Molecular Cell
Article

Global Analysis of Nascent RNA Reveals Transcriptional Pausing in Terminal Exons

Fernando Carrillo Oesterreich,[1] Stephan Preibisch,[1] and Karla M. Neugebauer[1,*]
[1]Max Planck Institute of Molecular Biology and Genetics, Pfotenhauerstrasse 108, 01307 Dresden, Germany
*Correspondence: neugebauer@mpi-cbg.de
DOI 10.1016/j.molcel.2010.11.004

SUMMARY

Pre-mRNA splicing is catalyzed by the spliceosome, which can assemble on pre-mRNA cotranscriptionally. However, whether splicing generally occurs during transcription has not been addressed. Indeed, splicing catalysis is expected to occur posttranscriptionally in yeast, where the shortness of terminal exons should leave insufficient time for splicing. Here, we isolate endogenous *S. cerevisiae* nascent RNA and determine gene-specific splicing efficiencies and transcription profiles, using high-density tiling microarrays. Surprisingly, we find that splicing occurs cotranscriptionally for the majority of intron-containing genes. Analysis of transcription profiles reveals Pol II pausing within the terminal exons of these genes. Intronless and inefficiently spliced genes lack this pause. In silico simulations of transcription and splicing kinetics confirm that this pausing event provides sufficient time for splicing before termination. The discovery of terminal exon pausing demonstrates functional coupling of transcription and splicing near gene ends.

INTRODUCTION

Pre-mRNA splicing is a two-step *trans*-esterification reaction that removes introns and ligates exons to produce a functional mRNA. This essential step in gene expression is carried out by the spliceosome, a megadalton complex comprised of the pre-mRNA, five spliceosomal small nuclear ribonucleoproteins (snRNPs), and numerous non-snRNP proteins (Wahl et al., 2009). Human and yeast spliceosomes contain more than 50 proteins associated with snRNPs and up to 100 non-snRNP proteins (Fabrizio et al., 2009; Wahl et al., 2009). Spliceosomal components interact with several intronic sequences in the pre-mRNA, which leads to assembly of the catalytically active spliceosome. At least eight ATP-dependent conformational rearrangements occur along the pathway to activation and catalysis. Thus, the spliceosome is not a simple enzyme that can bind quickly and react with its substrate; instead, spliceosome assembly and splicing must take time (Staley and Guthrie, 1998).

In the living cell, splicing can occur during transcription while the pre-mRNA is attached to chromatin by RNA polymerase II (Pol II), i.e., cotranscriptionally (Perales and Bentley, 2009). This introduces an additional temporal aspect into the process of spliceosome assembly, which is constrained by the 5′ to 3′ direction of transcription. The triggers for spliceosome assembly are the recognition of the 5′ splice site (5′SS) by U1 snRNP and the branchpoint near the 3′ splice site (3′SS) by U2 snRNP and other factors. It follows that U1 snRNP can interact with the 5′SS as transcription of the intron begins, but recognition of the 3′SS must wait until the whole intron is synthesized. These predictions of stepwise spliceosome assembly have been verified in vivo by chromatin immunoprecipitation (ChIP) in *Saccharomyces cerevisiae*, in which U1 snRNP accumulation peaks shortly after 5′SS synthesis and U2 snRNP accumulation peaks after 3′SS synthesis (Görnemann et al., 2005; Kotovic et al., 2003; Lacadie and Rosbash, 2005; Tardiff et al., 2006). Assembly of the active spliceosome occurs ~750 bp downstream of the 3′SS (Görnemann et al., 2005; Tardiff et al., 2006); splicing catalysis is completed ~1 kb downstream of the 3′SS (Lacadie et al., 2006; Tardiff et al., 2006). Because the median length of terminal exons in yeast is less than 500 nucleotides (nt), it was previously concluded that most splicing must be posttranscriptional (Tardiff et al., 2006). However, transcription elongation behavior may vary from gene to gene (Mason and Struhl, 2005; Zenklusen et al., 2008). Moreover, the prevalence of cotranscriptional splicing has never been directly tested. How many genes are completely or partially spliced cotranscriptionally? Are splicing and transcription of endogenous genes coordinated?

It is important to determine how general cotranscriptional splicing is and whether direct links between transcription and splicing exist. Coupling between transcription and splicing may lead to more efficient splicing and packaging of mRNPs for export (Moore and Proudfoot, 2009). Evidence that cotranscriptional splicing is more efficient comes from in vitro studies and is still controversial (Das et al., 2007; Hicks et al., 2006; Lazarev and Manley, 2007). In mammalian cells, intron removal often occurs on chromatin before cleavage and polyadenylation, or poly(A) (Listerman et al., 2006; Pandya-Jones and Black, 2009; Singh and Padgett, 2009; Wada et al., 2009). However, to date, only selected genes have been analyzed. An additional motivation for exploring connections between transcription elongation dynamics comes from findings that alternative splicing is influenced by changes in Pol II elongation rates. For example, placement of a pause site downstream of weak alternative exons promotes exon inclusion (Roberts et al., 1998); similarly, reducing elongation rate either through Pol II mutations or drug treatments enhances alternative exon inclusion and

cotranscriptional spliceosome assembly (de la Mata et al., 2003; Listerman et al., 2006). Physiological relevance was recently demonstrated by the observation that UV light treatment leads to changes in Pol II elongation rate and alternative splicing (Muñoz et al., 2009). However, the basic principles of coupling between transcription and splicing in the context of endogenous genes are still poorly understood.

A global approach is necessary to address precise relationships between transcription and splicing. We sought to obtain yeast nascent RNA by chromatin fractionation, a procedure developed for mammalian cells (Wuarin and Schibler, 1994). Here, we establish biochemical purification of a chromatin fraction from exponentially growing yeast cells. This preparation provided a source of endogenous nascent RNA for global analysis. Recently, high-density tiling microarrays have been shown to yield highly quantitative data for transcript profiling, including the detected loss of intronic sequences in mRNA pools and overrepresentation of intronic sequences when intron degradation is blocked (David et al., 2006; Huber et al., 2006; Zhang et al., 2007). Therefore, nascent RNA was subjected to analysis on high-density tiling arrays, from which the extent of cotranscriptional intron removal could be determined on a gene-by-gene basis. It was also possible to infer changes in polymerase density from the nascent RNA traces, providing insight into the splicing-coupled behavior of Pol II during elongation.

RESULTS

To isolate nascent RNA, we generated a chromatin fraction from *S. cerevisiae*. The ternary complex between genomic DNA, Pol II, and nascent RNA is resistant to high concentrations of detergents, salt, polyanions, and chaotropes (Cai and Luse, 1987; Wuarin and Schibler, 1994). This stability allowed for high-stringency washing of the chromatin fraction, which was recovered from each wash step by centrifugation (Figure 1A). Contaminating mRNA from translating ribosomes, which share a high sedimentation coefficient, was avoided by ribosomal dissociation with high EDTA concentrations in the high-stringency washing buffer (Buffer 2). The chromatin fraction (Pellet 3) was depleted of rRNA and tRNA (Figure 1B). At the protein level, histones (Hht1) and Pol II (Rbp1) were enriched in the chromatin fraction, whereas the ribosomal protein Rpl17a was lost (Figure 1C). A minor fraction of Rpb1 in the final supernatant may represent initiating ternary complexes, which are more sensitive to stringent washing (Cai and Luse, 1987). Mass spectrometry confirmed that the chromatin fraction was enriched in all histones and RNA polymerases as well as many chromatin remodeling factors (Tables 1 and S1). Interestingly, splicing factors were not detected, indicating that the affinity of splicing factors for nascent RNA is too weak to withstand high-stringency washing.

The establishment of chromatin fractionation in budding yeast facilitates global studies of cotranscriptional processing of RNAs transcribed by any polymerase, by enriching the small proportion of cellular RNA that is nascent. To analyze nascent transcripts, RNA was isolated from the chromatin fraction and depleted of contaminating mRNA with oligo dT Sepharose.

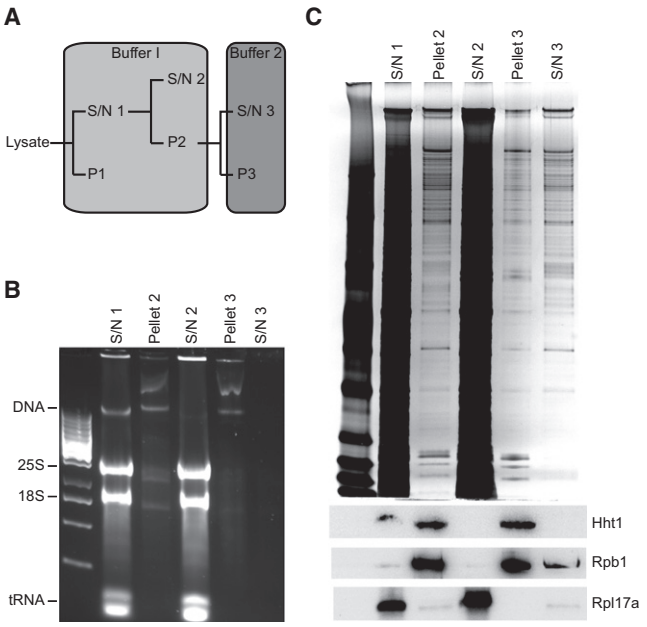

Figure 1. Establishment of *S. cerevisiae* Chromatin Fractionation

(A) Centrifugation scheme employed for chromatin fractionation.

(B) Nucleic acid analysis of fractions shown in (A). Genomic DNA (DNA, identified by its sensitivity to DNase I digestion, Figure S1) is quantitatively enriched in the chromatin fraction (Pellet 3). Abundant RNAs (25S and 18S rRNAs and tRNA) are quantitatively removed.

(C) Protein analysis of respective fractions: Silver-stained SDS-PAGE indicates a significant reduction of overall protein content during the fractionation process (upper panel). Western blot analysis indicates quantitative enrichment of histone H3 (Hht1) and RNA polymerase II (Rpb1) in the chromatin fraction. Ribosomes (Rpl17a) are removed during the purification process (lower panel). For quantification of nascent RNA enrichment, see Figure S1.

Quantification by quantitative RT-PCR, using either a reverse primer binding downstream of the poly(A) cleavage site or oligo dT, revealed that full-length nascent RNAs were ~40 times enriched in the preparation (Figures S1B–S1D). The actual nascent RNA enrichment is likely higher than this, because most nascent RNA is not full length. Moreover, nascent transcripts are rapidly cleaved at the poly(A) site (Boireau et al., 2007), thereby decreasing the amount of nascent RNA detectable in this assay. Underrepresentation of the 3′ end, especially downstream of the poly(A) site, makes this quantification of enrichment a conservative estimate.

The expectation for nascent RNA profiles is diagrammed in Figure 2A; while all nascent RNA molecules share a common 5′ end defined by the transcription start site, the distribution of 3′ ends reflects the distribution of Pol II as it progresses along the gene. As a consequence, the abundance of sequences present in the nascent RNA fraction is expected to decrease from 5′ to 3′ ends. This should lead to a 5′ to 3′ intensity decrease on the high-density tiling microarray, a prediction confirmed by the obtained data (Figures 2B and S2). Observed reductions in signal over annotated introns indicate the loss of introns from the chromatin fraction, which can be attributed to cotranscriptional splicing catalysis. Most nuclear genes in *S. cerevisiae*

Table 1. Proteomic Analysis of the Chromatin Fraction

Group	Complex	Detected Components	Missing Components
Nucleosome	Histones	H1, H2A, H2B, H3, H4, H2AZ	–
RNA polymerases	Pol I	Rpa190, Rpa135, Rpa49, Rpa43, Rpc40, Rpa34, Rpb5, Rpb19, Rpb8, Rpa12, Rpc10, Rpb10	Rpb6, Rpa14
	Pol II	Rpb1, Rpb2, Rpb4, Rpb5, Rpb7, Rpb8, Rpb9, Rpb10, Rbp11, Rpc10	Rpb3, Rpb6
	Pol III	Rpc160, Rpc128, Rpc82, Rpc53, Rpc40, Rpc37, Rpc34, Rpb5, Rpc25, Rpc19, Rpb8, Rpc10	Rpb6, Rpc17, Rpc11
DNA polymerases	Pol α	–	Pol1, Pol12, Pri1, Pri2, Hcs1
	Pol δ	–	Pol3, Pol31, Pol32
	Pol ε	Dpb4	Pol2, Dpb2, Dpb3, Dpb11
Splicing factors	–	–	all
Chromatin remodelers	RSC	Rsc1, Rsc2, Rsc3, Rsc4, Rsc58, Rsc6, Rsc7, Rsc8, Rsc9, Rsc11, Rsc12, Rsc14, Rtt102, Sth1,	–
	yINO80	Arp4, Arp5, Arp8, Ino80, Ies1, Ies5, Nhp10, Rvb1, Rvb2, Taf14	Ies2, Ies3, Ies4, Ies6
	SWI/SNF	Snf2, Swi3, Swp29, Swp59, Swp61, Swp73, Swp82, Rtt102	Swi1, Snf5, Snf6, Snf11
	ISWIa	Isw1, Ioc3	–
	ISWIb	Isw1, Ioc2, Ioc4	–
	ISW2	Isw2, Itc1, Dpb4	Dls1

Detection of proteins by mass spectrometry (MS) for selected complexes is shown. Grouping of complex (Group), complex name (Complex), proteins detected (Detected Components), and not detected (Missing Components) are indicated. The complete MS data set is provided in Table S1.

lack introns (296 intron-containing and 6575 intronless genes); among intron-containing genes, the majority (286) bear only a single intron. The high-density tiling arrays used here cover the Watson (plus) strand, which encodes 142 genes with single introns, of which 122 were expressed. The efficiency of cotranscriptional splicing was calculated for these 122 intron-containing genes, by fitting a linear function to probe intensities downstream of the 3′SS. Extrapolation of the linear fit to intronic positions permitted calculation of the differences between individual intronic probe intensities and the predicted values (Figure 2B). Further examples are shown in Figure S2. The median of all calculated values was computed to yield a composite value (cotranscriptional splicing efficiency, Δ), reflecting the fraction of intron removal in nascent RNA on a gene-by-gene basis. Note that the background level was defined empirically (Figure S2B). Computation of Δ for each of the 122 intron-containing genes revealed that the majority of introns are removed cotranscriptionally (Figure 2C). The median cotranscriptional splicing efficiency for all genes was 0.74; 43% of the genes analyzed were spliced nearly to completion (>0.8). Cotranscriptional splicing efficiencies determined by this method were independently validated for ten genes by quantitative RT-PCR, in which uncleaved nascent RNA was selected from total RNA with RT primers downstream of the poly(A) site and then amplified by splice-junction-specific quantitative PCR primers to determine splicing levels (Figure S2C). The striking agreement between the cotranscriptional splicing efficiencies determined by these two unrelated methods further indicates that the nascent RNA preparation is not contaminated by mRNA. We conclude that cotranscriptional splicing is widespread across the yeast genome.

Completion of spliceosome assembly and splicing catalysis cannot occur before the entire intron has been transcribed; therefore, cotranscriptional splicing can only occur during transcription of the terminal exon, imposing a time limit. This assumption is consistent with a previous study showing that reporter genes harboring short terminal exons are posttranscriptionally spliced (Tardiff et al., 2006). Our observation that most genes are cotranscriptionally spliced is surprising, because endogenous terminal exons in yeast are short. To predict how the efficiency of cotranscriptional splicing depends on terminal exon length, we developed an in silico simulation that assumes no coupling between splicing and transcription. Actual gene architectural features for the 122 genes studied were employed to query the expected outcome of high-density tiling array analysis; the terminal exon length for each gene was derived from recent RNA-Seq data (Nagalakshmi et al., 2008). In this simulation, we assume a uniform Pol II elongation rate over the entire gene, and splicing occurs 750 nt after synthesis of the 3′SS (Tardiff et al., 2006). A Poisson distribution around experimentally determined mean values was applied for all underlying rates, including promoter residence time and clearance, splicing, and elongation rate (see Supplemental Information for values used). For each gene studied, 3600 gene entities were simultaneously simulated to generate a pool of heterogeneous virtual nascent RNA molecules aligned along the length of the gene. Each simulation was run for 20 min; transcription and splicing reached steady state at ~5 min, yielding a nascent RNA profile analogous to that experimentally obtained. Figure 3A shows that predicted cotranscriptional splicing efficiency correlates positively with terminal exon length. In agreement with previous estimates, genes with very short terminal exons (<750 nt) are predicted to be spliced posttranscriptionally. However, the predicted values do not correlate with the observed cotranscriptional splicing efficiencies obtained in vivo (Figure 3A). Strikingly, the observed values are on average significantly higher than the predicted values, and genes with short terminal exons were

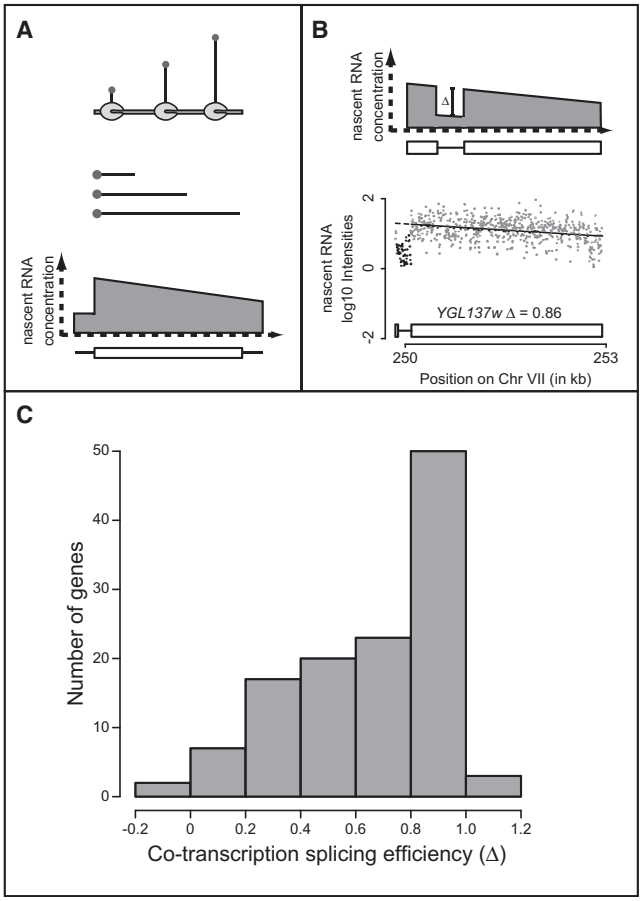

Figure 2. Quantification of Cotranscriptional Splicing by Microarray

(A) Expected profile of nascent RNA detected by high-density tiling microarray; generation of nascent RNA molecules by transcription (upper panel). The concentration of detected sequences correlates with distance from the transcriptional start site, as shown by schematic sequence alignment (middle panel). Plotting nascent RNA concentration versus gene position yields the expected tiling microarray hybridization profile (lower panel).

(B) Observed nascent RNA microarray traces and quantification of cotranscriptional intron removal. Schematic representation (upper panel) and actual nascent RNA tiling microarray traces for one representative gene (*YGL137w*, lower panel) are shown in logarithmic scale. Cotranscriptional splicing efficiency (Δ) is determined by deducing the relative concentration difference between intronic and exonic sequences. To this end, a linear function is fitted to terminal exon-delimited probe intensities (schematized by filled line). Differences of actual intronic probe intensities (black circles) and values gained by extrapolation of the fit (dotted line) are calculated for each intronic position. The median intensity difference for all intronic probes is used to determine cotranscriptional splicing efficiency (Δ). See Figure S2 for further examples and validation.

(C) Distribution of genes binned according to their cotranscriptional splicing efficiencies. All actual values are given in Table S2. n = 122 intron-containing genes analyzed.

remarkably well spliced. The discrepancy between the observed and predicted cotranscriptional splicing values suggests an unanticipated functional link between transcription and splicing.

In order to determine whether any gene-specific feature(s) correlate with the efficiency of cotranscriptional splicing, genes were divided into two groups. We found that the mean difference between the observed and predicted cotranscriptional splicing efficiencies for all genes analyzed was 0.5; therefore, genes with $\Delta_{Observed} - \Delta_{Predicted} \geq 0.5$ were classified as "high-efficiency," and the remaining genes were considered "low-efficiency" (Figure 3A). We first asked whether any aspect of gene architecture is unique to either group. Figure 3B shows the average lengths of gene regions for the low- and high-efficiency groups. As expected from Figure 3A, the average terminal exon length is shorter in the high-efficiency group. While intron lengths tend to be longer in the high-efficiency group, no significant differences were detected for exon 1 length or splice site strength between the two groups (Figures 3B and S3A). Simulation of transcription and splicing for a gene with the average gene architectural features of the low-efficiency group yields predicted cotranscriptional splicing efficiencies (Δ = 0.48) similar to the average experimental value (<Δ> = 0.36), indicating that the longer terminal exons in this group provide sufficient time for at least partial intron removal (Figure 3C, Table S2, and Movie S1). The correspondence between the predicted and observed cotranscriptional splicing efficiencies for the low-efficiency group suggests that the elongation and splicing rates used in the simulation approximate the in vivo situation. In contrast, simulation for the average gene in the high-efficiency group predicts a Δ of 0.0, which disagrees with the observed average value of 0.83 (Figure 3C, Table S2, and Movie S2). Second, we determined relative transcriptional frequencies from the nascent RNA tiling data. The median 5′ intensity values from nascent RNA tiling arrays correlated well with published data on transcription rates (Holstege et al., 1998) (Figure S3B). The high-efficiency group was characterized by ~2-fold higher transcriptional frequency compared to the low-efficiency group (Figure S3C). Taken together, the only gene architectural features that correlate with high-efficiency cotranscriptional splicing are long introns and, paradoxically, short terminal exons. The high-efficiency group is transcribed more frequently, suggesting that cotranscriptional splicing of these highly expressed genes may be important.

Functional coupling of splicing and transcription might explain the high degree of cotranscriptional splicing in the high-efficiency group of genes characterized by short terminal exon length. One possibility is that transcription elongation behavior in vivo does not conform to the assumed uniformity imposed on the simulation. We can deduce changes in Pol II elongation behavior from the nascent RNA traces within the body of genes of interest. The slope of the profile will be a constant negative value if transcription elongation along the length of the gene is uniform (Figure 2A). Deviations from a constant slope must reflect local changes in elongation, because the slope is proportional to Pol II density: a steeper negative slope indicates an increase in Pol II density over the gene region (see Supplemental Information for derivation). In the course of our analysis, we noticed a decrease in the slope of the nascent RNA profiles in terminal exons of high-efficiency genes (e.g., *YBL027w*, Figure 4A). These local changes suggested an increase in Pol II density near the ends of cotranscriptionally spliced genes. To test this, we systematically searched for changes in slope by fitting two linear functions to the nascent RNA intensities over terminal exons. The first function (Fit 1) was tethered to the

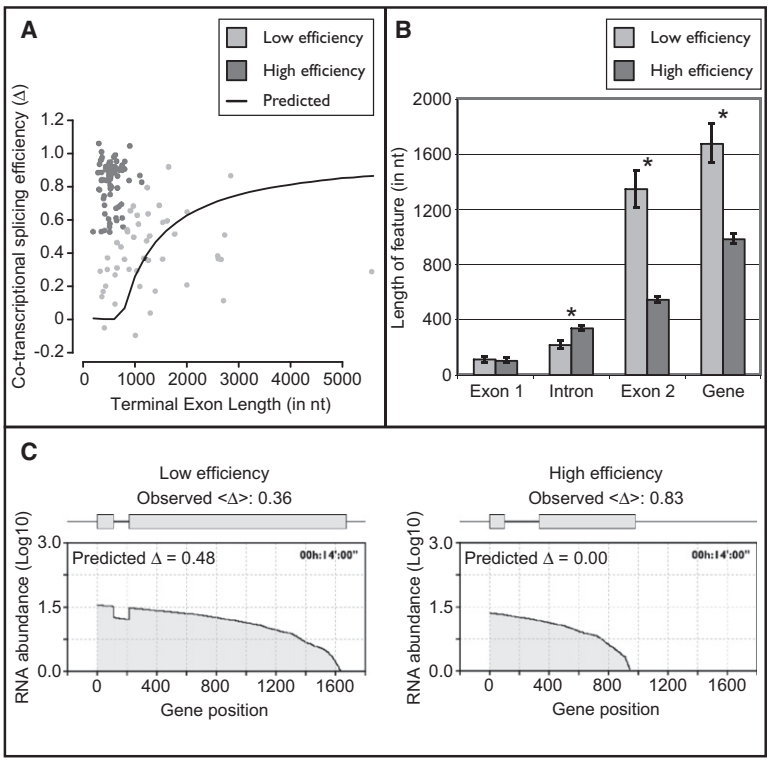

Figure 3. Highly Efficient Cotranscriptional Splicing Is Observed for Genes with Short Terminal Exons

(A) Observed (filled circles) and expected (black line) values of cotranscriptional intron removal for each gene analyzed are plotted against terminal exon length. Expected values are gained by in silico simulation of transcription and cotranscriptional splicing, assuming independent rates of the two reactions. Genes are assigned to high- (dark gray circles) and low- (light gray circles) efficiency groups, based on the difference between observed and predicted values ($\Delta_{Observed} - \Delta_{Predicted} \geq 0.5$).

(B) Comparison of gene architectural features of low- (light gray bars) and high- (dark gray bars) splicing-efficiency gene groups, designated as in (A). Lengths of gene regions are represented as mean ± SEM. Asterisks indicate significant differences between low- and high-efficiency genes (p values: Exon 1 > 0.5, Intron < 10^{-3}, Exon 2 < 10^{-6}, Gene < 10^{-4}). n = 48 and 74 genes in the low- and high-efficiency groups, respectively.

(C) Representative still images of simulated cotranscriptional splicing for genes with average gene architectural features of low- (left panel) or high- (right panel) efficiency genes. Corresponding gene features are diagrammed on top of each plot. Calculated abundance of nascent RNA (Log10) is plotted against gene position relative to the transcriptional start site. The still image represents the nascent RNA profile obtained after the profiles had reached steady state. Observed average (< Δ >) and predicted values (Δ) for cotranscriptional splicing efficiency are shown. See also Figure S3 and Movies S1 and S2.

3′SS and the second function (Fit 2) to the poly(A) site (Figure 4A); the fits were optimized by varying the position of the third tether (end of Fit 1 and beginning of Fit 2) and minimizing the global sum of squares (Figure S4). Genes in the low-efficiency group yielded similar slopes for Fits 1 and 2, suggesting a relatively uniform nascent RNA distribution (e.g., *YNL066w* in Figures 4A and S4A). In contrast, the slopes of Fits 1 and 2 were markedly different in the high-efficiency group of genes (e.g., *YBL027w* in Figures 4A and S4B). Figure 4B shows the average slope values for intronless and intron-containing genes in the low- and high-efficiency groups. Interestingly, the slopes of Fit 1 were indistinguishable among the gene groups, suggesting that transcription elongation is relatively constant in all promoter-proximal gene regions. The average slope of Fit 2 was steeper in each group, suggesting a tendency to increased Pol II density in promoter-distal gene regions. Strikingly, genes in the high-efficiency group displayed a ~5-fold steeper slope of Fit 2 compared to the low-efficiency genes. Thus, cotranscriptional splicing efficiency and short terminal exon length correlate with strong changes in Pol II elongation behavior toward the 3′ end.

Does the change in elongation behavior map to a specific position within genes? The onset of this change is marked by a point of discontinuity in the slopes of Fits 1 and 2 (dotted vertical line in Figure 4A), referred to here for simplicity as "discontinuity." We mapped the nucleotide position of each discontinuity and compared its location to gene architectural features: the transcriptional start site, 5′SS, 3′SS, and poly(A) site. Figure 4C shows that the position of the discontinuity correlates best with the poly(A) site for high-efficiency genes, exhibiting the least variation in distance. Discontinuities in intronless and low-efficiency genes did not correlate well with any gene feature (Figures 4C and S4C). On average, the discontinuity in high-efficiency genes is situated 246 ± 11 nt (mean ± SEM) upstream of the poly(A) site. In summary, analysis of nascent RNA profiles reveals that genes in the high-efficiency group specifically show altered elongation behavior well before transcription termination.

A steeper negative slope in the nascent RNA profile (Fit 2 versus Fit 1) predicts a relative increase in Pol II density over the gene region, as explained above. To test this, we analyzed published Pol II ChIP tiling microarray data (Venters and Pugh, 2009), in which the intensity of ChIP signals is a measure of Pol II density. Visual inspection of Pol II ChIP tiling microarray traces over individual genes in the low- and high-efficiency groups reveals differences in Pol II density near gene ends: Pol II density peaks downstream of the poly(A) site in genes of the low-efficiency group (e.g., *YNL066w* in Figures 5A and S5A), in agreement with Pol II pausing during termination (Brodsky et al., 2005; Core and Lis, 2008; Core et al., 2008; Fuda et al., 2009). In contrast, genes in the high-efficiency group exhibit a significant increase in Pol II density within the gene body, upstream of the poly(A) site (e.g., *YBL027w* in Figures 5A and S5B). To compare different gene groups globally, we calculated average Pol II ChIP tiling microarray traces for intronless as well as intron-containing genes in the low- and high-efficiency groups. Note that the 121 most highly transcribed intronless genes in the genome were selected for this analysis to best match the transcriptional properties of the high-efficiency group. Alignment of these traces to the mapped discontinuities reveals

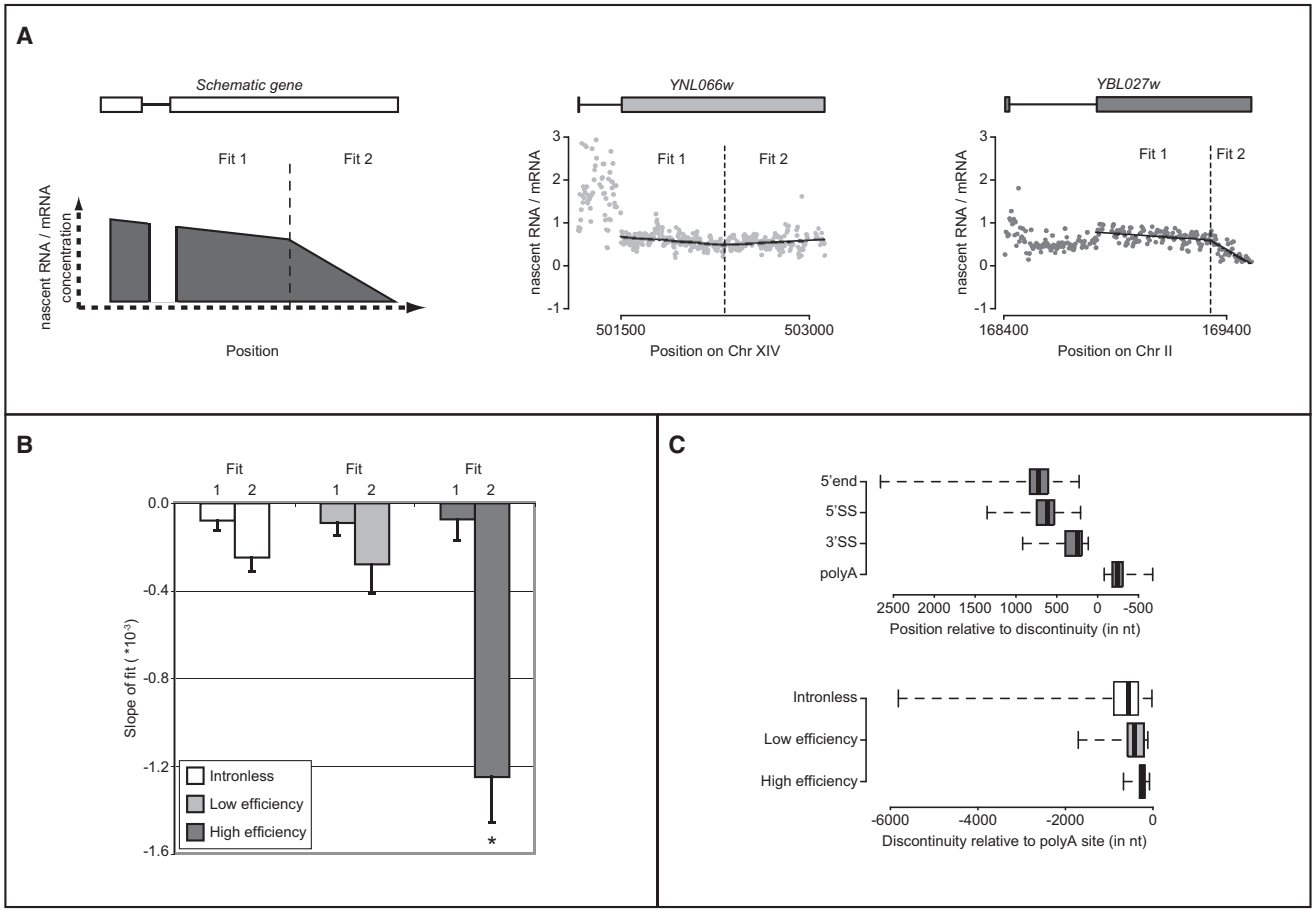

Figure 4. Quantification of Local Changes in Pol II Elongation Behavior

(A) Illustration of the method employed to detect local changes in elongation behavior. Two linear functions were fitted to the nascent RNA tiling array traces over terminal exons. When the slopes of Fits 1 and 2 are equal, a uniform distribution of Pol II along the gene is inferred; when the two slopes are unequal, a change in Pol II density is inferred. The upper panels show a schematic (left) and observed tiling microarray traces for a low-efficiency group gene (middle, *YNL066w*) with uniform Pol II density across the gene and high-efficiency group gene (right, *YBL027w*) with a local change in Pol II density towards the end of the gene. Tiling traces are shown in linear scale, and fits are represented by black lines. Vertical dotted lines mark the discontinuity between the two slopes. Normalization to mRNA was performed to account for hybridization bias. Gene name and architecture are indicated on top of the traces. For more examples and determination of the optimal fit, see Figure S4.

(B) Mean values for slopes of Fit 1 and Fit 2 for intronless (n = 2843), low-efficiency (n = 48), and high-efficiency (n = 74) genes grouped according to observed cotranscriptional splicing efficiencies (see Figure 3). Error bars represent the SEM. Asterisk indicates significant differences between slope of Fit 2 of different groups (intronless versus high-efficiency, $p < 10^{-4}$; low- versus high-efficiency, $p < 10^{-3}$) as well as significant differences between slopes of Fit 1 and Fit 2 (high-efficiency: slope Fit 1 versus slope Fit 2, $p < 10^{-5}$).

(C) Mapping the site of local change in Pol II behavior. The nucleotide distances of the mapped discontinuity from gene architectural landmarks (promoter [5′ end], 5′ splice site [5′SS], 3′ splice site [3′SS], and poly[A] site) are shown for high-efficiency genes (upper panel). Boxes represent values within three quartiles of the data, while whiskers indicate the range of the data set. Median values are shown by black horizontal lines in the respective boxes. The tightest correlation is observed between discontinuity and poly(A) site position (SDs: 5′ end, 303 nt; 5′SS, 208 nt; 3′SS, 145 nt; poly[A], 97 nt). Comparison of the discontinuity to poly(A) site distance for intronless, low-efficiency, and high-efficiency spliced genes are shown (lower panel). Discontinuity position and poly(A) site correlate best in the high-efficiency group of genes (SDs: intronless, 553 nt; low-efficiency, 320 nt; high-efficiency, 97 nt). SD, standard deviation. See also Figure S4.

that Pol II density specifically increases downstream of the discontinuity in the high-efficiency group (Figure 5B). This increase in Pol II ChIP signals downstream of the discontinuity validates the prediction that the slope of the nascent RNA trace correlates inversely with Pol II density. Only minor changes were detected near discontinuities in the remaining gene groups, indicating the robustness of our method for detecting changes in elongation behavior. One potential explanation for increased Pol II density upstream of poly(A) sites in high-efficiency genes could be a "pile-up" of Pol II as it encounters paused Pol II at the poly(A) site. High transcription frequencies in combination with short gene length of high-efficiency genes would favor this scenario. To test this, we generated a subset of 55 highly transcribed intronless genes with an average gene length matching that of the high-efficiency group of genes; the transcription frequency of these 55 intronless genes was significantly higher

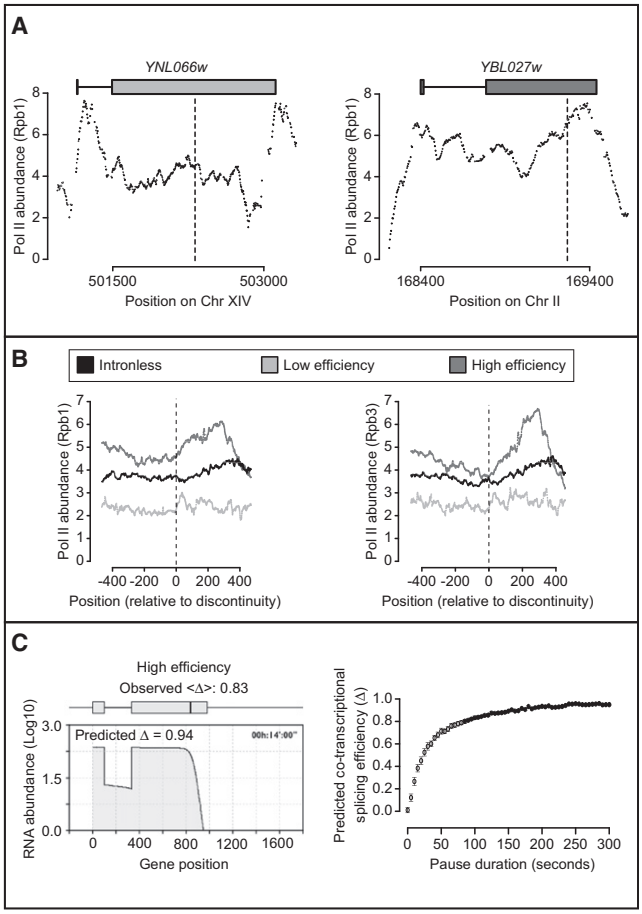

Figure 5. Pol II Density Increases before Termination in Cotranscriptionally Spliced Genes Only

(A) Pol II accumulates upstream of the poly(A) site in high-efficiency genes. Pol II (Rpb1) ChIP tiling microarray traces (Venters and Pugh, 2009) for representative low-efficiency (left panel, *YNL066w*) and high-efficiency (right panel, *YBL027w*) genes are shown. Gene name and architecture are displayed at the top. The position of the mapped discontinuity (see Figure 4) is indicated by a dotted vertical line. More examples are shown in Figure S5.

(B) Average Pol II (Rpb1, left panel; Rpb3, right panel) ChIP tiling microarray traces (Venters and Pugh, 2009) aligned to the mapped discontinuities (vertical dotted line) are shown for low- and high-efficiency genes as well as 121 highly transcribed intronless genes. The discontinuity marks the onset of an increase in overall Pol II density in the high-efficiency group, indicating changed elongation behavior in cotranscriptionally spliced genes only. For alignment to the poly(A) site, see Figure S5D.

(C) Pausing before termination recapitulates cotranscriptional splicing. Simulation of cotranscriptional splicing with incorporation of a pause site 123 nt upstream of the poly(A) site. Representative still image of simulation carried out on a gene with average gene architectural features of high-efficiency genes (left panel). Abundance of nascent RNA is plotted against gene position relative to the transcriptional start site. Gene architecture is diagrammed on top. A black bar indicates the pause site position, placed halfway between the poly (A) site and the mapped average location of the discontinuity. The pause duration (255 s) is derived from in vivo measurements (Darzacq et al., 2007). Observed average (< Δ >) and predicted values (Δ) for cotranscriptional splicing efficiency are shown. See also Movie S3. Dependence of cotranscriptional splicing levels on pause duration for genes with average architectural features of high-efficiency genes is shown in the right panel. Predicted cotranscriptional splicing efficiency (Δ) is plotted against pause duration in seconds.

than that of the high-efficiency group (Figure S5C). Compared to these short, highly transcribed intronless genes, the high-efficiency group displayed elevated Pol II density upstream of the poly(A) site (Figure S5D). Subtraction of the two traces reveals a discrete peak of Pol II upstream of the poly(A) site in the high-efficiency genes (Figure S5D), clearly distinguishing the two Pol II profiles from one another. We conclude that Pol II density specifically increases downstream of the discontinuity in high-efficiency genes only and that Pol II pile-up at the poly(A) site cannot account for this difference in elongation behavior.

Pol II density along the length of the gene reflects transcription elongation behavior. An increase in Pol II density in downstream regions can only occur if Pol II progresses more slowly across the gene region, a phenomenon generally referred to as "pausing." Transcriptional pausing downstream of the 3′SS and upstream of the poly(A) site should prolong the time available for cotranscriptional splicing catalysis. To test the effect of pausing before termination, we incorporated a discrete Pol II pausing event at an average position between the discontinuity and the poly(A) site (123 nt upstream of poly[A]) into our in silico simulation, using a published value for the pause duration (255 s) (Darzacq et al., 2007). In the absence of pausing, genes in the high-efficiency group were predicted to be entirely posttranscriptionally spliced (see Figure 3C). In contrast, addition of the pause site predicts high levels (Δ = 0.94) of cotranscriptional splicing (Figure 5C, left panel, and Movie S3). This value is similar to the average experimentally observed value in the high-efficiency group (< Δ > = 0.83). To address how pause duration might influence predicted cotranscriptional splicing, we simulated cotranscriptional intron removal for different pause durations (Figure 5C, right panel). Interestingly, pauses as short as ~60 s are sufficient to generate the average experimentally observed level of cotranscriptional splicing in the simulation. Thus, transcriptional pausing before poly(A) cleavage is capable of explaining the observed high cotranscriptional splicing efficiency determined experimentally for genes with short terminal exons.

DISCUSSION

It is well known that Pol II pauses at promoters and at poly(A) sites due to rate-limiting steps associated with promoter escape and termination, respectively (Fuda et al., 2009). Until now, whether additional sites of transcriptional pausing occur in eukaryotes has been unknown. Here, we report a distinct Pol II pausing event, which begins ~250 nt upstream of poly(A) sites within the terminal exons of cotranscriptionally spliced genes in yeast. The detected pause, characterized by depletion of nascent RNA sequences and increased Pol II density over the region, accounts for the unexpectedly high rate of cotranscriptional splicing detected by high-density tiling microarray analysis (Figure 6). Remarkably, the genes with the highest rates of cotranscriptional splicing are those with the shortest terminal exons. Simulation of transcription and splicing confirmed that, given published elongation and splicing rates, these short genes cannot be cotranscriptionally spliced unless Pol II pauses before termination. Genes with longer exons displayed average cotranscriptional splicing levels similar to predicted values; thus, transcription and splicing may occur in "kinetic competition"

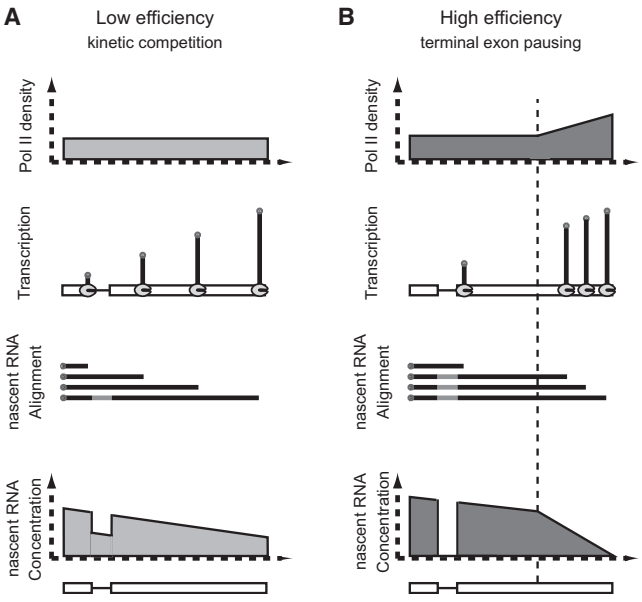

Figure 6. RNA Polymerase II Pauses for Splicing

(A and B) Schematic summary of our findings. Genes in the low-efficiency group show a uniform Pol II density along the gene, resulting in a constant decrease of nascent RNA abundance (A). Genes in the high-efficiency group are characterized by increased Pol II density toward gene ends, leading to a decrease in nascent 3′ end RNA sequences upstream of the poly(A) site (B). The slowing of transcription elongation begins at the mapped discontinuity (vertical dotted line) before poly(A) cleavage, increasing the time available for cotranscriptional spliceosome assembly and catalysis in these genes with short terminal exons.

for these genes (Mason and Struhl, 2005; Tardiff et al., 2006). Taken together, cotranscriptional splicing is the rule rather than the exception in budding yeast, an unexpected finding due to the shortness of yeast terminal exons (Tardiff et al., 2006). Because the terminal exon pausing described here was only detected in genes exhibiting a high degree of cotranscriptional splicing, its presence appears to ensure that transcripts harboring short exons are spliced before poly(A) cleavage and mRNA release from chromatin.

The observed change in Pol II behavior detected in the genes characterized by high-efficiency cotranscriptional splicing is designated terminal exon pausing, which appears to be specifically related to splicing for the following reasons. First, analysis of nascent RNA traces in the high-efficiency group of genes detected a highly significant change in elongation behavior within terminal exons. The gene position at which Pol II elongation changed was mapped to ~250 nt upstream of the poly(A) site by determining the point of discontinuity between two slopes fit to nascent RNA traces over terminal exons. Second, pausing is defined as a relative increase in Pol II density over a gene region (Fuda et al., 2009). Analysis of Pol II ChIP-on-chip data revealed that the mapped discontinuity in the high-efficiency group of genes marked the onset of increased Pol II density immediately downstream of the discontinuity. This demonstrates that Pol II density is elevated over the region where nascent RNA concentration is low, as expected. Discontinuities mapped in low-efficiency or intronless genes were associated with only minor changes in Pol II density, consistent with less robust changes in slope (~2-fold) observed for these gene groups as compared to the high-efficiency group (~10-fold). Third, to explain the high level of cotranscriptional splicing observed in the high-efficiency group of genes, Pol II elongation must slow down before termination. The only alternative explanation for the experimental data would be that splicing kinetics varies dramatically among genes, such that splicing of the high-efficiency group would take place much faster than in the low-efficiency group. To date, there is no evidence that rates of splicing catalysis differ from gene to gene. Because our data are well explained by terminal exon pausing, we do not currently have any reason to invoke differences in splicing rates.

A prominent concept in the field is that eukaryotic RNA processing steps—5′ end capping, splicing, and polyadenylation—are coupled to transcription (Moore and Proudfoot, 2009; Perales and Bentley, 2009). Yet evidence that splicing and transcription are mechanistically linked has been so far lacking. Pausing at transcriptional start sites is thought to reflect aspects of promoter escape, such as phosphorylation of the Pol II C-terminal domain (CTD) on serine 5 residues of the CTD's heptad repeat (Buratowski, 2009; Fuda et al., 2009). CTD phosphorylation is required for efficient recruitment of the capping machinery, ensuring that capping occurs immediately after transcription initiation. A corollary to this scenario is that promoter-proximal pausing may additionally contribute time necessary for capping to occur. Indeed, classically studied paused genes, such as *FOS* and heat shock genes, undergo transcription of the first ~50 nt of the transcript before pausing, and this short paused transcript is already capped (Listerman et al., 2006; Rasmussen and Lis, 1993). Similarly, pausing near the poly(A) site during transcription termination is associated with RNA processing, because poly(A) cleavage precedes Pol II release from the DNA template. In this case, poly(A) cleavage and mRNA release appear to occur much faster than transcription termination and Pol II release (Boireau et al., 2007). These tight links between pausing and RNA processing at gene ends have prompted speculation that regulation of transcription elongation rate may impact splicing outcome (Moore and Proudfoot, 2009; Perales and Bentley, 2009). The present demonstration that Pol II pauses at a consistent position in the majority of intron-containing genes in yeast provides evidence that splicing and transcription are directly coupled.

The precise mechanisms for Pol II pausing at any of these sites are unknown. In principle, Pol II pausing represents a net decrease in the rate of transcription elongation. Different mechanisms can in theory result in such a decrease in elongation speed: (1) Pol II can enter into backtracking behavior, in which the holoenzyme slips in reverse and can no longer elongate the nascent transcript until it is cleaved in the catalytic site, providing a new 3′ end for continued synthesis (Galburt et al., 2007). (2) Rearrangement of the catalytic center can cause decreased nucleotide addition, resulting in slower elongation rates (Landick, 2006). (3) Finally, *trans*-acting factors, like elongation factors, can regulate elongation and pausing (Buratowski, 2009). These pausing mechanisms can be caused or aided by chromatin environments hindering polymerase movement: Differences in the

stability of the RNA:DNA duplex in the catalytic core as well as positioned nucleosomes can induce backtracking of Pol II (Hodges et al., 2009; Nechaev et al., 2010). Nucleosomes tend to be positioned over internal exons in the human genome (Schwartz et al., 2009; Spies et al., 2009; Tilgner et al., 2009), raising the possibility that Pol II may pause over human exons, contributing to the time available for cotranscriptional RNA processing. Although positioning of yeast nucleosomes around the discontinuity described here was not detected (data not shown), a role for chromatin cannot be ruled out at present. Alternatively, protein-protein interactions among the transcription and RNA processing machineries involved could lead to conformational changes in or exert forces on Pol II. Regardless of the mechanistic explanation, this work establishes that all major mRNA processing steps are associated with transcriptional pausing at specific sites in genes.

The discovery of a transcriptional pause site within terminal exons demonstrates biologically relevant coupling between transcription and splicing and suggests that transcriptional mechanisms have evolved to ensure cotranscriptional splicing. In prokaryotes, coupling of transcription and translation allows for regulatory interactions that are crucial for both aspects of gene expression (Landick, 2006). In eukaryotes, translation is spatially and functionally separated from transcription, yet the structure and function of RNA polymerases are highly conserved. Perhaps transcriptional pausing has been co-opted by eukaryotes to facilitate RNA processing steps that are absent in prokaryotes. Interestingly, a well-conserved aspect of gene architecture among yeast and humans is the presence long terminal exons. In humans, typical internal exons are ~130 nt long, whereas typical terminal exons are ~950 nt long (Scherer, 2008). It is tempting to speculate that longer terminal exons generally facilitate cotranscriptional splicing. However, it is clear that terminal exon length alone is not sufficient to specify cotranscriptional splicing, given published elongation rates; many terminal exons in yeast and human are simply too short. The observation that Pol II pauses for splicing suggests that it is important that transcripts are spliced cotranscriptionally; cotranscriptional splicing may ensure proper mRNP formation, assembly of nuclear export machinery, and execution of surveillance mechanisms (Moore and Proudfoot, 2009; Perales and Bentley, 2009). It will be of interest to determine whether terminal exon pausing, described here in yeast, is present in higher eukaryotes.

EXPERIMENTAL PROCEDURES

Yeast Strains

All experiments conducted here were performed using the *S. cerevisiae* strain BY4741 (Mat a; his3D1; leu2D0; met15D0; ura3D0) obtained commercially (Euroscarf; Frankfurt, Germany).

Chromatin Fractionation

All steps were performed at 4°C, if not stated otherwise. A cell aliquot (see Supplemental Information) was resuspended in 1 ml buffer 1 (20 mM HEPES [pH 8.0], 60 mM KCl, 15 mM NaCl, 10 mM $MgCl_2$, 1 mM $CaCl_2$, 10 mM N-butyric acid, 0.8% Triton X-100, 0.25 M sucrose, 2.5 mM spermidine, 0.5 mM spermine) (Buchanan et al., 2009). The cell suspension was transferred to a 2 ml plastic tube containing 1.5 ml zirconia beads (Roth; Karlsruhe, Germany). Cells were lysed by 4 × 30" pulses of bead-beating interrupted by 30" pauses on ice. Beads were separated from the lysate by centrifugation. Residual intact cells were removed by centrifugation (5′ at 500 *g*). The supernatant was transferred into a fresh 50 ml plastic tube and centrifuged for 20′ at 2000 *g* (JLA 25.50). The pellet was resuspended in 1 ml buffer 1 and transferred to a 1.5 ml plastic tube. After centrifugation (20′ at 2000 *g*), the pellet was resuspended in 1 ml buffer 2 (20 mM HEPES [pH 7.6], 450 mM NaCl, 7.5 mM $MgCl_2$, 20 mM EDTA, 10% glycerol, 1% NP-40, 2 M urea, 0.5 M sucrose, 1 mM DTT, 0.125 mM PMSF) and centrifuged for 20′ at 20,000 *g*. The pellet was washed once with 1 ml buffer 2. The chromatin fraction was characterized for nucleic acid and protein components. For tiling microarray analysis, nascent RNA was isolated directly from the chromatin fraction. See Supplemental Information for further details.

Tiling Microarray Analysis

Strand-specific tiling microarray data were generated for nascent RNA, mRNA, and genomic DNA (see Supplemental Information). The median values for each probe of at least three independent biological replicates were determined. Nascent RNA and mRNA data sets were normalized to genomic DNA (Huber et al., 2006). Segmentation was performed on the normalized mRNA data set: The borders of initial segments were defined by neighboring local extrema extracted from a Gaussian blurred (σ = 40 nt) tiling microarray trace. Initial segments were grouped if the mean intensities were not different, judged by Student's t test. Gene-specific analysis was performed by matching segments with annotated features (http://www.yeastgenome.org/) (Nagalakshmi et al., 2008). The following analysis methods were performed on expressed genes only.

Cotranscriptional Splicing Efficiency Determination

Cotranscriptional splicing efficiency (Δ) was calculated for each expressed intron-containing gene on the Watson strand by fitting a linear function (linear regression) to the terminal exon probe intensities from the nascent RNA tiling microarray. To correct for hybridization bias, probe intensities were normalized to those of genomic DNA (Huber et al., 2006). To exclude bias from decreased intensity values toward gene ends, positions downstream of the mapped discontinuities (see below) were excluded. Extrapolated intensity values of the fit over intronic positions reflect the predicted intronic intensities, assuming no cotranscriptional splicing. Fraction of observed background-corrected intronic intensities relative to predicted values were determined for every intronic probe. For each gene, the median fraction was calculated from all the fractions determined at every probe position mapped to each annotated intron. Calculations were performed in linear space.

Grouping of Intron-Containing Genes

Grouping was based on observed versus predicted values for cotranscriptional splicing efficiencies (Δ). Predicted values were gained by in silico simulation of transcription and splicing (see Supplemental Information). Both transcription and splicing were simulated employing a discrete-event modeling framework (DESMO-J, http://desmoj.sourceforge.net/) assuming solely kinetic coupling (i.e., transcriptional termination and splicing compete kinetically). Genes with $\Delta_{Observed} \geq \Delta_{Predicted} + 0.5$ were classified in the high-efficiency group, whereas those not matching this criterion were classified in the low-efficiency group.

Local Change of Nascent RNA Density

Local changes of nascent RNA density in 2823 intronless genes and 122 terminal exons were calculated by fitting two linear functions (linear regression) to the linear probe intensity values of the nascent RNA tiling microarray normalized to mRNA to remove hybridization bias (Huber et al., 2006). The start of the first fit is restricted to the 5′ end of the gene or terminal exon, whereas the end of the second fit is restricted to the 3′ end of the gene. The fits were optimized by varying the position of the third tether (end of Fit 1 and beginning of Fit 2) and minimizing the global sum of squares (Figure S4). The point of discontinuity in the slopes of both fits was mapped and used for further analysis.

Pol II Profiling

To correlate abundance and distribution of Pol II holoenzyme components (Rpb1, Rpb3) with cotranscriptional splicing efficiencies, ChIP-on-chip data from high-density tiling microarrays (Venters and Pugh, 2009) were extracted for all genes analyzed. Genes analyzed belonged to three groups, including the 121 most highly expressed intronless genes on the Watson strand (our data based on the median 5′ intensities, see Supplemental Information), 74 high- and 48 low-efficiently spliced genes, according to the groupings shown in Figure 3 and Table S2. Pol II ChIP-on-chip tiling traces were aligned to different gene architectural features, including the discontinuity determined in this study. Alignments were normalized to position coverage of the underlying microarray and blurred (Gaussian kernel, $\sigma = 2$ nt). Mean values for each group are plotted versus nucleotide distance from positions of interest.

ACCESSION NUMBERS

The high-density tiling microarray data sets described here can be retrieved from Gene Expression Omnibus with the accession number GSE24040.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, Supplemental References, five figures, three movies, and three tables and can be found with this article online at doi:10.1016/j.molcel.2010.11.004.

ACKNOWLEDGMENTS

We gratefully acknowledge Julia Jarrells for expert assistance with microarray analysis and Luke Buchanan and Francis Stewart for helpful suggestions on the development of the chromatin preparation. We thank Jonathon Howard, Stephan Grill, and Pavel Tomancak for their insights into the statistical analysis and comments on the manuscript. Sabine Rospert generously provided antibodies. This work was supported by the Max Planck Society and a grant from the European Commission (EURASNET-518238 to K.M.N.).

Received: April 22, 2010
Revised: July 7, 2010
Accepted: September 22, 2010
Published: November 23, 2010

REFERENCES

Boireau, S., Maiuri, P., Basyuk, E., de la Mata, M., Knezevich, A., Pradet-Balade, B., Bäcker, V., Kornblihtt, A., Marcello, A., and Bertrand, E. (2007). The transcriptional cycle of HIV-1 in real-time and live cells. J. Cell Biol. *179*, 291–304.

Brodsky, A.S., Meyer, C.A., Swinburne, I.A., Hall, G., Keenan, B.J., Liu, X.S., Fox, E.A., and Silver, P.A. (2005). Genomic mapping of RNA polymerase II reveals sites of co-transcriptional regulation in human cells. Genome Biol. *6*, R64.

Buchanan, L., Durand-Dubief, M., Roguev, A., Sakalar, C., Wilhelm, B., Strålfors, A., Shevchenko, A., Aasland, R., Shevchenko, A., Ekwall, K., and Francis Stewart, A. (2009). The Schizosaccharomyces pombe JmjC-protein, Msc1, prevents H2A.Z localization in centromeric and subtelomeric chromatin domains. PLoS Genet. *5*, e1000726.

Buratowski, S. (2009). Progression through the RNA polymerase II CTD cycle. Mol. Cell *36*, 541–546.

Cai, H., and Luse, D.S. (1987). Transcription initiation by RNA polymerase II in vitro. Properties of preinitiation, initiation, and elongation complexes. J. Biol. Chem. *262*, 298–304.

Core, L.J., and Lis, J.T. (2008). Transcription regulation through promoter-proximal pausing of RNA polymerase II. Science *319*, 1791–1792.

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

Darzacq, X., Shav-Tal, Y., de Turris, V., Brody, Y., Shenoy, S.M., Phair, R.D., and Singer, R.H. (2007). In vivo dynamics of RNA polymerase II transcription. Nat. Struct. Mol. Biol. *14*, 796–806.

Das, R., Yu, J., Zhang, Z., Gygi, M.P., Krainer, A.R., Gygi, S.P., and Reed, R. (2007). SR proteins function in coupling RNAP II transcription to pre-mRNA splicing. Mol. Cell *26*, 867–881.

David, L., Huber, W., Granovskaia, M., Toedling, J., Palm, C.J., Bofkin, L., Jones, T., Davis, R.W., and Steinmetz, L.M. (2006). A high-resolution map of transcription in the yeast genome. Proc. Natl. Acad. Sci. USA *103*, 5320–5325.

de la Mata, M., Alonso, C.R., Kadener, S., Fededa, J.P., Blaustein, M., Pelisch, F., Cramer, P., Bentley, D., and Kornblihtt, A.R. (2003). A slow RNA polymerase II affects alternative splicing in vivo. Mol. Cell *12*, 525–532.

Fabrizio, P., Dannenberg, J., Dube, P., Kastner, B., Stark, H., Urlaub, H., and Lührmann, R. (2009). The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. Mol. Cell *36*, 593–608.

Fuda, N.J., Ardehali, M.B., and Lis, J.T. (2009). Defining mechanisms that regulate RNA polymerase II transcription in vivo. Nature *461*, 186–192.

Galburt, E.A., Grill, S.W., Wiedmann, A., Lubkowska, L., Choy, J., Nogales, E., Kashlev, M., and Bustamante, C. (2007). Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner. Nature *446*, 820–823.

Görnemann, J., Kotovic, K.M., Hujer, K., and Neugebauer, K.M. (2005). Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. Mol. Cell *19*, 53–63.

Hicks, M.J., Yang, C.R., Kotlajich, M.V., and Hertel, K.J. (2006). Linking splicing to Pol II transcription stabilizes pre-mRNAs and influences splicing patterns. PLoS Biol. *4*, e147.

Hodges, C., Bintu, L., Lubkowska, L., Kashlev, M., and Bustamante, C. (2009). Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II. Science *325*, 626–628.

Holstege, F.C., Jennings, E.G., Wyrick, J.J., Lee, T.I., Hengartner, C.J., Green, M.R., Golub, T.R., Lander, E.S., and Young, R.A. (1998). Dissecting the regulatory circuitry of a eukaryotic genome. Cell *95*, 717–728.

Huber, W., Toedling, J., and Steinmetz, L.M. (2006). Transcript mapping with high-density oligonucleotide tiling arrays. Bioinformatics *22*, 1963–1970.

Kotovic, K.M., Lockshon, D., Boric, L., and Neugebauer, K.M. (2003). Cotranscriptional recruitment of the U1 snRNP to intron-containing genes in yeast. Mol. Cell. Biol. *23*, 5768–5779.

Lacadie, S.A., and Rosbash, M. (2005). Cotranscriptional spliceosome assembly dynamics and the role of U1 snRNA:5′ss base pairing in yeast. Mol. Cell *19*, 65–75.

Lacadie, S.A., Tardiff, D.F., Kadener, S., and Rosbash, M. (2006). In vivo commitment to yeast cotranscriptional splicing is sensitive to transcription elongation mutants. Genes Dev. *20*, 2055–2066.

Landick, R. (2006). The regulatory roles and mechanism of transcriptional pausing. Biochem. Soc. Trans. *34*, 1062–1066.

Lazarev, D., and Manley, J.L. (2007). Concurrent splicing and transcription are not sufficient to enhance splicing efficiency. RNA *13*, 1546–1557.

Listerman, I., Sapra, A.K., and Neugebauer, K.M. (2006). Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. Nat. Struct. Mol. Biol. *13*, 815–822.

Mason, P.B., and Struhl, K. (2005). Distinction and relationship between elongation rate and processivity of RNA polymerase II in vivo. Mol. Cell *17*, 831–840.

Moore, M.J., and Proudfoot, N.J. (2009). Pre-mRNA processing reaches back to transcription and ahead to translation. Cell *136*, 688–700.

Muñoz, M.J., Pérez Santangelo, M.S., Paronetto, M.P., de la Mata, M., Pelisch, F., Boireau, S., Glover-Cutter, K., Ben-Dov, C., Blaustein, M., Lozano, J.J., et al. (2009). DNA damage regulates alternative splicing through inhibition of RNA polymerase II elongation. Cell *137*, 708–720.

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., and Snyder, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. Science *320*, 1344–1349.

Nechaev, S., Fargo, D.C., dos Santos, G., Liu, L., Gao, Y., and Adelman, K. (2010). Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. Science *327*, 335–338.

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

Perales, R., and Bentley, D. (2009). "Cotranscriptionality": the transcription elongation complex as a nexus for nuclear transactions. Mol. Cell *36*, 178–191.

Rasmussen, E.B., and Lis, J.T. (1993). In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes. Proc. Natl. Acad. Sci. USA *90*, 7923–7927.

Roberts, G.C., Gooding, C., Mak, H.Y., Proudfoot, N.J., and Smith, C.W. (1998). Co-transcriptional commitment to alternative splice site selection. Nucleic Acids Res. *26*, 5568–5572.

Scherer, S. (2008). A Short Guide to the Human Genome (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Schwartz, S., Meshorer, E., and Ast, G. (2009). Chromatin organization marks exon-intron structure. Nat. Struct. Mol. Biol. *16*, 990–995.

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

Spies, N., Nielsen, C.B., Padgett, R.A., and Burge, C.B. (2009). Biased chromatin signatures around polyadenylation sites and exons. Mol. Cell *36*, 245–254.

Staley, J.P., and Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. Cell *92*, 315–326.

Tardiff, D.F., Lacadie, S.A., and Rosbash, M. (2006). A genome-wide analysis indicates that yeast pre-mRNA splicing is predominantly posttranscriptional. Mol. Cell *24*, 917–929.

Tilgner, H., Nikolaou, C., Althammer, S., Sammeth, M., Beato, M., Valcárcel, J., and Guigó, R. (2009). Nucleosome positioning as a determinant of exon recognition. Nat. Struct. Mol. Biol. *16*, 996–1001.

Venters, B.J., and Pugh, B.F. (2009). A canonical promoter organization of the transcription machinery and its regulators in the Saccharomyces genome. Genome Res. *19*, 360–371.

Wada, Y., Ohta, Y., Xu, M., Tsutsumi, S., Minami, T., Inoue, K., Komura, D., Kitakami, J., Oshida, N., Papantonis, A., et al. (2009). A wave of nascent transcription on activated human genes. Proc. Natl. Acad. Sci. USA *106*, 18357–18361.

Wahl, M.C., Will, C.L., and Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell *136*, 701–718.

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

Zenklusen, D., Larson, D.R., and Singer, R.H. (2008). Single-RNA counting reveals alternative modes of gene expression in yeast. Nat. Struct. Mol. Biol. *15*, 1263–1271.

Zhang, Z., Hesselberth, J.R., and Fields, S. (2007). Genome-wide identification of spliced introns using a tiling microarray. Genome Res. *17*, 503–509.