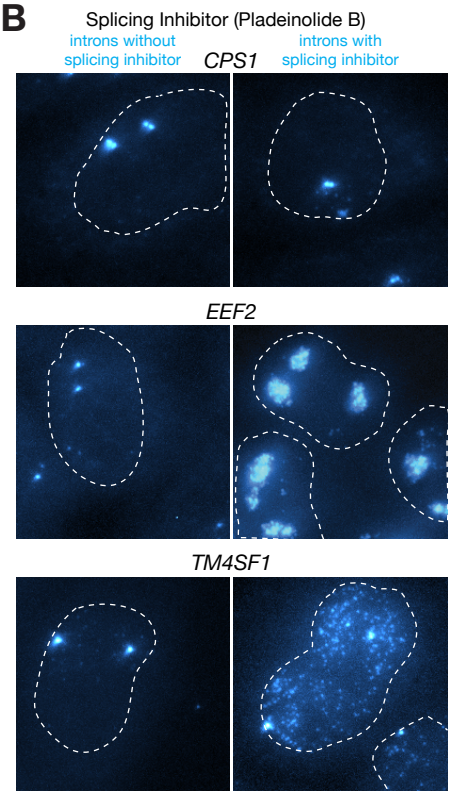

A
Time (seconds)
1
2
3
4
5
6
7
8
Computationally Identified Transcription Site
Binarized Image of Transcription Site
Eccentricity
Ec=0.258
Ec=0.518
Ec=0.746
Ec=0.670
Ec=0.760
Ec=0.773
Ec=0.847
Ec=0.877
B
Time (seconds)
1
2
3
4
5
6
7
8
Computationally Identified Transcription Site
Binarized Image of Transcription Site
Eccentricity
Ec=0.653
Ec=0.639
Ec=0.600
Ec=0.298
Ec=0.494
Ec=0.503
Ec=0.591
Ec=0.661
C
Ec=0.816
Ec=0.703
Ec=0.196
Ec=0
Ec=0.932
Ec=0.452
Cutoff for substantial eccentricity (Ec = 0.6)
Eccentricity
1.00
0.75
0.50
0.25
0.00
Transcription Sites
D
Trends Up Over Time
N = 11 Tx Sites
Stable Over Time
N = 14 Tx Sites
Trends Down Over Time
N = 11 Tx Sites
Eccentricity of Tx Site
0.00
0.25
0.50
0.75
2.5
5.0
7.5
Time Since Transcription Began

Figure 4

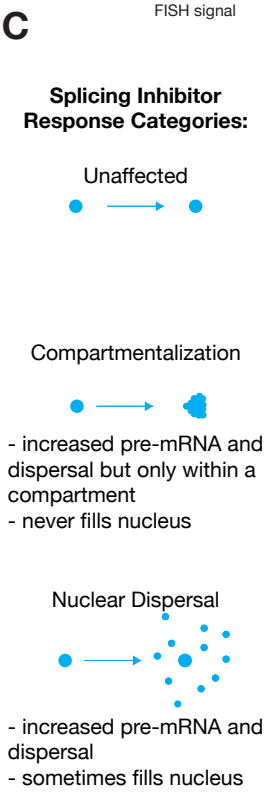
A Potential Pre-mRNA Behaviors after Leaving Transcription Site



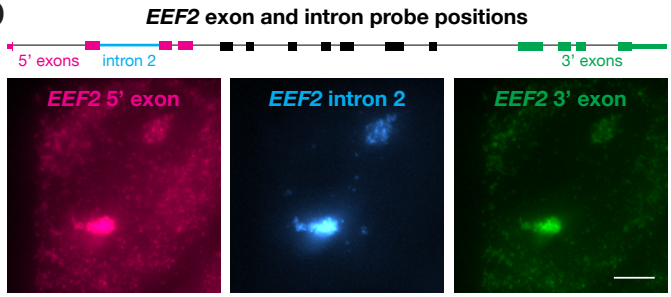
B



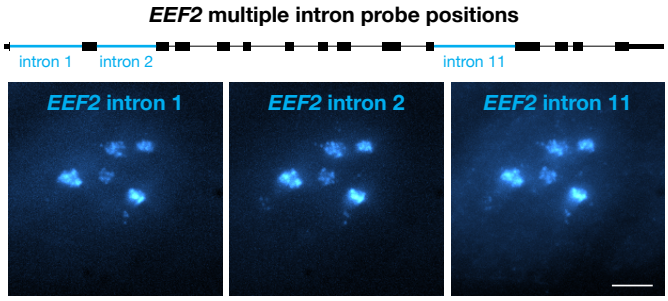
C



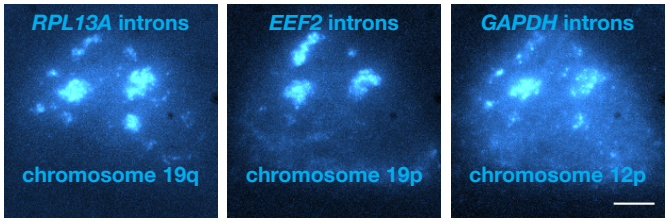
D



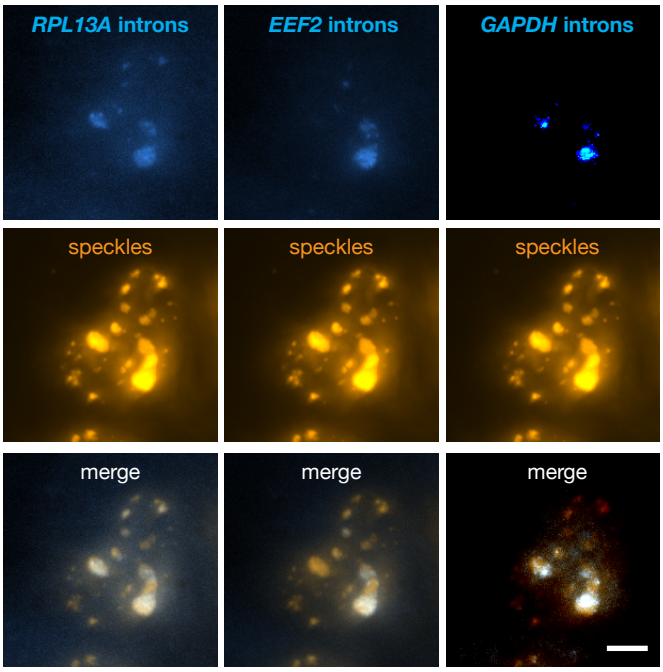
E



F



G



H

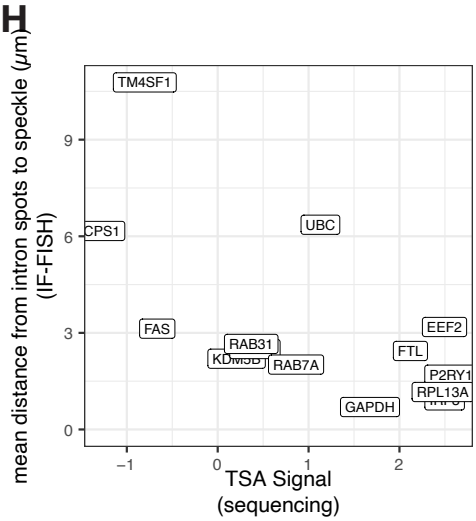
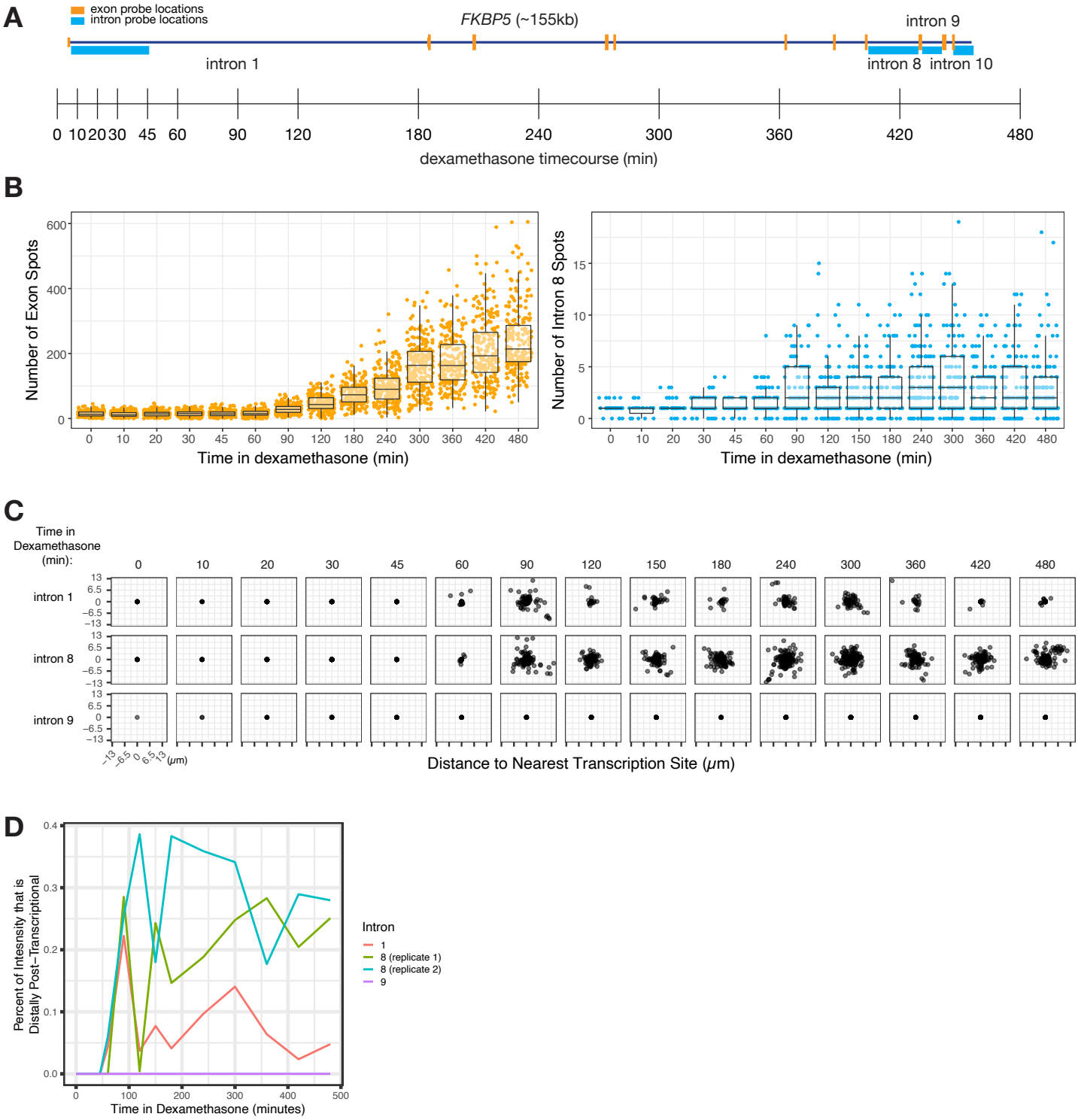


Figure 5


A
exon probe locations
intron probe locations
FKBP5 (~155kb)
intron 1
intron 9
intron 8
intron 10
0 10 20 30 45 60 90 120 180 240 300 360 420 480
dexamethasone timecourse (min)
B
Number of Exon Spots
Time in dexamethasone (min)
Number of Intron 8 Spots
Time in dexamethasone (min)
C
Time in Dexamethasone (min):
0 10 20 30 45 60 90 120 150 180 240 300 360 420 480
intron 1
intron 8
intron 9
Distance to Nearest Transcription Site (μm)
D
Percent of Intesnsity that is Distally Post−Transcriptional
Time in Dexamethasone (minutes)
Intron
1
8 (replicate 1)
8 (replicate 2)
9