SUPPLEMENTARY FIGURE LEGENDS

Supplementary Figure 1. Scatter plots of gene coverage comparing microreads from libraries generated from full-length, oligo d(T) priming (FL) and from priming using random hexamers (RP). Libraries were sequenced using individual Illumina 1G flowcells. Total 32-mer perfect matching tags were counted for 31,529 genes and both Pearson and Spearman (nonparametric) correlation coefficients were calculated. Correlation coefficients between independent technical replicates were 0.999 (Pearson) and 0.979 (Spearman) for the FL libraries and 0.998 (Pearson) and 0.988 (Spearman) for the RP libraries.

Supplementary Figure 2. Profiles of *A. thaliana* chromosome coverage by cDNA microreads derived from the full-length (FL) and randomly primed (RP) cDNA libraries. Plots were generated as described in Supplementary Methods. Each vertical blue bar represents the log2 of the frequency of perfectly matching unique single-copy 32-mer microreads. Chromosome size in million base pairs (Mbp) is shown to the right of each plot.

Supplementary Figure 3. Depth of chromosome coverage by unique single-copy (blue) and multicopy (black) perfectly matching 32-mer microreads versus DNA methylation profile (red) for each chromosome. For the microreads, each vertical bar represents the log2 of the frequency of perfectly matching 32-mers. Each red vertical red bar in the methylation profile represents the log2 signal ratio of the immunoprecipitated methylated DNA signal divided by the input control signal (total DNA) as described in Zilberman *et al*., 2007. The methylation plot was generated using *Arabidopsis* genome tiling array data (NCBI Gene Expression Omnibus GSE5974 record, GSM138296 data set, Zilberman *et al*., 2007). Illumina RNA-seq reads (~6.5 million) matching multiple locations in the Arabidopsis genome were identified using the ELAND program. All plots were generated in R using the same chromosome coordinates as described in Supplementary Methods. As expected, there is a positive relationship between the microread density in repetitive regions and DNA methylation in the centromeric, pericentromeric regions and heterochromatic knobs. Vertical arrows indicate instances of triple alignments (a gap in the unique single-copy microread coverage, peak in the repeat microreads profile and a high density of methylation). A schematic drawing of each chromosome and its features is shown below the methylation profiles. Approximate genetic boundaries of the centromers are indicated by grey. Centromeric gaps are indicated by red circles. Heterochromatic knobs are shown by violet ellipses. Chromosome portions corresponding to the telomeres and nucleolar organizing regions are not shown.

Supplementary Figure 4. Schematics of alternative splicing events in plants. *(A)* General classes of alternative splicing include intron retention (IntronR), exon skipping (ExonS) alternate donor (AltD) or acceptor (AltA), and alternate donor/acceptor position (AltP) splice sites. *(B)* Specific types of alternative splicing events that can alter particular features of the mature transcript. Depicted events include: nested intron retention (Nested IntronR), cassette and poison cassette exons (CassetteEx and Poison

CassetteEx), alternate 3' terminal exon (Alt TermEx), alternate 5' initiation exon (Alt InitEx), and alternate 5' untranslated region (UTR) (Alt 5'UTR) exon. Alternate 3' terminal exons introduce an alternate termination codon, depicted by a black upward pointing star. Poison cassette exons introduce an in-frame premature termination codon (PTC, black downward star) in the mature transcript (the red upward pointing star indicates the reference termination codon). Arrows indicate the 5' to 3' transcription direction, with the green arrow denoting the translation start site. Solid green lines join constitutive donor/acceptor splice sites in constitutively spliced isoforms. Black dashed lines join alternative splice sites. Constitutively spliced exons are depicted by dark rectangles. The portions of mRNA introduced in the mature transcript as a consequence of the alternative splicing event are shown by the hatched rectangles. 5' and 3' UTRs are shown by the light shaded rectangles and pointed rectangles, respectively.

Supplementary Figure 5. Breakdowns of the classifications of consensus and non-consensus introns. The combined TAIR8 and supersplat/TAU predicted introns were analyzed as described in Methods and classified as U2 or U12 using a Perl script. Introns that could not be assigned to a class using our search criteria are labeled 'unclassified'.

Supplementary Figure 6. Biological validation of predicted intron retention and novel SJ events in the *CCA1* locus. **a,** Agarose gel electrophoresis of RT-PCR products amplified using exon-flanking primers (left panel) and an IntronR4-specific primers (right panel). Lanes labeled HiLi, Heat, Cold, Salt and Dr correspond to high light, heat (42^{o}C), cold (4^{o}C), osmotic shock (0.25M sodium chloride) and dehydration (25% PEG) treatments, respectively. Ctrl indicates untreated control and lane M corresponds to a 100-bp ladder (sizes in base pairs are shown at the side). The RT-PCR fragment corresponding to a fully processed transcript is indicated by a "+". DNA fragments corresponding to the intron R4 and intron R4/4a (when nested intron 4a is spliced) events are marked by "*" and "**", respectively. RT-PCR products corresponding to pre-mRNA are indicated by the "-". All RT-PCR products were derived from the AT2G46830.1 transcript because both primer pairs A and B overlap the predicted 5' end of AT2G46830.2 cDNA. The expected sizes of all amplified products are given in Table S1. **b**, Schematic diagram of the *CCA1* gene model (AT2G46830.1) and primer design strategy for interrogation of IntronR4/4a event. Retained intron 4 and nested intron 4a are shown by turquoise and red filled boxes, respectively. The positions of forward (F) and reverse (R) primer pairs A and B are shown by black arrows. The positions of the reference and IntronR4/4a isoform-specific primers used for isoform quantification (Figure 6d) are shown by red and blue arrows, respectively. Coding regions of exons are indicated by filled rectangles and the 5' and 3' UTRs are indicated by open rectangles. Reference and premature termination codons are indicated by red (top) and black (bottom) symbols, respectively. Map is not drawn to scale. **c**, Alignment of nucleotide sequences obtained by Sanger sequencing of the individual RT-PCR fragments isolated from the gel shown in panel **a**. Splice donor and acceptor sites of intron 4a are shown in bold and underlined. SJ of the nested intron 4a is indicated by vertical arrow. Intronic and exonic sequences are in lower and upper case letters, respectively. Sequence of intron 4a is italicized. **d**, Microread coverage of a novel SJ

nested within retained intron 4. Microreads spanning SJ of intron 4a are aligned to the predicted SJ. The SJ is indicated by arrow and vertical lines in the sequence. Note: all perfectly matching microreads are shown; for supersplat predictions only microreads with at least six nucleotides overlapping the flanking exons were selected.

Supplementary Figure 7. Validation of AS events detected by RNA-seq and predicted by supersplat. **a**, Top panel: Primer design strategy for amplification of a predicted SJ. Forward oligonucleotide primers a and b were designed to partially overlap 5' and 3' portions of the predicted SJ. Melting temperatures of each primer segment alone were calculated to be insufficient for the PCR amplification under conditions described in Supplemental Methods. Reverse primers were designed when possible within the sequences of the closest adjacent exons. Flanking primer pairs were selected from the closest neighboring exonic sequences. Bottom panel: A representative example of microread coverage supporting the existence of a novel SJ in a transcript of the splicing factor ATRSP41. The SJ is indicated by an arrow and by vertical lines in individual microread sequences. **b**, Agarose gel electrophoresis of cDNA fragments amplified using SJ-specific (top and middle panels) or exon-flanking primer pairs (bottom). Lanes a and b for the At1g17870 transcript demonstrate SJ-specific PCR amplification using forward (a) and reverse (b) primer pairs. Multiple RT-PCR amplification products in exon flanking primers panel (bottom) are likely to result from several splice forms. A "*" denotes cDNA fragments corresponding to the predicted AS event product size. SJs At4g38770.1 and At3g05630.1 are annotated by TAIR (http://www.arabidopsis.org/) and were used as positive controls. The locus ID is shown above each lane. Lane M designates marker lanes (molecular sizes in nucleotide base pairs are indicated on left). A full list and detailed descriptions of selected novel AS events confirmed by RT-PCR are summarized in Supplementary Table 2. **c,** An example of microread coverage of the At1g10760.1 locus prepared in Gbrowse (http://athal.cgrb.oregonstate.edu). The novel AltD*/ExonS event confirmed by RT-PCR (panel B, lane At1g10760.1) is indicated by brown broken line above other SJs. Note significant variations of transcript coverage under different stress treatments suggesting that AS events are regulated. A maximum of 25 SJ matching microreads are shown even though more may have been obtained.

Supplementary Figure 8. Microread coverage and validation of stress-responsive intron retention events. **a**, Agarose gel electrophoresis of RT-PCR products of *ERD8* (At5g56030) and *HSP83* (AT5G52640) transcripts. Lanes labeled HiLi, Heat, Cold, Salt and Dr correspond to high light, heat shock (42^{o}C), cold (4^{o}C), osmotic shock and dehydration treatments, respectively. Ctrl indicates untreated control. RT-PCR products corresponding to the full-length unspliced transcripts are indicated by "-". DNA fragments encompassing retained intron 1 and spliced intron 2 are marked by "*" for both genes. Primers were designed as described in Supplementary Fig. 6. Forward and reverse primers were selected in the retained intron 1 and in downstream exon separated by intron 2, respectively. This design ensures that only cDNA corresponding to the retained intron 1/spliced intron 2 and pre-mRNA can be detected. The expected sizes of splicing products for At5g56030.1 transcripts are 464 bp for the isoform with retained intron 1 and spliced intron 2 and 570 bp for pre-mRNA. The expected sizes for At5g52640.1 transcripts are 414 bp with retained intron 1 and spliced intron 2 and 499

bp for pre-mRNA. Actin 2 (AT5G09810) was used as a positive control. M indicates a 100 bp ladder (sizes in base pairs are shown on the left). **b**, Retained intron 1 has dense microread coverage under the heat stress condition (shown by horizontal brackets).

Supplementary Figure 9. Validation of novel stress-responsive SJ. **a**, Agarose gel electrophoresis of SJ-specific RT-PCR products obtained from *Arabidopsis* seedlings treated under various stress conditions. A transcript of a zinc finger C2H2-type family protein gene (At2g36480) was amplified by RT-PCR using SJ-specific oligonucleotide primers. Ctrl indicates untreated sample; lanes labeled HiLi, Heat, Cold, Salt and Dr indicate high light, 42°C, 4°C, osmotic shock and dehydration, respectively. As a positive control actin 2 (At5g09810) was amplified. Lane M is a 100-bp ladder (sizes in base pairs are shown on left). **b**, Map of the stress-related supersplat-predicted and empirical splice junctions of At2g36480 in Gbrowse (http://athal.cgrb.oregonstate.edu). Insert shows details of the novel stress-responsive SJ sequence. Donor/acceptor sites are boxed. Previously annotated and novel donor sites are highlighted by green and yellow, respectively. Microread portions encompassing SJ are shaded. Dense microread coverage under the high light treatment (bracketed) suggests that this AS event is light-responsive.

Supplementary Figure 10. Stress-associated alternative splicing of *Arabidopsis* splicing regulators SR30 and SR34. **a**, Map of stress-related splice and supersplat SJs in the *SR30* gene. SJs corresponding to the untreated control, high light, heat, cold, salt and dehydration treatments are shown in gray, yellow, red, blue, green and brown, respectively. A maximum of 20 identical SJs is shown. **b**, Microread density coverage of *SR30* gene features under different stress conditions. A maximum of 20 overlapping microreads are shown. **c**, Structures of SR30 and SR34 splice isoforms. Locations of splice isoform-specific primer pairs and corresponding RT-PCR products are shown by arrows and bars, respectively. A novel poison exon cassette is depicted in red. **d**, RT-PCR of SR30 splice isoforms 1 (top) and 4 (middle). EF1α mRNA (bottom) was used as an equal sample amplification control. PCR reactions were performed as described in Supplementary Information online. M indicates molecular weight markers. **e**, Stress-responsive accumulation of productive (1) and PTC+ unproductive (2 and 3) SR34 isoforms. Relative expression levels were determined by qRT PCR and calculated using -$\Delta\Delta$Ct method (Livak, Schmittgen, 2001) and EF1αmRNA as an internal reference control. Note that both PTC+ isoforms accumulate at similar levels under identical stress conditions.

Supplementary Figure 11. HTS-based transcriptional unit modeling supports an experimentally validated functionally important regulatory AS event. The aberrant SOC1 transcripts with the retained sixth intron are direct targets for binding of the ELF9 regulatory protein that controls SOC1 expression levels possibly through an NMD pathway (Song *et al*., 2009). Note that the TAU assembly in this case is based only on Illumina microread coverage data and is not supported by Sanger cDNA evidence. Arrow shows position of the retained sixth intron in TAU model.

Supplementary Tables

Table S1. Classifications and frequencies of supersplat predicted non-consensus introns.

Table S2. Validation of AS events predicted by RNA-seq.

Table S3. Validation of novel introns predicted by supersplat using RNA-seq data.

Table S4. RNA-seq survey of constitutive and alternative splicing of SR splicing factor pre-mRNAs.