the plant journal

SEB Society for Experimental Biology

The Plant Journal (2017) **91,** 518–533 doi: 10.1111/tpj.13571

TECHNICAL ADVANCE

Proteogenomic analysis reveals alternative splicing and translation as part of the abscisic acid response in Arabidopsis seedlings

Fu-Yuan Zhu[1,2,3,†], Mo-Xian Chen[2,3,†], Neng-Hui Ye[3,4,†], Lu Shi[2,5,†], Kai-Long Ma[6,†], Jing-Fang Yang[7], Yun-Ying Cao[2,8], Youjun Zhang[9], Takuya Yoshida[9,10], Alisdair R. Fernie[9], Guang-Yi Fan[6], Bo Wen[6], Ruo Zhou[6], Tie-Yuan Liu[2], Tao Fan[1], Bei Gao[2], Di Zhang[2], Ge-Fei Hao[7], Shi Xiao[11], Ying-Gao Liu[1,*] and Jianhua Zhang[2,3,*]

[1]*State Key Laboratory of Crop Biology, College of Life Science, Shandong Agricultural University, Taian, Shandong, China,*
[2]*School of Life Sciences and State Key Laboratory of Agrobiotechnology, The Chinese University of Hong Kong, Shatin, Hong Kong,*
[3]*Shenzhen Research Institute, The Chinese University of Hong Kong, Shenzhen, China,*
[4]*Southern Regional Collaborative Innovation Center for Grain and Oil Crops in China, Hunan Agricultural University, Changsha 410128, China,*
[5]*State Key Laboratory of Membrane Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing, China,*
[6]*BGI-Shenzhen, Shenzhen, China,*
[7]*College of Chemistry, Central China Normal University, Wuhan, China,*
[8]*College of Life Sciences, Nantong University, Nantong, Jiangsu, China,*
[9]*Max-Planck-Institut für Molekulare Pflanzenphysiologie, Am Mühlenberg 1, 14476 Potsdam-Golm, Germany,*
[10]*Laboratory of Plant Molecular Physiology, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo 113-8657, Japan, and*
[11]*State Key Laboratory of Biocontrol and Guangdong Provincial Key Laboratory of Plant Resources, School of Life Sciences, Sun Yat-sen University, Guangzhou, China*

Received 1 March 2017; revised 5 April 2017; accepted 7 April 2017; published online 13 April 2017.
*For correspondence (e-mails liuyg@sdau.edu.cn and jhzhang@cuhk.edu.hk).
†These authors contributed equally to this work.

SUMMARY

In eukaryotes, mechanisms such as alternative splicing (AS) and alternative translation initiation (ATI) contribute to organismal protein diversity. Specifically, splicing factors play crucial roles in responses to environment and development cues; however, the underlying mechanisms are not well investigated in plants. Here, we report the parallel employment of short-read RNA sequencing, single molecule long-read sequencing and proteomic identification to unravel AS isoforms and previously unannotated proteins in response to abscisic acid (ABA) treatment. Combining the data from the two sequencing methods, approximately 83.4% of intron-containing genes were alternatively spliced. Two AS types, which are referred to as alternative first exon (AFE) and alternative last exon (ALE), were more abundant than intron retention (IR); however, by contrast to AS events detected under normal conditions, differentially expressed AS isoforms were more likely to be translated. ABA extensively affects the AS pattern, indicated by the increasing number of non-conventional splicing sites. This work also identified thousands of unannotated peptides and proteins by ATI based on mass spectrometry and a virtual peptide library deduced from both strands of coding regions within the Arabidopsis genome. The results enhance our understanding of AS and alternative translation mechanisms under normal conditions, and in response to ABA treatment.

Keywords: abscisic acid, alternative splicing, *Arabidopsis thaliana*, proteogenomics, splicing factor, translation, technical advance.

INTRODUCTION

Eukaryotic organisms possess intron-containing genes that facilitate comprehensive mechanisms at the post-transcriptional level. Using next-generation sequencing, the human transcriptome has been suggested to consist of greater than 80 000 transcripts and 250 000 proteins from approximately 20 000 genes (de Klerk and t Hoen, 2015). These estimates provided important insight concerning the greater than previously anticipated level of complexity of eukaryotic genomes, and their embedded genetic information. Gene expression is affected on four levels, including alternative transcription initiation, splicing, polyadenylation and translation initiation, which interdependently determines protein diversity in eukaryotic cells (de Klerk and t Hoen, 2015). The precursor RNA of genes undergoes splicing, which greatly increases the coding potential and functionality of multi-exonic genes in both animals and plants (James *et al.*, 2012; Ruhl *et al.*, 2012; Chang *et al.*, 2014; Feng *et al.*, 2015). Splicing machinery can be classified into two categories, which are U2-(GT-AG) and U12-type (AT-AC), according to the differential assembly of the spliceosome (Lorkovic *et al.*, 2005; Will and Luhrmann, 2011). Amongst them, GT/AT is referred to as a 5′ donor splicing site, whereas AG/AC is referred to as a 3′ acceptor site. In addition to the middle branch point sequence, both motifs serve as binding sites for the components of the spliceosome (Lorkovic *et al.*, 2005; Will and Luhrmann, 2011). It has been estimated that greater than 95% of intron-containing genes are alternatively spliced in mammalian cells (Eckardt, 2013). Mutations that affect splicing mechanisms or the abundance of spliced isoforms account for approximately 15% of human genetic diseases (Eckardt, 2013), indicating their pivotal role in cellular function. Recent studies proposed that greater than 60% of intron-containing genes undergo alternative splicing (AS) in plants, such as Arabidopsis, *Glycine max* (soybean) and *Zea mays* (maize) (Filichkin *et al.*, 2010; Aghamirzaie *et al.*, 2013; Thatcher *et al.*, 2014).

Alternative translation initiation (ATI) was discovered in the early 1990s, and has drawn considerable attention recently (Sonenberg and Hinnebusch, 2009). Using combinatorial approaches of ribosome sequencing and proteomic identification, a large number of alternative protein initiation sites have been identified (Ingolia *et al.*, 2011; Lee *et al.*, 2012; Menschaert *et al.*, 2013). Surprisingly, a considerable portion of these sites do not use the conventional AUG as a start signal (Sonenberg and Hinnebusch, 2009; Ingolia *et al.*, 2011; Lee *et al.*, 2012), suggesting that unidentified mechanisms in protein translation initiation may exist. In addition, the identification of upstream open reading frames (uORFs) and other small ORFs in traditional non-coding regions further expand our understanding of genome coding ability (Slavoff *et al.*, 2013; Mackowiak *et al.*, 2015).

Alternative splicing (AS) mediates plant biological processes from plant development to stress responses (James *et al.*, 2012; Ruhl *et al.*, 2012; Chang *et al.*, 2014; Feng *et al.*, 2015; Wang *et al.*, 2015; Zhan *et al.*, 2015; Thatcher *et al.*, 2016). Crucial developmental processes such as the circadian clock and flowering time are strictly affected by AS (James *et al.*, 2012; Rosloski *et al.*, 2013). For instance, the intron-containing AS isoform of the *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*) transcript was observed to be non-functional because of the presence of a premature termination codon (PTC). This AS isoform was specifically increased under high light levels and was reduced under cold conditions (Filichkin *et al.*, 2010). Similarly, another recent study indicated that the transcription of *FLOWERING LOCUS C* (*FLC*) was modulated by its antisense lncRNA AS isoforms (Marquardt *et al.*, 2014). In addition, AS has been documented to participate in stress responses. For example, polypyrimidine tract binding protein homologs (PTBs) were found to be key components of AS in response to stress treatments. A total of 452 AS events from 307 independent genes were observed to be affected by PTB1 and PTB2 (Ruhl *et al.*, 2012). Two HYPERSENSITIVE TO ABA 1 (HAB1) AS isoforms exhibited opposite roles in abscisic acid (ABA) signal transduction during seed germination through competitive interaction with OPEN STOMATA 1 (OST1) (Wang *et al.*, 2015). Furthermore, an SNW/Ski-interacting protein (SKIP) domain protein was reported to be crucial in the splicing of some salt-responsive genes in Arabidopsis (Feng *et al.*, 2015). Genome-wide analysis indicated that SKIP was responsible for the mRNA maturation of several salt-tolerance genes. The splicing variant of FLOWERING LOCUS M (FLM) has been demonstrated as a developmentally- and environmentally-specific isoform to control flowering in a temperature-dependent mode (Pose *et al.*, 2013). Non-sense-mediated mRNA decay and AS were considered major factors that affect the transcript level of its splicing isoforms (Sureshkumar *et al.*, 2016). Recent discoveries thus reveal extensive underlying mechanisms at the post-transcriptional level, indicating that a large currently hidden network in plants remains to be elucidated.

Although stress-induced genome-wide AS changes have been extensively studied, the function of these AS isoforms at the protein level are seldom reported. In addition, it is difficult to get full-length AS transcripts through next-generation sequencing techniques such as short-read RNA sequencing (sr_RNA seq.). Proteogenomics, the parallel analysis of genomic and proteomic data to refine existing gene models based on prediction (Castellana *et al.*, 2008; Kumar *et al.*, 2016), represents an attractive approach to gain deeper insight into the functional importance of AS. In the current study, a single molecule long-read sequencing

(lr_RNA seq.) assisted proteogenomic approach has been used to simultaneously analyse both transcriptome and proteome data sets. Given that ABA is a crucial phytohormone that participates in various stress responses, we treated seedlings with ABA to study its action on AS in Arabidopsis. For this purpose, a combined experimental and bioinformatics approach was taken in order to compare the relative abundances of splice variants and their corresponding proteins in plants in the presence and absence of exogenously applied ABA.

RESULTS

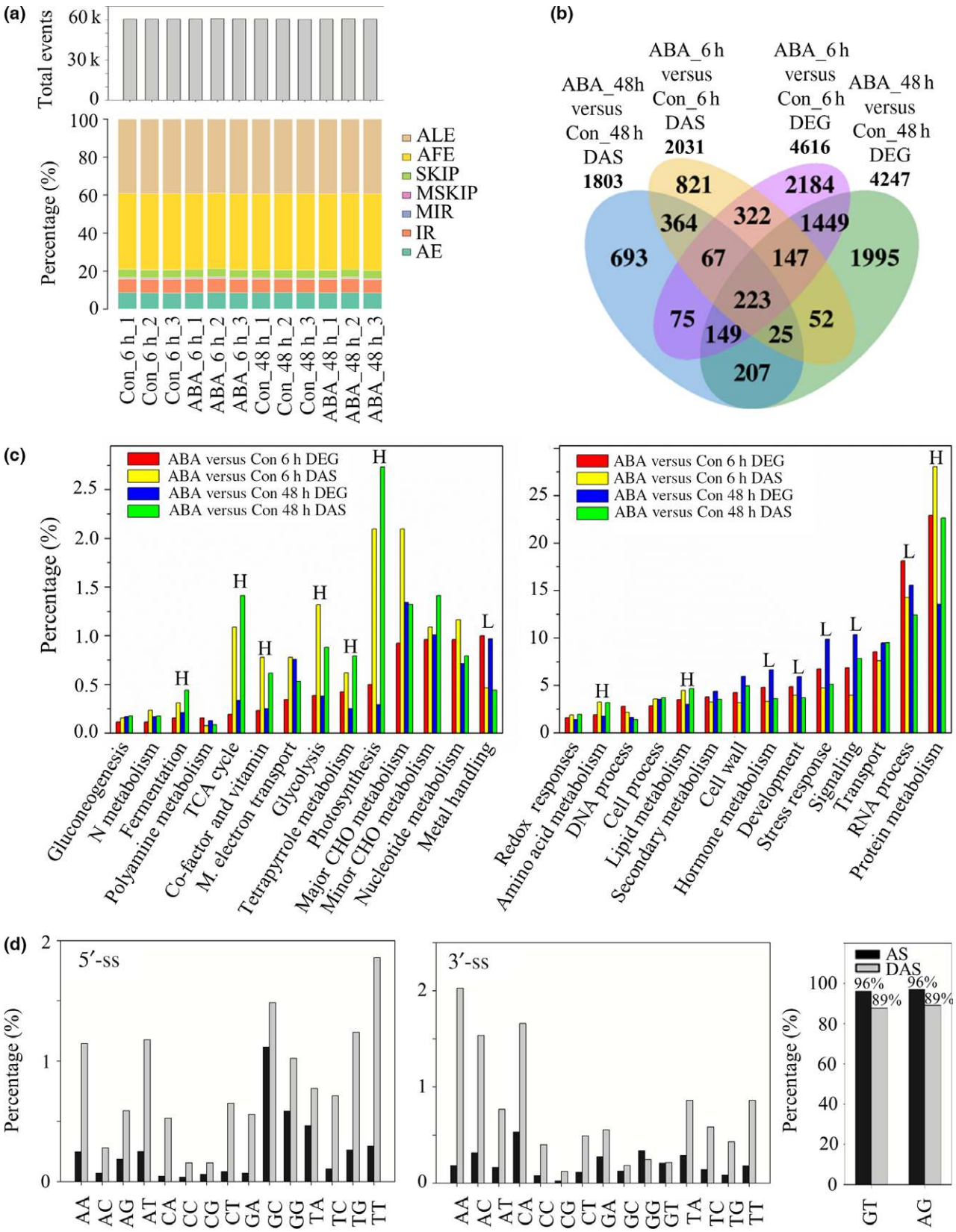
Short-read RNA sequencing reveals extensive AS changes subject to ABA treatment

The overall analytical pipeline is presented as a flowchart (Figure S1). Initially, we performed transcriptomic analysis for AS event identification using 12-day-old Arabidopsis seedlings treated with or without ABA (50 μM) at 6 and 48 h (Figure 1a; Table S1). The obtained sequencing data were of high quality (i.e. over 40 million reads and approximately 90% mapping percentage per sample, three replicates per treatment), and exhibit distinct categories (Figure S2). Approximately 78% (22 785/29 103) of intron-containing genes exhibited AS events. Amongst them, alternative first exon (AFE) and alternative last exon (ALE) were detected as the most abundant AS events in the 12 samples analysed (Figures 1a and S3a). Computational analysis estimates that at least 49% of human transcripts have alternative 3′ ends, which may subsequently affect functional aspects such as subcellular localization, stability and translational efficiency (Andreassi and Riccio, 2009; Sonmez *et al.*, 2011; Jenal *et al.*, 2012). Therefore, in this study, we used software designed for animal research, which led to the discovery of two types of AS events at a much higher abundance than previously claimed in Arabidopsis (Filichkin *et al.*, 2010; James *et al.*, 2012). Subsequently, a comparison was made between differentially expressed genes (DEGs) and differentially expressed AS (DAS) genes in response to exogenous ABA (Figure 1b). The data set was examined statistically to confirm its validity under treatment with ABA (50 μM; Figure S3b–d). Both the cluster and correlation analysis successfully classified the three replicates of each treatment into subgroups (Figure S3c,d). The transcript levels were consistent amongst three replicates of each sample, and certain pathways were enriched (Figure S4a,b). Genes involved in the ABA pathways (Nemhauser *et al.*, 2006; Hirayama and Shinozaki, 2007; Goda *et al.*, 2008) were altered at the transcript level ($Log_2FC > 2$ and q value; false discovery rate, FDR < 5%) in 50 μM ABA-treated samples for 6 and 48 h, compared with the water-treated control (Figure S4c). The induction of several major marker genes in ABA-treated samples was validated by real-time quantitative PCR (qRT-PCR; Figure S4d). Interestingly, approximately 40% (i.e. 821/2031 in ABA_6 h versus Con_6 h DAS; 693/1803 in ABA_48 h versus Con_48 h DAS) of the AS isoforms were differentially expressed despite the fact that the total abundance of transcripts from the genes that encode both isoforms was invariant, suggesting an independent mechanism of the AS transcripts from the conventional group of the differentially expressed genes (Figure 1b). Moreover, gene ontology classification revealed that certain categories, such as the tricarboxylic acid (TCA) cycle, glycolysis, photosynthesis, etc., were highly enriched in DAS compared with DEG, whereas some categories such as metal handling, hormone metabolism, development, etc., were downregulated, indicating that AS may play an important role in the ABA response by affecting gene groups other than conventional DEG (Figure 1c). Furthermore, categories enriched or depleted during treatment (i.e. 48-h versus 6-h samples) may represent long-term AS attenuation in response to ABA (Figure S5a). In addition, the conservation of splicing sites was also analysed (Figure 1d). Conventional splicing sites (5′-GT-AG-3′) accounted for 96% of AS sites identified from all the samples in our data set (Figure S5b,c), which is consistent with previous findings (Will and Luhrmann, 2011). The percentage of GT/AG decreased to 89% following 50-μM ABA treatment (i.e. ABA_6 h versus Con_6 h and ABA_48 h versus Con_48 h), however, whereas the percentage of other non-canonical AS sites reached 130–1110% in comparison with the water control (Figure 1d), suggesting that a portion of differentially expressed AS events under ABA treatment may result from the increased usage of the non-canonical splicing sites.

Single molecule long-read sequencing is complementary to short-read RNA sequencing to facilitate full-length AS isoform identification

In sr_RNA seq., several AS events have been identified in one locus, but how are they organized in transcripts

Figure 1. Bioinformatic comparison between the differentially expressed genes (DEGs) and differentially expressed alternative splicing (DAS) gene data sets. (a) Summary of the identified AS quantity and types. The total number of AS events identified in each replicate was 60 473 (Con_6h_1), 60 502 (Con_6h_2), 60 488 (Con_6h_3), 60 519 (ABA_6h_1), 60 781 (ABA_6h_2), 60 781 (ABA_6h_2), 60 374 (ABA_6h_3), 60 576 (Con_48 h_1), 60 572 (Con_48 h_2), 60 483 (Con_48 h_3), 60 490 (ABA_48 h_1), 60 670 (ABA_48h_2) and 60 512 (ABA_48h_3). ALE, alternative last exon; AFE, alternative first exon; SKIP, exon skipping; MSKIP, multiple exon skipping; MIR, multiple intron retention; IR, intron retention; AE, alternative exon. (b) Venn diagram presents common and unique genes between the DEG and DAS gene data sets. Significant changes of DEG and DAS were determined using $Log_2FC > 2$ and q-value (false discovery rate, FDR < 5%). (c) Gene ontology comparison between the DEG and DAS genes. M. (mitochondria) electron transport; H (higher) and L (lower) represent the increased or decreased percentage of GO terms at both ABA_6 and 48 h versus Con_6 and 48 h, respectively. (d) Frequency of the splicing sites (ss) selection between total AS events (103 740) identified from all the samples (AS: black bars) and the 50 μM ABA-affected events (5020), i.e. ABA_6 h versus Con_6 h and ABA_48 h versus Con_48 h (DAS: grey bars).



(a)

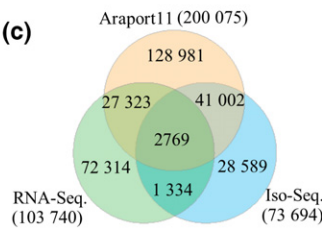
	Total transcripts	Known transcripts	New transcripts in known gene	New transcripts in known coding gene	Novel transcripts
Con_6 h	42 480	8387	27 846	25 691	34 871
ABA_6 h	35 530	9071	21 325	19 384	27 219

	Fusion transcripts	Antisense transcripts	Intergenic transcripts
Con_6 h	2098	506	5210
ABA_6 h	1343	530	3444

(b)

AS Events	Match/Include	Overlap	Coverage
AFE	8.93	59.34	68.27
ALE	9.23	60.34	69.57
AE	28.96	53.37	82.32
IR	31.10	54.15	85.26
SKIP	36.36	25.87	62.22
Total	13.43	62.57	76.00

(c)



(d)

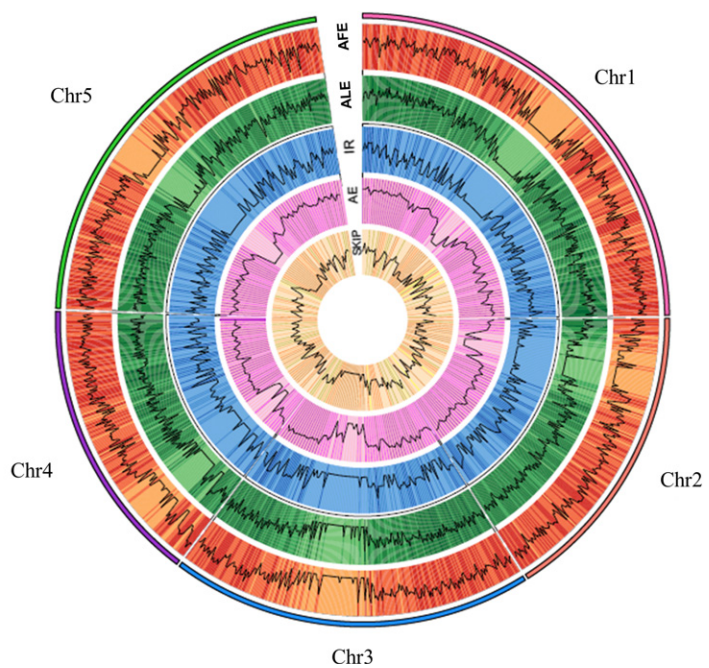


Figure 2. Comparison of alternative splicing identification between the short-read RNA sequencing and single-molecule long-read sequencing approaches. (a) Summary of full-length transcripts identified by single-molecule long-read sequencing. Known gene, transcribed gene annotated in Araport11; known coding gene, transcribed gene annotated in Araport11 with protein coding ability. (b) Coverage comparison of identified alternative splicing (AS) event types between short-read RNA sequencing (sr_RNA seq.) and single-molecule long-read sequencing (lr_RNA seq.). (c) Venn diagram presenting common and unique AS events identified by two sequencing technologies in comparison with Araport11. (d) CIRCOS visualization of the frequency of five AS event types at the genome-wide level.

remains unknown. Single-molecule long-read sequencing technology was therefore applied to identify full-length AS isoforms from two 6-h treated samples (Con_6 h and ABA_6 h). A total of 844 003 reads of the insert were identified after sequencing four libraries (i.e. 1–2, 2–3, 3–6, and 5–10 k) of both samples. Amongst these reads, 62.99% of them were classified as full-length non-chimeric reads, 32.94% of them were non-full-length reads. Short and chimeric reads accounted for less than 5% of the total reads (Table S2). The sequencing generated 42 480 and 35 530 high-quality transcripts from two Arabidopsis samples, respectively. Approximately 35% of the genes annotated in Araport11 were detected in our analysis, with an average number of 38.6 and 33.7 reads of insert per gene in Con_6 h and ABA_6 h samples, respectively (Figure 2a). Amongst them, a considerable proportion of unannotated transcripts were identified (Figure 2a). AS identification revealed 73 694 AS events (6.65 and 5.38 events per gene) from the Con_6 h and ABA_6 h samples, which is considerably fewer than the total number of AS events (103 740) identified in the sr_RNA seq. data set. There is an overlapping 67.6% of AS events identified by both technologies (Figure 2b,c and Appendix S1, S2). Interestingly, a proportion of AS events were uniquely identified either by sr_RNA seq. or by lr_RNA seq. in comparison with Araport11 (Cheng *et al.*, 2017), suggesting that combining both technologies will enhance AS identification. Combining the data of the two sequencing methods, over 80% of intron-containing genes revealed AS in Arabidopsis seedlings. Furthermore, frequency analysis using CIRCOS

indicated that some genomic regions harbour AS events with higher frequency (Figure 2d). Given the comparatively high cost and low throughput of single-molecule sequencing, this approach is not attractive to quantify differentially expressed AS transcripts or events. Therefore, we continued the following AS and proteomic analysis based on sr_RNA seq. data.

Classical components in the ABA pathway exhibit distinct AS isoforms

The 5020 differentially expressed AS events detected (i.e. ABA_6 h versus Con_6 h and ABA_48 h versus Con_48 h) were subjected to further investigation (Figures S6 and S7a). Amongst them, 2819 events were consistently upregulated, and 2007 events were consistently downregulated, at both 6 and 48 h (Figures S6 and S7b). As in the total AS data set, ALE, AFE and AE (other alternative exons) were the three most abundant AS types (Figure S7c). In addition, isoforms were randomly picked up and assembled using bioinformatic pipelines and validated using RT-PCR (Marquez *et al.*, 2012). Specifically, amongst 26 randomly selected AS isoforms, 22 were detectable at the mRNA level (Figure S7d). Moreover, a total of 48 AS events from 20 genes in the ABA signalling pathway were extensively affected when subjected to ABA (50 μM) treatment (Figures 3, S8, S9a–c; Table S3), and four (approximately 8.3%) of these events were not annotated by Araport11 (Figure S9d; Table S3). The results from the sr_RNA seq. data (44 of 48) were confirmed by RT-PCR using primers designed specifically to detect each isoform (Figure 3). The ABA effects on AS observed for some genes probably only reflect changes in gene transcript abundance, and the differences in their protein level remain to be elucidated. Additionally, protein isoforms translated from splicing variants shown alternative translation initiation, termination and domain composition, suggesting that ABA may affect the protein functions through AS. Given that AFE and ALE events were identified as the most common AS types, representative AFE of *ABI1* and ALE of *GBF3* were selected and validated through the rapid amplification of cDNA ends (RACE; Figure S9e). The PCR fragments matching the predicted size were subsequently confirmed in the DNA sequencing analysis.

Differentially expressed AS events are more likely to be translated

To further understand the fate of the AS events identified, we conducted an additional large-scale proteomic identification using tandem mass spectrometry (MS/MS) for all four samples used in sr_RNA seq. (Alfaro *et al.*, 2014; Tavares *et al.*, 2015). Proteomic profiling generated approximately 800 000–1 100 000 spectra per sample for peptide identification with high quality (Figures S10 and S11). Approximately 6000–8000 proteins per sample were identified in the UniProt database (Figure 4a; Table S4). An AS junction database was constructed following the approach of Alfaro *et al.* (2014), and used to search against spectra signals for AS peptide identification (Alfaro *et al.*, 2014). In total, 13 400 AS peptides and 2241 DAS peptides were identified (Figure 4b). Conventional thinking dictates that a large portion of transcribed mRNA will be degraded using RNA surveillance mechanisms, such as nonsense-mediated mRNA decay (Nicholson *et al.*, 2010; Drechsel *et al.*, 2013). Consistent with this in our data set, only 12.9% of the AS events were translated into peptides, suggesting that the majority of the AS isoforms may be processed at the post-transcriptional level; however, a considerably greater proportion (44.6%) of AS events induced by ABA (50 μM) treatment are translated into peptides, especially for event types such as ALE, AFE, exon skipping (SKIP), multi-exon skipping (MSKIP) and intron retention (IR) (Figures 4b–d and S12a,b). This observation suggests that a considerable number of induced AS proteins are generated in response to ABA, and these may well function in parallel with conventional components of the ABA response. Artefactual reasons cannot be ruled out to explain the majority of AS isoforms not being 'translated', however. For example, insufficient identification because of the low throughput of the proteomic study may lead to the non-identification of translated AS peptides. Concurrently with this general increase in AS proteins, ALE became considerably more abundant than AFE following ABA (50 μM) treatment, whereas the number of AE events was twofold increased compared with IR (Figure 4c). Given that these events (AFE versus ALE and AE versus IR) were comparable in quantity at the mRNA level, the reduction in AFE and AE events may result from a loss in their ability to be translated into proteins. For example, these two AS types may generate a greater proportion of transcripts containing premature stop codons (Nicholson *et al.*, 2010; Drechsel *et al.*, 2013). Moreover, a large proportion of the AS events identified were marked as unannotated features in comparison with Araport11 annotation. For example, 71.0% (RNA-seq.) and 59.4% (Iso-seq.) of the AS events cannot be identified in the Araport11 annotation (Figures 2c and 3b). Amongst the DAS events, 72.2% events were not annotated. Additionally, 70.2% of the DAS peptides were not present in the Araport11 annotation (Figure 3b), which demonstrates that these unannotated AS events may be involved in ABA responses. Selected DAS protein isoforms such as STT7 HOMOLOG 7 (STN7), PHOTOTROPIN 2 (PHOT2), DE-ETIOLATED 1 (DET1) and PHYTOCHROME INTERACTING FACTOR 3 (PIF3) were assembled and confirmed at both the mRNA and protein levels using RT-PCR and western blot, respectively (Figure 4e). Interestingly, all of protein-confirmed AS events were validated using RT-PCR, which demonstrates the validity of the proteogenomic approach for identifying

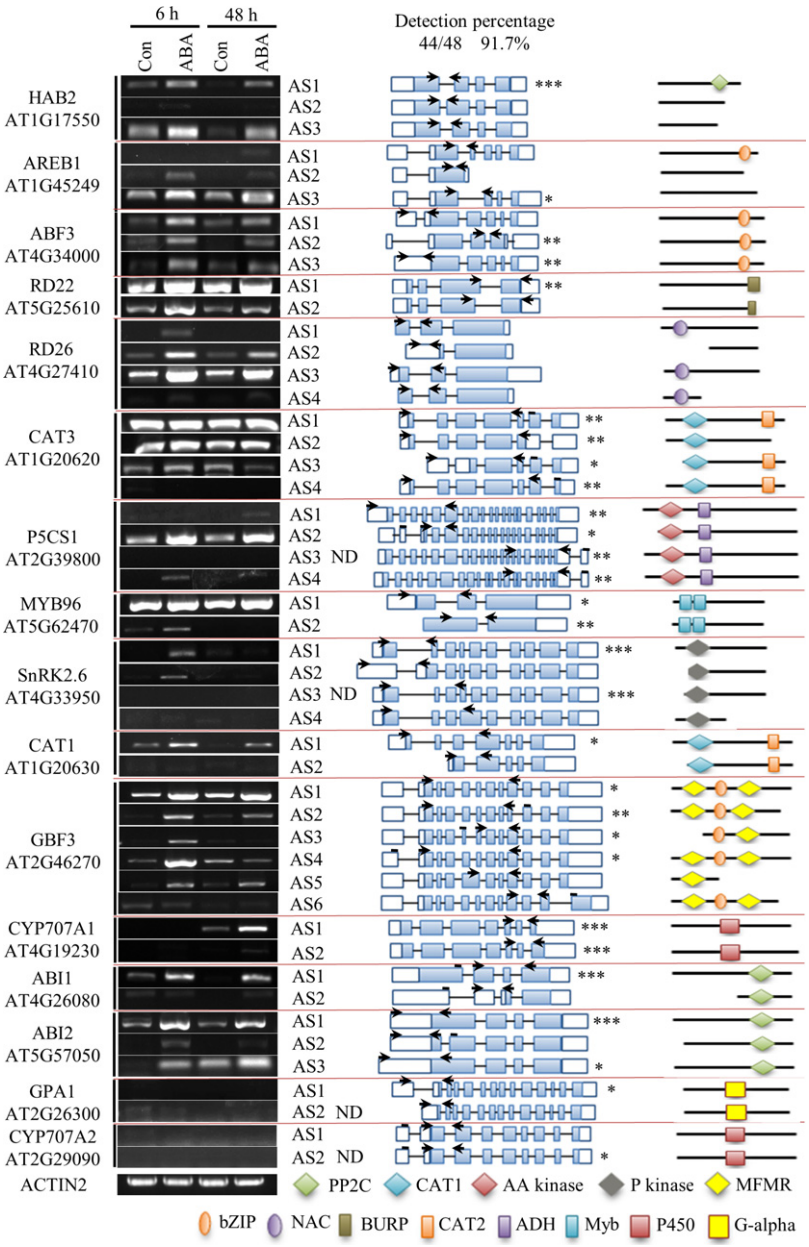
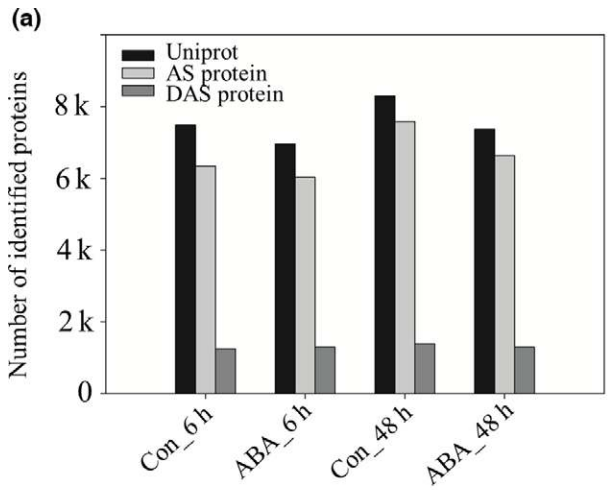
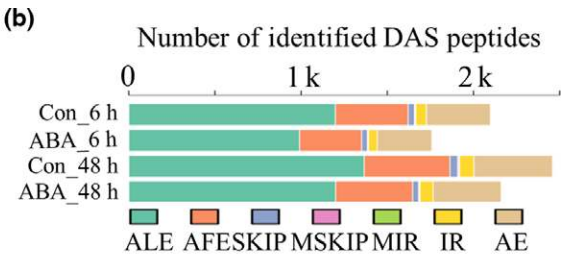


Figure 3. Validation of the differentially expressed alternative splicing (DAS) events in the abscisic acid (ABA) pathway. RT-PCR validation of the selected AS events. Gene models of each isoform are indicated (blue, coding region; white, non-coding untranslated regions; not drawn to scale). Arrows represent primers designed for each isoform. *, ** and *** represent 90, 95 and 99% mapped isoforms from the single-molecule long-read sequencing data set, respectively. ND, not detected. Primers used in the experiment are listed in Table S9. AA kinase, amino acid kinase domain; ADH, aldehyde dehydrogenase domain; BURP, BURP domain; bZIP, basic leucine zipper; CAT1, catalase domain 1; CAT2, catalase domain 2; G-alpha, G-protein alpha subunit; MFMR, domain of a G-box binding protein; Myb, MYB domain; NAC, NAM, ATAF1/2 and CUC2 domain; P450, cytochrome P450; P kinase, protein kinase domain; PP2C, protein phosphatase 2C domain.

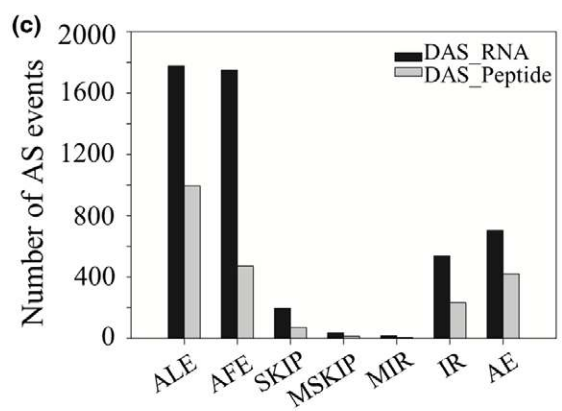
protein isoforms in plants (Figure 4e). In addition, a unique group of peptides was identified in the ABA-treated data sets from the 900 peptides/sample (Figure S12c), which may serve as good candidates for further investigation. Among them, the proteins involved in metabolism, such as fatty-acid and amino-acid metabolism, were over-represented, whereas the proteins functioning in transcription were under-represented (Table S5).

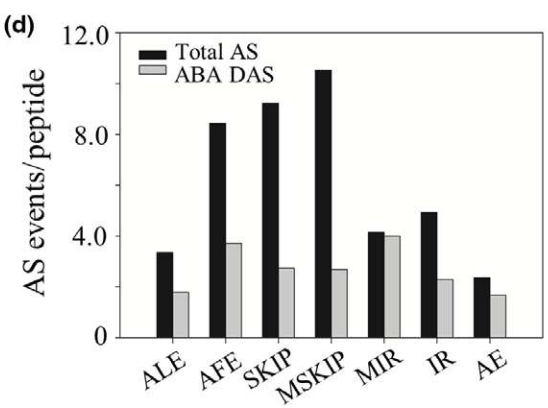
Figure 4. Proteomic identification of alternative splicing (AS) events. (a) Number of identified proteins from each database [Uniprot, AS and differentially expressed AS (DAS)]. (b) Statistical analysis of the AS events identified in the proteomic data set (upper panel). Summary of the AS events identified from the sr_RNA seq. and the AS peptides identified from the proteomic data (lower panel). (c) Statistics with respect to each type of AS event identified in the sr_RNA seq. and proteomic data sets. (d) Number of AS events per peptide. (e) Western blot and RT-PCR validation of the selected AS events identified in the proteomic data set. Arrowheads represent the protein isoforms identified using the corresponding antibodies. The red line represents the peptide region used to generate the polyclonal antibodies. The remainder of genes used antibodies generated from whole proteins. The arrows represent primers designed for each isoform. The gene models of each isoform are denoted (blue, coding region; white, non-coding untranslated regions; not drawn to scale). The arrows represent primers designed for each isoform. All validation experiments were repeated at least twice. 'x' represents the stop codon introduced in the AS isoforms. *, ** and *** represent 90, 95 and 99% mapped isoforms from the single-molecule long-read sequencing data set, respectively.

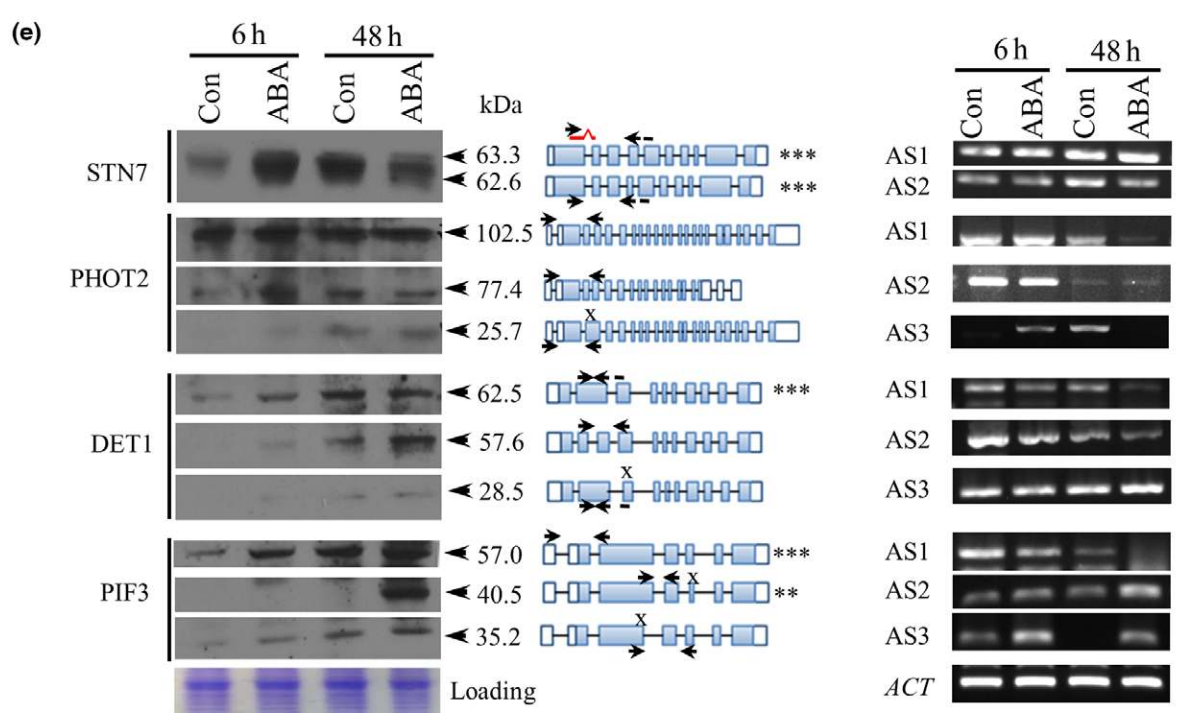




Class	Araport11	New	Total	Percentage
AS_RNA	30 092	73 648	103 740	
AS_peptide	3612	9788	13 400	12.9%
DAS_RNA	1394	3626	5020	
DAS_peptide	667	1574	2241	44.6%







Unannotated proteins involved in the ABA response were identified by customized libraries

An intriguing finding in the proteomic data set was that the spectra usage for identification was not greater than 40% when using the UniProt and the AS junction databases as input files, indicating that a large quantity of peptides cannot be identified in the spectra data (Kim *et al.*, 2014) because of a lack of reference annotation (Figure 5a). Surprisingly, a certain number of peptides were identified in the second frame using the AS junction library generated at six frames. Although it is possibly part of a protein isoform because of the use of the second frame (Brown *et al.*, 2015), it may also suggest that a single transcript has the ability to encode more than one protein through the use of alternative translation initiation sites (Brar and Weissman, 2015). Therefore, we constructed a modified virtual peptide library (Castellana *et al.*, 2008) using all transcripts annotated in TAIR 10. Putative proteins greater than 80 amino acids were selected from loci transcribed at six frames, followed by subtracting existing proteins in the TAIR 10 annotation. Interestingly, 1957 potential protein-coding genes were identified in other ORFs in either the sense or the antisense strand of transcribed regions. Amongst them, 895 protein-coding antisense genes were identified (Table S6), indicating that the coding ability of the genome can be further increased using the ATI sites on both strands of the same locus. In addition, a group of proteins were present uniquely in the ABA-treated (50 μM) groups (Figure 5b), suggesting that a considerable number of proteins of unknown function may be involved in the ABA response. Currently, it is widely accepted that gene expression in response to ABA is mainly regulated by conserved *cis*-acting elements, called ABA-responsive elements (ABREs), bZIP transcription factors and ABRE-binding protein/ABRE-binding factors (AREB/ABFs) (Fujita *et al.*, 2013). Examples of new proteins efficiently using the coding ability of genome sequences were illustrated (Figure 5c; Table S7), demonstrating the synergistic effect of ATI, and that antisense genes extensively contribute to protein diversity in response to ABA.

Isoforms of splicing factors in response to ABA treatment

In order to further understand the function of splicing factors (SFs) in the ABA response, we examined AS changes of putative splicing factors in Arabidopsis. In our data set, 31 events from 21 genes were differentially expressed under ABA treatment, and approximately 50% of these AS events were not present in the TAIR 10 annotation (Figure S13; Table S8). Thirty-nine of the 44 isoforms were validated using RT-PCR (Figures 6 and S14), indicating that a considerable number of splicing components were affected by the ABA treatment.

DISCUSSION

Transcriptomic and proteomic approaches have been applied in plant research to study AS (Castellana *et al.*, 2008; Yang *et al.*, 2015). The sr_RNA seq. and single-molecule long-read sequencing data reveal AS changes at the transcript level; however, which transcript isoforms will be translated into proteins is difficult to determine because of the presence of a variety of mRNA degradation mechanisms (Nicholson *et al.*, 2010; Drechsel *et al.*, 2013). By contrast, pure proteomic methods are limited by the fact that currently available databases were generated on the basis of incompletely annotated genome information. This

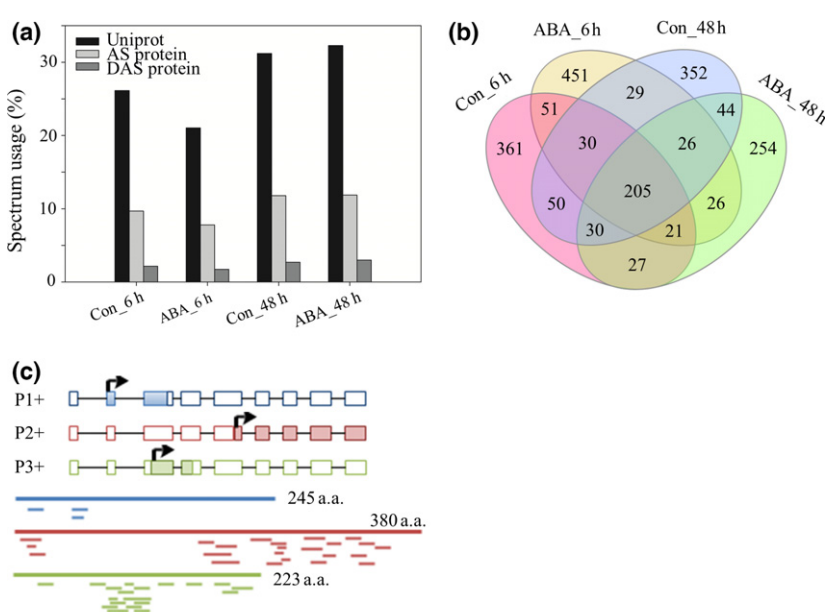


Figure 5. Characterization of proteins identified from the virtual peptide library. (a) Summary of the spectra used for the proteomic data. (b) Venn diagram depicting the overlapping and unique proteins identified from the virtual peptide library between the water control and the ABA-treated (50 μM) groups. (c) Schematic diagram of the three proteins identified, presented in the same translation direction by peptides (not drawn to scale).

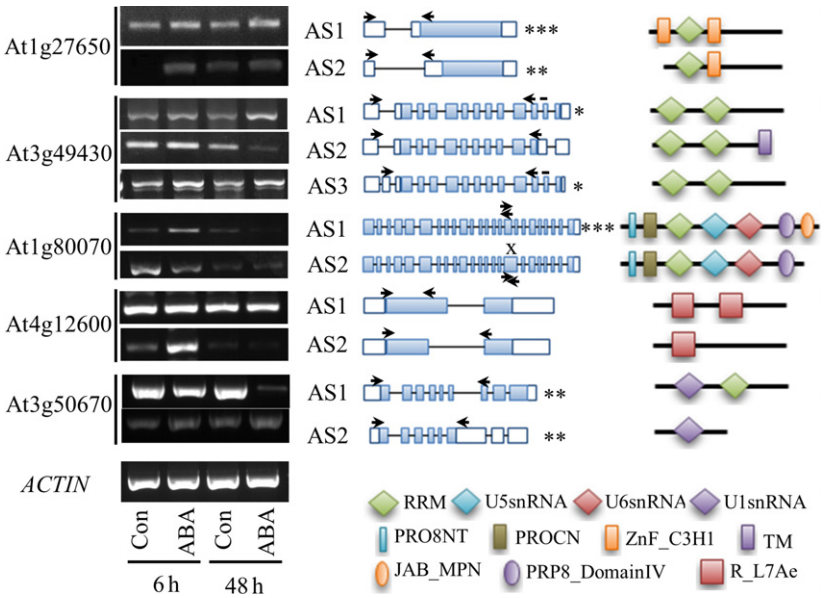


Figure 6. Regulation of splicing of known and predicted splicing factors in response to abscisic acid (ABA). RT-PCR validation and domain analysis on the differentially expressed splicing factor isoforms. Gene models of each isoform are indicated (blue, coding region; white, non-coding untranslated regions; not drawn to scale). Arrows represent primers designed for each isoform. All validation experiments were repeated at least twice. 'x' represents the stop codon introduced in the alternative splicing (AS) isoforms. *, ** and *** represent 90, 95 and 99% mapped isoforms from the single-molecule long-read sequencing data set, respectively. JAB_MPN, PAD1-like domain; PRO8NT, N-terminal domain of PRO8; PROCN, central domain of PRO8; PRP8_DomainIV, essential domain of PRP8 that can interact with U5snRNA and U6snRNA; R_L7Ae, ribosome-associated binding sites; RRM, RNA recognition motif; TM, transmembrane domain; ZnF_C3H1, zinc finger.

study represents a customized analytical pipeline developed to study the involvement of AS in the response of plants to ABA.

Consistent with the fact that endogenous ABA levels increase in response to abiotic stresses, such as drought and high salinity, transcriptome studies have shown that exogenous ABA induces a myriad of genes responsive to the stresses (Seki *et al.*, 2002; Nemhauser *et al.*, 2006; Kilian *et al.*, 2007; Goda *et al.*, 2008). Accordingly, transcriptional regulation in ABA signalling has been extensively studied, and a lot of key regulators, including transcription factors, have been identified. In contrast, less attention has been paid to post-transcriptional regulation in ABA signalling. Our proteogenomic study using Arabidopsis treated with ABA suggests that AS plays important roles in ABA signalling. Not only AS, but RNA processing and metabolism has been thought to be essential in ABA signalling, as evidenced by the mutants harbouring altered ABA sensitivity. For example, increased sensitivity to ABA was observed in the mutants of HYL1 and ABH1, which encode a double-strand RNA binding protein and an mRNA cap binding protein, respectively, and the expression of several genes in ABA signalling was altered in each mutant (Fedoroff, 2002; Kuhn and Schroeder, 2003; Hirayama and Shinozaki, 2007). Given that splicing factors themselves are affected via ABA (Figures 6 and S14), RNA processing and metabolism in ABA signalling might be affected by complex mechanisms consisting of a large number of yet uncharacterized genes. As in previous studies, we showed that the genes involved in the stress response and signalling are likely to be affected by expression rather than AS (Figure 1c). By contrast, the genes involved in photosynthesis and the TCA cycle seem to be affected by AS, and the isoform proteins involved in light signalling, such as PHOT2, PIF3 and DET1, were differentially accumulated in response to ABA (Figure 4e). Whereas PHOT2 functions as a blue light receptor in guard cells to mediate stomatal opening as well as PHOT1 (Shimazaki *et al.*, 2007), PIF3 and DET1, a transcription factor and a nuclear-localized protein involved in chromatin modification, respectively, play roles in light-related gene expression (Jiao *et al.*, 2007). The expression of many photosynthesis-related genes is decreased in response to ABA or abiotic stresses (Seki *et al.*, 2002). Some of the protein isoforms, such as STN7 (62.6 kDa, 6 h), PHOT2 (77.4 kDa, 6 h) and DET1 (62.5 kDa, 6 h, 57.6 kDa, 48 h), were elevated under ABA treatment, whereas their transcript abundance remained unchanged or was reduced (Figure 4e). As most of the ABA-responsive genes are inducible (Seki *et al.*, 2002; Nemhauser *et al.*, 2006), however, our results imply that the involvement of AS may be more important than transcriptional repression to modulate the abundance of proteins involved in growth and development. Further evidence is needed to corroborate this hypothesis, however.

The pattern of AS was extensively changed under ABA treatment (Figure 1), ranging from the ABA-responsive genes to splicing factors (Figures 3, 4 and 6). Splicing factors have long been demonstrated to participate in spliceosome assembly and AS site selection (Kalyna *et al.*, 2006; Will and Luhrmann, 2011). In particular, the U1snRNA and U2AF small units are major players in 5′ and 3′ splicing site recognition, respectively (Golovkin and Reddy, 1996; Krummel *et al.*, 2009; Kondo *et al.*, 2015; Yoshida *et al.*, 2015). Moreover, splicing factors such as the polypyrimidine tract binding proteins (PTBs) and SKIPs (Ruhl *et al.*, 2012; Feng *et al.*, 2015) are involved in hormonal signalling and stress responses. In this study, the splicing factors (SFs) were

extensively affected by AS in response to ABA (Figures 6 and S14). They may participate in the ABA response by increasing the number of non-conventional splicing sites (Figure 1d,e). Simulations of the binding between SF isoforms and ssRNA imply the possibility of other SF splicing sites (Figures S16 and 17), suggesting that this alternation of the splicing sites under ABA treatment may result from the recruitment of different SF isoforms in the spliceosome. Increasing evidence implies that isoforms generated from AS can rewire protein interaction networks with the presence or absence of protein binding motifs (Buljan *et al.*, 2012; Ellis *et al.*, 2012). Therefore, important protein isoforms were assembled according to information provided by both sequencing technologies, and subject to structural analysis. For example, SUS2 (At1g80070) is a central component in spliceosome, and the loss of the JAB_MPN domain in isoform 2 resulted in the abortion of the protein interactions between SUS2 and At1g66510, demonstrating its role in remodelling protein interaction networks (Figure S15a). Furthermore, At4g12600 and At5g20160 are two auxiliary proteins that can bind to U4snRNA. The alternation of different isoforms may cause a decrease (Figure S15b) or even a complete loss of the binding ability between the protein and U4SnRNA or between protein–protein interactions with the third component, PRP31, in the complex (Figure S16). Moreover, significant changes in binding energy were not observed amongst the conventional AG and the other 3′ AS sites when calculated using U2AF35a (At1g27650) (Figure S17), suggesting that the possible selection of non-canonical splicing sites may occur when subject to ABA treatment. Further experiments must be conducted to reveal the underlying mechanism in this induced splicing site selection, however. Subsequent phenotypic analysis using Arabidopsis T-DNA insertion mutants (Figure S18a,b) revealed that the knock-out of two splicing-related proteins contributed to differential sensitivity in response to ABA treatment (Figure S18c). This observation indicates that the alternation of the protein level of splicing factors may affect the ABA sensitivity in Arabidopsis seedlings. Therefore, ABA may also induce compositional or conformational changes of the spliceosome.

Our combined analyses indicate that differentially expressed AS events are more likely to undergo protein translation. RNA surveillance mechanisms such as nonsense-mediated decay are considered to be one of the major determinants of the Arabidopsis transcriptome (Drechsel *et al.*, 2013). The majority of AS events detected in our study were not observed in the proteomic data set, however (Figure 4b). That said, a greater percentage of AS events were translated into peptides under ABA treatment (Figure 4b), indicating their potential roles in the response to ABA. Assembled AS isoforms were further validated using lr_RNA seq. (Figures 3, 4, 6, S7, S9 and S14), which is an innovative technology that can aid in the accurate identification of full-length transcripts for subsequent functional study. This sequencing technology is hampered by its high cost and low throughput, however. The sr_RNA seq. analysis detected more AS events than the lr_RNA seq. analysis (Figure 2c). Only 17.7% of the AS events identified by this method were matched with sr_RNA seq. data (Figure 2c); by contrast, approximately 67.6% of the AS events found in sr_RNA seq. were matched or overlapped by the lr_RNA seq., suggesting that each method has its own special advantages and disadvantages. The lr_RNA seq. cannot supersede sr_RNA seq. technology in time at this moment. Furthermore, the emergence of genome editing technologies (Bortesi and Fischer, 2015) is likely to make the functional analysis of the different spliced variants considerably easier.

Using sr_RNA seq. and lr_RNA seq. techniques, we have increased the total number of identified AS events in Arabidopsis and have additionally characterized two AS types (AFE and ALE). AFE and ALE are two event types identified in mammals that have great potential to generate different mRNA transcripts from the same transcription unit (Yan and Marr, 2005; de Klerk and t Hoen, 2015). It has been proposed that an average of four AFE events occur per gene in humans, and most of the AFE events are regulated in a tissue-specific manner (Consortium *et al.*, 2014, Taj *et al.*, 2014; de Klerk and t Hoen, 2015). Conversely, ALE resulting from both AS and polyadenylation generates transcripts differing at their 3′ end (Yan and Marr, 2005; de Klerk and t Hoen, 2015), which constitutes an extra layer to the regulation of protein diversity. In Arabidopsis, a considerable number of proteins (1957) was discovered using customized virtual peptide libraries, which extends the current annotation of the Arabidopsis genome. Their reliability needs to be validated using an independent approach such as western blot analysis, however. Pervasive transcription has long been proposed as a common phenomenon in both eukaryotic and prokaryotic species (Jensen *et al.*, 2013; Wade and Grainger, 2014). Its significance within biological processes remains unclear, however. Previous profiling analysis has estimated that at least 38% of annotated human transcripts have antisense expression, referred to as non-coding RNAs (Pelechano and Steinmetz, 2013). Such antisense expression may subsequently affect the expression of the corresponding sense transcripts (Song *et al.*, 2013; Marquardt *et al.*, 2014; Balbin *et al.*, 2015). Recent findings have, however, demonstrated that some of these antisense genes were able to encode proteins (Suenaga *et al.*, 2014). That said, the coding ability of antisense genes is rarely reported in plants. Here, using reverse complementation translated protein databases, 960 potential protein-coding antisense gene fragments were identified, 895 of which were not annotated in TAIR 10, implying that the antisense genes also contribute to protein diversity,

both under normal growth conditions and in response to ABA treatment, and with a considerably higher genome coding capacity than had previously been anticipated.

In conclusion, this study expands our knowledge of the genome coding ability of Arabidopsis under normal growth conditions and following the exogenous application of ABA. AS and ATI are two crucial factors that influence protein diversity. Frame usage in the same transcript further increases the genetic complexity. Importantly, ABA extensively affects AS and ATI patterns, and can lead to the discovery of the underlying regulatory mechanisms in its signal transduction (Figure 7). The components in ABA signal transduction were found to have multiple DAS events in response to ABA treatment. Substantial use of non-canonical splicing sites when subject to ABA treatment may result from the alternation of the SF isoforms. In comparison to AS events detected under normal conditions, DAS events were more likely to be translated into proteins, indicating a large underground network of ABA signalling at the post-transcriptional level. This study demonstrates, using the ABA response as a case study, how proteogenomic methods may contribute to elucidate the coding capability of the genome.

EXPERIMENTAL PROCEDURES

Plant material, growth conditions and ABA treatment

Arabidopsis T-DNA insertion mutants (SALK_081292C and CS420602) were obtained from the Arabidopsis Biological Resource Center (ABRC; https://www.arabidopsis.org). Seeds of *Arabidopsis thaliana* (Col-0) were surface-sterilized with 20% bleach and 0.05% Tween-20. They were then sown on half-strength MS medium plates (Murashige and Skoog, 1962) supplemented with 1.0% (w/v) agar and 1.5% (w/v) sucrose. After 2 days of stratification, plates were incubated under 16-h light (23°C)/8-h dark (21°C) cycles. Twelve-day-old seedlings were treated with water control or 50 μM ABA. Seedling samples were harvested at 6 and 48 h, and then subjected to further transcriptomic and proteomic analysis. For the measurement of primary root length, 3-day-old seedlings of Col-0 and the T-DNA insertion mutants grown on MS plates were transferred to MS control plates and MS medium supplemented with 50 μM ABA. The seedlings were grown vertically under light/dark cycles of 16 h/8 h at 23°C. Photos were taken 10 days after transplantation.

RT-PCR and western blot analysis

Total RNA (5 μg) was reverse-transcribed into cDNA by using Superscript First-Strand Synthesis System (Invitrogen, now ThermoFisher Scientific, https://www.thermofisher.com), following the manufacturer's protocol. RT-PCR was conducted as described previously (Zhu *et al.*, 2013): the specific primers used for AS isoform identification are listed in Table S9. For T-DNA mutant identification, gene-specific primers (Table S9) were used to amplify the target area in both wild-type (WT) and mutant backgrounds. Western blot was carried out according to a previous description (Xu *et al.*, 2002), with minor modifications. Commercially available antibodies including Anti-STN7 (ab65483, Abcam, http://www.abcam.com/), Anti-PHOT2 (R3518, Abiocode, http://www.abiocode.com/), Anti-DET1 (R1241, Abiocode, http://www.abiocode.com/) and Anti-PIF3 (R3392, Abiocode) were used for western blot analysis. In general, proteins after SDS-PAGE separation were subsequently transferred to Hybond-C membrane (Amersham, now GE Healthcare Life Sciences, http://www.gelifesciences.com). The membrane was blocked in blocking solution (5% w/v non-fat milk) for 1 h and incubated for 2 h with corresponding primary antibodies (1 : 2000). After washing, the blot was incubated in anti-rabbit antibody (1 : 5000) for another 1 h. At last, the ECL Western Blotting Detection Kit (Amersham, now GE Healthcare Life Sciences) was used for band detection.

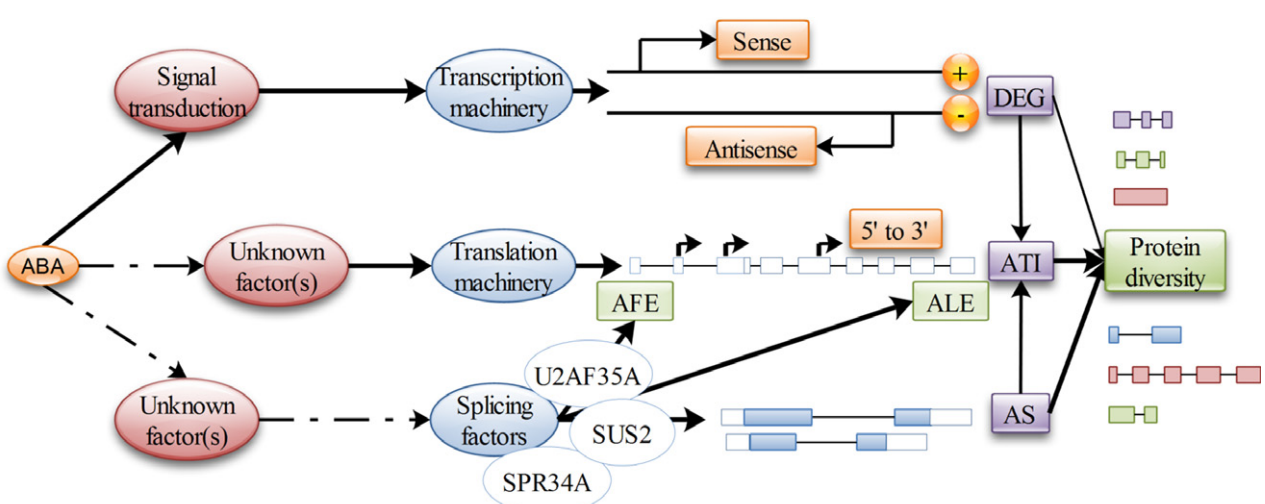


Figure 7. Model of abscisic acid (ABA) responses affecting protein diversity in Arabidopsis. Three pathways are proposed to participate in generating protein diversity. The conventional DEG pathway can generate differential regulated genes at either the sense or the antisense strand. The second pathway (alternative splicing, AS) generates different splicing isoforms through splicing factors like U2AF35A, SUS2 and SPR34A. These two pathways can be further expanded using alternative translation initiation (ATI) in the third pathway. sense, transcription in positive strand; antisense, transcription in negative strand. AFE, alternative first exon; ALE, alternative last exon; ATI, alternative translation initiation; DEG, differentially expressed genes.

Plant RNA extraction and sr_RNA seq

Plant total RNA was extracted by using the RNeasy Mini Kit (Qiagen, https://www.qiagen.com), according to the manufacturer's instructions. The extracted RNA was qualified using the kaiaoK5500® spectrophotometer (Kaiao, China, https://drawell-lab.en.alibaba.com/?spm=a2700.8304367.0.0.vTAjZT) and 2100 Agilent Bioanalyzer (Agilent, https://www.agilent.com) before library construction. A total quantity of 3 μg RNA per sample was used for paired-end library construction. Sequencing libraries were generated using NEBNext®Ultra™ RNA Library Prep Kit for Illumina® (#E7530L, NEB, USA), following the manufacturer's recommendations. Briefly, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads (Beckman Coulter, https://www.beckmancoulter.com). Fragmentation was carried out using divalent cations under elevated temperatures in NEBNext First Strand Synthesis Reaction Buffer (5X). First-strand cDNA was synthesized using random hexamer primer and RNase H. Second-strand cDNA synthesis was subsequently performed using buffer, dNTPs, DNA Polymerase I and RNase H. The library fragments were purified with QIAQuick PCR kits (Qiagen) and eluted with Elution buffer. After terminal repair, poly(A) sequence and adapter were implemented. The resulting cDNA libraries were run on a 2% agarose gel and bands of approximately 250 bp were excised and used for paired-end (2 × 125 bp) sequencing on an Illumina HiSeq 2500 platform by Annoroad Gene Technology Co. Ltd. (http://www.annoroad.com). Raw reads (567 729 742) from 12 samples (three replicates for each time point) were trimmed to obtain clean reads for subsequent analysis (Figure S2; Table S1).

Single-molecule lr_RNA seq. and data analysis

The library construction and subsequent sequencing was described previously (Wang *et al.*, 2016). Four libraries (i.e. 1–2, 2–3, 3–6, 5–10 k) were constructed according to the manufacturer's instructions (https://pacbio.secure.force.com/SamplePrep). A total of 16 SMRT cells (eight for Con_6 h and eight for ABA_6 h) were then sequenced by the Pacific Biosciences (http://www.pacb.com) RSII platform in Beijing Genomics Institute. The raw reads were processed by ToFu scripts as described previously [https://github.com/PacificBiosciences/cDNA_primer/wiki/tofu-Tutorial-(optional).-Removing-redundant-transcripts]. Mapping of high-quality reads and AS identification were performed by using GMAP with default settings (Abdel-Ghany *et al.*, 2016). To examine the overlap between sr_RNA seq. and lr_RNA seq., BLASTN was run with the cut-off criteria at minimal coverage (90, 95 and 99%) and minimal identity (90, 95 and 99%). CIRCOS was used to build an AS frequency map at the genome level with a 0.5-Mb sliding window (Wang *et al.*, 2016).

Bioinformatic analysis on sr_RNA seq. data

Arabidopsis reference genome annotation files were downloaded from ABRC. Clean reads were mapped to the reference genome by TOPHAT 2.0.12, with default parameters (Table S1). The analytical pipeline was similar to that used in a previous study (Yang *et al.*, 2015), with minor modifications. Generally, the unique mapped reads were used for transcript assembly. The transcripts were assembled by CUFFLINKS according to TAIR 10 annotation (Kumar *et al.*, 2016). The counts of each gene were quantified by HTSEQ 0.6.0 and the quantitative estimation of each transcript was achieved using fragments per kilobase of exon model per million mapped reads (FPKM). DEGs were analyzed by the DESEQ package. Significant changes were determined using $Log_2FC > 2$ and the *q*-value (false discovery rate, FDR < 5%) from multiple testing adjustment as the cut-off. Gene ontology (GO, http://geneontology.org) and the Kyoto encyclopedia of genes and genomes (KEGG, http://www.kegg.jp) enrichment analysis were conducted amongst differentially expressed genes using BLAST2GO and KAAS, respectively. AS events were identified and quantified by using ASPROFILE (http://ccb.jhu.edu/software/ASprofile; Florea *et al.*, 2013), and AS events with no expression data were filtered out before subsequent analysis. The DAS events were identified using the same criteria in the DEG determination ($Log_2FC > 2$ and *q*-value; false discovery rate, FDR < 5%). The conservation of splicing sites were presented using WEBLOGO 3 (http://weblogo.threeplusone.com; Crooks *et al.*, 2004).

Plant protein extraction, digestion and MS/MS analysis

Plant total proteins were extracted and digested as described previously (Chen *et al.*, 2014), with minor modifications. In total, approximately 10 g of Arabidopsis seedlings were used for protein extraction. The obtained pellets were subjected to trypsin digestion and desalted by Sep-Pak C_{18} column (Waters, http://www.waters.com; Chen *et al.*, 2014). The resulting peptides were separated and detected using TripleTOF 5600+ (AB SCIEX, https://sciex.com) coupled with the splitless Ultra 1D Plus (Eksigent, http://www.eksigentllc.com) system (Andrews *et al.*, 2011). A total of 3 886 816 spectra of high quality were detected for subsequent proteogenomic analysis (Table S4).

Database construction and mass spectrometry data set searching

The AS junction database (141 289 entries) was constructed as described previously (Sheynkman *et al.*, 2013; Castellana *et al.*, 2014; Walley and Briggs, 2015), with minor modifications. Briefly, six-frame translations (i.e. three on the forward strand and three on the reverse strand) were applied to corresponding junction sequences. The same method was used to construct the DAS junction library (30 120 entries) from 5020 DAS events. A virtual peptide library (203 517 entries) was constructed using all transcripts annotated in TAIR 10 translated in all six reading frames. The resulting sequences were then subtracted from existing protein entries in TAIR 10 annotation. Entries over 80 amino acids were further selected to form the final virtual peptide library. Database searching (Table S4) was conducted according to a previous description (Chen *et al.*, 2014). In general, raw spectra data were searched with PROTEINPILOT 5.0 (AB SCIEX) against the Uniprot *A. thaliana* complete proteome database (release-2015_10; 31480 entries), customized AS/DAS junction and frame databases, using preset parameters. The FDR analysis was performed by using the tool integrated in PROTEINPILOT. All data were filtered at 1% FDR.

Structural analysis and homology modelling

Conserved domains were predicted using PFAM 29.0 (http://pfam.xfam.org). Homology modelling was carried out as described previously. Homologous protein structures carrying the same domain complex and same residues on the binding surface were obtained by searching the MODexplorer server (http://modorama.org/modexplorer/index; Barbato *et al.*, 2012; Kosinski *et al.*, 2013). Their protein partner sequences were obtained by using ACCELRYS DISCOVERY STUDIO 2.5 (http://accelrys.com/events/webinars/discovery-studio-25/abstracts.html) and subsequently searched for homologous sequences (At1g66510 and At1g60170) in *A. thaliana* by BlastP (National Center for Biotechnology Information NCBI, http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins). On the basis of the sequence alignment (Barbato *et al.*, 2012; Kosinski *et al.*, 2013),

accurate 3D structures of corresponding protein isoforms were built using the SWISSMODEL server (http://www.swissmodel.expasy.org). The X-ray crystallographic structures (PDB ID: 4YH8, At1g27650; 3JB9, At1g80070; 3SBT, AT1G66510; 1E7K, At4g12600; 3SIU, At1g60170; 1ZWZ, At5g20160.2; 1E7K, At5g20160.1; 3SIV, At5g20160.3) were selected as templates.

Molecular dynamics simulation and docking

Molecular dynamics simulation was carried out by superimposing the structures from homology modelling to the crystal structure. Energy minimizations and molecular dynamics (MD) simulations were performed with the sander and PMEMD modules of AMBER 14 (Tan *et al.*, 2015). In detail, the ff14SB force field for the protein, RNA and non-bonded interactions were truncated at 10 Å without positional restraints. A total of 1500 cycles of steepest descent and conjugate gradient energy minimizations were performed, respectively. Two 100-ps equilibration MD runs were subsequently performed, in which the system was first heated gradually to 298.15 K at constant volume and then equilibrated at a constant pressure of 1.01325×10^5 Pa. The following unrestrained equilibration (2 ns) and production runs were conducted at constant temperature (298.15 K). Free binding energies of the corresponding interaction complexes were calculated using the molecular mechanics/Poisson–Boltzmann surface area (MM/PBSA) method (Wang *et al.*, 2014). For docking experiments between ssRNA strands and At1g27650, the structure of ssRNA was constructed using ACCELRYS DISCOVERY STUDIO 2.5. The centre is assigned to the middle of two zinc atoms ($x = 80$, $y = 33$, $z = 60$), and the size of the box is pre-determined ($x = 22.5$, $y = 22.5$, $z = 21.0$). AUTODOCK VINA (Trott and Olson, 2010) was then used to calculate the docking scores as free binding energy. All figures with structural representation were drafted in the PYMOL molecular graphics system (http://www.pymol.org).

Data submission

The data from next-generation sequencing and single-molecule long-read sequencing has been uploaded to the Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra) under Bioproject PRJNA371677. We have submitted our proteomic raw data into the PRIDE (PRoteomics IDEntifications) database. Please refer to accession number PXD005888.

ACKNOWLEDGEMENTS

This work was supported by the Natural Science Foundation of Shandong Province (BS2015NY002), the National Natural Science Foundation of China (NSFC31301251), the Science and Technology Programme of Nantong (MS12016044), the National Natural Science Foundation of China (NSFC31101099), and the Training Programme for University Prominent Young and Middle-aged Teachers and Presidents, National Basic Research Programme of China (2012CB114300), JSPS 'Strategic Young Researcher Overseas Visits Programme for Accelerating Brain Circulation' (S2503), Shenzhen Overseas Talents Innovation and Entrepreneurship Funding Scheme (The Peacock Scheme) (KQTD201101) and Hong Kong Research Grant Council (AoE/M-05/12, CUHK 14122415, CUHK 14160516).

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Schematic view of analytical pipeline of AS identification and validation.
Figure S2. Quality control of the short-read RNA sequencing data set.
Figure S3. Gene expression quantification and cluster analysis.
Figure S4. Analysis of differentially expressed genes (DEGs).
Figure S5. Characterization of AS types and splicing sites (ss) in *Arabidopsis thaliana*.
Figure S6. Heat-map representation of the 5020 differentially expressed AS events.
Figure S7. Summary of the DAS events under ABA treatment.
Figure S8. Heat-map representation of the DAS events in the ABA pathway.
Figure S9. Characterization of AS events involved in ABA responses.
Figure S10. Quality control of the large-scale proteomic identification using the Uniprot database as an input file.
Figure S11. Quality control of the large-scale proteomic identification using AS and DAS databases as input files.
Figure S12. Comparison between the proteomic and transcriptomic data sets.
Figure S13. Heat-map representation of the DAS events belongs to the category of splicing components.
Figure S14. Validation of the DAS events belong to the splicing components.
Figure S15. Molecular simulation of putative splicing factor isoforms.
Figure S16. Molecular mechanism simulation of the potential splicing factor isoforms.
Figure S17. Docking simulation amongst protein and selected ssRNAs.
Figure S18. Mutant phenotype validation of selected splicing factors.
Table S1. Summary of the basic parameters in the short-read RNA sequencing data set.
Table S2. Classification of reads of insertions in single-molecule long-read sequencing.
Table S3. List of the differentially expressed AS events in the ABA signalling pathway.
Table S4. Summary of proteins identified at 1% FDR search against different databases.
Table S5. Over- and under-represented biological processes of the DAS proteins commonly detected in four samples.
Table S6. Features of the identified proteins encoded by different frames of the same transcripts.
Table S7. List of the distinct peptides (confidence > 90%) detected in the proteomic analysis in Figure 4.
Table S8. List of differentially expressed AS events for potential splicing components in Arabidopsis.
Table S9. Primers used in this study.
Appendix S1. Splicing junction identified by short-read RNA sequencing.
Appendix S2. Splicing junction identified by long-read RNA sequencing.

REFERENCES

Abdel-Ghany, S.E., Hamilton, M., Jacobi, J.L., Ngam, P., Devitt, N., Schilkey, F., Ben-Hur, A. and Reddy, A.S. (2016) A survey of the sorghum transcriptome using single-molecule long reads. *Nat. Commun.* **7**, https://doi.org/10.1038/ncomms11706.

Aghamirzaie, D., Nabiyouni, M., Fang, Y., Klumas, C., Heath, L.S., Grene, R. and Collakova, E. (2013) Changes in RNA Splicing in Developing Soybean (Glycine max) Embryos. *Biology*, **2**, 1311–1337.

Alfaro, J.A., Sinha, A., Kislinger, T. and Boutros, P.C. (2014) Onco-proteogenomics: cancer proteomics joins forces with genomics. *Nat. Methods*, **11**, 1107–1113.

Andreassi, C. and Riccio, A. (2009) To localize or not to localize: mRNA fate is in 3′ UTR ends. *Trends Cell Biol.* **19**, 465–474.

Andrews, G.L., Simons, B.L., Young, J.B., Hawkridge, A.M. and Muddiman, D.C. (2011) Performance characteristics of a new hybrid quadrupole time-of-flight tandem mass spectrometer (TripleTOF 5600). *Anal. Chem.* **83**, 5442–5446.

Balbin, O.A., Malik, R., Dhanasekaran, S.M. *et al.* (2015) The landscape of antisense gene expression in human cancers. *Genome Res.* **25**, 1068–1079.

Barbato, A., Benkert, P., Schwede, T., Tramontano, A. and Kosinski, J. (2012) Improving your target-template alignment with MODalign. *Bioinformatics*, **28**, 1038–1039.

Bortesi, L. and Fischer, R. (2015) The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnol. Adv.* **33**, 41–52.

Brar, G.A. and Weissman, J.S. (2015) Ribosome profiling reveals the what, when, where and how of protein synthesis. *Nat. Rev. Mol. Cell Biol.* **16**, 651–664.

Brown, J.W., Simpson, C.G., Marquez, Y., Gadd, G.M., Barta, A. and Kalyna, M. (2015) Lost in translation: pitfalls in deciphering plant alternative splicing transcripts. *Plant Cell*, **27**, 2083–2087.

Buljan, M., Chalancon, G., Eustermann, S., Wagner, G.P., Fuxreiter, M., Bateman, A. and Babu, M.M. (2012) Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. *Mol. Cell*, **46**, 871–883.

Castellana, N.E., Payne, S.H., Shen, Z., Stanke, M., Bafna, V. and Briggs, S.P. (2008) Discovery and revision of Arabidopsis genes by proteogenomics. *Proc. Natl Acad. Sci. USA*, **105**, 21034–21038.

Castellana, N.E., Shen, Z., He, Y., Walley, J.W., Cassidy, C.J., Briggs, S.P. and Bafna, V. (2014) An automated proteogenomic method uses mass spectrometry to reveal novel genes in Zea mays. *Mol. Cell Proteomics*, **13**, 157–167.

Chang, C.Y., Lin, W.D. and Tu, S.L. (2014) Genome-wide analysis of heat-sensitive alternative splicing in *Physcomitrella patens*. *Plant Physiol.* **165**, 826–840.

Chen, X., Chan, W.L., Zhu, F.Y. and Lo, C. (2014) Phosphoproteomic analysis of the non-seed vascular plant model *Selaginella moellendorffii*. *Proteome Sci.* **12**, 16.

Cheng, C.Y., Krishnakumar, V., Chan, A.P., Thibaud-Nissen, F., Schobel, S. and Town, C.D. (2017) Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. *Plant J.*, **89**, 789–804.

Crooks, G.E., Hon, G., Chandonia, J.M. and Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190.

Drechsel, G., Kahles, A., Kesarwani, A.K., Stauffer, E., Behr, J., Drewe, P., Rätsch, G. and Wachter, A. (2013) Nonsense-mediated decay of alternative precursor mRNA splicing variants is a major determinant of the Arabidopsis steady state transcriptome. *Plant Cell*, **25**, 3726–3742.

Eckardt, N.A. (2013) The plant cell reviews alternative splicing. *Plant Cell*, **25**, 3639.

Ellis, J.D., Barrios-Rodiles, M., Colak, R. *et al.* (2012) Tissue-specific alternative splicing remodels protein-protein interaction networks. *Mol. Cell*, **46**, 884–892.

FANTOM Consortium and the RIKEN PMI and CLST (DGT), Forrest, A.R., Kawaji, H., Rehli, M., Baillie, J.K., de Hoon, M.J., Haberle, V., Lassmann, T. *et al.* (2014) A promoter-level mammalian expression atlas. *Nature*, **507**, 462–470.

Fedoroff, N.V. (2002) RNA-binding proteins in plants: the tip of an iceberg? *Curr. Opin. Plant Biol.* **5**, 452–459.

Feng, J., Li, J., Gao, Z. *et al.* (2015) SKIP confers osmotic tolerance during salt stress by controlling alternative gene splicing in Arabidopsis. *Mol. Plant*, **8**, 1038–1052.

Filichkin, S.A., Priest, H.D., Givan, S.A., Shen, R., Bryant, D.W., Fox, S.E., Wong, W.K. and Mockler, T.C. (2010) Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. *Genome Res.* **20**, 45–58.

Florea, L., Song, L. and Salzberg, S.L. (2013) Thousands of exon skipping events differentiate among splicing patterns in sixteen human tissues. *F1000Research*, **2**, https://doi.org/10.12688/f1000research.2-188.v2.

Fujita, Y., Yoshida, T. and Yamaguchi-Shinozaki, K. (2013) Pivotal role of the AREB/ABF-SnRK2 pathway in ABRE-mediated transcription in response to osmotic stress in plants. *Physiol. Plant.* **147**, 15–27.

Goda, H., Sasaki, E., Akiyama, K. *et al.* (2008) The AtGenExpress hormone and chemical treatment data set: experimental design, data evaluation, model data analysis and data access. *Plant J.* **55**, 526–542.

Golovkin, M. and Reddy, A.S. (1996) Structure and expression of a plant U1 snRNP 70K gene: alternative splicing of U1 snRNP 70K pre-mRNAs produces two different transcripts. *Plant Cell*, **8**, 1421–1435.

Hirayama, T. and Shinozaki, K. (2007) Perception and transduction of abscisic acid signals: keys to the function of the versatile plant hormone ABA. *Trends Plant Sci.* **12**, 343–351.

Ingolia, N.T., Lareau, L.F. and Weissman, J.S. (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*, **147**, 789–802.

James, A.B., Syed, N.H., Bordage, S., Marshall, J., Nimmo, G.A., Jenkins, G.I., Herzyk, P., Brown, J.W. and Nimmo, H.G. (2012) Alternative splicing mediates responses of the Arabidopsis circadian clock to temperature changes. *Plant Cell*, **24**, 961–981.

Jenal, M., Elkon, R., Loayza-Puch, F. *et al.* (2012) The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. *Cell*, **149**, 538–553.

Jensen, T.H., Jacquier, A. and Libri, D. (2013) Dealing with pervasive transcription. *Mol. Cell*, **52**, 473–484.

Jiao, Y., Lau, O.S. and Deng, X.W. (2007) Light-regulated transcriptional networks in higher plants. *Nat. Rev. Genet.* **8**, 217–230.

Kalyna, M., Lopato, S., Voronin, V. and Barta, A. (2006) Evolutionary conservation and regulation of particular alternative splicing events in plant SR proteins. *Nucleic Acids Res.* **34**, 4395–4405.

Kilian, J., Whitehead, D., Horak, J., Wanke, D., Weinl, S., Batistic, O., D'Angelo, C., Bornberg-Bauer, E., Kudla, J. and Harter, K. (2007) The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. *Plant J.* **50**, 347–363.

Kim, M.S., Pinto, S.M., Getnet, D. *et al.* (2014) A draft map of the human proteome. *Nature*, **509**, 575–581.

de Klerk, E. and t Hoen, P.A. (2015) Alternative mRNA transcription, processing, and translation: insights from RNA sequencing. *Trends Genet.* **31**, 128–139.

Kondo, Y., Oubridge, C., van Roon, A.M. and Nagai, K. (2015) Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5′ splice site recognition. *Elife*, **4**, e04986.

Kosinski, J., Barbato, A. and Tramontano, A. (2013) MODexplorer: an integrated tool for exploring protein sequence, structure and function relationships. *Bioinformatics*, **29**, 953–954.

Krummel, D.A.P., Oubridge, C., Leung, A.K., Li, J. and Nagai, K. (2009) Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. *Nature*, **458**, 475–480.

Kuhn, J.M. and Schroeder, J.I. (2003) Impacts of altered RNA metabolism on abscisic acid signaling. *Curr. Opin. Plant Biol.* **6**, 463–469.

Kumar, D., Yadav, A.K., Jia, X., Mulvenna, J. and Dash, D. (2016) Integrated transcriptomic-proteomic analysis using a proteogenomic workflow refines rat genome annotation. *Mol. Cell Proteomics*, **15**, 329–339.

Lee, S., Liu, B., Lee, S., Huang, S.X., Shen, B. and Qian, S.B. (2012) Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc. Natl Acad. Sci. USA*, **109**, 2424–2432.

Lorkovic, Z.J., Lehner, R., Forstner, C. and Barta, A. (2005) Evolutionary conservation of minor U12-type spliceosome between plants and humans. *RNA*, **11**, 1095–1107.

Mackowiak, S.D., Zauber, H., Bielow, C. *et al.* (2015) Extensive identification and analysis of conserved small ORFs in animals. *Genome Biol.* **16**, 179.

Marquardt, S., Raitskin, O., Wu, Z., Liu, F., Sun, Q. and Dean, C. (2014) Functional consequences of splicing of the antisense transcript COOLAIR on FLC transcription. *Mol. Cell*, **54**, 156–165.

Marquez, Y., Brown, J.W., Simpson, C., Barta, A. and Kalyna, M. (2012) Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis. *Genome Res.* **22**, 1184–1195.

Menschaert, G., Van Criekinge, W., Notelaers, T., Koch, A., Crappe, J., Gevaert, K. and Van Damme, P. (2013) Deep proteome coverage based on ribosome profiling aids mass spectrometry-based protein and peptide

discovery and provides evidence of alternative translation products and near-cognate translation initiation events. *Mol. Cell Proteomics*, **12**, 1780–1790.

Murashige, T. and Skoog, F. (1962) A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiol. Plant.* **15**, 473–497.

Nemhauser, J.L., Hong, F. and Chory, J. (2006) Different plant hormones regulate similar processes through largely nonoverlapping transcriptional responses. *Cell*, **126**, 467–475.

Nicholson, P., Yepiskoposyan, H., Metze, S., Zamudio Orozco, R., Kleinschmidt, N. and Muhlemann, O. (2010) Nonsense-mediated mRNA decay in human cells: mechanistic insights, functions beyond quality control and the double-life of NMD factors. *Cell. Mol. Life Sci.* **67**, 677–700.

Pelechano, V. and Steinmetz, L.M. (2013) Gene regulation by antisense transcription. *Nat. Rev. Genet.* **14**, 880–893.

Pose, D., Verhage, L., Ott, F., Yant, L., Mathieu, J., Angenent, G.C., Immink, R.G. and Schmid, M. (2013) Temperature-dependent regulation of flowering by antagonistic FLM variants. *Nature*, **503**, 414–417.

Rosloski, S.M., Singh, A., Jali, S.S., Balasubramanian, S., Weigel, D. and Grbic, V. (2013) Functional analysis of splice variant expression of MADS AFFECTING FLOWERING 2 of *Arabidopsis thaliana*. *Plant Mol. Biol.* **81**, 57–69.

Ruhl, C., Stauffer, E., Kahles, A., Wagner, G., Drechsel, G., Ratsch, G. and Wachter, A. (2012) Polypyrimidine tract binding protein homologs from Arabidopsis are key regulators of alternative splicing with implications in fundamental developmental processes. *Plant Cell*, **24**, 4360–4375.

Seki, M., Ishida, J., Narusaka, M., Fujita, M., Nanjo, T., Umezawa, T., Kamiya, A., Nakajima, M., Enju, A. and Sakurai, T. (2002) Monitoring the expression pattern of around 7,000 Arabidopsis genes under ABA treatments using a full-length cDNA microarray. *Funct. Integr. Genomics*, **2**, 282–291.

Sheynkman, G.M., Shortreed, M.R., Frey, B.L. and Smith, L.M. (2013) Discovery and mass spectrometric analysis of novel splice-junction peptides using RNA-Seq. *Mol. Cell Proteomics*, **12**, 2341–2353.

Shimazaki, K., Doi, M., Assmann, S.M. and Kinoshita, T. (2007) Light regulation of stomatal movement. *Annu. Rev. Plant Biol.* **58**, 219–247.

Slavoff, S.A., Mitchell, A.J., Schwaid, A.G., Cabili, M.N., Ma, J., Levin, J.Z., Karger, A.D., Budnik, B.A., Rinn, J.L. and Saghatelian, A. (2013) Peptidomic discovery of short open reading frame-encoded peptides in human cells. *Nat. Chem. Biol.* **9**, 59–64.

Sonenberg, N. and Hinnebusch, A.G. (2009) Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell*, **136**, 731–745.

Song, L., Liberman, L.M., Mukherjee, N., Benfey, P.N. and Ohler, U. (2013) Integrated detection of natural antisense transcripts using strand-specific RNA sequencing data. *Genome Res.* **23**, 1730–1739.

Sonmez, C., Baurle, I., Magusin, A., Dreos, R., Laubinger, S., Weigel, D. and Dean, C. (2011) RNA 3′ processing functions of Arabidopsis FCA and FPA limit intergenic transcription. *Proc. Natl Acad. Sci. USA*, **108**, 8508–8513.

Suenaga, Y., Islam, S.R., Alagu, J., Kaneko, Y., Kato, M., Tanaka, Y., Kawana, H., Hossain, S., Matsumoto, D. and Yamamoto, M. (2014) NCYM, a Cis-antisense gene of MYCN, encodes a de novo evolved protein that inhibits GSK3β resulting in the stabilization of MYCN in human neuroblastomas. *PLoS Genet.* **10**, e1003996.

Sureshkumar, S., Dent, C., Seleznev, A., Tasset, C. and Balasubramanian, S. (2016) Nonsense-mediated mRNA decay modulates FLM-dependent thermosensory flowering response in Arabidopsis. *Nat. Plants*, **2**, 16055.

Taj, M., Jalean, P., Corcoran, D.L., Song, L., Winter, C.M., Alexa, C., Benfey, P.N., Uwe, O. and Molly, M. (2014) Paired-end analysis of transcription start sites in Arabidopsis reveals plant-specific promoter signatures. *Plant Cell*, **26**, 2746–2760.

Tan, Y.S., Spring, D.R., Abell, C. and Verma, C.S. (2015) The Application of Ligand-Mapping Molecular Dynamics Simulations to the Rational Design of Peptidic Modulators of Protein–Protein Interactions. *J. Chem. Theory Comput.* **11**, 3199–3210.

Tavares, R., Scherer, N.M., Ferreira, C.G., Costa, F.F. and Passetti, F. (2015) Splice variants in the proteome: a promising and challenging field to targeted drug discovery. *Drug Discov. Today*, **20**, 353–360.

Thatcher, S.R., Zhou, W., Leonard, A., Wang, B.B., Beatty, M., Zastrow-Hayes, G., Zhao, X., Baumgarten, A. and Li, B. (2014) Genome-wide analysis of alternative splicing in Zea mays: landscape and genetic regulation. *Plant Cell*, **26**, 3472–3487.

Thatcher, S.R., Danilevskaya, O.N., Meng, X., Beatty, M., Zastrow-Hayes, G., Harris, C., Van Allen, B., Habben, J. and Li, B. (2016) Genome-Wide Analysis of Alternative Splicing during development and Drought Stress in Maize. *Plant Physiol.* **170**, 586–599.

Trott, O. and Olson, A.J. (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461.

Wade, J.T. and Grainger, D.C. (2014) Pervasive transcription: illuminating the dark matter of bacterial transcriptomes. *Nat. Rev. Microbiol.* **12**, 647–653.

Walley, J.W. and Briggs, S.P. (2015) Dual use of peptide mass spectra: protein atlas and genome annotation. *Curr. Plant Biol.* **2**, 21–24.

Wang, X., Li, C., Wang, Y. and Chen, G. (2014) Interaction of classical platinum agents with the monomeric and dimeric Atox1 proteins: a molecular dynamics simulation study. *Int. J. Mol. Sci.* **15**, 75–99.

Wang, Z., Ji, H., Yuan, B., Wang, S., Su, C., Yao, B., Zhao, H. and Li, X. (2015) ABA signalling is fine-tuned by antagonistic HAB1 variants. *Nat. Commun.* **6**, https://doi.org/10.1038/ncomms11708.

Wang, B., Tseng, E., Regulski, M., Clark, T.A., Hon, T., Jiao, Y., Lu, Z., Olson, A., Stein, J.C. and Ware, D. (2016) Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. *Nat. Commun.* **7**, 11708.

Will, C.L. and Luhrmann, R. (2011) Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* **3**, a003707.

Xu, L., Liu, F., Lechner, E., Genschik, P., Crosby, W.L., Ma, H., Peng, W., Huang, D. and Xie, D. (2002) The SCFCOI1 ubiquitin-ligase complexes are required for jasmonate response in Arabidopsis. *Plant Cell*, **14**, 1919–1935.

Yan, J. and Marr, T.G. (2005) Computational analysis of 3′-ends of ESTs shows four classes of alternative polyadenylation in human, mouse, and rat. *Genome Res.* **15**, 369–375.

Yang, M., Xu, L., Liu, Y. and Yang, P. (2015) RNA-Seq Uncovers SNPs and Alternative Splicing Events in Asian Lotus (*Nelumbo nucifera*). *PLoS ONE*, **10**, e0125702.

Yoshida, H., Park, S.Y., Oda, T. *et al.* (2015) A novel 3′ splice site recognition by the two zinc fingers in the U2AF small subunit. *Genes Dev.* **29**, 1649–1660.

Zhan, X., Qian, B., Cao, F., Wu, W., Yang, L., Guan, Q., Gu, X., Wang, P., Okusolubo, T.A. and Dunn, S.L. (2015) An Arabidopsis PWI and RRM motif-containing protein is critical for pre-mRNA splicing and ABA responses. *Nat. Commun.* **6**, https://doi.org/10.1038/ncomms9139.

Zhu, F.Y., Li, L., Lam, P.Y., Chen, M.X., Chye, M.L. and Lo, C. (2013) Sorghum extracellular leucine-rich repeat protein SbLRR2 mediates lead tolerance in transgenic Arabidopsis. *Plant Cell Physiol.* **54**, 1549–1559.

