



# Cotranscriptional and Posttranscriptional Features of the Transcriptome in Soybean Shoot Apex and Leaf

Jiafu Zhu<sup>1</sup>, Han Zhao<sup>2</sup>, Fanjiang Kong<sup>1</sup>, Baohui Liu<sup>1</sup>, Min Liu<sup>1\*</sup> and Zhicheng Dong<sup>1\*</sup>

<sup>1</sup> Guangzhou Key Laboratory of Crop Gene Editing, Innovative Center of Molecular Genetics and Evolution, School of Life Sciences, Guangzhou Higher Education Mega Center, Guangzhou University, Guangzhou, China, <sup>2</sup> Provincial Key Laboratory of Agrobiolology, Institute of Crop Germplasm and Biotechnology, Jiangsu Academy of Agricultural Sciences, Nanjing, China

## OPEN ACCESS

### Edited by:

Mingli Xu,  
University of South Carolina,  
United States

### Reviewed by:

Rui Xia,  
South China Agricultural University,  
China  
Yalong Guo,  
Institute of Botany, Chinese Academy  
of Sciences, China

### \*Correspondence:

Min Liu  
minl@gzhu.edu.cn  
Zhicheng Dong  
zc\_dong@gzhu.edu.cn

### Specialty section:

This article was submitted to  
Plant Development and EvoDevo,  
a section of the journal  
Frontiers in Plant Science

**Received:** 05 January 2021

**Accepted:** 02 March 2021

**Published:** 09 April 2021

### Citation:

Zhu J, Zhao H, Kong F, Liu B,  
Liu M and Dong Z (2021)  
Cotranscriptional  
and Posttranscriptional Features  
of the Transcriptome in Soybean  
Shoot Apex and Leaf.  
Front. Plant Sci. 12:649634.  
doi: 10.3389/fpls.2021.649634

Transcription is the first step of central dogma, in which the genetic information stored in DNA is copied into RNA. In addition to mature RNA sequencing (RNA-seq), high-throughput nascent RNA assays have been established and applied to provide detailed transcriptional information. Here, we present the profiling of nascent RNA from trifoliolate leaves and shoot apices of soybean. In combination with nascent RNA (chromatin-bound RNA, CB RNA) and RNA-seq, we found that introns were largely spliced cotranscriptionally. Although alternative splicing (AS) was mainly determined at nascent RNA biogenesis, differential AS between the leaf and shoot apex at the mature RNA level did not correlate well with cotranscriptional differential AS. Overall, RNA abundance was moderately correlated between nascent RNA and mature RNA within each tissue, but the fold changes between the leaf and shoot apex were highly correlated. Thousands of novel transcripts (mainly non-coding RNA) were detected by CB RNA-seq, including the overlap of natural antisense RNA with two important genes controlling soybean reproductive development, *FT2a* and *Dt1*. Taken together, we demonstrated the adoption of CB RNA-seq in soybean, which may shed light on gene expression regulation of important agronomic traits in leguminous crops.

**Keywords:** soybean, chromatin-bound RNA, co-transcriptional splicing, non-coding RNA, nascent RNA

## INTRODUCTION

Transcription, the first step of gene expression, is accomplished by the multisubunit protein complex RNA polymerase. In eukaryotic cells, RNA polymerase II (RNA Pol II) is involved in protein-coding gene transcription and some non-coding gene transcription. Before maturation, messenger RNA precursors (pre-mRNAs) are subjected to multiple processing steps, including 5' capping, splicing of introns, 3' cleavage and polyadenylation, and editing (Bentley, 2014). These steps are known as posttranscriptional processing. However, increasing evidence suggests that most processes are cotranscriptional. For example, introns can be either co- or posttranscriptionally spliced, which is supported by the splicing loops of nascent RNA observed by electron microscopy in *Drosophila melanogaster* and *Chironomus tentans* (Beyer and Osheim, 1988; Baurén and Wieslander, 1994). In addition, high-throughput sequencing of nascent RNA revealed genome-wide cotranscriptional splicing (Khodor et al., 2011; Nojima et al., 2015; Drexler et al., 2020). Studies

from budding yeast, flies, and mammals indicated that cotranscriptional splicing frequencies are similarly high, ranging from 75 to 85% (Neugebauer, 2019).

Since Core et al. (2008) published a method wherein the nuclei run on RNA were affinity purified followed by high-throughput sequencing, nascent RNA sequencing (RNA-seq) technologies have significantly improved our ability to analyze transcription at each step across the genome. Rather than steady-state mRNA, nascent RNA-seq detects pre-mRNAs, divergent transcripts, enhancer-derived RNA (eRNA), etc., which are usually unstable and not polyadenylated. Recently, we and another laboratory have reported cotranscriptional splicing in the model plant *Arabidopsis* using genome-wide nascent RNA-seq approaches, plant native elongating transcript sequencing (pNET-seq), and plaNET-seq (Zhu et al., 2018; Kindgren et al., 2020). pNET-seq and plaNET-seq detect nascent RNA through enrichment of transcriptionally engaged RNA Pol II complexes, and splicing intermediates can also be observed when some spliceosomes are copurified with Pol II complexes (Zhu et al., 2018). Moreover, three recent publications directly sequenced the chromatin-bound RNA (CB RNA) of *Arabidopsis* and found genome-wide cotranscriptional splicing (Jia et al., 2020; Li et al., 2020; Zhu et al., 2020). However, the *Arabidopsis* genome is the first plant genome to be sequenced and is compact (140 million base/haploid genome), with an average gene length of 2,000 bp and an average intron length of 180 bp (Arabidopsis Genome Initiative, 2000). While harboring thousands to tens of thousands of genes, plant genome size ranges from approximately 0.1 to 100 gigabases (Pellicer and Leitch, 2020). Therefore, knowledge of transcription obtained from *Arabidopsis* may not be applicable to other plant genomes, especially some complicated crop genomes.

As one of the most important crops, soybean provides protein and oil for humans and livestock. During the past decades, great progress has been made in soybean genome research (Shen et al., 2018; Xie et al., 2019; Liu Y. et al., 2020). Furthermore, many important genes involved in agronomic traits have been characterized *via* genetic, cellular biology, and biochemical approaches (Kasai et al., 2007; Lu et al., 2017, 2020). For example, *Dt1*, which controls soybean growth habits, has been cloned as a *TFL1* homolog encoding a 173-amino-acid peptide (Liu et al., 2010; Tian et al., 2010). *FT2a* and *FT5a*, two distant homologous genes of *Dt1* within the same family, have been shown to play a conserved role in controlling flowering time (Kong et al., 2010; Takeshima et al., 2019; Wu et al., 2017).

Soybean [*Glycine max* (L.) Merr.] is a paleopolyploid derived from two whole genome duplication events approximately 59 and 13 million years ago. It has a relatively complicated and large genome, with a size of approximately 1.1 gigabases (Schmutz et al., 2010). The average gene length is approximately 4,000 bp, and the average intron length is approximately 539 bp in soybean (Shen et al., 2014), which are longer than those in *Arabidopsis*. Despite the considerable transcriptomic analyses of various soybean tissues using mature RNA-seq (Libault et al., 2010; Severin et al., 2010; Shen et al., 2014; Wang et al., 2014; Gazara et al., 2019), genome-wide analysis of nascent RNA from soybean has not yet been reported. In addition to capturing

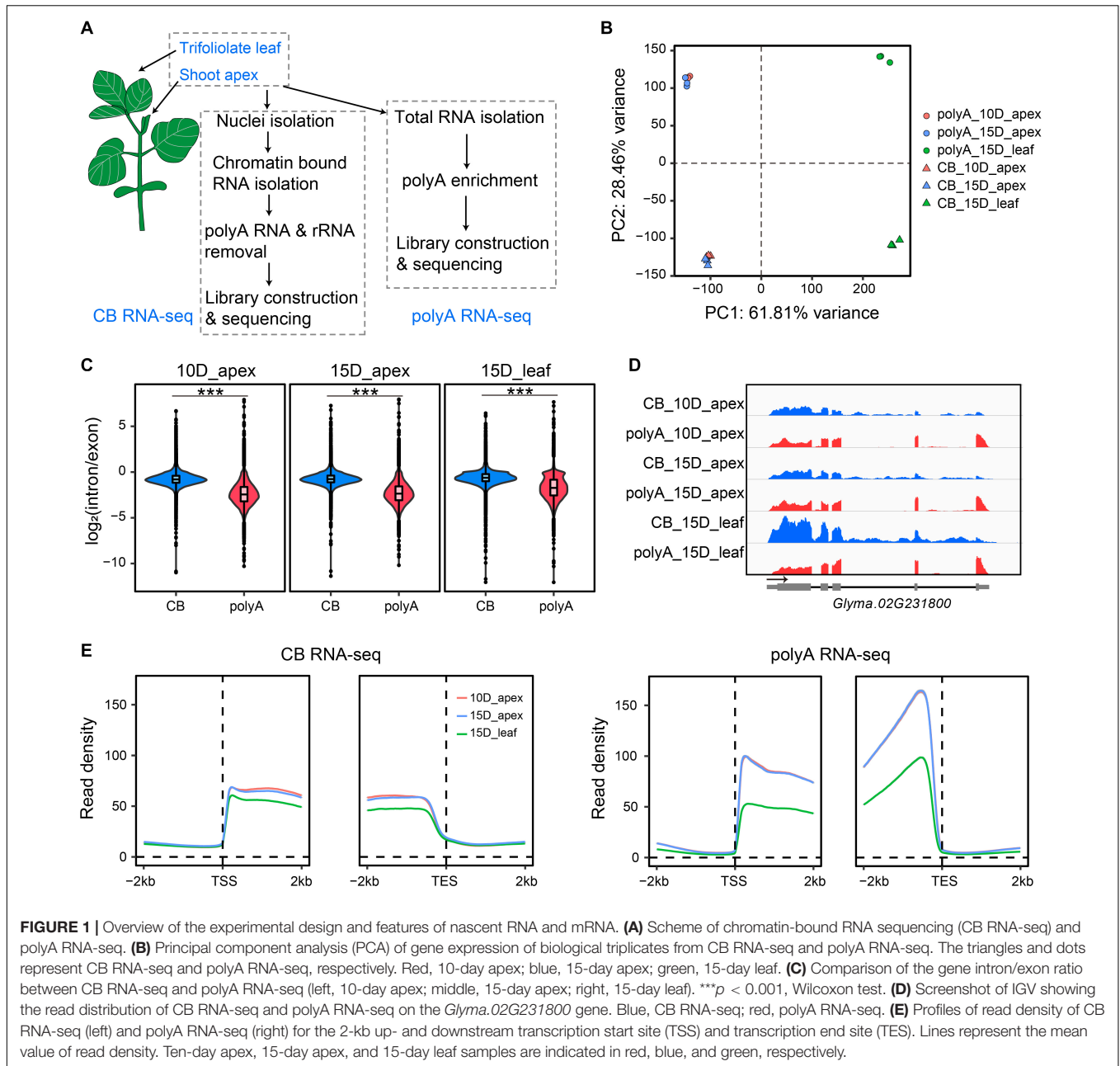
cotranscriptional features, nascent RNA is very sensitive to the detection of unstable regulatory RNAs, such as long non-coding RNAs (ncRNAs). Therefore, the investigation on nascent RNA in soybean would provide a comprehensive description of cotranscriptional characteristics in leguminous crops. Here, we report for the first time the analysis of nascent RNA from the shoot apex and leaf tissues of the soybean cv. Williams 82.

## RESULTS

### Nascent RNA Profiling of Soybean by CB RNA-Seq

The spatial and temporal expression of genes in the shoot apex largely determines the architecture of crop plants, including the numbers of branches, flowers, and nodes, which finally affect the yield per plant. Specifically, mRNA of *Dt1* was detected in the shoot apex at 15 days after emergence under a long-day condition (Liu et al., 2010); therefore, we set to investigate the transcriptome of the shoot apex from 10- to 15-day-old plants (Figure 1A, see section “Materials and Methods”). To gain insights of the shoot apex-specific gene, we chose the first trifoliolate leaves from 15-day-old plants as control. For nascent RNA, CB RNA was isolated, and the rRNA and polyA RNA it contained were depleted prior to library construction and high-throughput sequencing as described by Zhu et al. (2020). To further reveal cotranscriptional and posttranscriptional processes, we also conducted parallel mature polyA RNA-seq by enriching polyA RNA from total RNA, and these RNAs were constructed into libraries. Three biological replicates were sequenced and analyzed for each tissue. Principal component analysis (PCA) and Pearson correlation analysis of gene expression indicated high reproducibility of biological replication (Figure 1B and Supplementary Figure 1). In addition, the first two components of PCA explained more than 90% of the variation, indicating that the tissue difference (apex vs. leaf, 61.81% of variance) and methodological difference (CB RNA-seq vs. polyA RNA-seq, 28.46% of variance) were the dominant factors for intersample differentiation (Figure 1B).

As expected, the read distribution of nascent RNA shows two characteristics compared with that of polyA RNA. First, CB RNA-seq detected more intron signals than polyA RNA-seq because more unspliced reads were sequenced at the nascent RNA level. Approximately 25% of unique mapped reads were located in the intron region with CB RNA-seq, while less than 4% of unique mapped reads were located in the intron region with polyA RNA-seq (Supplementary Figure 2). In addition, the read density ratio of introns to exons in CB RNA was significantly higher than that in polyA RNA (Figure 1C). Second, the read density on the gene decreased gradually from the 5' end to the 3' end, while there was no such phenomenon in polyA RNA (Figures 1D,E). For example, the read signal of the gene *Glyma.02G231800* declined from 5' to 3' in CB RNA-seq but not in polyA RNA-seq. Furthermore, an intron signal was evident in CB RNA but absent from polyA RNA (Figure 1D). These characteristics were consistent with the results from previous studies and confirmed

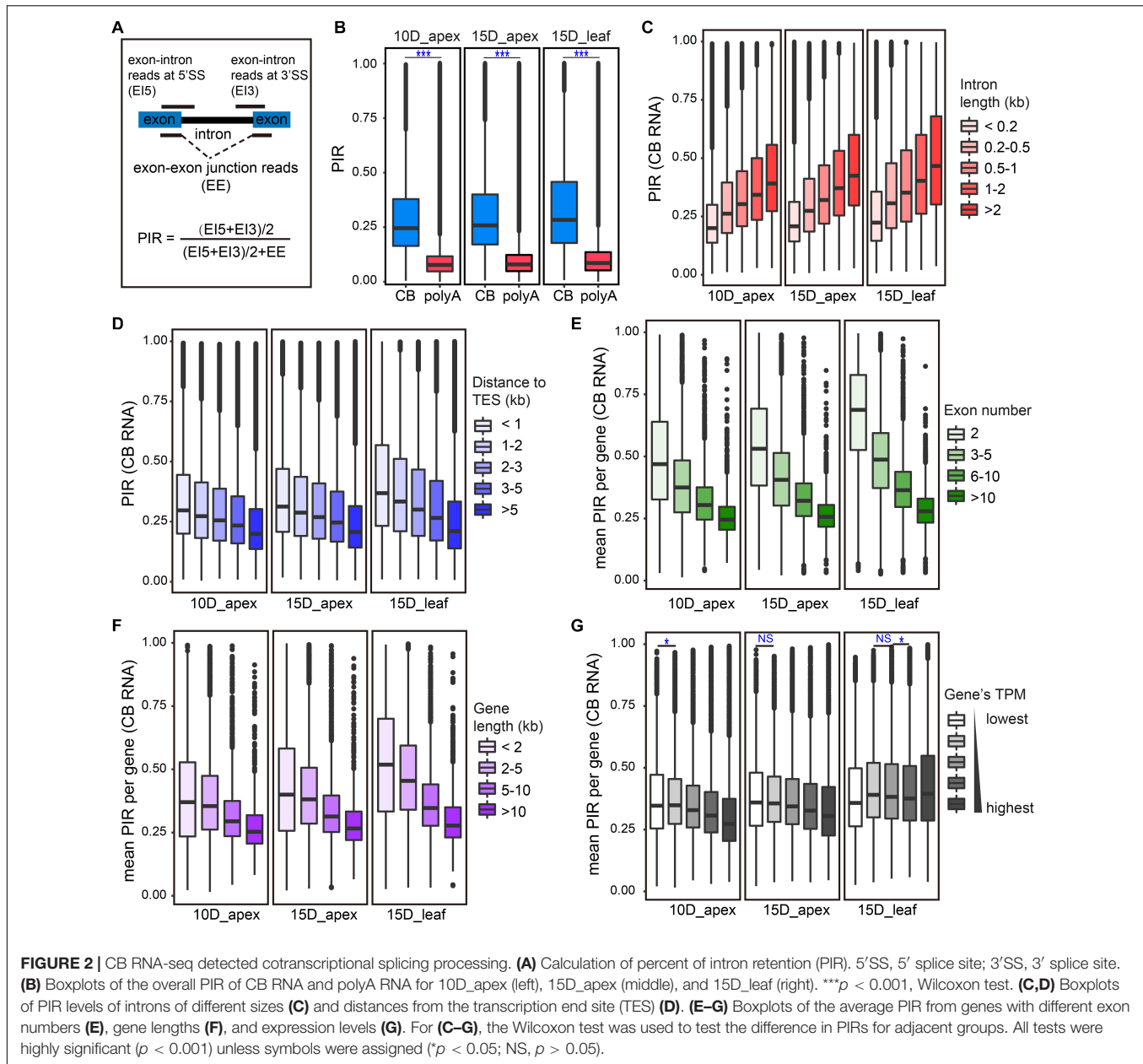


that the CB RNAs obtained here were bona fide transcriptional processing nascent RNAs (Li et al., 2020; Zhu et al., 2020).

## Multiple Factors Regulate Cotranscriptional Splicing Efficiency

Cotranscriptional splicing has been widely found in eukaryotic cells. We wondered whether splicing coupled with transcription is widespread in the soybean genome. The intron retention ratio is an indicator of intron splicing efficiency. Thus, we adopted an index for the percent of intron retention (PIR) to measure the extent of cotranscriptional splicing (Braunschweig et al., 2014). In short, the PIR of an intron was calculated as the ratio of unspliced

exon–intron junction reads to the total junction reads (unspliced exons–introns and spliced exons–exons). Since each unspliced exon–intron read from one RNA molecule has the chance to be sequenced twice in high-throughput sequencing, the average count of exon–intron reads at the 5' splice site (EI5) and of exon–intron reads at the 3' splice site (EI3) was considered an intron's unspliced exon–intron read count (**Figure 2A**). Introns with lower PIR values are more efficient for splicing. Constitutive introns of active genes (TPM > 1) were calculated for PIR both in CB RNA and polyA RNA. As expected, the intron retention levels of CB RNA were significantly higher than those of polyA RNA, both in the apex and leaf (**Figure 2B**). Most introns in polyA RNA have a very low PIR, usually smaller than 0.1. The



median PIR was close to 0.25 (in the apex) or above 0.25 (in the leaf) in CB RNA. These results were similar to those of a previous study of *Arabidopsis* (Li et al., 2020). The PIR of most introns in CB RNA was lower than 0.5 (PIR = 1 means completely unspliced), indicating the existence of genome-wide cotranscriptional splicing in soybean.

Although most introns undergo cotranscriptional splicing, the extent of intron retention is highly variable. Studies in *Drosophila* and *Arabidopsis* have indicated that multiple factors, such as intron characteristics, gene expression level, and number of introns, are related to cotranscriptional splicing efficiency (Khodor et al., 2011; Li et al., 2020; Zhu et al., 2020). To examine how these factors affect the splicing efficiency in soybean, we first divided introns into five groups by length and found that

intron retention became more prominent as the intron length increased (Figure 2C).

In addition to intron length, the intron position is also supposed to influence splicing efficiency. According to the “first come, first served” model, there may be more splicing chances for introns transcribed first (Aebi et al., 1986). Based on the distance to transcription end sites (TES), introns were divided into five groups, and the PIR was compared among groups. Introns more distant from TES are transcribed early and thus are more likely to be spliced first. As expected, the PIR index gradually declined as the intron distance to TES decreased (Figure 2D).

In addition, the cotranscriptional splicing efficiency was positively correlated with exon number (Figure 2E) and gene length (Figure 2F). These patterns were consistent between the

apex and leaf tissues. However, a weak positive correlation of cotranscriptional splicing and gene expression was detected in the apex instead of in the leaf (Figure 2G).

## Cotranscriptional Splicing Efficiency Is Correlated With Certain Histone Modifications

Specific histone modifications have been shown to regulate cotranscriptional splicing by either directly recruiting spliceosomes or indirectly influencing transcriptional elongation (Luco et al., 2010; Hu et al., 2020). To test whether cotranscriptional splicing is associated with certain histone modifications in soybean, we used ChIP-seq data of several histone modifications (H3K27me3, H3K4me1, H3K4me3, H3K36me3, H3K56ac, and H2A.Z) in leaf tissue collected from a previous study (Supplementary Table 1; Lu et al., 2019). We then quantified the level of different histone modifications around introns in different groups based on the retention rates (Figure 3). PIR is positively correlated with the levels of H3K27me3, H3K4me3, H3K56ac, and H2A.Z-marked histone, which means that introns with higher cotranscriptional splicing efficiency have lower levels of those histone modifications. PIR is negatively correlated with the level of H3K4me1-marked histones. Notably, H3K27me3, H3K4me3, H3K56ac, H3K36me3, and H2A.Z showed a higher modification level at the upstream exon than at the downstream exon, while H3K4me1 showed a higher modification level at the downstream exon. It is most likely that these histone modifications, H3K27me3, H3K4me3, H3K56ac, H3K36me3, and H2A.Z, preferentially locate at the gene's 5' end, except for H3K4me1 (Supplementary Figure 3).

## Alternative Splicing Events Are Likely Determined Cotranscriptionally

In higher eukaryotes, alternative splicing (AS), as an important regulatory step of gene expression, plays a critical role in the development and stress response of organisms (Baralle and Giudice, 2017; Laloum et al., 2018). Previous studies in mammalian cells and *Arabidopsis* showed that AS events occur co- or post-transcriptionally (Jia et al., 2020). Thus, we wondered to what extent AS is determined cotranscriptionally. We adopted percent spliced-in (PSI) (Wang et al., 2008) to describe the relative abundance of splicing events. We focused on four AS events: alternative 3' splice sites (A3SS), alternative 5' splice sites (A5SS), exon skipping (ES), and retained introns (RI) (Figure 4A). The PSI values of AS events from CB RNA and polyA RNA were significantly correlated, suggesting that AS events are likely determined cotranscriptionally for all AS types (Figure 4B). This was true for both shoot apex and leaf tissues (Figure 4 and Supplementary Figure 4). However, the overall PSI value was higher in CB RNA (Figure 4B, insets). For AS events with a higher PSI in CB RNA than in polyA RNA, there are two possible explanations. First, some highly abundant transcripts in CB RNA with AS events may likely be rapidly degraded. For example, coupling of AS and nonsense-mediated mRNA decay (NMD) has been reported to fine-tune gene expression (McGlinchy and Smith, 2008). Second,

posttranscriptional splicing may lead to a higher PSI in CB RNA, especially for RI events.

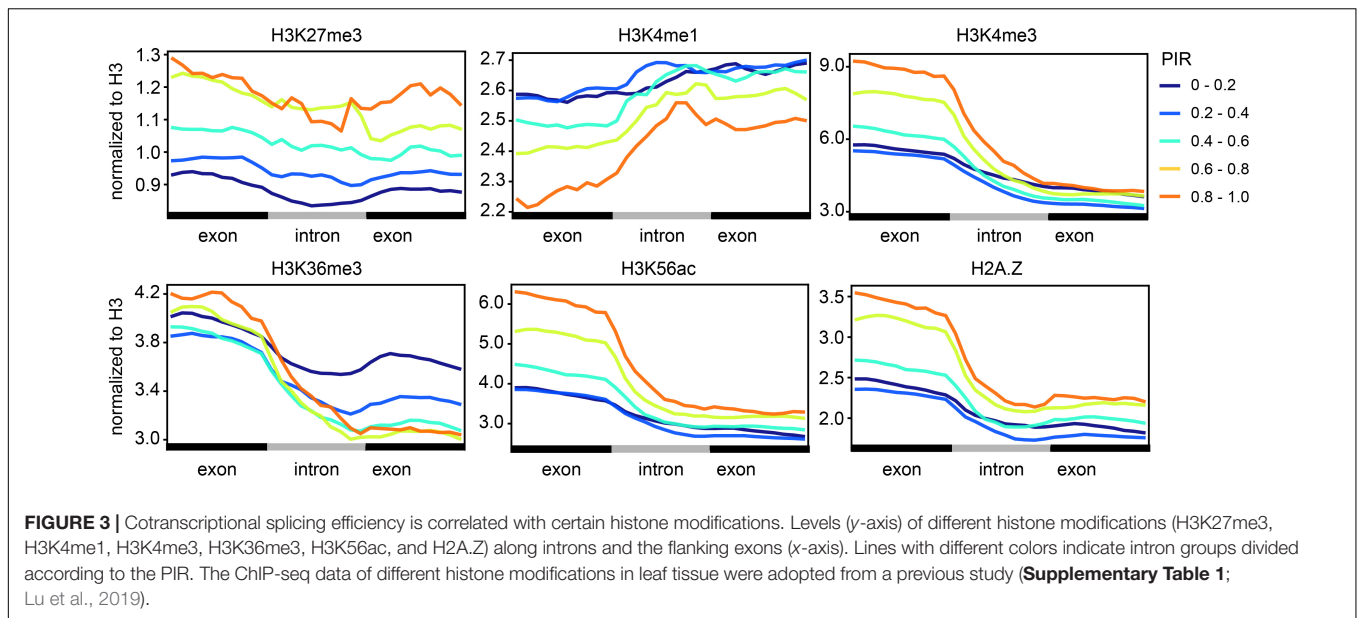
## Differential Alternative Splicing Between Leaf and Shoot Apex Tissues Is Not Determined Merely by Cotranscriptional Splicing

Given that most AS events are determined cotranscriptionally, we then asked whether differences in AS between the shoot apex and leaf tissues detected by CB RNA-seq and polyA RNA-seq are consistent. Thus, we compared the AS difference of both CB RNA and polyA RNA between the 15-day apex and leaf tissues. Differential splicing events were analyzed by the program SUPPA2 (Trincado et al., 2018). A splicing event was considered differential when the absolute value of the PSI difference ( $\Delta$ PSI) between tissues  $>0.1$  and the  $p$ -value  $< 0.05$ . A small number of the different splicing events between the leaf and shoot apex tissues were detected by both CB RNA and polyA RNA (Figure 5A).  $\Delta$ PSI<sub>mRNA</sub> and  $\Delta$ PSI<sub>CB</sub> were barely correlated (Spearman correlation ranged from 0.22 to 0.35) (Figure 5B). Furthermore, genes with different splicing events detected by CB RNA and polyA RNA were not concordant (Supplementary Figure 5A). Although overall AS events are highly correlated at the cotranscriptional level and posttranscriptional level within the same tissue, tissue-specific mRNA processing, such as degradation and posttranscriptional splicing, may result in the differential AS events that are detected by polyA RNA but not by CB RNA. For those differential AS events detected by CB RNA but not by polyA RNA, it was probably caused by the differentially cotranscriptional splicing efficiency between the shoot apex and leaf tissues and further corrected at the posttranscriptional splicing step, exemplified by the first intron of *Glyma.07G206100* (Supplementary Figure 6).

Genes associated with intertissue differential splicing events detected by CB RNA and polyA RNA were also different (Supplementary Figure 5A). To explore the biological function of genes with different AS events, we conducted Gene Ontology (GO) enrichment analysis. Interestingly, genes with different splicing events between the 15-day apex and leaf tissues were significantly enriched in mRNA splicing and RNA processing, which somehow explains the differential splicing efficiency between the shoot apex and leaf tissues (Supplementary Figure 5B).

## The Level of Steady-State mRNA Is Moderately Correlated With the Biogenesis of Nascent RNA

Chromatin-bound RNA-seq is applied to detect transcribed RNAs, which are subject to multiple steps of mRNA processing, including cotranscriptional and posttranscriptional processes prior to maturation. Thus, there might be discordance in the abundance at the nascent RNA and mRNA levels. To test this hypothesis, we compared the TPM values of nascent RNA and mature RNA. Overall, the levels of nascent RNA and mature RNA were moderately correlated (Spearman correlation = 0.71–0.73) (Figure 6A and Supplementary Figures 7A–C). There are two



types of discordant genes. One is a gene that is highly transcribed with a low level of mature RNA, which might result from a high turnover of mRNA and is designated unstable RNA. The other is a gene with relatively low transcription activity but a high level of mature RNA, which might be due to the high RNA stability and is called stable RNA.

To select unstable and stable RNA transcripts, we first established a linear regression model of the log<sub>2</sub> values of TPM genes obtained with CB RNA-seq and polyA RNA-seq. Then, the predicted TPM values of genes in polyA RNA were calculated based on the linear regression model. If the actual TPM of a gene was threefold higher (or lower) than the predicted TPM, the gene was considered to be stable (or unstable) (Figure 6A and Supplementary Figures 7B,C). To investigate whether the stability of RNA is associated with specific biological functions, we performed GO enrichment analysis. For unstable RNAs, defense response, protein phosphorylation, and signal transduction were the most enriched terms. Stable RNAs were mainly associated with translation, photorespiration, ribosome biogenesis, and glycolytic processes (Figure 6B and Supplementary Figures 7C,E).

## Differentially Expressed Genes Are Consistent at the Nascent and Mature RNA Levels

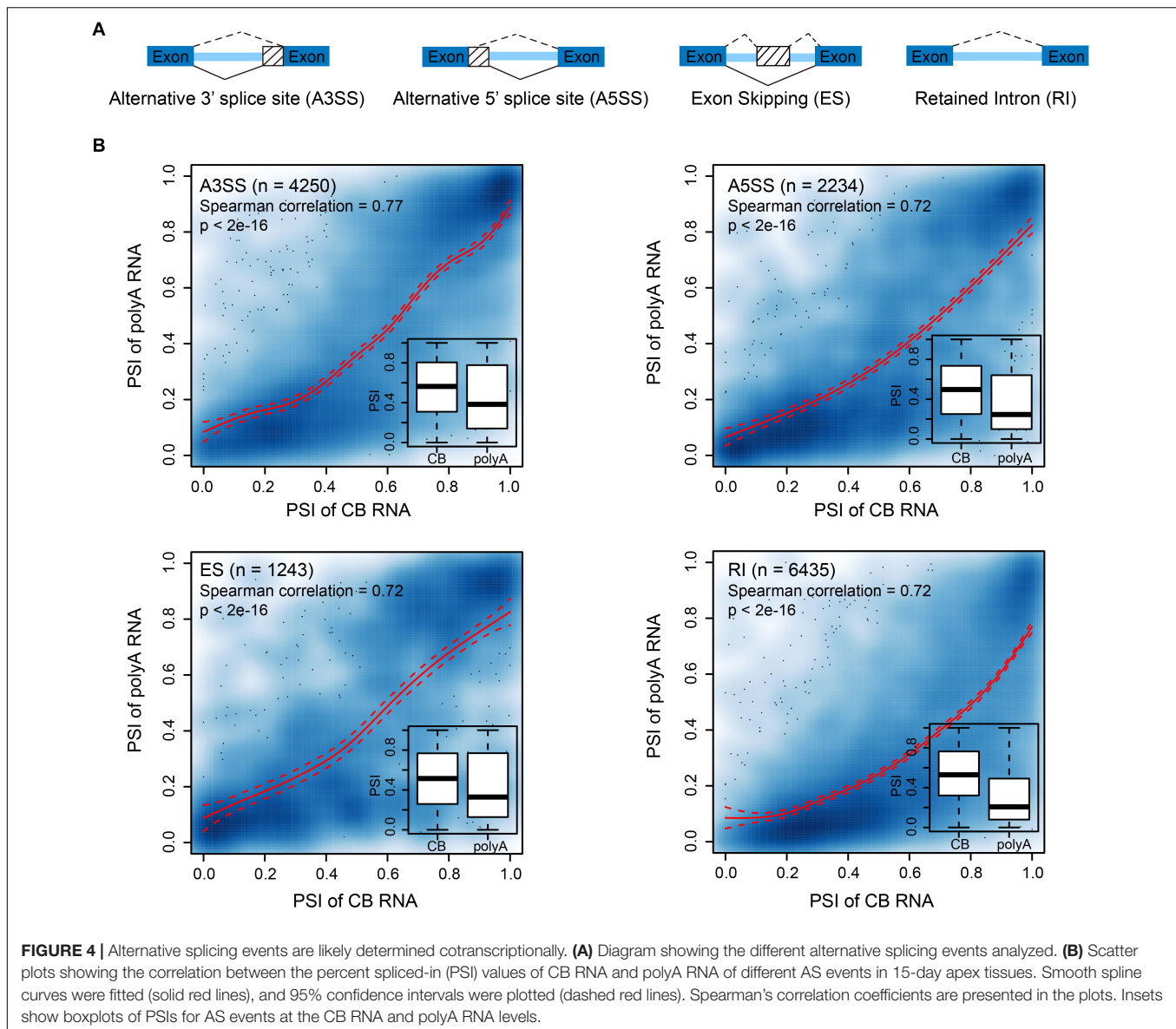
We then identified differentially expressed genes (DEGs) between 15-day apex and 15-day leaf tissues at both nascent and mature RNA levels. More than 10,000 genes were expressed more in the apex than in the leaf, and vice versa (Supplementary Figure 8A and Supplementary Table 2). Most of these DEGs detected by CB RNA-seq and polyA RNA-seq overlapped (Figure 6C and Supplementary Figure 8B). Furthermore, fold changes at the CB RNA level and polyA RNA level were highly correlated (Spearman correlation = 0.93) (Figure 6D).

Gene Ontology enrichment analysis was performed to determine the biological functions of the DEGs. Genes with higher expression in the apex were mainly associated with RNA methylation, histone methylation, translation, DNA replication, and meristem initiation and maintenance. Genes with higher expression levels in the leaves were mainly related to photosynthesis and plastid organization (Supplementary Figure 8C).

In addition, only a small number of genes were called DEGs between the 15-day apex and 10-day apex (Supplementary Figure 9A), and they had concordant changes at the nascent RNA and mRNA levels (Supplementary Figure 9B). GO enrichment indicated that genes highly expressed in the 10-day apex were involved in the response to stress, circadian rhythm, etc., and genes highly expressed in the 15-day apex were involved in long-day photoperiodism flowering, response to hormones, and circadian rhythm (Supplementary Figure 9C).

## More Non-coding RNAs Were Identified by CB RNA-Seq Than PolyA RNA-Seq

Considering that unstable transcripts are readily detected at the nascent RNA level, we calculated the expression level of ncRNA as defined in a previous study (Lin et al., 2020). As expected, more active ncRNA genes were detected by CB RNA-seq than polyA RNA-seq (Figure 7A). Furthermore, we determined the antisense transcription of annotated mRNAs by counting reads mapped to the opposite strand, and there were more active antisense transcriptional signals at the nascent RNA level (Figure 7B, left). These results indicate that some non-coding transcripts were unstable or not polyadenylated. For example, a transcript encoded from the antisense strand of *FT2a*, the essential gene involved in flowering timing, was identified in 15-day leaves by CB RNA-seq. *Dt1*, the key gene controlling growth habit, overlapped with another strong antisense transcript at the nascent RNA level in the apex (Figure 7B, right).

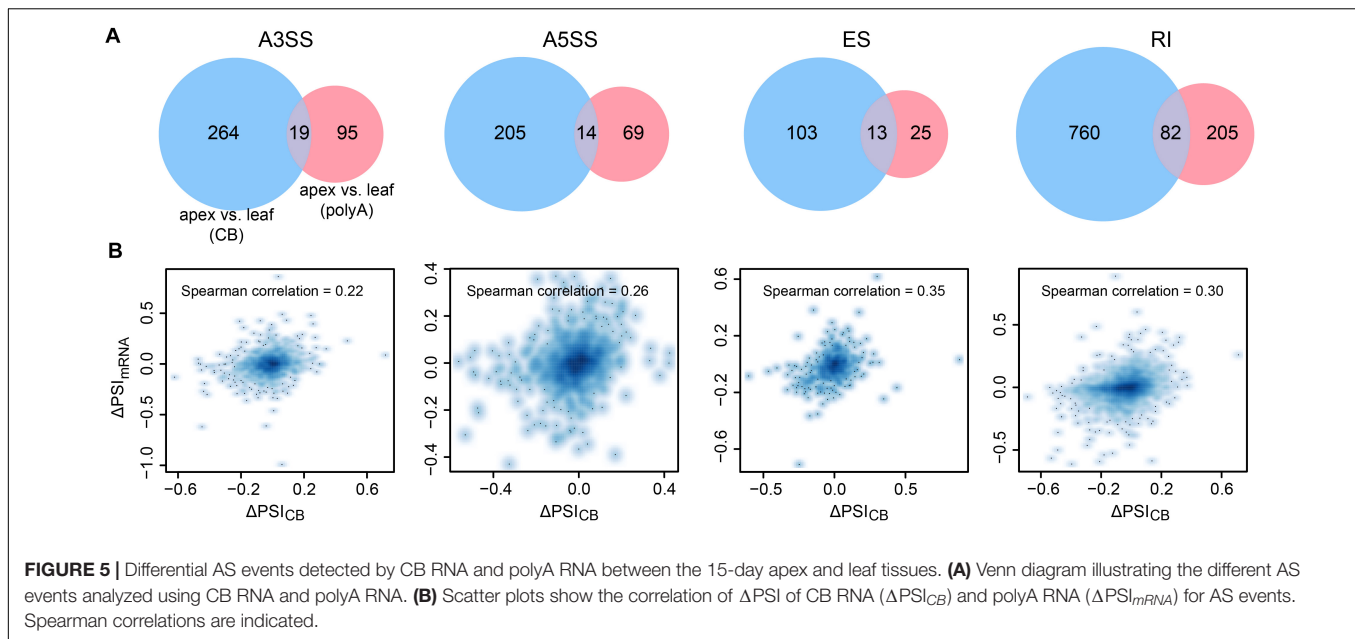


To identify novel transcripts, we assembled transcripts from nascent RNA and polyA RNA of each tissue separately. Then, all transcripts were merged and compared based on reference annotations (see section “Materials and Methods”). Only intergenic transcripts were included for further analysis. In total, there were 5,927 and 1,515 active intergenic transcripts from CB RNA-seq and polyA RNA-seq, respectively, with 1,326 transcripts overlapping (Figure 7C, upper panel; Supplementary Table 3). These transcripts were encoded from 4,835 loci, of which 1,142 were shared by CB RNA and polyA RNA (Figure 7C, bottom panel).

We then applied two tools, CNCI and FEELnc, to evaluate the protein-coding potential of these new transcripts. In total, 4,001 and 974 active new transcripts of CB RNA and polyA RNA were considered non-coding transcripts by both methods, respectively

(Figure 7D), and more ncRNAs were observed in the leaves at the nascent RNA level (Figure 7E).

Non-coding RNA detected only at the nascent RNA level might be unstable or unpolyadenylated. ncRNAs detected only at the polyA RNA level might be very stable and accumulate by slow transcription. Different types of ncRNAs may be regulated differently at the transcriptional level. To gain insight into the effects of histone modifications on ncRNA expression, we compared the metaprofiles of histone modifications for three groups of ncRNAs from the leaf tissue (group I: only detected by CB RNA; group II: detected by both; group III: only detected by polyA RNA) (Figure 7F). Group II and III ncRNA genes were associated with H3K56ac, H3K4me3, and histone variant H2A.Z (Figure 7G).



## DISCUSSION

Although nascent RNA-seq has been extensively used to detect cotranscriptional regulation in yeast, fly, and mammalian cells, its application in plants is still lagging behind. Recently, several methods have been developed to detect nascent RNA and reveal plant-specific transcriptional features (Hetzl et al., 2016; Zhu et al., 2018). However, with the exception of one maize publication using GRO-seq (Erhard et al., 2015), all studies have focused on the model plant *Arabidopsis*. Here, we describe the soybean transcriptome using CB RNA-seq. As expected, CB RNA isolation greatly enriched the nascent RNA by removing the abundant cytosolic mRNAs and nucleoplasmic RNAs. We demonstrated that CB RNA-seq successfully detected nascent RNA biogenesis and cotranscriptional processing of pre-mRNA from the leaves and growing apex tissues. This method can be applied to other tissues at various developmental stages and/or under different environmental conditions, which may further shed light on the transcriptional regulation of the soybean genome.

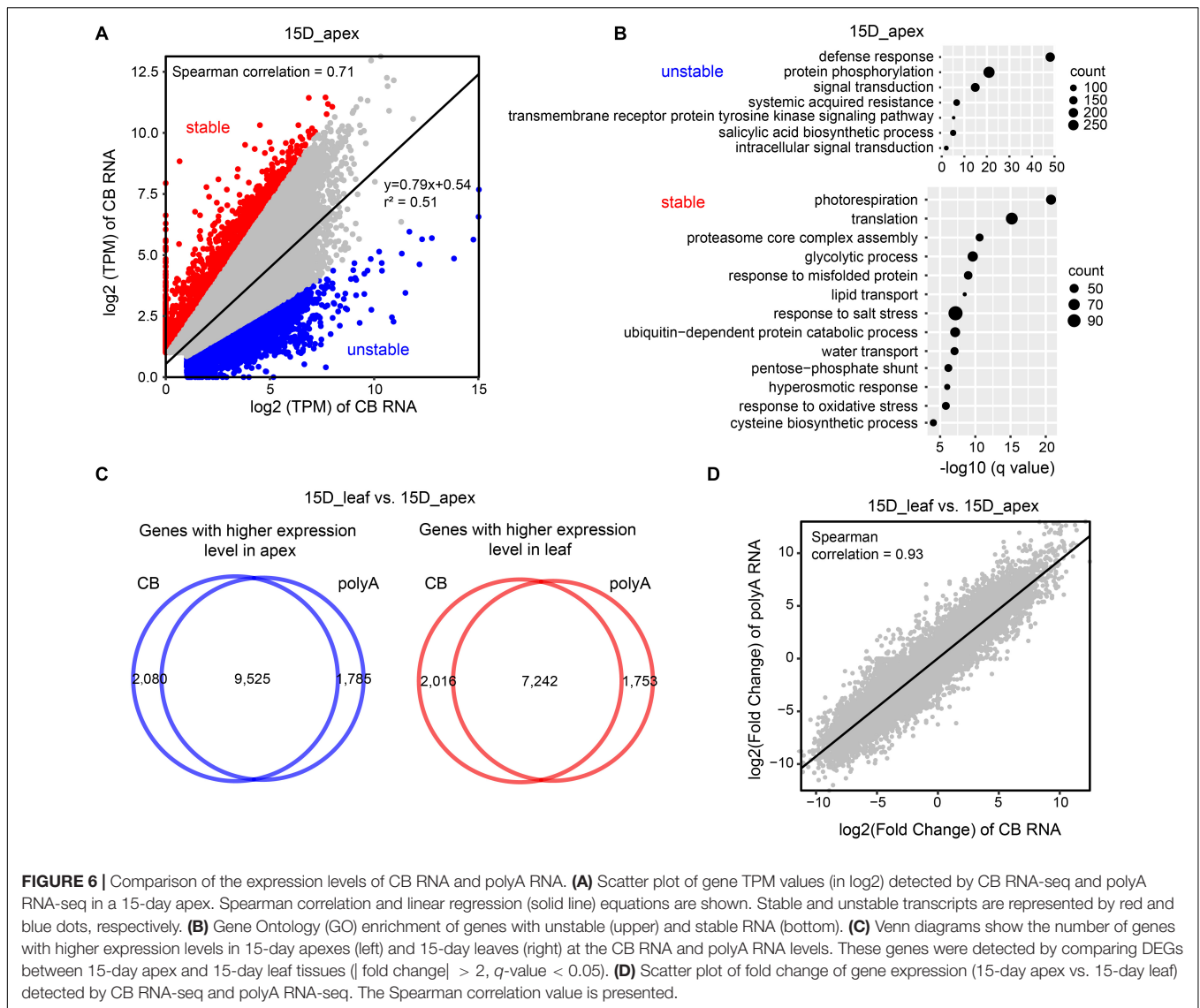
We found genome-wide cotranscriptional splicing in soybean. Cotranscriptional splicing efficiency is related to intron length, distance from TES, intron number, and gene length. These characteristics are similar to those previously observed in yeast, fly, mammalian, and *Arabidopsis* cells, indicating a conserved mechanism that controls cotranscriptional splicing in eukaryotic cells (Khodor et al., 2011; Kindgren et al., 2020; Li et al., 2020). Interestingly, we found that both active (H3K4me3 and H3K56ac) and inactive (H3K27me3) histone markers are negatively related to cotranscriptional splicing efficiency. The elongation rate of RNA Pol II can affect splicing efficiency by fine-tuning the timing of the spliceosome search for splice sites, as the spliceosome is physically recruited by the carboxyl terminal domain of the largest subunit of RNA Pol II (Nojima et al., 2018).

The inverse correlation between elongation speed and splicing efficiency was proven in yeast *in vivo* (Carrillo Oesterreich et al., 2016; Aslanzadeh et al., 2018). Moreover, the RNA Pol II elongation rate is regulated by transcription elongation factors and chromatin structural barriers such as nucleosomes. Thus, factors that affect transcription elongation also affect splicing efficiency. Active histone markers are thought to be related to a higher transcription elongation rate. Therefore, it is reasonable that introns with higher H3K4me3 or H3K56ac contents are less efficiently spliced. In addition, the pattern described in this study and a previous study on *Arabidopsis* revealed that the retained introns are derived from genes with low H3K4me1 and high H3K27me3 signatures (Mahrez et al., 2016). However, further studies of mutants with impaired histone modification are needed to verify their function in cotranscriptional splicing. Actually, these effects are not unidirectional. Cotranscriptional splicing can in turn influence the elongation rate and establishment of histone modifications (Kim et al., 2011).

Alternative splicing is an important part of gene regulation. In our study, a highly correlated relative AS event (PSI) was observed between CB RNA and polyA RNA, suggesting that most AS events are determined cotranscriptionally. This agrees with a previous study in *Arabidopsis* (Zhu et al., 2020). However, when comparing intertissue AS events, differential AS events detected at the cotranscriptional and posttranscriptional levels only partially overlapped. Thus, differential AS events cannot be predicted at the nascent RNA level, indicating the complexity of AS regulation. These regulations may be attributed to different degradation rates and/or posttranscriptional splicing among various tissues.

Gene expression is regulated at multiple levels, including transcription, post-transcription, and translation. Steady-state mRNA is the output of transcriptional activity and RNA degradation. Thus, there might be some discordance in gene





activity detected by nascent RNA-seq and polyA RNA-seq. As expected, we found that gene activity at these two levels was moderately correlated. However, when comparing different tissues, the changes in gene activity at both levels were highly consistent, indicating that tissue-specific gene expression was mainly associated with transcription. The stability of RNA might contribute to the discordance in gene activity at the nascent and mature RNA levels. It is meaningful for stable mRNA genes to be involved in housekeeping biological processes. Moreover, under normal conditions, keeping regulatory genes at low mRNA levels and relatively high transcription by fast turnover of mRNA is an effective way to ensure rapid responses to potential stimuli. As we have previously reported in *Arabidopsis*, genes induced highly and quickly by short-term heat shock usually exhibit basic transcription under normal temperature (Liu M. et al., 2020).

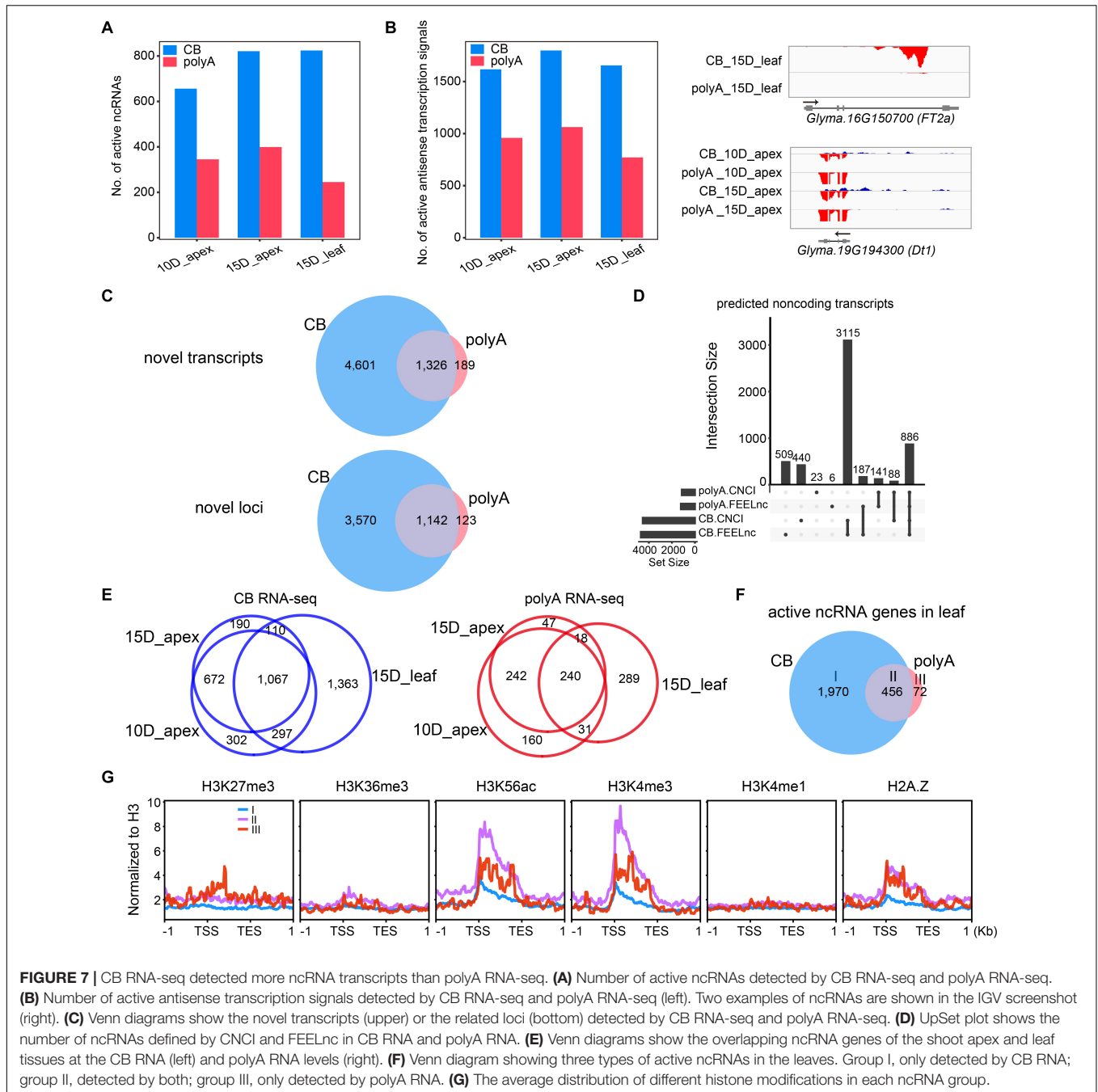
Since some ncRNAs are unstable or unpolyadenylated, such as enhancer RNAs and antisense RNAs, more transcripts are expected to be detected by CB RNA-seq. However, this does

not rule out the possibility that some transcripts detected only in CB RNA are not nascent RNA but rather chromatin-bound transcripts. To further elucidate the biological significance of these ncRNAs, approaches such as RNA interference and gene editing are needed. It will be interesting to apply CB RNA-seq to various tissues and build a transcriptional regulatory network at the nascent RNA level in the future.

## MATERIALS AND METHODS

### Plant Materials and Growth Conditions

Soybean Wm82 plants were grown under long light day conditions (16 h light, 8 h dark) with a constant 25°C temperature in a growth chamber. Shoot apices from 10- to 15-day seedlings were collected in three biological replicates, with each replicate collected from approximately 20 plants. For the leaves, the first trifoliate leaves of two 15-day-old plants were collected as one



biological replicate. All samples were frozen with liquid nitrogen immediately after collection.

## RNA Isolation, Transcriptome Library Preparation, and Sequencing

The chromatin RNA extraction protocol was modified from a previously published method (Zhu et al., 2020). Briefly, tissues were ground into a fine powder with liquid nitrogen and solubilized in cold nuclei isolation buffer (20 mM/KOH pH

7.4, 0.44 M sucrose, 1.25% Ficoll, 2.5% Dextran T40, 10 mM MgCl<sub>2</sub>, 0.75% Triton X-100, 0.5 mM EDTA, 1 mM DTT, 8 mM β-mercaptoethanol, 1 μg/ml pepstatin A, 1 μg/ml aprotinin, and 1 mM PMSF). The crude nuclei were precipitated at 3,500 rpm and washed with resuspension buffer (50% glycerol, 25 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 100 mM NaCl, 1 mM DTT, 0.4 U/μl RNase inhibitor, 1 μg/ml pepstatin A, 1 μg/ml aprotinin, and 8 mM β-mercaptoethanol) once, followed by washing buffer (25 mM Tris-HCl pH 7.5, 300 mM NaCl, 1 M urea, 0.5 mM EDTA, 1 mM DTT, 1% Tween-20, 0.4 U/μl RNase inhibitor, 1 μg/ml pepstatin A, 1 μg/ml aprotinin, and 8 mM

$\beta$ -mercaptoethanol) twice. Chromatin RNA was extracted from washed nuclei using TRIzol reagent (Life Technologies).

After degrading genomic DNA by TURBO DNase (Life Technologies), CB RNA was subjected to rRNA depletion using a riboPOOL kit (siTOOLS Biotech, PanPlant-10 nmol) and polyA RNA removal by oligo(dT) beads (NEB, S1419). Poly(A) RNA was enriched from total RNA by oligo(dT) beads. Both CB RNA and polyA RNA were transformed into cDNA libraries using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (NEB #E7765) and sequenced on an Illumina NovaSeq platform.

## CB RNA and mRNA Data Processing

Raw reads of CB RNA and polyA RNA were first evaluated by FastQC<sup>1</sup>, and then Cutadapt was used to remove adapters and low-quality reads (Martin, 2011). Clean reads were subsequently aligned to the genome Wm82.a2.v1 by STAR (Dobin et al., 2013). Only uniquely mapped reads were retained for the following analysis. Read distribution on genomic features was evaluated by RSeQC with the subcommand “read\_distribution.py” (Wang et al., 2012). To calculate the ratio of introns vs. exons of each gene, featureCounts was used to quantify the read counts on introns and exons separately (Liao et al., 2014). Read density was normalized by the length of introns and exons.

## Calculating the Percent of Intron Retention

The proportion of intron-retained reads across an intron is usually used to evaluate the splicing efficiency of the intron. To quantitatively evaluate the genome-wide cotranscriptional splicing efficiency in soybean, we calculated the PIR value for constitutive introns as described previously (Braunschweig et al., 2014). Briefly, three types of reads on an intron were counted: (1) exon–intron junction reads across the 5′SS (EI5), (2) exon–intron junction reads across the 3′SS (EI3), and (3) spliced exon–exon junction reads (EE) (Figure 2A). The PIR of an intron was calculated by dividing the intron-retained reads by the sum of intron-retained reads and intron-skipping reads (Figure 2A). Constitutive introns from the annotation Wm82.a2.v1 were subjected to PIR calculations.

## Alternative Splicing Analysis

Mapped reads were assembled into putative transcripts based on a reference guided assembly strategy using the single-sample transcript assembly tool StringTie v2.1.2 (Pertea et al., 2015). Multiple putative transcripts were merged into a unified set of transcripts using the meta-assembly tool TACO v0.7.3, which was considered to be superior to Cuffmerge and StringTie merge (Niknafs et al., 2017). Then, the merged transcripts were compared with the reference gene GTF file using GffCompare v0.11.2 (Pertea and Pertea, 2020). Since CB RNA was nascent RNA with no full splicing, AS analysis was based on transcripts merged from polyA RNA data. AS events were quantified based on the PSI in the program SUPPA2 (Trincado et al., 2018). Since SUPPA2 estimated the PSI based on transcript abundance,

we first used salmon for alignment-free transcript abundance estimates (Patro et al., 2017). Transcripts with TPM > 1 in at least three samples were used for analysis. For detection of differential splicing between two samples, we chose  $\Delta$ PSI > 0.1 and  $p$ -value < 0.05 as cut-offs.

## Detection of Differentially Expressed Genes

For detecting genes with differential expression, mapped reads in each gene were quantified using featureCounts. Then, differential gene expression was evaluated by the R package DESeq2 (Love et al., 2014). DEGs were defined by the following criteria: they had to show more than twofold up- or downregulation, and the false discovery rate (FDR)-adjusted  $q$ -value calculated by DESeq2 had to be less than 0.05. The read density for each gene was calculated by normalizing the read count to the library size and mappable length (TPM).

## Gene Ontology Enrichment Analysis

Gene Ontology annotation of genes was extracted from the annotation file for Wm82.a2.v1. A hypergeometric test was explored for the statistical test, and the Benjamini and Hochberg method (1995) was used to adjust the  $p$ -value to control the FDR. All analysis was done in R software.

## Detection of New Non-coding RNA Genes

To detect new ncRNA genes at the nascent RNA and polyA RNA levels, transcripts were assembled in CB RNA and polyA RNA data separately and merged by TACO as described above in the AS event analysis. Then, annotation GTF files of transcripts were compared with reference annotation GTF files using GffCompare (with the -r option). For each putative transcript, its relationship to the closest reference transcript was described by a “class code” value. For example, the code “=” indicates that the introns of a transcript completely match the introns of the reference transcript. We chose only unknown, intergenic transcripts that were assigned the code “u” and estimated their protein-coding potential by two software programs, CNCI and FEELnc (Sun et al., 2013; Wucher et al., 2017).

## Reanalysis of ChIP-Seq Data of Histone Modifications

ChIP-seq raw data of histone modifications were downloaded from NCBI (Supplementary Table 1). The raw data were first processed with adapter removal by Cutadapt and mapping to the genome by STAR. Then, the average distribution of different histone modifications on genomic features was plotted using deepTools by normalization to histone 3 (Ramírez et al., 2016).

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: NCBI website under Bioproject accession PRJNA689321.

<sup>1</sup><http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

## AUTHOR CONTRIBUTIONS

ZD, ML, and JZ designed the research. JZ and ML performed the research. ML, HZ, FK, and BL analyzed the data. ML and ZD wrote the manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by grants from the National Transgenic Major Project of China (2019ZX08010003-002-015),

## REFERENCES

- Aebi, M., Hornig, H., Padgett, R. A., Reiser, J., and Weissmann, C. (1986). Sequence requirements for splicing of higher eukaryotic nuclear pre-mRNA. *Cell* 47, 555–565. doi: 10.1016/0092-8674(86)90620-3
- Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815. doi: 10.1038/35048692
- Aslanzadeh, V., Huang, Y., Sanguinetti, G., and Beggs, J. D. (2018). Transcription rate strongly affects splicing fidelity and cotranscriptionality in budding yeast. *Genome Res.* 28, 203–213. doi: 10.1101/gr.225615.117
- Baralle, F. E., and Giudice, J. (2017). Alternative splicing as a regulator of development and tissue identity. *Nat. Rev. Mol. Cell Biol.* 18, 437–451. doi: 10.1038/nrm.2017.27
- Baurén, G., and Wieslander, L. (1994). Splicing of Balbiani ring 1 gene pre-mRNA occurs simultaneously with transcription. *Cell* 76, 183–192. doi: 10.1016/0092-8674(94)90182-1
- Bentley, D. L. (2014). Coupling mRNA processing with transcription in time and space. *Nat. Rev. Genet.* 15, 163–175. doi: 10.1038/nrg3662
- Beyer, A. L., and Osheim, Y. N. (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev.* 2, 754–765. doi: 10.1101/gad.2.6.754
- Braunschweig, U., Barbosa-Morais, N. L., Pan, Q., Nachman, E. N., Alipanahi, B., Gonatopoulos-Pournatzis, T., et al. (2014). Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res.* 24, 1774–1786. doi: 10.1101/gr.177790.114
- Carrillo Oesterreich, F., Herzel, L., Straube, K., Hujer, K., Howard, J., and Neugebauer, K. M. (2016). Splicing of nascent RNA coincides with intron exit from RNA Polymerase II. *Cell* 165, 372–381. doi: 10.1016/j.cell.2016.02.045
- 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. doi: 10.1126/science.1162228
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635
- Drexler, H. L., Choquet, K., and Churchman, L. S. (2020). Splicing kinetics and coordination revealed by direct nascent RNA sequencing through nanopores. *Mol. Cell* 77, 985.e8–998.e8. doi: 10.1016/j.molcel.2019.11.017
- Erhard, K. F., Talbot, J. E. R. B., Deans, N. C., McClish, A. E., and Hollick, J. B. (2015). Nascent transcription affected by RNA polymerase IV in *Zea mays*. *Genetics* 199, 1107–1125. doi: 10.1534/genetics.115.174714
- Gazara, R. K., de Oliveira, E. A. G., Rodrigues, B. C., Nunes da Fonseca, R., Oliveira, A. E. A., and Venancio, T. M. (2019). Transcriptional landscape of soybean (*Glycine max*) embryonic axes during germination in the presence of paclobutrazol, a gibberellin biosynthesis inhibitor. *Sci. Rep.* 9:9601. doi: 10.1038/s41598-019-45898-2
- Hetzl, J., Duttke, S. H., Benner, C., and Chory, J. (2016). Nascent RNA sequencing reveals distinct features in plant transcription. *Proc. Natl. Acad. Sci. U.S.A.* 113, 12316–12321. doi: 10.1073/pnas.1603217113
- Hu, Q., Greene, C. S., and Heller, E. A. (2020). Specific histone modifications associate with alternative exon selection during mammalian development. *Nucleic Acids Res.* 48, 4709–4724. doi: 10.1093/nar/gkaa248
- the National Key Research and Development Program of China (2016YFD010190401), the National Natural Science Foundation of China (32090061), and Guangdong University Innovation Team Project (2019KCXTD010).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.649634/full#supplementary-material>

- Jia, J., Long, Y., Zhang, H., Li, Z., Liu, Z., Zhao, Y., et al. (2020). Post-transcriptional splicing of nascent RNA contributes to widespread intron retention in plants. *Nat. Plants* 6, 780–788. doi: 10.1038/s41477-020-0688-1
- Kasai, A., Kasai, K., Yumoto, S., and Senda, M. (2007). Structural features of GmIRCHS, candidate of the I gene inhibiting seed coat pigmentation in soybean: implications for inducing endogenous RNA silencing of chalcone synthase genes. *Plant Mol. Biol.* 64, 467–479. doi: 10.1007/s11103-007-9169-4
- Khodor, Y. L., Rodriguez, J., Abruzzi, K. C., Tang, C. H. A., Marr, M. T., and Rosbash, M. (2011). Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in *Drosophila*. *Genes Dev.* 25, 2502–2512. doi: 10.1101/gad.178962.111
- Kim, S., Kim, H., Fong, N., Erickson, B., and Bentley, D. L. (2011). Pre-mRNA splicing is a determinant of histone H3K36 methylation. *Proc. Natl. Acad. Sci. U.S.A.* 108:13564. doi: 10.1073/pnas.1109475108
- Kindgren, P., Ivanov, M., and Marquardt, S. (2020). Native elongation transcript sequencing reveals temperature dependent dynamics of nascent RNAPII transcription in *Arabidopsis*. *Nucleic Acids Res.* 48, 2332–2347. doi: 10.1093/nar/gkz1189
- Kong, F., Liu, B., Xia, Z., Sato, S., Kim, B. M., Watanabe, S., et al. (2010). Two coordinately regulated homologs of FLOWERING LOCUS T are involved in the control of photoperiodic flowering in soybean. *Plant Physiol.* 154, 1220–1231. doi: 10.1104/pp.110.160796
- Laloum, T., Martin, G., and Duque, P. (2018). Alternative splicing control of abiotic stress responses. *Trends Plant Sci.* 23, 140–150. doi: 10.1016/j.tplants.2017.09.019
- Li, S., Wang, Y., Zhao, Y., Zhao, X., Chen, X., and Gong, Z. (2020). Global co-transcriptional splicing in *Arabidopsis* and the correlation with splicing regulation in mature RNAs. *Mol. Plant* 13, 266–277. doi: 10.1016/j.molp.2019.11.003
- Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. doi: 10.1093/bioinformatics/btt656
- Libault, M., Farmer, A., Joshi, T., Takahashi, K., Langley, R. J., Franklin, L. D., et al. (2010). An integrated transcriptome atlas of the crop model *Glycine max*, and its use in comparative analyses in plants. *Plant J.* 63, 86–99. doi: 10.1111/j.1365-313X.2010.04222.x
- Lin, X., Lin, W., Ku, Y.-S., Wong, F.-L., Li, M.-W., Lam, H.-M., et al. (2020). Analysis of soybean long non-coding RNAs reveals a subset of small peptide-coding transcripts. *Plant Physiol.* 182:1359. doi: 10.1104/pp.19.01324
- Liu, B., Watanabe, S., Uchiyama, T., Kong, F., Kanazawa, A., Xia, Z., et al. (2010). The soybean stem growth habit gene Dt1 is an ortholog of *Arabidopsis* TERMINAL FLOWER1. *Plant Physiol.* 153, 198–210. doi: 10.1104/pp.109.150607
- Liu, M., Zhu, J., and Dong, Z. (2020). Immediate transcriptional responses of *Arabidopsis* leaves to heat shock. *J. Integr. Plant Biol.* 63, 468–483. doi: 10.1111/jipb.12990
- Liu, Y., Du, H., Li, P., Shen, Y., Peng, H., Liu, S., et al. (2020). Pan-genome of wild and cultivated soybeans. *Cell* 182, 162.e13–176.e13. doi: 10.1016/j.cell.2020.05.023
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550–550. doi: 10.1186/s13059-014-0550-8

- Lu, S., Dong, L., Fang, C., Liu, S., Kong, L., Cheng, Q., et al. (2020). Stepwise selection on homeologous PRR genes controlling flowering and maturity during soybean domestication. *Nat. Genet.* 52, 428–436. doi: 10.1038/s41588-020-0604-7
- Lu, X., Xiong, Q., Cheng, T., Li, Q.-T., Liu, X.-L., Bi, Y.-D., et al. (2017). A PP2C-1 allele underlying a quantitative trait locus enhances soybean 100-seed weight. *Mol. Plant* 10, 670–684. doi: 10.1016/j.molp.2017.03.006
- Lu, Z., Marand, A. P., Ricci, W. A., Ethridge, C. L., Zhang, X., and Schmitz, R. J. (2019). The prevalence, evolution and chromatin signatures of plant regulatory elements. *Nat. Plants* 5, 1250–1259. doi: 10.1038/s41477-019-0548-z
- Luco, R. F., Pan, Q., Tominaga, K., Blencowe, B. J., Pereira-Smith, O. M., and Misteli, T. (2010). Regulation of alternative splicing by histone modifications. *Science* 327, 996–1000. doi: 10.1126/science.1184208
- Mahrez, W., Shin, J., Muñoz-Viana, R., Figueiredo, D. D., Trejo-Arellano, M. S., Exner, V., et al. (2016). BRR2a affects flowering time via FLC splicing. *PLoS Genet.* 12:e1005924. doi: 10.1371/journal.pgen.1005924
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17, 10–12. doi: 10.14806/ej.17.1.200
- McGlinchy, N. J., and Smith, C. W. J. (2008). Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense? *Trends Biochem. Sci.* 33, 385–393. doi: 10.1016/j.tibs.2008.06.001
- Neugebauer, K. M. (2019). Nascent RNA and the coordination of splicing with transcription. *Cold Spring Harb. Perspect. Biol.* 11:a032227. doi: 10.1101/cshperspect.a032227
- Niknafs, Y. S., Pandian, B., Iyer, H. K., Chinnaiyan, A. M., and Iyer, M. K. (2017). TACO produces robust multisample transcriptome assemblies from RNA-seq. *Nat. Methods* 14, 68–70. doi: 10.1038/nmeth.4078
- Nojima, T., Gomes, T., Grosso, A. R. F., Kimura, H., Dye, M. J., Dhir, S., et al. (2015). Mammalian NET-seq reveals genome-wide nascent transcription coupled to RNA processing. *Cell* 161, 526–540. doi: 10.1016/j.cell.2015.03.027
- Nojima, T., Rebelo, K., Gomes, T., Grosso, A. R., Proudfoot, N. J., and Carmo-Fonseca, M. (2018). RNA polymerase II phosphorylated on CTD serine 5 interacts with the spliceosome during co-transcriptional splicing. *Mol. Cell* 72, 369.e4–379.e4. doi: 10.1016/j.molcel.2018.09.004
- Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14, 417–419. doi: 10.1038/nmeth.4197
- Pellicer, J., and Leitch, I. J. (2020). The Plant DNA C-values database (release 7.1): an updated online repository of plant genome size data for comparative studies. *New Phytol.* 226, 301–305. doi: 10.1111/nph.16261
- Perte, G., and Perte, M. (2020). GFF Utilities: GffRead and GffCompare. *F1000Res.* 9:ISCB Comm J-304. doi: 10.12688/f1000research.23297.1
- Perte, M., Perte, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33, 290–295. doi: 10.1038/nbt.3122
- Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., et al. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* 44, W160–W165. doi: 10.1093/nar/gkw257
- Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., et al. (2010). Genome sequence of the palaeopolyploid soybean. *Nature* 463, 178–183. doi: 10.1038/nature08670
- Severin, A. J., Woody, J. L., Bolon, Y.-T., Joseph, B., Diers, B. W., Farmer, A. D., et al. (2010). RNA-seq atlas of Glycine max: a guide to the soybean transcriptome. *BMC Plant Biol.* 10:160. doi: 10.1186/1471-2229-10-160
- Shen, Y., Liu, J., Geng, H., Zhang, J., Liu, Y., Zhang, H., et al. (2018). *De novo* assembly of a Chinese soybean genome. *Sci. China. Life Sci.* 61, 871–884. doi: 10.1007/s11427-018-9360-0
- Shen, Y., Zhou, Z., Wang, Z., Li, W., Fang, C., Wu, M., et al. (2014). Global dissection of alternative splicing in paleopolyploid soybean. *Plant Cell* 26, 996–1008. doi: 10.1105/tpc.114.122739
- Sun, L., Luo, H., Bu, D., Zhao, G., Yu, K., Zhang, C., et al. (2013). Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res.* 41:e166. doi: 10.1093/nar/gkt646
- Takeshima, R., Nan, H., Harigai, K., Dong, L., Zhu, J., Lu, S., et al. (2019). Functional divergence between soybean FLOWERING LOCUS T orthologues FT2a and FT5a in post-flowering stem growth. *J. Exp. Bot.* 70, 3941–3953. doi: 10.1093/jxb/erz199
- Tian, Z., Wang, X., Lee, R., Li, Y., Specht, J. E., Nelson, R. L., et al. (2010). Artificial selection for determinate growth habit in soybean. *Proc. Natl. Acad. Sci. U.S.A.* 107, 8563–8568. doi: 10.1073/pnas.1000088107
- Trincado, J. L., Entizne, J. C., Hysenaj, G., Singh, B., Skalic, M., Elliott, D. J., et al. (2018). SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. *Genome Biol.* 19:40. doi: 10.1186/s13059-018-1417-1
- Wang, E. T., Sandberg, R., Luo, S., Khrebtkova, I., Zhang, L., Mayr, C., et al. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470–476. doi: 10.1038/nature07509
- Wang, L., Cao, C., Ma, Q., Zeng, Q., Wang, H., Cheng, Z., et al. (2014). RNA-seq analyses of multiple meristems of soybean: novel and alternative transcripts, evolutionary and functional implications. *BMC Plant Biol.* 14:169. doi: 10.1186/1471-2229-14-169
- Wang, L., Wang, S., and Li, W. (2012). RSeQC: quality control of RNA-seq experiments. *Bioinformatics* 28, 2184–2185. doi: 10.1093/bioinformatics/bts356
- Wu, F., Sedivy, E. J., Price, W. B., Haider, W., and Hanzawa, Y. (2017). Evolutionary trajectories of duplicated FT homologues and their roles in soybean domestication. *Plant J.* 90, 941–953. doi: 10.1111/tpj.13521
- Wucher, V., Legeai, F., Hédan, B., Rizk, G., Lagoutte, L., Leeb, T., et al. (2017). FEELnc: a tool for long non-coding RNA annotation and its application to the dog transcriptome. *Nucleic Acids Res.* 45:e57. doi: 10.1093/nar/gkw1306
- Xie, M., Chung, C. Y., Li, M. W., Wong, F. L., Wang, X., Liu, A., et al. (2019). A reference-grade wild soybean genome. *Nat. Commun.* 10:1216. doi: 10.1038/s41467-019-09142-9
- Zhu, D., Mao, F., Tian, Y., Lin, X., Gu, L., Gu, H., et al. (2020). The features and regulation of co-transcriptional splicing in *Arabidopsis*. *Mol. Plant* 13, 278–294. doi: 10.1016/j.molp.2019.11.004
- Zhu, J., Liu, M., Liu, X., and Dong, Z. (2018). RNA polymerase II activity revealed by GRO-seq and pNET-seq in *Arabidopsis*. *Nat. Plants* 4, 1112–1123. doi: 10.1038/s41477-018-0280-0

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Zhu, Zhao, Kong, Liu, Liu and Dong. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.