Inventory of Supplementary Materials:
Supplementary Figures:
Figure S1, related to Figure 1
Figure S2, related to Figure 2
Figure S3, related to Figure 2
Figure S4, related to Figure 4
Figure S5, related to Figure 4
Figure S6, related to Figure 5
Figure S7, related to Figure 6

Table S1, related to Figure 1. Tab1. Names of cell types. Tab2. Reads mapped in each library. Tab3. Correlation between array and RNAseq, and between cufflinks and stringtie.

Table S2, related to Figure 2. Tab1. List of local alternative acceptor events. Tab2. List of local alternative donor events. Tab3. List of local exon skipping events. Tab4. List of type I and type II intron retention events.

Table S3, related to Figure 7. Tab1. Expression levels of significantly differentially expressed lincRNAs.

Table S4, related to Figure 5. Tab1. Peptide quantification. Tab2. Quantification of protein abundance. Tab3. Peptides uniquely mapped to alternative spliced isoforms.

Table S5, related to Figure 4. Average expression level of all isoforms.

File S6, related to Figure 1. Newly assembled protein coding gene models.

File S7, related to Figure 7. Genomic locations for the newly identified and modified lincRNAs.

External Databases: Raw read data were deposited in NCBI SRA database under BioProject PRJNA323955.

Supplemental Experimental Procedures

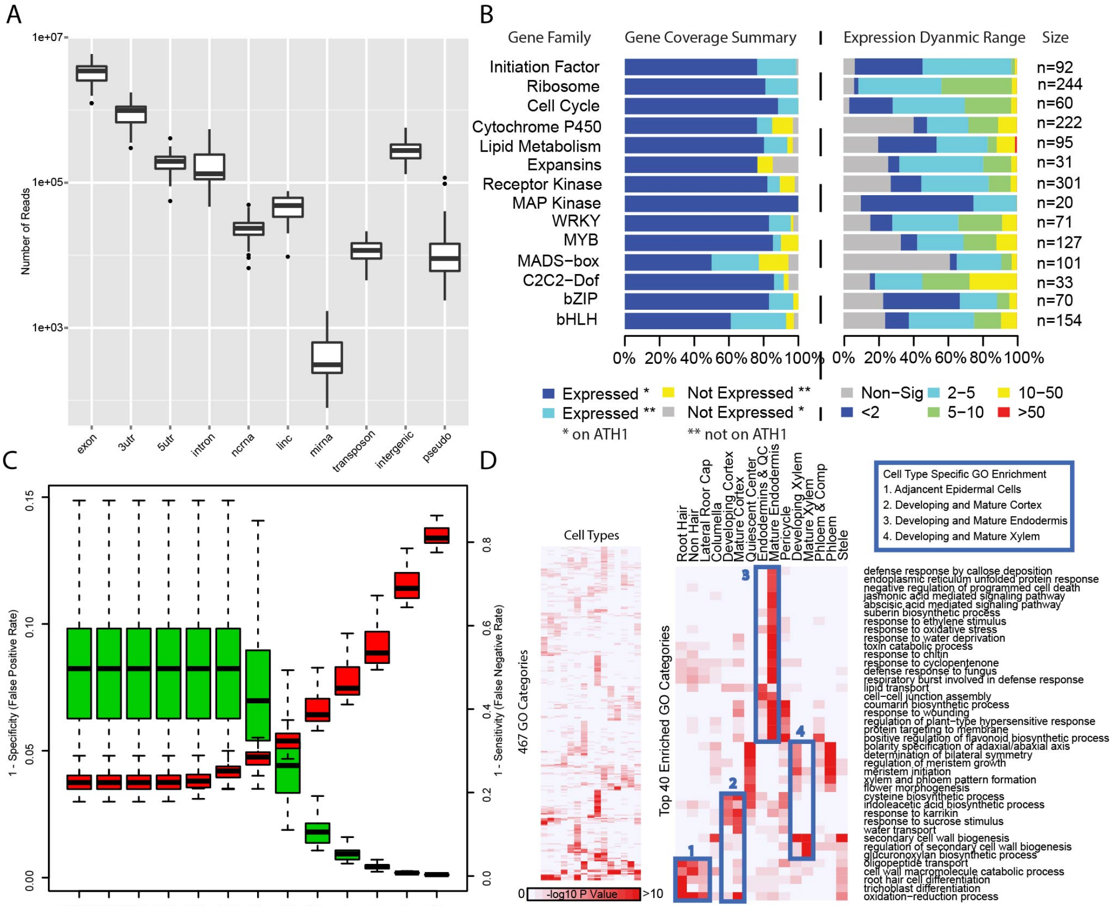

Figure S1. Gene expression analysis. **(A)** Number of reads mapped to each type of genomic feature. The distribution of read counts from 15 cell types and whole root are plotted as a boxplot. The genomic locations are annotated according to TAIR10 and lincRNA databases. The ncRNA category is according to TAIR10 annotation, and the lincRNA category excludes those lincRNAs that are annotated in TAIR10 as ncRNA. **(B)** Expression coverage and dynamic range of gene families. **(C)** Simulation study shows the threshold of detection for expressed protein coding genes is 0.05 FPKM. Known protein coding genes are used as true positives. Randomly sampled intergenic regions are used as true negatives. Boxplot shows the false positive rate (green) and false negative rate (red) of 15 cell types and whole root samples. **(D)** GO analysis for cell type specific genes. **Related to Figure 1.**

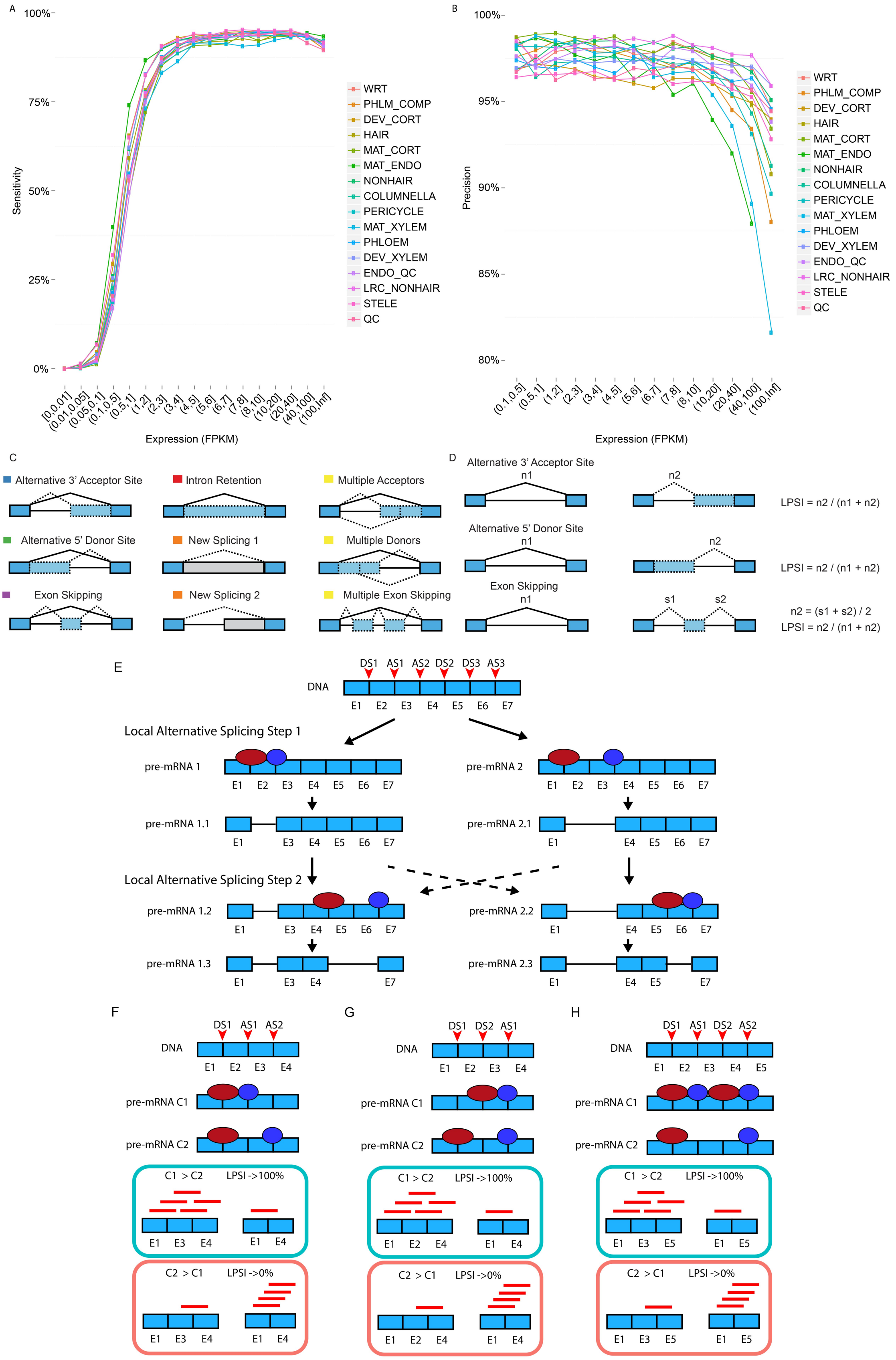
A
100%
75%
50%
25%
0%
Sensitivity
[0,0.01]
(0.01,0.05]
(0.05,0.1]
(0.1,0.5]
(0.5,1]
(1,2]
(2,3]
(3,4]
(4,5]
(5,6]
(6,7]
(7,8]
(8,10]
(10,20]
(20,40]
(40,100]
(100,inf]
Expression (FPKM)
WRT
PHLM_COMP
DEV_CORT
HAIR
MAT_CORT
MAT_ENDO
NONHAIR
COLUMNELLA
PERICYCLE
MAT_XYLEM
PHLOEM
DEV_XYLEM
ENDO_QC
LRC_NONHAIR
STELE
QC
B
100%
95%
90%
85%
80%
Precision
(0.1,0.5]
(0.5,1]
(1,2]
(2,3]
(3,4]
(4,5]
(5,6]
(6,7]
(7,8]
(8,10]
(10,20]
(20,40]
(40,100]
(100,inf]
Expression (FPKM)
WRT
PHLM_COMP
DEV_CORT
HAIR
MAT_CORT
MAT_ENDO
NONHAIR
COLUMNELLA
PERICYCLE
MAT_XYLEM
PHLOEM
DEV_XYLEM
ENDO_QC
LRC_NONHAIR
STELE
QC
C
Alternative 3' Acceptor Site
Intron Retention
Multiple Acceptors
Alternative 5' Donor Site
New Splicing 1
Multiple Donors
Exon Skipping
New Splicing 2
Multiple Exon Skipping
D
Alternative 3' Acceptor Site
n1
n2
LPSI = n2 / (n1 + n2)
Alternative 5' Donor Site
n1
n2
LPSI = n2 / (n1 + n2)
Exon Skipping
n1
s1
s2
n2 = (s1 + s2) / 2
LPSI = n2 / (n1 + n2)
E
DS1 AS1 AS2 DS2 DS3 AS3
DNA
E1 E2 E3 E4 E5 E6 E7
Local Alternative Splicing Step 1
pre-mRNA 1
E1 E2 E3 E4 E5 E6 E7
pre-mRNA 2
E1 E2 E3 E4 E5 E6 E7
pre-mRNA 1.1
E1 E3 E4 E5 E6 E7
pre-mRNA 2.1
E1 E4 E5 E6 E7
Local Alternative Splicing Step 2
pre-mRNA 1.2
E1 E3 E4 E5 E6 E7
pre-mRNA 2.2
E1 E4 E5 E6 E7
pre-mRNA 1.3
E1 E3 E4 E7
pre-mRNA 2.3
E1 E4 E5 E7
F
DS1 AS1 AS2
DNA
E1 E2 E3 E4
pre-mRNA C1
pre-mRNA C2
C1 > C2 LPSI ->100%
E1 E3 E4
E1 E4
C2 > C1 LPSI ->0%
E1 E3 E4
E1 E4
G
DS1 DS2 AS1
DNA
E1 E2 E3 E4
pre-mRNA C1
pre-mRNA C2
C1 > C2 LPSI ->100%
E1 E2 E4
E1 E4
C2 > C1 LPSI ->0%
E1 E2 E4
E1 E4
H
DS1 AS1 DS2 AS2
DNA
E1 E2 E3 E4 E5
pre-mRNA C1
pre-mRNA C2
C1 > C2 LPSI ->100%
E1 E3 E5
E1 E5
C2 > C1 LPSI ->0%
E1 E3 E5
E1 E5

Figure S2. (A) Sensitivity and **(B)** precision plot for junction identification for genes at different expression levels. Most known splice junctions were identified for genes expressed above 3 FPKM. **(C)** Types of alternative splicing. Color codes are same as Figure 2B. **(D)** How we calculated longer isoform percentage spliced in (LPSI) using spliced reads. In the figure, n1, n2, s1, s2 are number of spliced reads. **(E)** Mature mRNAs are generated by local splicing events. DS: donor site. AS: acceptor site. For pre-mRNA 1, DS1 and AS1 are used to splice out the intron in the pre-mRNA. In pre-mRNA 1.1, there is no available donor site for AS2. DS2 and AS3 are used to splice out the intron in pre-mRNA 1.2. Solid arrows indicate potential splicing path. Dashed arrows indicate other alternative splicing paths. **(F)** For an alternative acceptor event, there are two different configurations. In configuration C1, the upstream acceptor is used. In configuration C2, the downstream acceptor is used. **(G)** An alternative donor event. In configuration C1, downstream donor is used. In configuration C2, upstream donor is used. **(H)** An exon skipping-event. In configuration C1, all acceptors and donor sites are used. In configuration C2, only distal donor and acceptor sites are used. In all cases, if most isoforms were generated by configuration C1, the LPSI values are close to 100%. If most isoforms were generated by configuration C2, the LPSI values are close to 0%. **Related to Figure 2.**

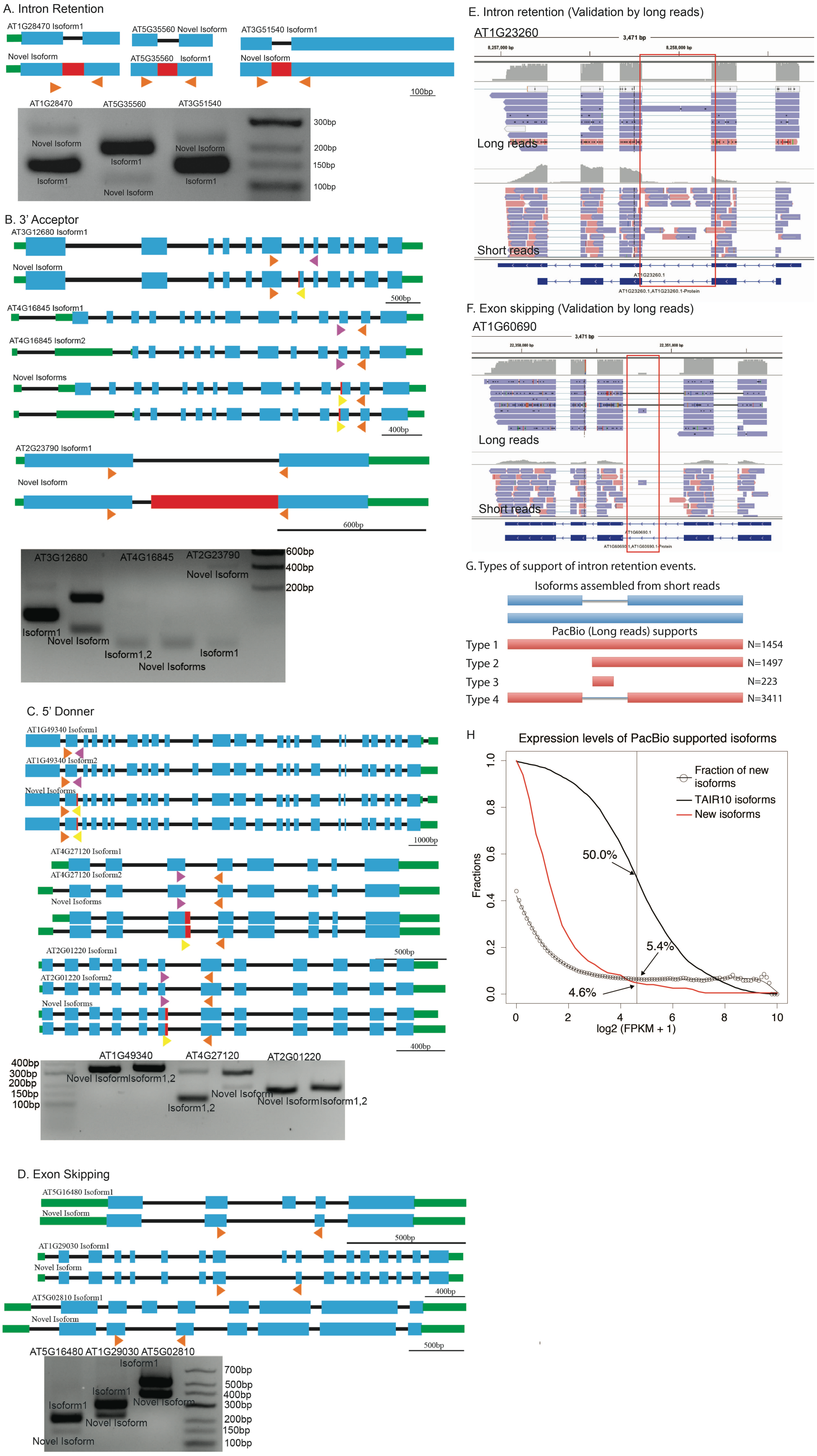
A. Intron Retention
AT1G28470 Isoform1
Novel Isoform
AT5G35560 Novel Isoform
AT5G35560 Isoform1
AT3G51540 Isoform1
Novel Isoform
100bp
AT1G28470
AT5G35560
AT3G51540
Novel Isoform
Isoform1
Isoform1
Novel Isoform
Novel Isoform
Isoform1
300bp
200bp
150bp
100bp
B. 3' Acceptor
AT3G12680 Isoform1
Novel Isoform
500bp
AT4G16845 Isoform1
AT4G16845 Isoform2
Novel Isoforms
400bp
AT2G23790 Isoform1
Novel Isoform
600bp
AT3G12680
AT4G16845
AT2G23790
Novel Isoform
600bp
400bp
200bp
Isoform1
Novel Isoform
Isoform1,2
Novel Isoforms
Isoform1
C. 5' Donner
AT1G49340 Isoform1
AT1G49340 Isoform2
Novel Isoforms
1000bp
AT4G27120 Isoform1
AT4G27120 Isoform2
Novel Isoforms
AT2G01220 Isoform1
500bp
AT2G01220 Isoform2
Novel Isoforms
400bp
400bp
300bp
200bp
150bp
100bp
AT1G49340
AT4G27120
AT2G01220
Novel Isoform
Isoform1,2
Isoform1,2
Novel Isoform
Novel Isoform
Isoform1,2
D. Exon Skipping
AT5G16480 Isoform1
Novel Isoform
500bp
AT1G29030 Isoform1
Novel Isoform
400bp
AT5G02810 Isoform1
Novel Isoform
500bp
AT5G16480 AT1G29030 AT5G02810
Isoform1
Isoform1
Novel Isoform
Isoform1
Novel Isoform
Novel Isoform
700bp
500bp
400bp
300bp
200bp
150bp
100bp
E. Intron retention (Validation by long reads)
AT1G23260
3,471 bp
8,257,000 bp
8,258,000 bp
Long reads
Short reads
AT1G23260.1
AT1G23260.1,AT1G23260.1-Protein
F. Exon skipping (Validation by long reads)
AT1G60690
3,471 bp
22,350,000 bp
22,351,000 bp
Long reads
Short reads
AT1G60690.1
AT1G60690.1,AT1G60690.1-Protein
G. Types of support of intron retention events.
Isoforms assembled from short reads
PacBio (Long reads) supports
Type 1 N=1454
Type 2 N=1497
Type 3 N=223
Type 4 N=3411
H
Expression levels of PacBio supported isoforms
Fraction of new isoforms
TAIR10 isoforms
New isoforms
50.0%
5.4%
4.6%
Fractions
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
log2 (FPKM + 1)

Figure S3. Validation of four different types of alternative splicing isoforms: **(A)** Intron Retention, **(B)** Alternative 3'Acceptor, **(C)** Alternative 5'Donor, and **(D)** Exon Skipping by RT-PCR. Green Boxes: 3' or 5' UTR. Blue boxes: Exons. Black lines: Introns. Triangles mark the location of the primers used to detect alternatively spliced isoforms. Orange triangles mark the primers used to detect both isoforms. Magenta and Yellow triangles mark the primers used to detect the specific isoform. Scale indicates the length of mRNAs. cDNAs were generated from polyA selected RNA of whole roots. **(E)** and **(F)** example of isoforms supported by both long reads and short reads. **(G)** For type 1, the PacBio read fully supports the intron retention isoform. For type 2 and type 3, the PacBio read partially supports the intron retention isoform. For type 4, the PacBio read supports the spliced isoform. We found 1454 type 1 support and 3411 type 4 support. We also found 109 cases where PacBio reads support both spliced isoforms and intron retention isoforms. **(H)** Cumulative distribution of expression levels for TAIR10 and new isoforms. **Related to Figure 2.**

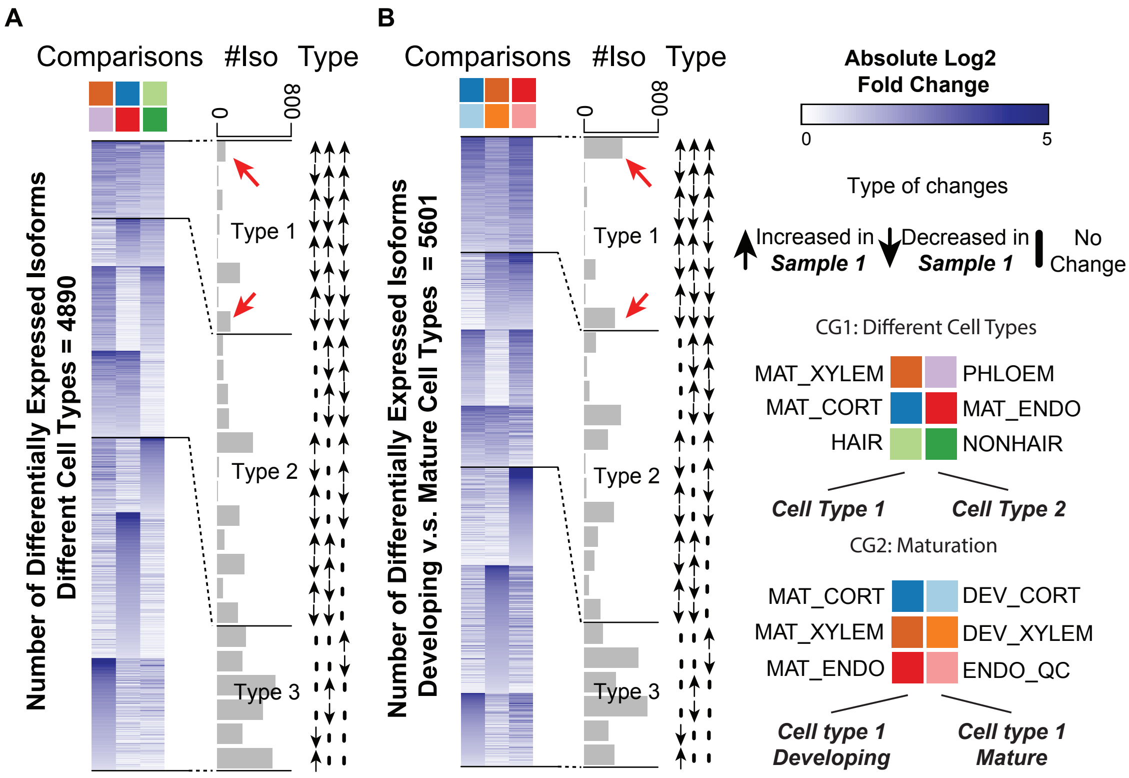

Figure S4. Heat map for differentially expressed isoforms in six pairwise comparisons. **(A)** Comparisons between different cell types. **(B)** Comparisons between the same cell type at different developmental stages. Bar plots show the number of isoforms found in each category according to the expression patterns. Black arrows indicate whether the isoform is increased in sample 1, increased in sample 2 or no change between the two samples. Red arrows highlight the number of genes found in two specific patterns: increased or decreased expression in all three comparisons. **Related to Figure 4.**

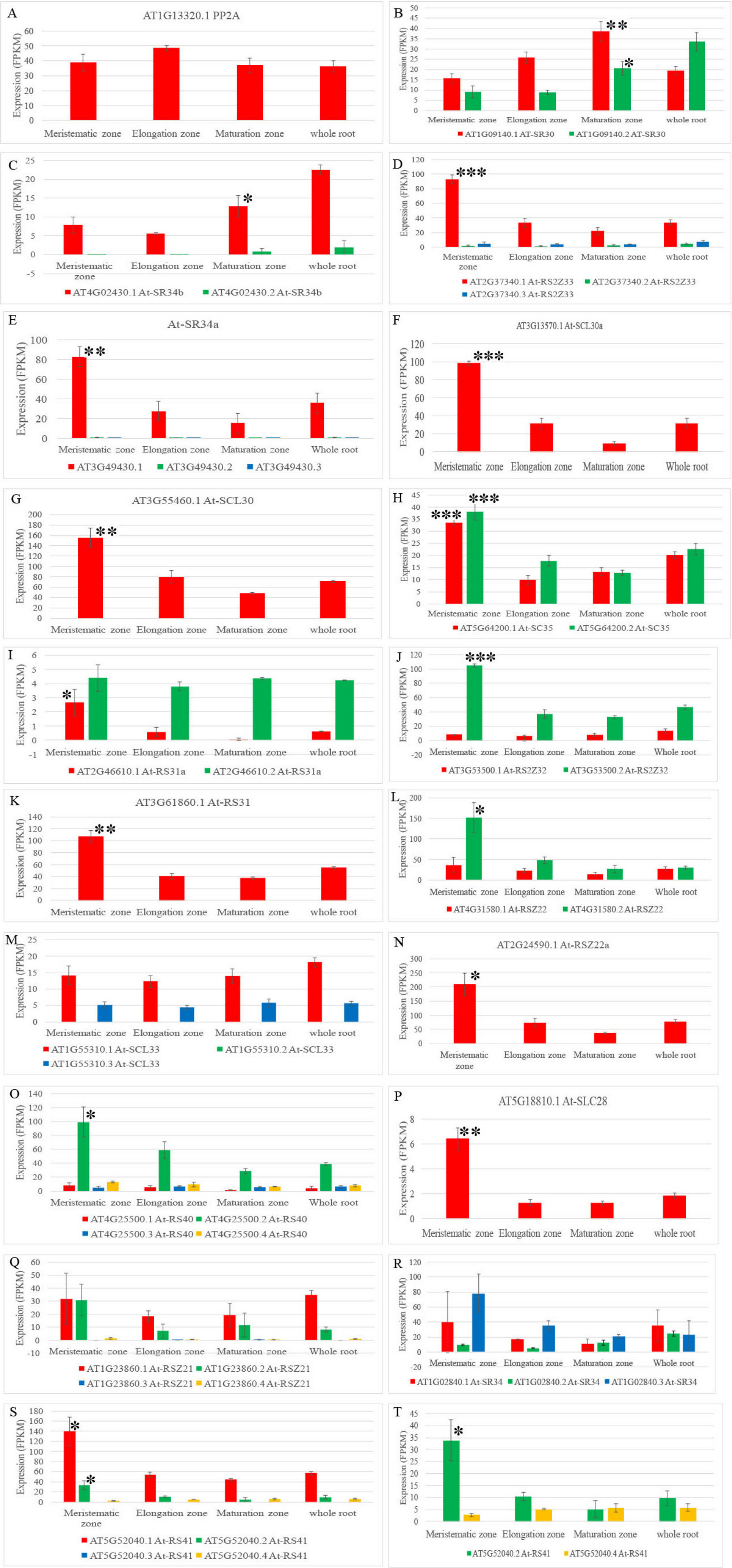
A
AT1G13320.1 PP2A
Expression (FPKM)
Meristematic zone
Elongation zone
Maturation zone
whole root
B
**
*
AT1G09140.1 AT-SR30
AT1G09140.2 AT-SR30
C
*
AT4G02430.1 At-SR34b
AT4G02430.2 At-SR34b
D
***
AT2G37340.1 At-RS2Z33
AT2G37340.2 At-RS2Z33
AT2G37340.3 At-RS2Z33
E
At-SR34a
**
AT3G49430.1
AT3G49430.2
AT3G49430.3
Whole root
F
AT3G13570.1 At-SCL30a
***
G
AT3G55460.1 At-SCL30
**
H
***
***
AT5G64200.1 At-SC35
AT5G64200.2 At-SC35
I
*
AT2G46610.1 At-RS31a
AT2G46610.2 At-RS31a
J
***
AT3G53500.1 At-RS2Z32
AT3G53500.2 At-RS2Z32
K
AT3G61860.1 At-RS31
**
L
*
AT4G31580.1 At-RSZ22
AT4G31580.2 At-RSZ22
M
AT1G55310.1 At-SCL33
AT1G55310.2 At-SCL33
AT1G55310.3 At-SCL33
N
AT2G24590.1 At-RSZ22a
*
O
*
AT4G25500.1 At-RS40
AT4G25500.2 At-RS40
AT4G25500.3 At-RS40
AT4G25500.4 At-RS40
P
AT5G18810.1 At-SLC28
**
Q
AT1G23860.1 At-RSZ21
AT1G23860.2 At-RSZ21
AT1G23860.3 At-RSZ21
AT1G23860.4 At-RSZ21
R
AT1G02840.1 At-SR34
AT1G02840.2 At-SR34
AT1G02840.3 At-SR34
S
*
*
AT5G52040.1 At-RS41
AT5G52040.2 At-RS41
AT5G52040.3 At-RS41
AT5G52040.4 At-RS41
T
*
AT5G52040.2 At-RS41
AT5G52040.4 At-RS41

Figure S5 Expression of isoforms of the SR genes in three developmental zones and in whole roots. N=3, *p<0.01, **p<0.001, ***p<0.0005. **Related to Figure 4.**

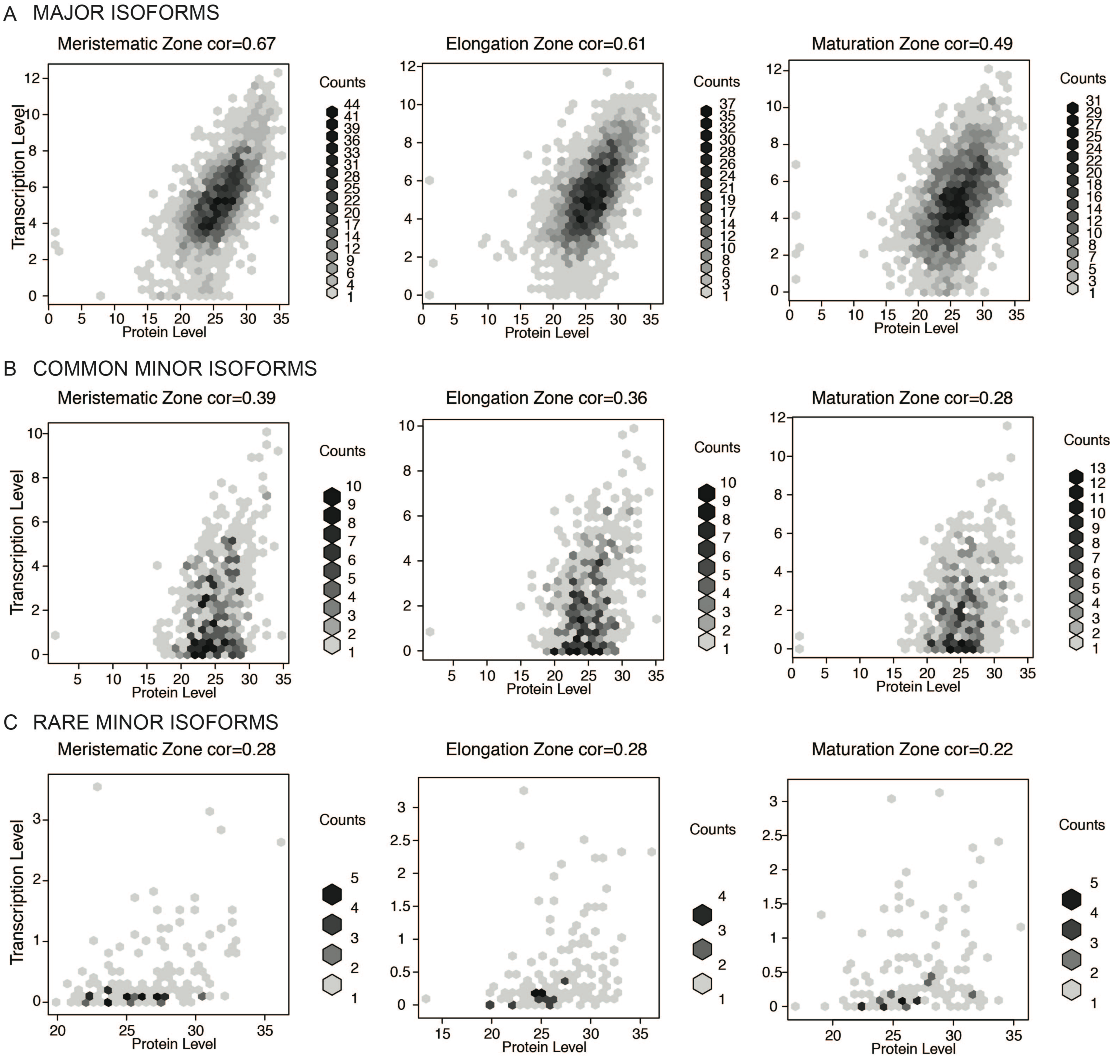

Figure S6. Correlation between all major isoforms (A), common minor isoforms (B), rare minor isoforms (C) and their corresponding protein levels. **Related to Figure 5.**

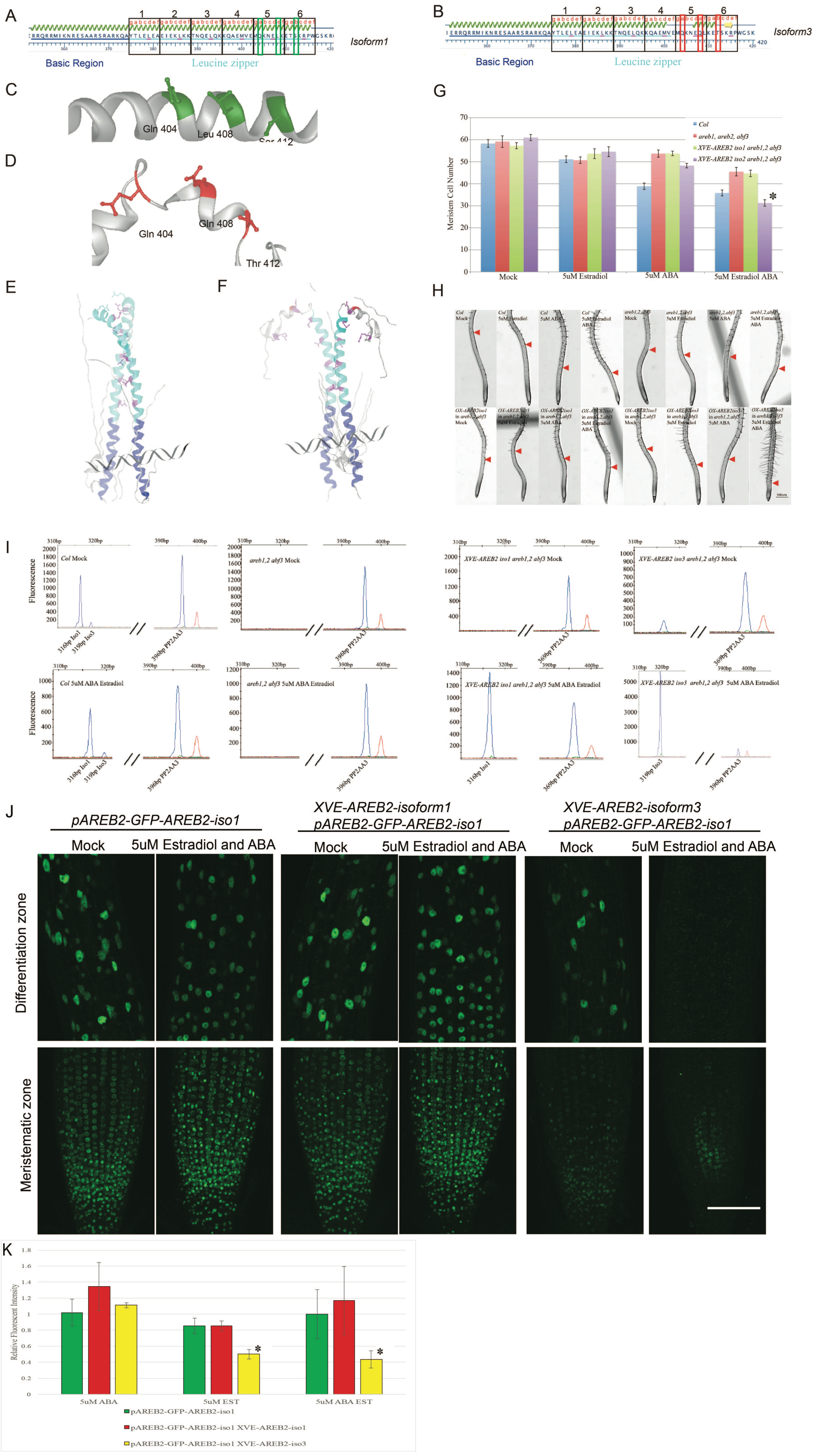
A
B
1
2
3
4
5
6
Isoform1
Isoform3
Basic Region
Leucine zipper
C
Gln 404
Leu 408
Ser 412
D
Gln 404
Gln 408
Thr 412
E
F
G
Meristem Cell Number
Col
areb1, areb2, abf3
XVE-AREB2 iso1 areb1,2 abf3
XVE-AREB2 iso2 areb1,2 abf3
Mock
5uM Estradiol
5uM ABA
5uM Estradiol ABA
*
H
I
Fluorescence
Col Mock
areb1,2 abf3 Mock
XVE-AREB2 iso1 areb1,2 abf3 Mock
XVE-AREB2 iso3 areb1,2 abf3 Mock
Col 5uM ABA Estradiol
areb1,2 abf3 5uM ABA Estradiol
XVE-AREB2 iso1 areb1,2 abf3 5uM ABA Estradiol
XVE-AREB2 iso3 areb1,2 abf3 5uM ABA Estradiol
316bp Iso1
319bp Iso3
390bp PP2AA3
J
pAREB2-GFP-AREB2-iso1
XVE-AREB2-isoform1
pAREB2-GFP-AREB2-iso1
XVE-AREB2-isoform3
pAREB2-GFP-AREB2-iso1
Mock
5uM Estradiol and ABA
Differentiation zone
Meristematic zone
K
Relative Fluorescent Intensity
5uM ABA
5uM EST
5uM ABA EST
*
pAREB2-GFP-AREB2-iso1
pAREB2-GFP-AREB2-iso1 XVE-AREB2-iso1
pAREB2-GFP-AREB2-iso1 XVE-AREB2-iso3

Figure S7. (A, B) Amino acid sequences and domain structures of *AREB2-Iso1* and *–Iso3*. Black boxes with numbers above indicate a heptad repeat domain of *AREB2-Iso1* and *–Iso3*. Red letters (a-g) show the location of the amino acid in each heptad. Green boxes mark Glutamine 404, Leucine 408, and Serine 412. Red boxes mark Glutamine 404, Glutamine 408, and Threonine 412. **(C, D)** Predicted alpha helical structures of the leucine zippers of the *AREB2-iso1* and *–iso3*. The green and red amino acid residues indicate the same amino acids marked by the green and red boxes in A and B. **(E, F)** Predicted three dimensional structures of *AREB2-Iso1* and *–Iso3*. The dark blue alpha helix is the Basic region for DNA binding. The light blue alpha helix is the Leucine zipper for dimerization. The magenta amino acid residues are required for dimerization. **(G)** Average cell number in the root meristem of *Col-0*, the *areb1,2* and *abf3* triple mutant *XVE-AREB2-Iso1*, and *-Iso3* in the triple mutant on MS medium under the same conditions as Figure 4B. (N>7, ±SD, *p<0.001) **(H)** Morphologies of Root hairs of *Col-0*, the *areb1, 2, abf3* triple mutant, *XVE-AREB2-Iso1* and *–Iso3* on the same condition as Figure 4B. Red arrowheads indicate the first visible root hairs. Scale bar 300um **(I)** Capillary electrophoresis analysis of RT-PCR products (See Supplementary Methods, primers used in RT-PCR validation). to show transcript levels of two isoforms of *AREB2* in *Col-0*, the *areb1,2, abf3* mutant, *XVE-AREB2-Iso1*, and *–Iso3* in the triple mutant five days after treatment with 5uM Estradiol, 5uM ABA, 5uM Estradiol and ABA. PCR products were obtained after 20 cycles. *PP2AA3* was used as an internal control. **(J)** Confocal images of *pAREB2-GFP-AREB2-iso1, XVE-AREB2-Iso1* in *pAREB2-GFP-AREB2-iso1,* and *XVE-AREB2-Iso3* in *pAREB2-GFP-AREB2-iso1*with Mock and 5uM Estradiol and ABA for 24h. 6-day-old seedlings were grown on MS medium for 5 days before transfer. Scale bar, 100um. **(K)** Relative fluorescent intensity of *pAREB2-GFP-AREB2-iso1* in wild type, *XVE-AREB2-Iso1*and *XVE-AREB2-Iso3* after treatments with Mock, 5uM ABA, 5uM Estradiol (EST), 5uM ABA and EST for 24h. **Related to Figure 6.**

Supplemental Tables

Table S1, related to Figure 1. Tab1. Names of cell types. Tab2. Number of mapped reads and mapping rate for each sample. Tab3. Correlation between array and RNAseq. Tab 4. Correlation between cufflinks and stringtie.

Table S2, related to Figure 2. Tab1. List of local alternative acceptor events. Tab2. List of local alternative donor events. Tab3. List of local exon skipping events. Tab4. List of type I and type II intron retention events.

Table S3, related to Figure 4 and Figure 7. Tab1. List of primers used in the RT-PCR validation. Tab2. Expression levels of significantly differentially expressed lincRNAs.

Table S4, related to Figure 5. Tab1. Peptide quantification. Tab2. Quantification of protein abundance. Tab3. Peptides uniquely mapped to alternative spliced isoforms.

Table S5, related to Figure 4. Average expression level of all isoforms as calculated by cufflinks.

File S6, related to Figure 1. Newly assembled protein coding gene models.

File S7, related to Figure 7. Genomic locations for the newly identified and modified lincRNAs.

Supplemental Experimental Procedures

Experimental Methods

Plant material and Growth conditions

All 15 GFP marker lines (WOL, SCR, S4, S17, S32, S18, GL2, CO2, CORTEX 315, WOX5, PET111, WER, E30, COBL9, and APL) are in the Columbia-0 (Col-0) background (Brady et al., 2007, Heidstra et al., 2004, Nawy et al., 2005, Birnbaum et al., 2005, Lee et al., 2006, Bonke et al., 2003). Seeds were surface sterilized using 50% (vol/vol) bleach and 0.1% Tween20 (Sigma) for 10 min and then rinsed five times with sterile water. All seeds were plated on mesh (Nitex 03-100/44, opening 100um) on standard MS media (1× Murashige and Skoog salt mixture, Caisson Laboratories), 0.5 g/L MES, 1% Sucrose, and 1% Agar (Difco) and adjusted to pH 5.7 with KOH. All plated seeds were stratified at 4 °C for 2 d before germination on vertically positioned square plates in a Percival incubator with 16 h of daily illumination at 22 °C.

Total RNA preparation, RNA amplification and library preparation for RNA-seq

Approximately 50,000 GFP positive cells were isolated for each of the marker lines. Sorting conditions were as previously described (Birnbaum et al., 2005). Total RNA was extracted with the RNeasy Micro Kit (Qiagen) from sorted and unsorted whole roots as control and 15 marker lines. All total RNA samples were treated with DNaseI during RNA extraction. RNA quality was examined using a 2100 Bioanalyzer (Agilent). The RNA Integrity Number (RIN) was over 8.0 in all of the cell populations except for the WOX5 marker line for which it was 4.5. The concentration of total RNA was measured using a Qubit (Invitrogen) instrument. From each of the cell populations, 50ng total RNA was amplified using the Ovation RNA-seq System V2 (NuGEN), except for the RNA isolated from the WOX5 line from which 10ng was amplified. Following amplification, 3 μ g of cDNA was fragmented using the Covaris S-Series System, then 400ng of the fragmented cDNA with an average size of 400bp was used for library preparation using the Encore NGS Library Multiplex System I (NuGEN). Illumina sequencing was performed at the Duke Genome Sequencing Shared Resource and New York Genome Center. The libraries for three biological replicates of each marker line and sorted and unsorted whole roots were sequenced on an Illumina HiSeq2000 (100/125 base-paired reads) with 8 samples loaded in each flow cell.

The meristematic, elongation, and maturation zones of seedling roots were collected under a dissecting microscope (Axio Zoom, Zeiss). The meristematic zone was dissected using GFP expression from a *HIGH PLOIDY2* (*HYP2*)-*GFP* line (Ishida et al., 2009). The elongation and maturation zones were dissected using as landmarks changes in cell size and presence of root hairs, respectively. The RNA-seq libraries were generated from total RNA of 50 root sections for each developmental zone as described above. Three replicates of each developmental zone were sequenced on an Illumina HiSeq2500 (125 base-paired reads) with 4 samples loaded in each flow cell.

Cycloheximide treatment

Cycloheximide (CHX) treatment was performed as previously described (Drechsel et al., 2013). 6 day-old seedlings were transferred to liquid MS medium with 10ug/ml CHX (Sigma) or water as mock treatment and incubated for 4h at room temperature. Total RNA was extracted from whole roots after CHX or mock treatment. The libraries were prepared as described in the previous section. Three biological replicates of each treatment were sequenced.

Library preparation for Pacific Biosciences sequencing

Total RNA was extracted from whole roots using the RNeasy Micro Kit (Qiagen). Poly-A plus RNA was isolated using the Poly(A) purist MAG kit (Life Technologies , AM1922) from 48.6 ug total RNA (RIN 9.1 in a bioanalyzer). First strand cDNA was synthesized from 6.4 ng poly-A plus RNA using the SMRTer PCR cDNA synthesis kit (Clontech). Second strand synthesis was performed with the KAPA HiFi PCR kit (Kapa Biosystems) at 30 PCR cycles using SMRTer II A (5'AAG CAG TGG TAT CAA CGC AGA GTA C 3', Clonetech) as a primer. The SMRTbell libraries were generated with the PacBio Template Prep Kit (Pacific Biosciences) without size selection.

Sample preparation for Mass Spectrometry analysis

At least 230 sections of each root developmental zone from a HYP2-GFP expressing line were dissected under a dissecting microscope. Nine samples (3 developmental zones x 3 replicates) were each pooled from 5 tubes per sample and resuspended in a total of 125 μl of 50 mM ammonium bicarbonate, pH 8 (AmBic) followed by addition of 25 μl of 1% acid labile surfactant (ALS-1) in

AmBic. Each sample was sonicated for 3 x 5s with cooling on ice between sonication steps, followed by centrifugation. Supernatants were collected, and protein concentrations were quantified by Bradford assay. After normalization of protein concentration with lysis buffer, samples were denatured and reduced by addition of 10 mM DTT and heating at 80 ºC for 15 min, followed by alkylation with 25 mM iodoacetamide in the dark for 30 min. Sequencing grade modified trypsin (Promega; 1:50 w/w trypsin:protein) was added, and proteins were digested at 37 ºC overnight. Samples were then acidified with 1% trifluoroacetic acid (TFA), 2% acetonitrile (ACN), followed by heating at 60 ºC for 2 h, to inactivate trypsin and degrade ALS-1. Samples were lyophilized and resuspended in 20 mM ammonium formate, pH 10, containing trypsinized yeast alcohol dehydrogenase 1 (MassPrep, Waters) at 25 fmol per µg as an internal standard to each sample. Following centrifugation, supernatants were transferred to Maximum Recovery LC vials (Waters). QC pools were prepared by mixing equal volumes of all samples.

Quantitative Mass Spectrometry

Quantitative two-dimensional liquid chromatography, tandem mass spectrometry (2D-LC-MS/MS) was performed on 1.5 µg of the peptide digests per sample in singlicate, with additional QC and conditioning analyses. Samples were analyzed using a nanoACQUITY 2D UPLC system (Waters) coupled to a Q-Exactive Plus high-resolution accurate mass tandem mass spectrometer (Thermo) via a nanoelectrospray ionization source. Peptides were first trapped at 2 µl/min at 97/3 v/v H2O/MeCN in 20 mM ammonium formate (pH 10) on a 5 µm XBridge BEH130 C18 300 um °— 50 mm column(Waters). A series of step-wise elutions of MeCN at 2 µl/min was used to elute peptides from the 1st dimension column. Elutions utilized 10.8%, 14.0%, 16.7%, 20.4% and 50.0% MeCN for the five fractions; these percentages were optimized for delivery of an approximately equal load to the 2nd dimension column for each fraction. For 2nd dimension separation, the eluent from the 1st dimension was first diluted 10-fold online with 99.8/0.1/0.1 v/v/v H2O/MeCN/formic acid and trapped on a 5 µm Symmetry C18 180 µm °— 20 mm trapping column (Waters). The 2nd dimension separations were performed on a 1.8 µm Acquity HSS T3 C18 75 µm °—150 mm column (Waters) using a linear gradient of 7 to 35% MeCN with 0.1% formic acid over 37 min, at a flow rate of 0.5 µl/min and column temperature of 35 °C. Data collection on the Q-Exactive Plus mass spectrometer was performed in data-dependent acquisition (DDA) mode of acquisition with Rs = 70,000 (@ m/z 200) full MS scan from m/z 375 to 1600 with a target AGC value of 1e6 ions followed by 10 MS/MS scans at Rs = 17,500 (@ m/z 200) at a target AGC value of 5e4 ions. A 20 s dynamic exclusion was employed. The total analysis cycle time for each sample injection was approximately 5 h. Following 12 total analyses (including triplicate QC injections), data was imported into Rosetta Elucidator v3.3 (Rosetta Biosoftware, Inc), and all LC-MS runs were aligned based on the accurate mass and retention time of detected ions (“features”) using PeakTeller algorithm in Elucidator. Relative peptide abundance was calculated based on area-under-the-curve (AUC) of aligned features across all runs. The overall dataset had 365,243 quantified features, and 705,947 high collision energy (peptide fragment) spectra. These spectra were exported by Elucidator and subjected to database searching. This MS/MS data was searched against The Arabidopsis Information Resource (TAIR10) database (downloaded on 03/25/14; 32,784 entries) with addition of yeast ADH1 as well as an equal number of reversed-sequence “decoys” for false discovery rate determination (65,570 total entries). Mascot Server (v2.5, Matrix Sciences) was utilized to perform the database searches, with 5ppm precursor tolerance and 0.02 Da product ion tolerance. Included in the database searches were fixed modification on Cys (carbamidomethyl) and variable modifications on Met (oxidation) and Asn/Gln (deamidation). After individual peptide scoring using PeptideProphet (Ma et al., 2012) algorithm in Elucidator, the data was annotated at a 1.0% peptide false discovery rate. For quantitative processing, the data was first curated to contain only high quality peptides with appropriate chromatographic peak shape and the dataset was intensity scaled to the robust median across all samples analyzed.

Plasmid Construction

The coding sequences of *AREB2-isoform1* and *–isoform3* of *AREB2* (AT3G19290) were amplified using the Phusion High-Fidelity DNA polymerase (NEB) with primers 5’-CACCATGGGAACTCACATCAATTTCA-3’ and 5’- TCACCATGGTCCGGTTAAT-3’ from a wild-type cDNA library, and then subcloned into the pENTR/D/TOPO vector (Invitrogen). The sequences of *AREB2-isoform1* and *–isoform3* in pENTR/D/TOPO vector were confirmed by Sanger sequencing. Both clones were recombined with the pMDC7 vector (Curtis and Grossniklaus, 2003) using LR clonase II (Invitrogen) to fuse the estradiol inducible promoter (XVE) and the *AREB2* coding region.

Assays for Root Elongation

The *XVE-AREB2-iso1* and *–iso3* constructs were transformed into triple mutants of *areb1*, *2* and *abf3*. To perform root elongation assays, two independent lines of *XVE-AREB2-iso1*, *–iso3* in the *areb1, 2, and abf3* triple mutant*,* wild-type and the triple mutant control were grown on MS media for six days, then transferred to MS media containing 5µM ABA (Sigma), 5µM 17β-estradiol (Sigma), 5µM ABA and 17β-estradiol or DMSO and EtOH as the carrier control. Images of roots were acquired with a scanner (Epson) four days after transfer and measured using ImageJ. The root elongation rate was calculated by comparing the root length on the mock treatment versus treatments with ABA, 17β-estradiol, and both.

Microscopy

For root morphology, images were collected using a dissecting microscope (Axio Zoom, Zeiss) with a digital camera. To count the number of cells in the meristem and examine the GFP signal of *pAREB2-GFP-AREB2-iso1 (Yoshida et al., 2010)* seedlings were mounted in 0.1mg/mL propidium iodide (10 mg/mL; Sigma) and observed under a Zeiss LSM510 confocal microscope. Cell numbers were counted from three independent experiments. Seedling ages and numbers of seedlings are described in the figure legend. GFP signal intensity was measured using Image J (Schneider et al., 2012).

RT-PCR

For validation of novel alternative spliced isoforms, mRNAs were extracted from total RNA of whole roots using Dynabeads Oligo (dT) 25 (Life Technologies). Ten micrograms of poly-A selected RNA were used to make cDNA with the Ovation RNA-seq V2 kit (NuGEN). The gene for a housekeeping protein, PROTEIN PHOSPHATASE 2A SUBUNIT A3 (PP2AA3, AT1G13320) and splice isoforms were amplified by PCR with specific primers. To determine the linear range of amplification, preliminary PCR reactions were carried out with different cycle numbers and 30 to 45 cycles were chosen for DNA detection, depending on the gene expression levels.

Capillary electrophoresis analysis of RT-PCR for detecting *AREB2* isoforms

PCR was conducted at 94ºC for 2 min followed by linear phase amplification of 20 cycles at 98ºC for 30 sec, 60ºC for 30 sec, and 72ºC for 30 sec. All samples were then extended at 72ºC for 4 min and, finally, cooled to 4ºC. The following gene specific PCR primers were used. AREB2 F (6FAM-CCA GGA ACA AGC AGC GCA), AREB2 R (TCC TTC TCA AGC ATT GCC TTT), PP2AA3 F (6FAM-ATG TCT ATG GTT GAT GAG CCT), and PP2AA3 R (CTT AGC CAG AGG AGT GAA AT). Capillary electrophoresis analysis of the PCR products was performed on an ABI 3130 X1. The data were analyzed using GeneMapper 4.0 software.

Primers used for the validation by RT-PCR

Intron retention:
AT1G28470 F (CTCCCATCACGACTTGTCC), R (CTCCAAATGCATAAGGATCTC).
AT5G35560 F (CTCTGCCTTGCTTAATAATTG), R (CACCCCAACAAAGAGACTCT).
AT3G51540 F (GCTTTATACCATTGCAGAAGTT), R (GTTGGTCTTGAAGATGGATCA).
Alternative 3' Acceptor:
AT3G12680 F (CAGTCCTCTCAGGATGATTG), R (CAGCTACCATCATCGAGTC), NF (TCCACTTCACCCTGAAAAAG).
AT4G16845 12F (GATCGCCAGATGCTTGATGA), isoF (GCCAGCTGCAGATGCTTG), R (CTGCTATAACCCTTTGTTTTCTTA).
AT2G23790 F (GTTTGTATTAGACCTGATCAG), R (CTTTAAACTCGATTCTCCGAG).
Alternative 5' Donner:
AT1G49340 F (TCAAGAAAGCGAGACTGGAA), NR (TTGATCCGGCAGTGGTTTAAG), 12R (TTGATCCGGCAGTGCACAG).
AT4G27120 F (CAACTTCAGCTCTTCTGCCT), R (GAAGATGAAGCTGGTGGAAC).
AT2G01220 F (AACTTACGGAGCCGAGAAGA), R (GTCAAGCCAGAGTCAAGTGT)
Exon Skipping:
AT5G16480: F (CTTTCCGCGCTTACAGTGG), R (TTATCTGTGTCCTGAGCCATA).
AT1G29030: F (CTGGTCAAAAGCAGGCATG), R (AGCCTAGGGGATGTAACT).
AT5G02810: F (CTGAGCGCCACTCCCATC), R (GGTGATCATGCCTTACTTATC).

Computational Methods

Read mapping and analysis.

The paired-end (2 x 100 bases or 2 x 125 bases) libraries were mapped to the Arabidopsis genome

(TAIR10) and transcriptome using **STAR** version 2.4.2 (Dobin et al., 2013) and a GTF annotation file from the TAIR website (http://arabidopsis.org). Raw reads and alignment files will be uploaded to NCBI-SRA. For the STAR alignment, we used the following parameters: “--outFilterMultimapNmax 20 --alignSJoverhangMin 8 --alignSJDBoverhangMin 8 --outFilterMismatchNmax 8 --alignIntronMin 35 --alignIntronMax 2000 --alignMatesGapMax 100000” and other parameters were kept as default. The “–alignIntronMax” option defines the maximum intron length and was set to 2000. The “–alignIntronMin” parameter defines the minimum intron length and was set to 35. These two parameters were selected such that less than 0.1% of known junctions in TAIR10 annotation are longer or shorter than these thresholds. We chose this threshold to reduce false positive rates in junction discovery and we obtained a high recall rate of the known splice junctions. In the alignment files, only properly paired reads that mapped to unique genomic locations were kept. On average, 40 million reads were found in each library, 85% of reads aligned to either one or two locations in the genome, 24 million reads mapped to one unique location in the genome, 17% of reads were spliced reads (**Table S1**). We analyzed the read distribution in different annotated genomic locations, and found that on average, 3.4 million read pairs mapped to exon regions per library. 0.93 million reads mapped to 3’UTR, 0.19 million reads mapped to 5’UTR and 186,000 reads mapped to annotated intron regions (**Figure S1**).

Gene expression analysis

A simulation (Ramskold et al., 2009) was performed to determine the detection threshold in FPKM (fragments per kilobase per million reads) for protein coding genes in our data. For known coding regions, the number of read pairs in each annotated gene region was counted using Rsamtools (http://bioconductor.org/packages/release/bioc/html/Rsamtools.html) and gene expression levels were summarized as FPKM. For genes with multiple isoforms, the longest isoform was used in the FPKM calculation. For intergenic regions, a random sampling was carried out to sample regions from known intergenic regions in the TAIR10 annotation according to the following criteria: 1) the region is 500 bp away from neighboring annotated transcripts; 2) the length of the region is equal to 1 kb. For intergenic regions that are larger than 2kb, only one random region was selected. We then counted the number of read pairs that mapped to these randomly sampled intergenic regions and calculated FPKM for these intergenic regions. The known coding regions were used as true positives and the sampled intergenic regions were used as true negatives. For each biological sample, we calculated the change of the false positive rate and false negative rate at different FPKM thresholds. We found 4% false negative rates for protein coding genes below a threshold of 0.05 FPKM in all samples, suggesting a fixed fraction (4%) of genes are not detected in any of our samples **(Figure S1C)**. We also found an average 8% of intergenic regions are identified below 0.05 FPKM, and the number of detected intergenic regions starts to drop at 0.1 FPKM. This suggests that a reasonable threshold to determine expressed protein coding genes lies between 0.02 and 0.1 FPKM. We choose 0.05 as our threshold to determine the number of expressed genes in each sample. For the gene family coverage analysis (**Figure S1B**), gene family annotations were downloaded from TAIR10.

Correlation between RNAseq and microarray analyses

For 13 of the cell types in the RNAseq analysis, microarray data from the same cell types were available (Brady et al., 2007) (**Table S1**). Cell types (CO2 for developing cortex and WOX5 for QC cells) without corresponding microarray data were not included in this comparison. Microarray expression data were normalized using RMA (Gautier et al., 2004). Probes mapping to multiple genes were discarded. Expression levels of biological replicates were averaged. For RNAseq data, we calculated the expression levels using FPKM and averaged between replicates in each biological sample. For each sample, we calculated both R^2 for linear regression and Spearman rank correlation between the expression levels of protein coding genes. Only those genes found in both the microarray and RNAseq datasets were used in the analysis (**Table S1**).

Identification of tissue specific genes

Differential expression analysis was carried out using edgeR (Robinson et al., 2010). Our data contains samples from 15 cell types and whole roots. The linear model has 16 coefficients: each cell type has a cell type-specific coefficient and the last coefficient is the grand mean, which corresponds to the mean expression in the whole root sample. We tested the hypothesis that at least one cell type-specific coefficient is not equal to zero, and the test was carried out using a likelihood ratio test implemented in edgeR (Robinson et al., 2010). Differentially expressed genes were further categorized as single cell type-specific genes, cell type-selective genes or constitutively expressed genes. A single cell type-specific gene is highly expressed in one cell type, but not highly expressed in other cell types. A cell type-selective gene is expressed at a high level in more than one cell type, but not in more than half of all cell types assayed. A constitutively expressed gene is detected in all cell types assayed and does not show significant differential expression between any pair of cell types. An entropy based method (ROKU (Kadota et al., 2006)) was applied to identify tissue-specific genes. The ROKU method calculates entropy for each gene across different cell types using centered absolute expression levels (equations 1, 2 and 3) and has been shown to perform well in identifying tissue-specific genes in microarray experiments.

$$\tilde{y}_i = abs((y_i - median(y))/std(y)) \tag{1}$$

$$p(\tilde{y}_i) = \tilde{y}_i/\sum \tilde{y}_i \tag{2}$$

$$H_{roku} = -\sum p(\tilde{y}_i)\log(p(\tilde{y}_i)) \tag{3}$$

A grand mean of gene expression levels for each gene was calculated using the linear model. To estimate p values for the H_{roku}, genes with expression that is within a 2 fold change from the grand mean were selected and used to calculate the background distribution of H_{roku}. For each gene in the genome, H_{roku} was calculated and compared to the background distribution. The p value was estimated as the number of genes in the background distribution that have H_{roku} larger than the gene specific H_{roku} divided by the total number of background genes. A gene is selectively activated/repressed in a cell type if the H_{roku} p value is smaller than 0.001. Gene ontology analysis was carried out using the eliminate count method in the topGO package (Alexa et al., 2006). Gene ontology annotation was downloaded from TAIR10 (June 2013). The p-values calculated by topGO account for the dependency in multiple comparisons based on the GO topology. For GO categories that are enriched (p <0.01) in at least one cell type, the negative log p-values were used to construct a heat map. The top 40 most enriched GO categories (excluding those enriched in all cell types) were plotted.

Splice junction detection and alternative splicing discovery

We identified unannotated splice junctions using spliced reads by requiring a minimum of 8 nt mapped to the neighboring exon (Kim et al., 2013). We further required that a splice junction has to be supported by at least one read in two biological replicates for each sample to ensure that the putatively novel splice junctions are supported by biological replicates.

To identify local splicing events, the spliced reads from all libraries were transformed into a list of potential splice junctions. To discover novel alternative splicing events, splice junctions were grouped based on their spliced-out and spliced-in locations according to the following rules: 1) one splice junction can only belong to one group, 2) for each group, all splice junctions were within the 3' and 5' boundaries of the group and 3) in each group, any splice junction has to overlap with at least one other splice junction in the same group. Newly identified junctions, as well as known junctions, supported by RNA-seq reads from our data were exhaustively searched for these junction groups. For simple alternative donor and alternative acceptor events (**Figure S2C**), we first identified junction groups with only two splice junctions, and then required that the two splice junctions are spliced-out (or spliced-in) at the same genomic location and spliced-in (or spliced-out) at different genomic locations. For an exon-skipping event, there are three exons involved: a left exon, a right exon, and a middle exon. The left and right exons were included in both transcripts and the middle exon was not included in one of the alternative isoforms (**Figure S2C**). To identify exon skipping events, we required a combination of three splice junctions: a long splice junction (n1) that connects the left exon and the right exon, and two shorter splice junctions (s1 and s2) that connect the left exon to the middle exon and the middle exon to the right exon. Complex splicing patterns, such as tri-acceptors and tri-donors (**Figure S2C**), where three splice junctions spliced-out (or spliced-in) at the same location, were found to be rare (**Figure 2B**).

Whole gene model construction and expression analysis of major isoforms

Cufflinks version 2.1.1 (Trapnell et al., 2013) and StringTie 1.0.4 (Pertea et al., 2015) were used to assemble novel and known transcripts in each of the biological samples. The transcripts from each biological sample were combined into a unified set of transcripts using Cuffmerge. Cuffnorm, cuffdiff and stringtie were then used on each library to quantify the expression of individual isoforms and gene expression levels. The Cuffnorm estimation of gene expression is highly consistent with the FPKM estimated by counting reads mapped to each gene ($R^2 > 0.97$ for all comparisons). The assembled gene models that cover more than one known protein coding gene locus (from TAIR10) were excluded. The longest coding region was determined for each transcript and transcripts. A premature termination codon (PTC) is defined as the presence of a splice junction (SJ) greater than 50 nucleotides downstream of the stop codon, exactly as defined in (Drechsel et al., 2013). PSI (percentage spliced in) was calculated using FPKM values estimated by cufflinks. For each multi-isoform gene, the PSI is calculated as:

$$PSI_{ij} = \frac{y_{ij}}{\sum_j y_{ij}} \qquad (4)$$

where y_{ij} is the expression of isoform j in cell type i.

For alternative acceptor sites and alternative donor sites, the number of reads supporting the shorter junction was divided by the total number of reads supporting both splice junctions (**Figure S2D**). Because a shorter splice junction implies a longer isoform, the number of reads supporting the shorter splice junction was used as numerator to calculate the longer isoform PSI (LPSI) for each alternative acceptor or donor event (**Figure 3C, Figure S2D**). For an exon skipping event, there are three types of splice junctions: one long junction across the skipped exon, and two shorter junctions define the boundary of the middle exon (**Figure S2D**). The PSI value is calculated using the average number of reads mapped to the two shorter junctions ((s1+s2)/2) divided by the sum of the total number of reads (n1 + (s1+s2)/2), as suggested (Barbosa-Morais et al., 2012).

Gene expression and splicing of SR proteins

Ser/Arg-rich (SR) proteins are a family of RNA binding proteins that play an important role in splicing (Reddy and Shad Ali, 2011). The Arabidopsis genome possesses 18 SR genes (Barta et al., 2010) most of which have alternatively spliced isoforms (Barta et al., 2010) These isoforms exhibit differential expression across organs and developmental ages Ectopic expression of SRs results in pleiotropic changes during development (Kalyna et al., 2003, Lopato et al., 1999) suggesting that alternatively spliced isoforms of SRs control organ development. We detected splice-isoforms of all SRs in the three developmental zones (**Figure S5**). As a control, the PP2A gene expressed at almost the same levels (**Figure S5A**). Most splice isoforms of the SR genes were highly expressed in the meristematic zone, and decreased in expression in the differentiation zone (**Figure S5D-T**). Taken together our results indicate that alternative splicing is coordinately regulated during root differentiation and that splice isoforms of SR genes may play a role in this regulation.

Intron retention and evolutionary analysis

Intron retention events were extracted from the cufflinks assembled whole transcripts by comparing common minor isoforms to the corresponding dominant isoform. Type I intron retention events were found when the intron was retained in the common minor isoform, whereas the Type II intron retention events were found when intron retention occurs in the dominant isoform. Multiple sequence alignment of the 9 crucifier species (Yeo et al., 2012) was downloaded from http://mustang.biol.mcgill.ca:8885. Only intron retention events with a length in a multiple of three were used for the analysis. To determine the protein coding sequences for the intron retention events, the phases of codons for the dominant isoforms were determined from the TAIR annotation, and the phases of the codons in the intron retention isoform were determined according to the phase of the dominant isoform. For the intron sequences in other species, we used Arabidopsis as a template to translate the aligned nucleotide sequences into protein sequences.

LincRNA identification and expression

Known Arabidopsis lincRNAs (RepTAS predicted lincRNAs, RNAseq predicted lincRNAs and TAIR10 annotated lincRNAs) were downloaded from a lincRNA database (Jin et al., 2013) (http://chualab.rockefeller.edu/gbrowse2/homepage.html). To identify intergenic transcribed regions from our RNAseq data, we performed the following analyses. We first used Cufflinks and Stringtie to construct full-length putative transcripts for each biological samples. We then used Cuffmerge to combine putative transcripts constructed in the first step into a unified set of transcripts. In this set of transcripts, those that overlap (as short as 1 base pair) with known protein coding genes were removed. The remaining transcripts were used as a preliminary set of predicted lincRNAs. These predicted lincRNAs were compared to the three types of lincRNAs found in the lincRNA database. The predicted lincRNAs from our data were merged with the known lincRNAs by extending the boundaries of overlapping lincRNAs to the furthest genomic location covered by either the lincRNA database predictions or by our RNAseq data. We then filtered the resulting lincRNAs using similar criteria as the lincRNA database: 1) transcribed regions longer than 200bp; 2) the longest open reading frame does not encode more than 100 amino acids; 3) transcribed regions are 500 bp away from any protein coding gene and 4) these transcribed regions do not overlap with any transposable elements. Our analysis identified 4200 lincRNAs (average FPKM>0.05 in one or more cell types), of which, 430 are newly identified in our dataset and 203 overlap with known lincRNAs. LincRNA expression was summarized in FPKM and differentially expressed LincRNAs were identified using edgeR (FDR<0.001). Differentially expressed LincRNAs with expression higher than 0.05 FPKM in more than half of our samples were used to calculate co-expression networks between mRNA (n=5000, genes with highest variation between samples) and lincRNA (n= 877). The WGCNA package was used to construct a co-expression network between mRNAs and lincRNAs using FPKM values across all cell types. Default parameters from WGCNA package were used in this analysis.

Supplemental References

Alexa, A., Rahnenfuhrer, J. and Lengauer, T. 2006. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics,* 22**,** 1600-7.

Barbosa-Morais, N. L., Irimia, M., Pan, Q., Xiong, H. Y., Gueroussov, S., Lee, L. J., Slobodeniuc, V., Kutter, C., Watt, S., Colak, R., et al. 2012. The evolutionary landscape of alternative splicing in vertebrate species. *Science,* 338**,** 1587-93.

Barta, A., Kalyna, M. and Reddy, A. S. 2010. Implementing a rational and consistent nomenclature for serine/arginine-rich protein splicing factors (SR proteins) in plants. *Plant Cell,* 22**,** 2926-9.

Birnbaum, K., Jung, J. W., Wang, J. Y., Lambert, G. M., Hirst, J. A., Galbraith, D. W. and Benfey, P. N. 2005. Cell type-specific expression profiling in plants via cell sorting of protoplasts from fluorescent reporter lines. *Nat Methods,* 2**,** 615-9.

Bonke, M., Thitamadee, S., Mahonen, A. P., Hauser, M. T. and Helariutta, Y. 2003. APL regulates vascular tissue identity in Arabidopsis. *Nature,* 426**,** 181-6.

Brady, S. M., Orlando, D. A., Lee, J. Y., Wang, J. Y., Koch, J., Dinneny, J. R., Mace, D., Ohler, U. and Benfey, P. N. 2007. A high-resolution root spatiotemporal map reveals dominant expression patterns. *Science,* 318**,** 801-6.

Curtis, M. D. and Grossniklaus, U. 2003. A gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant Physiol,* 133**,** 462-9.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T. R. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics,* 29**,** 15-21.

Drechsel, G., Kahles, A., Kesarwani, A. K., Stauffer, E., Behr, J., Drewe, P., Ratsch, 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**,** 3726-42.

Gautier, L., Cope, L., Bolstad, B. M. and Irizarry, R. A. 2004. affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics,* 20**,** 307-15.

Heidstra, R., Welch, D. and Scheres, B. 2004. Mosaic analyses using marked activation and deletion clones dissect Arabidopsis SCARECROW action in asymmetric cell division. *Genes Dev,* 18**,** 1964-9.

Ishida, T., Fujiwara, S., Miura, K., Stacey, N., Yoshimura, M., Schneider, K., Adachi, S., Minamisawa, K., Umeda, M. and Sugimoto, K. 2009. SUMO E3 ligase HIGH PLOIDY2 regulates endocycle onset and meristem maintenance in Arabidopsis. *Plant Cell,* 21**,** 2284-97.

Jin, J., Liu, J., Wang, H., Wong, L. and Chua, N. H. 2013. PLncDB: plant long non-coding RNA database. *Bioinformatics,* 29**,** 1068-71.

Kadota, K., Ye, J., Nakai, Y., Terada, T. and Shimizu, K. 2006. ROKU: a novel method for identification of tissue-specific genes. *BMC Bioinformatics,* 7**,** 294.

Kalyna, M., Lopato, S. and Barta, A. 2003. Ectopic expression of atRSZ33 reveals its function in splicing and causes pleiotropic changes in development. *Mol Biol Cell,* 14**,** 3565-77.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. and Salzberg, S. L. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol,* 14**,** R36.

Lee, J. Y., Colinas, J., Wang, J. Y., Mace, D., Ohler, U. and Benfey, P. N. 2006. Transcriptional and posttranscriptional regulation of transcription factor expression in Arabidopsis roots. *Proc Natl Acad Sci U S A,* 103**,** 6055-60.

Lopato, S., Kalyna, M., Dorner, S., Kobayashi, R., Krainer, A. R. and Barta, A. 1999. atSRp30, one of two SF2/ASF-like proteins from Arabidopsis thaliana, regulates splicing of specific plant genes. *Genes Dev,* 13**,** 987-1001.

Ma, K., Vitek, O. and Nesvizhskii, A. I. 2012. A statistical model-building perspective to identification of MS/MS spectra with PeptideProphet. *BMC Bioinformatics,* 13 Suppl 16**,** S1.

Nawy, T., Lee, J. Y., Colinas, J., Wang, J. Y., Thongrod, S. C., Malamy, J. E., Birnbaum, K. and Benfey, P. N. 2005. Transcriptional profile of the Arabidopsis root quiescent center. *Plant Cell,* 17**,** 1908-25.

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T. and Salzberg, S. L. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol,* 33**,** 290-5.

Ramskold, D., Wang, E. T., Burge, C. B. and Sandberg, R. 2009. An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. *PLoS Comput Biol,* 5**,** e1000598.

Reddy, A. S. and Shad Ali, G. 2011. Plant serine/arginine-rich proteins: roles in precursor messenger RNA splicing, plant development, and stress responses. *Wiley Interdiscip Rev RNA,* 2**,** 875-89.

Robinson, M. D., Mccarthy, D. J. and Smyth, G. K. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics,* 26**,** 139-40.

Schneider, C. A., Rasband, W. S. and Eliceiri, K. W. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods,* 9**,** 671-5.

Trapnell, C., Hendrickson, D. G., Sauvageau, M., Goff, L., Rinn, J. L. and Pachter, L. 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol,* 31**,** 46-53.

Yeo, Z. X., Chan, M., Yap, Y. S., Ang, P., Rozen, S. and Lee, A. S. 2012. Improving indel detection specificity of the Ion Torrent PGM benchtop sequencer. *PLoS One,* 7**,** e45798.

Yoshida, T., Fujita, Y., Sayama, H., Kidokoro, S., Maruyama, K., Mizoi, J., Shinozaki, K. and Yamaguchi-Shinozaki, K. 2010. AREB1, AREB2, and ABF3 are master transcription factors that cooperatively regulate ABRE-dependent ABA signaling involved in drought stress tolerance and require ABA for full activation. *Plant J,* 61**,** 672-85.