Supplementary Materials for

Direct roles of SPEECHLESS in the specification of stomatal self-renewing cells

On Sun Lau, Kelli A. Davies, Jessica Chang, Jessika Adrian, Matthew H. Rowe, Catherine E. Ballenger, Dominique C. Bergmann*

*Correspondence author. E-mail: dbergmann@stanford.edu

This PDF file includes:

Materials and Methods
Figs. S1 to S15
Tables S1, S2 and S6 to S8
References

Other Supplementary Materials for this manuscript includes the following:

Table S3 to S5 (Excel files)

Materials and Methods

Plant materials and growth conditions

Arabidopsis thaliana ecotype Columbia-0 (Col) was used as the wild type control. The following mutants and transgenic lines (in Col) used in the study were described previously: *spch3* and SPCHpro:nucGFP (*7*); SPCHpro:SPCH2-4A-YFP (*30*); ML1pro:mCherry-RCI2A (mCherry variant of the reporter in (*31*)); *bes1* RNAi (*27*) and *ice1-2* and *scrm2-1* (*14*). Seedling were grown on ½ strength Murashige and Skoog (MS) agar media at 22°C under long-day conditions (16h-light/8h-dark) and were harvested, treated or examined at the indicated time.

Vector construction and plant transformation

To construct SPCHpro:SPCH2-4A-MYC, gateway entry vectors that contain the 2.5 kb promoter of *SPCH* (in pDONR^{TM}P4-P1R) and the SPCH2-4A cDNA (in pENTR/D-TOPO®) were recombined into the plant binary vector R4pGWB519 (*30*, *32*). The estrogen inducible SPCH construct was generated by recombining SPCH1-4A cDNA (in pENTR/D-TOPO®) into pMDC7 (*30*, *33*). The YFP-transcriptional reporter of SPCH was generated by a tripartite recombination of the 2.5 kb promoter of *SPCH* (in pDONR^{TM}P4-P1R), pENTR-YFP-NLS and the binary vector R4pGWB501 (*32*). The YFP-transcriptional reporter of *ICE1* and a version with the SPCH-binding motifs altered were constructed in a similar fashion with pENTR5' that contains either the wild-type or the mutated form of the 2.5 kb promoter of *ICE1*. The mutant form of *ICE1* promoter (mICE1pro) has mutations in two of the SPCH-binding motifs (shown in Fig. 2H) and the mutations were introduced through DNA synthesis (Invitrogen's GeneArt® Strings^{TM} DNA fragment). For transcriptional reporters of At2g34510 and *BIM2*, 1.5 and 1.6 kb promoter of the respective genes were cloned into pENTR5'-TOPO® and pDONR^{TM}P4-P1R, respectively, and recombined with pENTR-YFP into R4pGWB540 (*32*). The At2g34510 protein is predicted to be a GPI-anchored protein ((*34*) and big-PI Predictor http://mendel.imp.ac.at/sat/gpi/gpi_server.html), so translational reporters of At2g34510 were made by inserting an YFP (Citrine) after the predicted N-terminal cleavage site (SignalP, http://www.cbs.dtu.dk/services/SignalP). The N-terminal portion of At2g34510 (up to the cleavage site) was cloned in front of the YFP in pENTR-Citrine and the remaining sequences of either the CDS or the intron-containing genomic fragment of At2g34510 were inserted after it. The resulting pENTR constructs were then recombined with the pENTR5' At2g34510 promoter construct into R4pGWB501 (*32*). To generate the translational reporter of *ARK3*, the genomic fragment starting from its 1.9 kb promoter through the last coding sequence were amplified and cloned into pENTR/D-TOPO® and was recombined into pHGY (*35*). For the SPCHpro:amiR-ark3 construct, the artificial microRNA sequence TCTATTAATAGTGATTTCCCC was designed with the Web MicroRNA Designer tool (http://wmd3.weigelworld.org). The microRNA sequence was engineered using the pRS300 plasmid, subcloned into pENTR/D-TOPO®, and together with the pDONR^{TM} SPCH promoter construct, was recombined into the binary vector R4pGWB501 (*32*). The BASLpro:*bin2-1*-YFP was constructed by recombining pENTR-*bin2-1* (a gain-of-function allele of *BIN2* (*36*)), pDONR^{TM} BASL promoter construct into R4pGWB440 (*32*). The BES1pro:*bes1*-D-GFP construct was

previously described (*26*). The binary vectors were transformed into Arabidopsis with the floral-dip method (*37*) and transgenic plants were selected on plates containing appropriate antibiotic-containing ½ MS agar media. The construct SPCHpro:SPCH2-4A-MYC was transformed into Col and *spch3*/+, and transgenic lines homozygous for *spch3* were recovered in subsequent generations and assayed for expression and rescue (fig. S1). All other vectors were transformed into Col, into Col carrying the epidermal plasma membrane marker ML1pro:mCherry-RCI2A or into the indicated genetic background.

Chromatin Immunoprecipitation (ChIP) assays

The large-scale MOBE-ChIPs for ChIP-Seq were performed with a modified protocol based on general procedures in (*38*, *39*). Briefly, a much larger amount of starting materials was used in our assay than in standard ChIP protocols (~24 g vs. ~1.5 g of plant materials). However, samples were processed in smaller-sized aliquots starting from cross-linking, through nuclei isolation and chromatin fragmentation to maintain efficacy in these steps. The size distribution of the fragmented chromatin was also validated on a 2100 Bioanalzyer (Agilent) and is on average ~400 bp. The fragmented chromatin extracts from the smaller aliquots were then combined for immunoprecipitation. Four-day-old seedlings of SPCHpro:SPCH2-4A-MYC (in *spch3*) and Col were used in the experiments. The Myc epitope was chosen because of the availability of high-quality monoclonal antibodies that may facilitate high signal enrichment at larger scales (Fig. 1B). Chromatin fragmentation was performed on a Bioruptor® sonicator (Diagenode) with settings at high intensity for 3 x 7.5 min and 1 x 2.5 min (cycles of 30 s on and 30 s off) at 4°C. Immunoprecipitation was carried out overnight with a monoclonal anti-MYC antibody 71D10 (Cell Signaling Technology), followed by an 1h incubation with Dynabeads® Protein A magnetic beads (Invitrogen). Immunoprecipitated DNA was reverse cross-linked overnight at 65°C and purified by the ChIP DNA Clean & Concentrator™ (Zymo). A total of 6 large-scale ChIPs (total starting materials: 144 g per genotype) were performed for constructing the ChIP-Seq library (see below).

For subsequent quantitative PCR validation, reactions were performed with Ssofast™ EvaGreen® Supermix (Bio-Rad) and primers specific to the promoter or other specified region of the target genes (table S8) on a CFX96 Real-Time PCR detection system (Bio-Rad). Signals from the ChIPed DNA were first normalized to their input DNA. Fold enrichment was then calculated by dividing the normalized value of SPCHpro:SPCH2-4A-MYC with that of Col (SPCHpro:nucGFP was used as a control for SPCHpro:SPCH2-4-YFP). Three technical replicates were assayed for each sample.

ChIP-Seq library preparation and high-throughput sequencing

The DNA yield from the large-scale ChIPs was very low (<1 ng per ChIP, quantified by the Qubit® fluorometer), which is likely due to the low background achieved with the use of monoclonal antibody and magnetic beads and the high purity of ChIPed DNA. To construct ChIP-Seq libraries, the final purified ChIPed DNAs from the 6 independent large-scale ChIPs on SPCHp:SPCH2-4A-Myc and Col described above were pooled. The high number of biological replicates (6 vs. 2-3 in most studies) and a pooling strategy were chosen to ensure enough starting DNA for unbiased sequencing library

construction. Sequencing libraries were constructed with the Ovation® Ultralow Library System (NuGEN), according to manufacturer's protocol. Suitable size distribution of the libraries was validated by analysis on a 2100 Bioanalyzer (Agilent). Each library (SPCHp:SPCH2-4A-Myc and Col) was sequenced on a lane of Illumina HiSeq 2000 (single-read 50-bp run), whose higher sequencing output was found to be beneficial to peak discovery (fig. S15).

ChIP-Seq data analysis

In order to generate a stringent set of SPCH-enriched regions, two peak calling methods were performed in parallel to independently detect areas of SPCH binding. In the first approach, reads were aligned to the TAIR10.18 genome assembly via SOAP2 (*40*) with default settings. Only uniquely mapped reads were then retained for peak calling with CSAR (*9*), where enriched regions between SPCHp:SPCH2-4A-Myc and Col are detected through a Poisson distribution. CSAR permutes computation of enrichment scores for read-enriched regions to calculate false discovery rate (FDR) thresholds, and only those peaks meeting a FDR < $1x10^{-6}$ and a score ≥ 3 were retained. In a second method, reads were aligned to the Arabidopsis genome via Bowtie 2 (*41*) with default settings, and peaks were called by BayesPeak (*42*), which utilizes a hidden Markov model to detect enriched regions. Peaks with a posterior probability greater than 95% were of interest. The resulting list of 8,327 peaks consists of the enriched regions that were identified by both methods. Peaks with an enrichment score of at least 10 (guided by SPCH targets that are known stomatal regulators) and that were non-intergenic (see below), were considered high-confidence SPCH-binding peaks (n=1,217).

Association of peaks to genes was initially accomplished via CSAR, where peaks are annotated 3kb upstream of a transcriptional start site (TSS) and 1kb downstream. The distribution of peaks along a gene was further refined using the TAIR10 GFF3 gene annotation file as belonging to one of the following categories: promoter (< 0.5 kb to TSS), distal promoter (0.5-3kb to TSS), 5'UTR, exon, intron, 3'UTR, downstream (1kb to TSS), or intergenic. A total of 8,958 genes are associated with the 8,327 SPCH-binding peaks. The 1,517 genes associated with the high-confidence SPCH-binding peaks are defined as high-confidence SPCH targets.

For the comparison in Fig. 1D, the ChIP-Seq datasets were analyzed through DNAnexus (http://classic.dnanexus.com) with default settings.

De-novo motif discovery and Gene Ontology (GO) terms analysis

To determine the *in vivo* binding motif of SPCH, 100 bp extensions from the summit of each SPCH-enriched region (total: 200 bp; n=8,327) defined input sequences for de novo motif discovery by MEME-ChIP with a maximum width of 10 (http://meme.nbcr.net/meme/cgi-bin/meme-chip.cgi) (*43*). Motifs with an expected value (E-value) << 0.05 were deemed significant.

Gene set enrichment analysis was carried out by agriGO (http://bioinfo.cau.edu.cn/agriGO/index.php) (*44*), where significance in enrichment of

gene ontology (GO) terms in the high-confidence SPCH-binding targets is relative to background enrichment of the Arabidopsis genome (TAIR10 assembly) as an input. Following multiple hypothesis testing correction (Benjamini & Hochberg method), GO terms with FDR < 0.05 are called significantly enriched.

SPCH induction and RNA extraction for RNA-Seq

The estrogen receptor-based XVE system (*45*) as introduced into vector pMDC7 (*33*) was used to induce SPCH expression. Transgenic lines homozygous for pMDC7-SPCH1-4A-YFP and Col were grown on ½ MS agar media for 4 days and the seedlings were bathed in a solution containing 10 μM beta-estradiol (diluted with water). Two replicates per genotype were collected at time zero (not treated), and at 6 and 8 hours post treatment. Total RNA from the seedlings was extracted using RNeasy Plant Mini Kit with DNaseI digestion (QIAGEN).

RNA-Seq library preparation and high-throughput sequencing

RNA quality of each sample was assessed with a 2100 Bioanalyzer (Agilent). Barcoded mRNA libraries were constructed by Duke University's Genomics Core Facility using Illumina Poly-A Purification TruSeq library reagents and protocols. Samples were sequenced on Illumina HiSeq 2000 (single-read 50-bp run) with the 6 samples described above multiplexed on two lanes.

RNA-Seq data analysis

Reads were aligned to the TAIR10 genome assembly via Tophat (*46*), where an initial alignment is followed by mapping to known exon junctions. Differential expression between SPCH induction samples and their WT counterparts in either 6 or 8 h post-treatment was identified via DESeq (*47*), where gene expression is modeled by a negative binomial distribution, which adopts a mean-variance relationship to account for the variation between biological replicates. Due to 3' bias of reads in one WT 0-hour sample, this sample was not considered in the analysis. *P*-values were adjusted via the Benjamini & Hochberg method, and differentially expressed transcripts with a FDR < 0.05 were deemed significant.

Identification of cell-type specific transcripts

Gene expression profiles derived from Fluorescence-Activated Cell Sorting (FACS)-isolated populations of specific stomatal lineage cell types by Dr. Jessika Adrian were used to compare with SPCH-binding targets. Briefly, this stomatal expression map contains microarray-based expression profiles of four fluorescence reporter-marked cell types in the lineage, as well as the epidermal cells as a control (fig. S7). Fuzzy k-means clustering was employed to identify dominant expression pattern and relevant clusters were used in the comparison.

RT-qPCR experiments

Total RNA from seedlings was purified as described above. Reverse transcription was carried out with 700 ng of total RNA using the iScript™ cDNA synthesis kit (Bio-Rad),

according to the manufacturer's protocol. Quantitative PCR with gene specific primers (Table S8; designed with QuantPrime (*48*)) were performed with the SsoAdvanced™ SYBR® Green Supermix (Bio-Rad) on a CFX96 Real-Time PCR detection system (Bio-Rad). Signals were normalized to the reference gene *ACTIN2* using the Δ^{CT} method and relative expression of a target gene was calculated from the ratio of test samples to Col. For each genotype, three biological replicates were assayed in two qPCR replicates.

Analysis of transcriptional and translational reporters and mutants

For confocal microscopy of reporters and mutants, fluorescence images were captured on a Leica SP5 microscope and were processed with ImageJ (National Institutes of Health). Cell outlines were visualized by either the plasma membrane marker ML1pro:mCherry-RCI2A or with propidium iodide (Molecular Probes, P3566; 0.1 mg/ml).

For quantification of stomatal phenotypes, seedlings were first cleared in 7:1 ethanol:acetic acid solution and mounted in Hoyer's medium. For a given genotype, differential contrast interference (DIC) images of the abaxial epidermis of cotyledons were captured at 20X (0.320 mm^{-2} field of view) on a Leica DM2500 microscope. More than 20 cotyledons were examined per test and scored blind (scorer did not know genotype or treatment of sample).

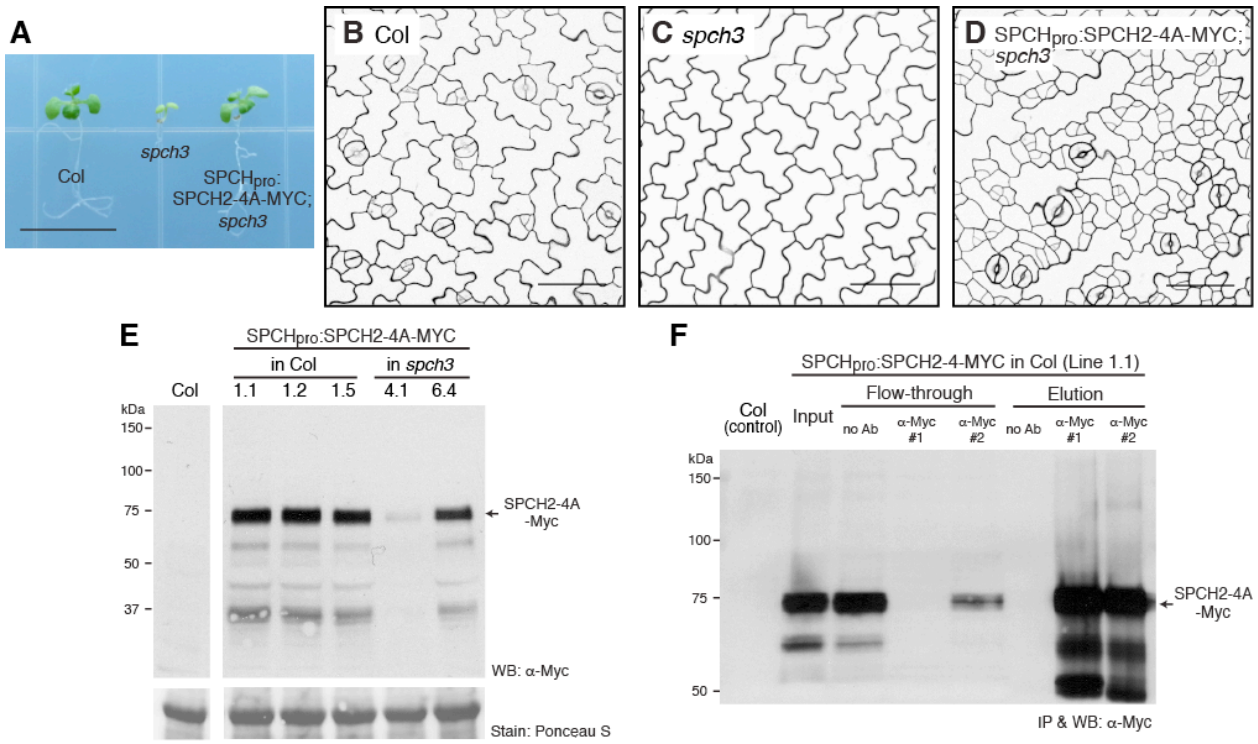


Fig. S1. Generation of the complementing SPCHpro:SPCH2-4A-MYC transgenic plants and *in vivo* immunoprecipitation of SPCH2-4A-Myc.

(**A**) Complementation of the seedling lethality of a *spch* mutant by SPCHpro:SPCH2-4A-MYC. 11-dpg seedlings of the indicated genotypes are shown together. (**B** to **D**) Confocal images of the abaxial cotyledons of 4-dpg seedlings of the indicated genotypes. Note in (D) that there is a modest increase in meristemoid divisions, but these divisions are asymmetric, well-spaced and there are no errors in mature stomatal pattern. Cell outlines were visualized with propidium iodide. Scale bar, 10 mm (A) and 50 μm (B to D). (**E**) Western detection of SPCH2-4A-Myc in transgenic plants harboring the SPCHpro:SPCH2-4A-MYC construct in either Col (3 transgenic lines, 1.1, 1.2 & 1.5) or *spch3* (2 transgenic lines, 4.1 & 6.4) backgrounds. Total protein was extracted from 4-dpg seedlings of the indicated genotypes, probed with anti-Myc antibody and detected by Western analysis with an enhanced chemiluminescent substrate (Thermo's SuperSignal West Dura substrate). Recombinant SPCH2-4A-Myc has a calculated M.W. of 57 kDa. The larger detected size was likely due to phosphorylations of the protein as reported in (*22*, *30*). (**F**) *In vivo* pull-down assay of *SPCH* promoter-driven SPCH2-4A-Myc from transgenic seedlings. Total soluble protein from SPCHpro:SPCH2-4A-MYC (in Col; line 1.1) was incubated with one of two commercial anti-Myc antibodies (#1: GenScript's THE^{TM}; #2: Cell Signaling Technology's 71D10). 71D10 was used for subsequent experiments. Flow-through and precipitated samples (Elution) were analyzed by Western blotting with anti-Myc antibody. WB: Western blot, IP: Immunoprecipitation.

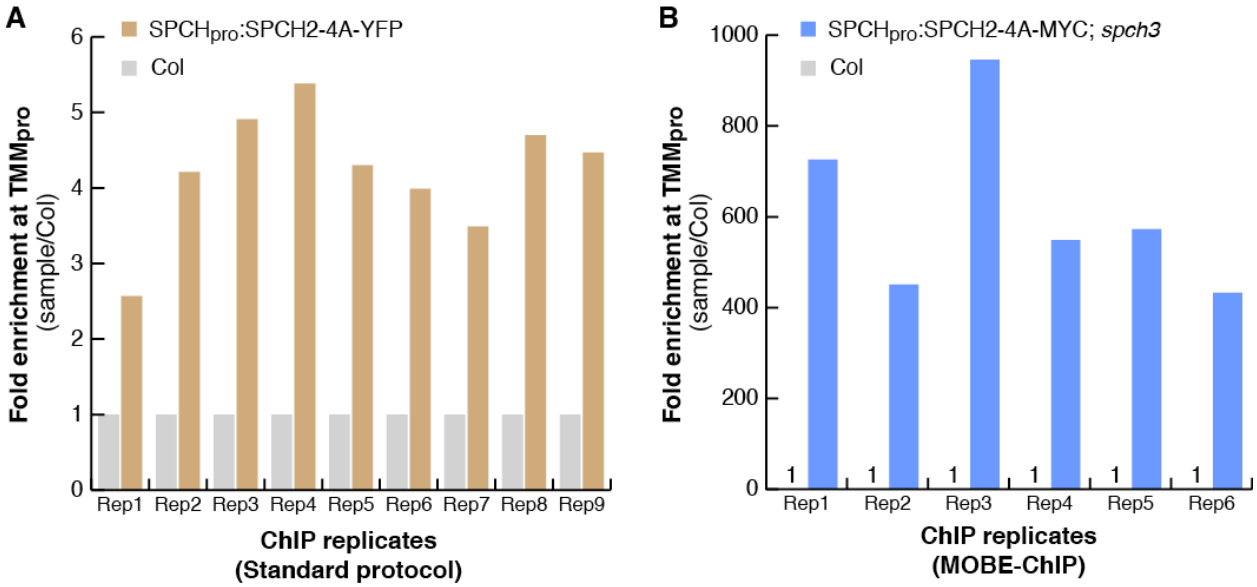


Fig. S2. Quality assessment of ChIP replicates used for sequencing library construction.

Multiple large-scale ChIPs were performed for the construction of sequencing library (See methods). The quality of each of the ChIP experiments was assessed by ChIP-qPCR for the enrichment of the promoter region of *TMM*. Shown are the qPCR results of 9 ChIP replicates for our initial SPCH sequencing library performed with standard protocol (A) and of the 6 ChIP replicates for our high-quality sequencing library performed with MOBE-ChIP (see main text and Methods) (B). Note the difference in scale in (A) and (B). Rep: replicate.

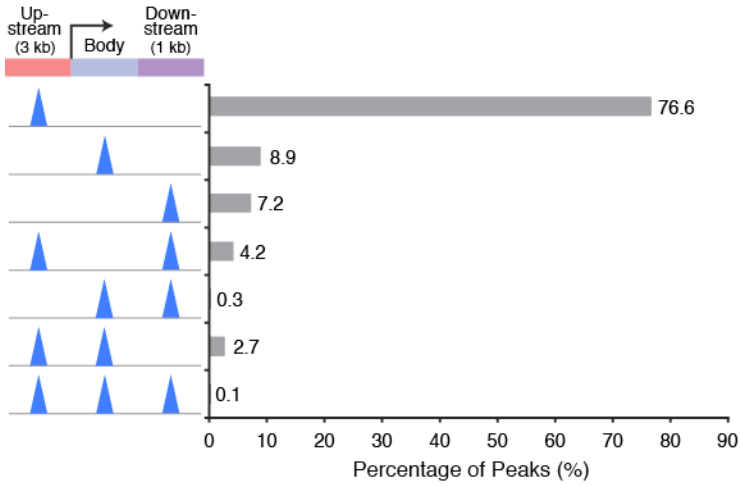


Fig. S3. Binding pattern of SPCH relative to gene models.

The binding pattern of SPCH-binding peaks relative to an associated gene model can be divided into 7 categories (y-axis), based on the division of a gene model into 3 regions. The 3 regions are “Upstream” (3 kb upstream of the transcriptional start site, TSS), “Body” (between TSS and end of transcript), and “Downstream” (1 kb downstream of the end of transcript). Each binding pattern was counted (multiple peaks can occur at one region) and the percentage of peaks that fall in each category is shown.

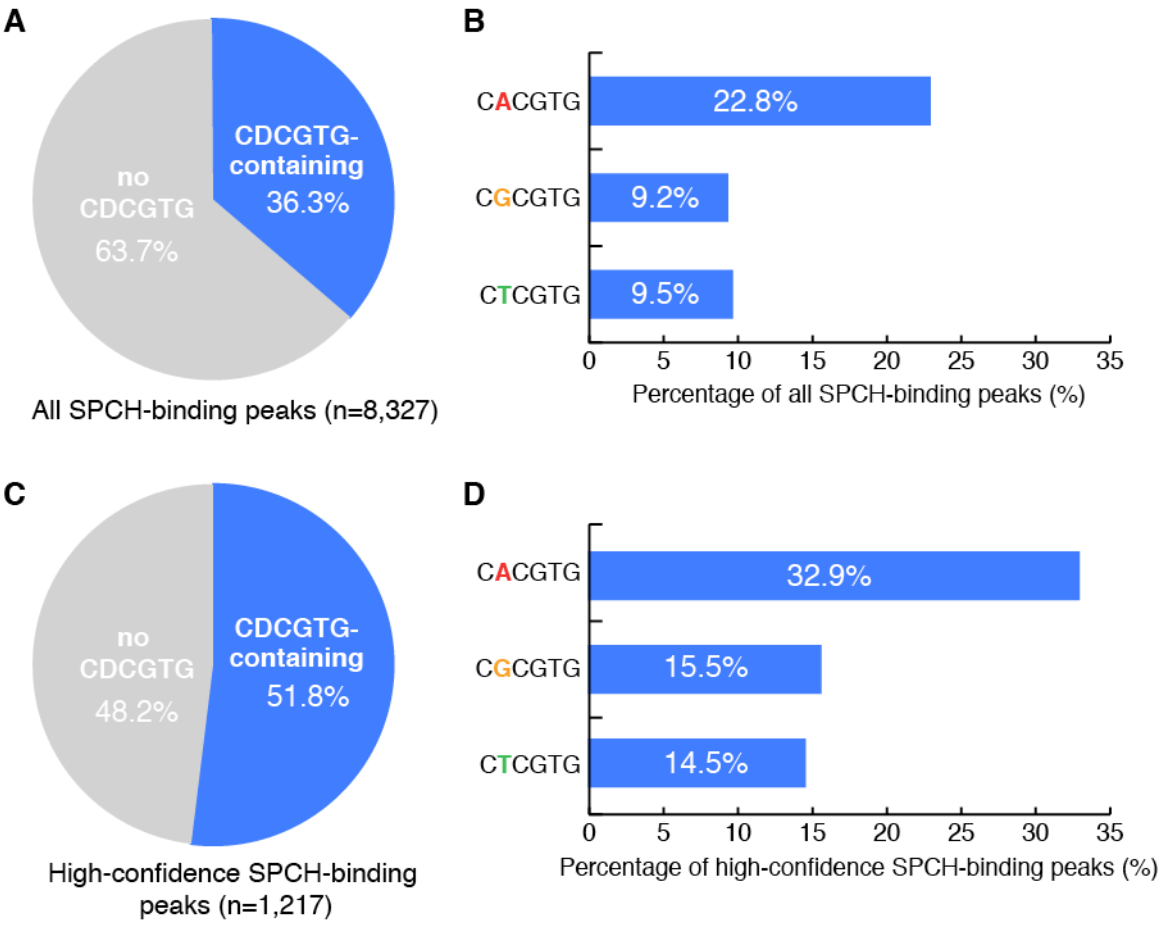


Fig. S4. Percentage of SPCH-binding peaks that contain the SPCH-binding motif (CDCGTG).

(**A** to **D**) The E-box-like motif, CDCGTG, where D denotes A, G or T, is the most enriched motif among the 8,327 SPCH-binding peaks (within the +/- 100 bp of summit window). Shown here are the proportion of CDCGTG-containing peaks in the complete data set (peak score ≥ 3) (A), and the proportion of peaks in the high-confidence data set (score ≥ 10) (C). The percentage of all (B) or the high-confidence (D) SPCH-binding peaks with a particular motif variant is also shown.

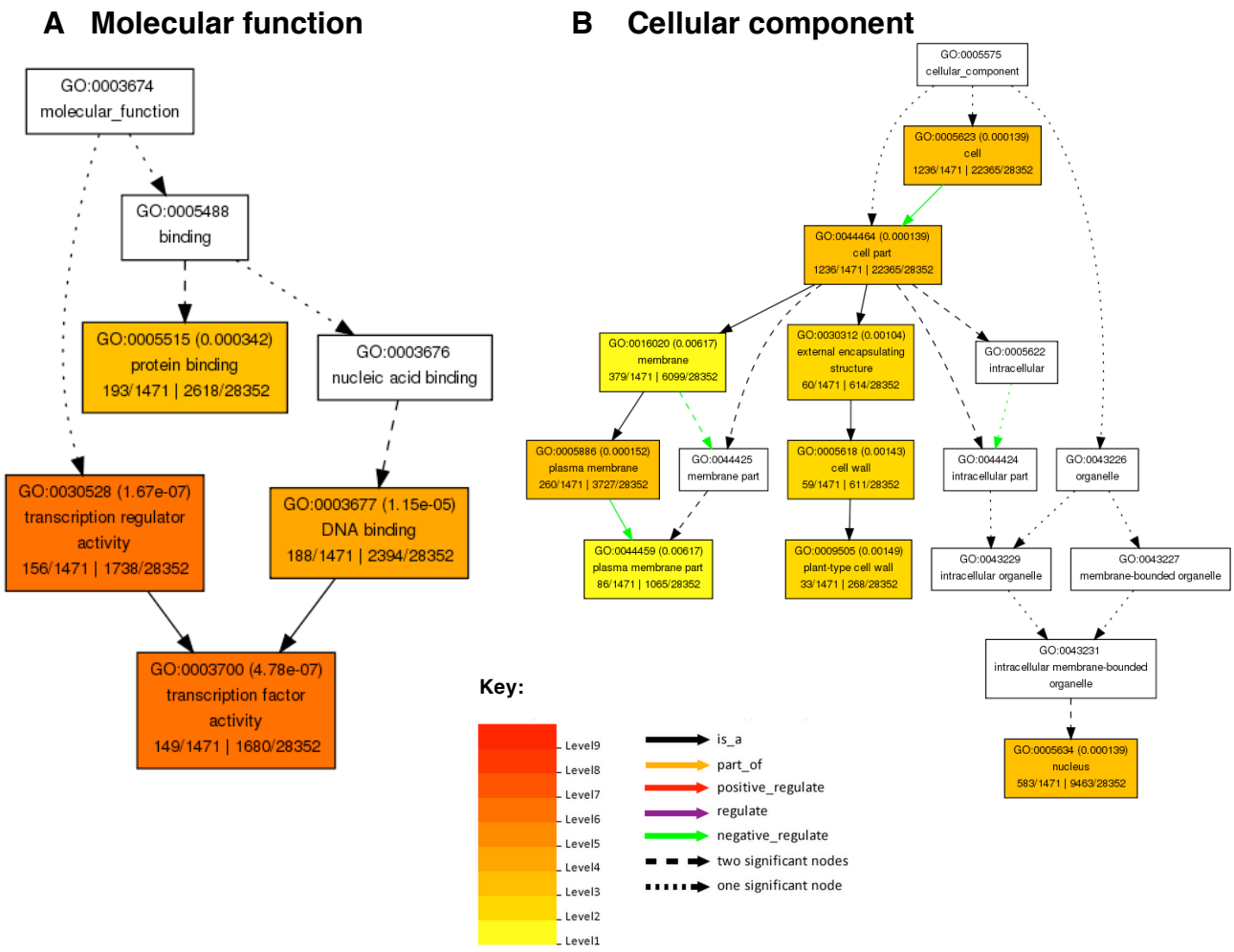


Fig. S5. Additional gene ontology (GO) categories enriched among high-confidence SPCH target genes.

(**A** & **B**) Significantly enriched GO terms (FDR < 0.05) among the 1,517 high-confidence SPCH target genes by Molecular Function (A) or Cellular Component (B). Darker colors represent higher significance levels. Analysis was performed with AgriGO (*44*).

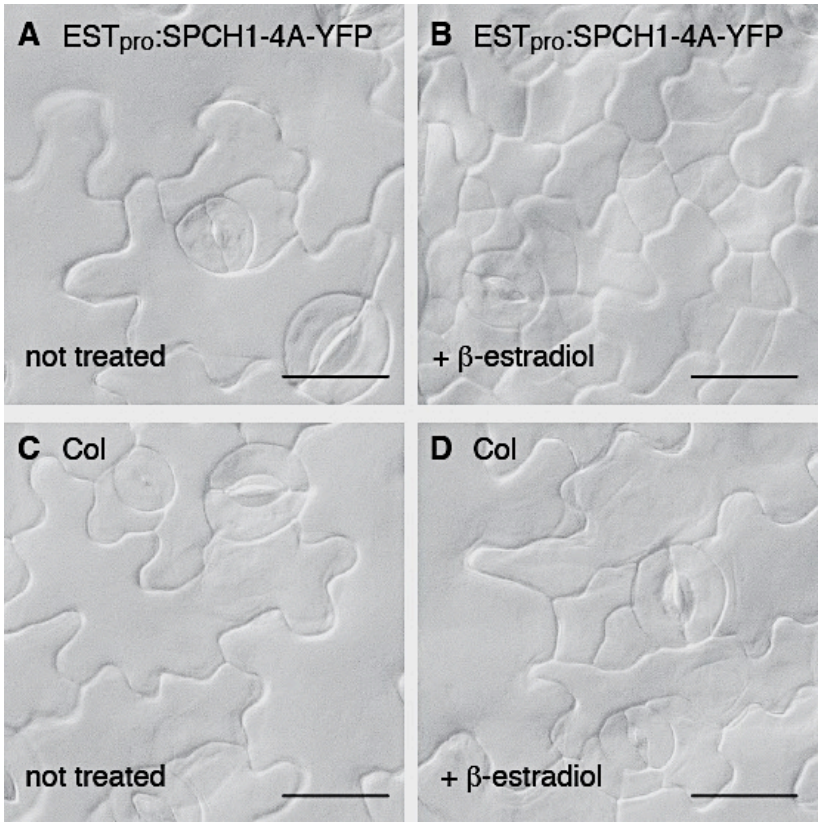
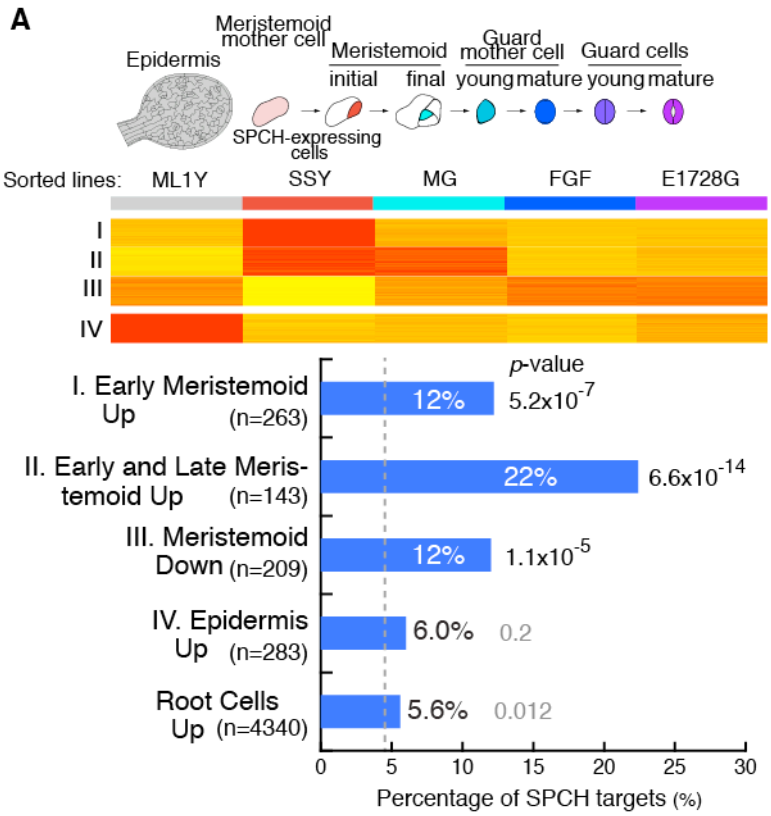


Fig. S6. Functional assessment of the inducible SPCH expression system.

(**A** to **D**) Four-day-old seedlings carrying the estrogen-inducible construct, ESTpro:SPCH1-4A-YFP (A and B), and Col (C and D) were incubated with (B and D) or without (A and C) β-estradiol. Images from abaxial cotyledons of seedlings 3 days post-treatment. Note the extra divisions on the epidermis of ESTpro:SPCH1-4A-YFP (B) induced by β-estradiol incubation. Scale bar, 25 μm.



B

AFFY_ID	AGI	Gene symbol	ML1Y_ median	SSY_ median	MG_ median	FGF_ median	E1728G_ median
Select genes showing cell-type specific expression in the lineage							
248247_at	AT5G53210	SPCH	2.3766	9.1470	4.0043	2.2968	2.4437
262040_at	AT1G80080	TMM	1.7999	5.6276	4.8452	2.3196	2.4784
247599_at	AT5G60880	BASL	2.1356	5.4508	3.1781	2.1056	2.3052
256395_at	AT3G06120	MUTE	1.5436	1.6883	2.4100	1.5045	1.6569
257247_at	AT3G24140	FAMA	3.1504	2.3239	6.5029	10.0803	7.2886
260633_at	AT1G62400	HT1	3.0215	2.7005	3.0916	5.9003	5.9993

Fig. S7. Percentage of SPCH targets among differential expressed genes in meristemoid-enriched cell populations.

(**A**) Comparison between high-confidence SPCH targets and differential expressed genes in microarray (ATH1) analysis of FACS-sorted stomatal lineage cells. Significant enrichment of the SPCH targets was found among genes differentially expressed in the early stomatal lineage (I, II and III), but not those in epidermal (IV) or root cells (*49*). Grey dash line indicates the predicted percentage by chance (1517 targets/33602 Arabidopsis genes). *P*-values are calculated by Fisher's exact test. (**B**) Expression values on the Affymetrix ATH1 array of select stomatal lineage genes, demonstrating the validity of the expression dataset.

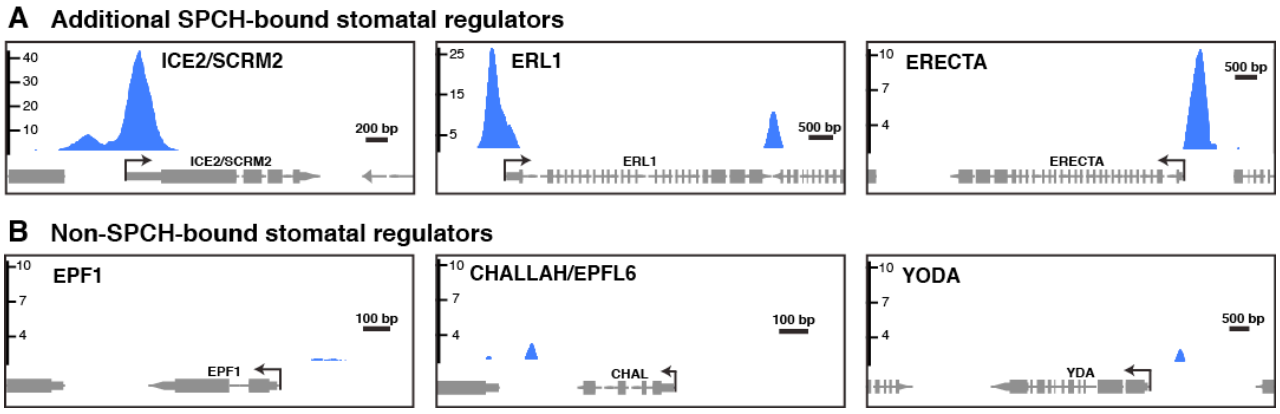


Fig. S8. SPCH ChIP-Seq profiles at additional stomatal regulators.

SPCH binds to *INDUCER OF CBF EXPRESSION 2/SCREAM2 (ICE2/SCRM2)*, *ERECTA-LIKE 1 (ERL1)* and *ERECTA* (A) (refer to main text for details), but not to *EPIDERMAL PATTERNING FACTOR 1 (EPF1)*, *CHALLAH* and *YODA* (B). EPF1 and CHALLAH, like EPF2, are secreted peptide ligands of the ERECTA family of receptor-like kinases (*4*). However, *EPF1* is expressed at a later stage in the stomatal lineage, while *CHALLAH* is expressed predominantly in internal tissues of hypocotyls and stems (*50*, *51*). *YODA*, which encodes a mitogen-activated triple kinase involved in the post-translational repression of SPCH, is expressed at a low level throughout the plants (*30*, *52*). The y-axis represents the enrichment score computed by CSAR (note different scales in ICE2/SCRM2 and ERL1 relative to others) and arrows indicate transcriptional start sites and orientation of genes.

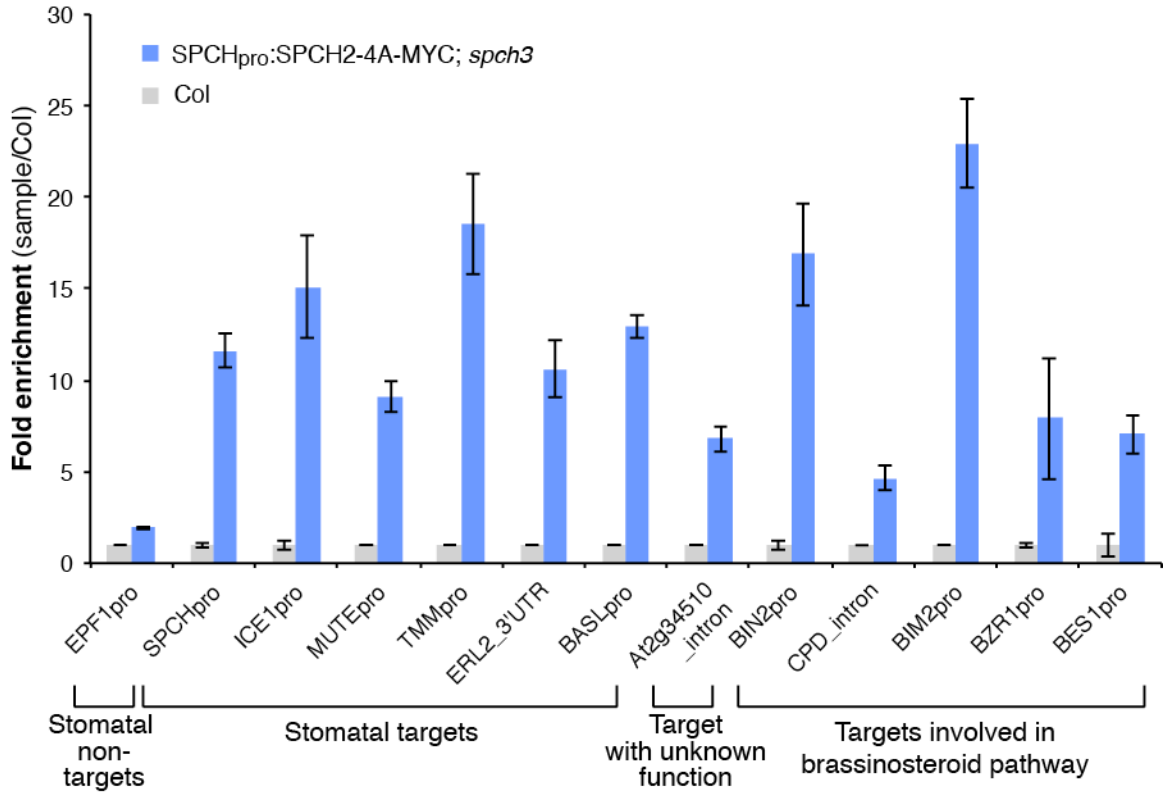
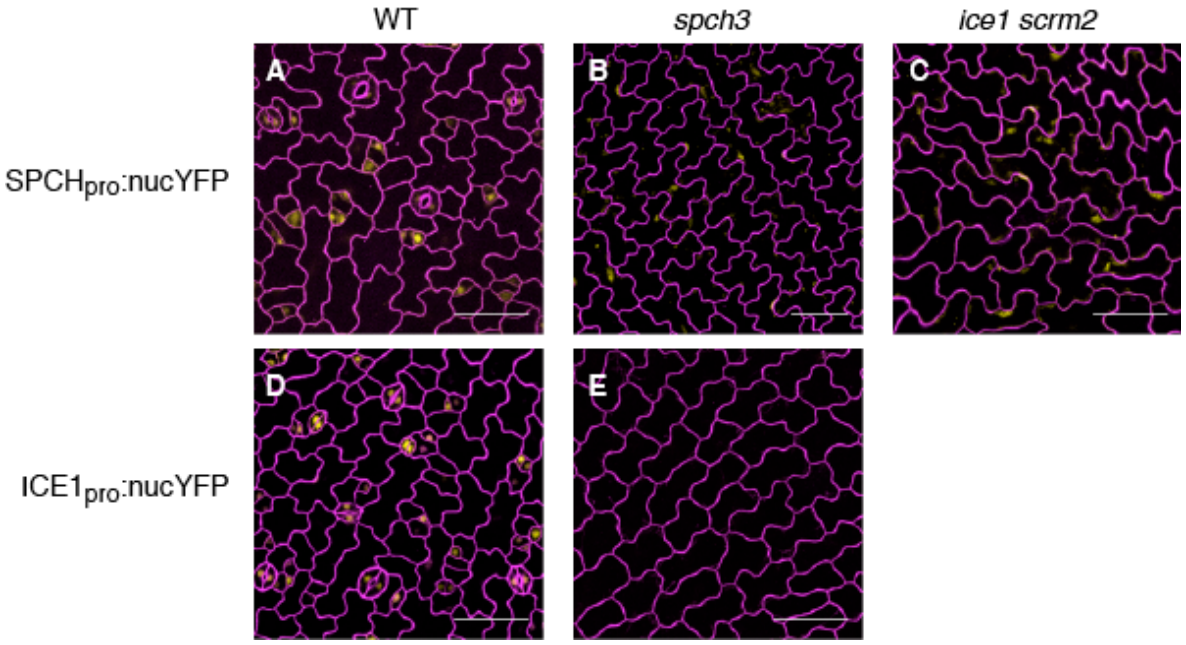


Fig. S9. ChIP-qPCR validation of select SPCH ChIP-Seq targets.

ChIP-qPCR assays were performed on SPCHpro:SPCH2-4A-MYC; *spch3* and Col at 4x scale (see main text) using the 71D10 anti-Myc antibody (Cell Signaling Technology). Twelve select SPCH-binding targets were tested for enrichment using primers specific to the indicated regions of the gene. The tested regions were within the width of the SPCH-binding peak of the targets. *EPF1*, which has no significant enrichment in our SPCH ChIP-Seq data, was used as the negative control. List of primers are shown in Table S8. Fold enrichment of an amplified regions was calculated by dividing the input-adjusted signals of the Myc-tagged line over that of Col. Values are mean +/- SEM. Key: pro, promoter.



F

Other published observations about SPCH targets and regulation	Reference(s)
SPCH transcript could be detected in *spch-1* (nonsense null mutation) by RT-PCR	(*7*)
Transcriptional reporter of *ICE1* was not expressed in cotyledon epidermis of *spch-3 (T-DNA null).* No *SPCH* transcripts were detected in *ice1 scrm2* by RT-PCR	(*14*)
Transcriptional reporters of *EPF2* and *TMM*, which are targets of SPCH, were not expressed in *spch-3*	(*53*)
Overexpression of SPCH led to extra cell divisions	(*7*, *8*)

Fig. S10. Initial transcriptional activation of *SPCH* is likely independent of the presence of a functional SPCH-ICE heterodimer.

(**A** to **C**) Weak but detectable expression of a *SPCH* transcriptional reporter (SPCHpro:nucYFP) in 4-day old abaxial cotyledons of *spch* (B) and *ice1 scrm2* (C) mutants, compared to WT (Col; A). Larger spindle-shaped yellow structures in B and C are nuclei expressing SPCH, but note the background signal from chloroplasts (smaller yellow structures) that results from the higher laser power and gain required to visualize SPCH expression. (**D** and **E**) Transcription of ICE1pro:nucYFP *ICE1* in the epidermis of 3-day old cotyledons depends on *SPCH*. Cell outlines (purple) are visualized with propidium iodide. Scale bar, 50 μm. (**F**) Table summarizing relevant published data on self-regulation of SPCH.

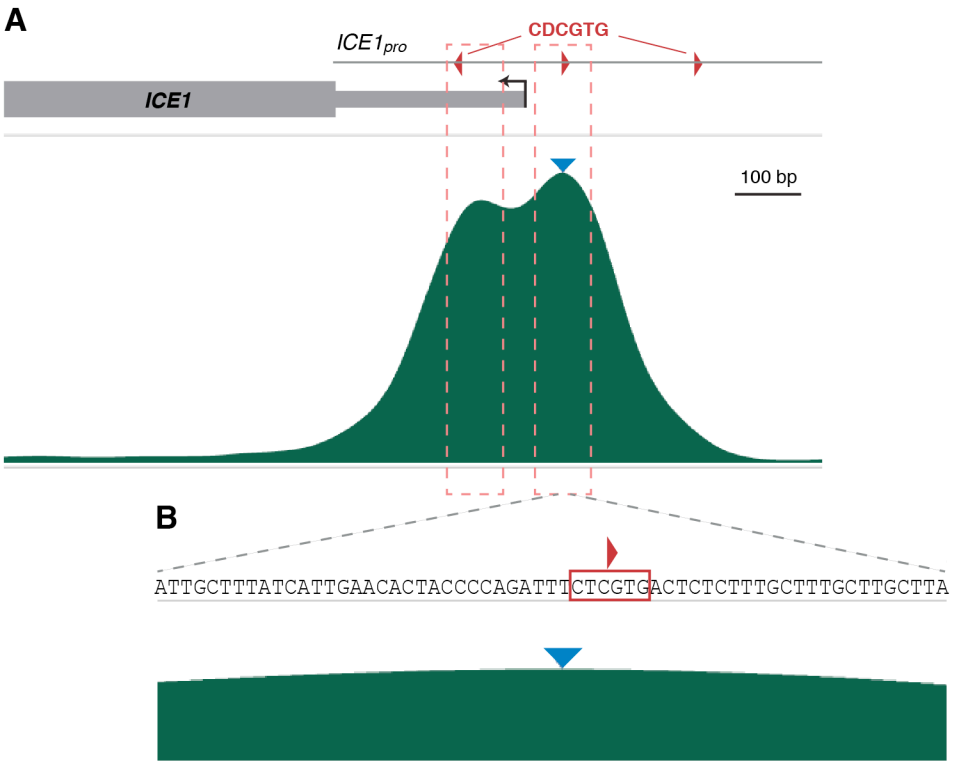


Fig. S11. SPCH binds to the region of the *ICE1* promoter that contains the SPCH-binding motifs (CDCGTG).

(**A**) SPCH ChIP-Seq profile at the promoter region of *ICE1*. The single promoter peak at *ICE1* (Fig. 2F) can be resolved into two at higher resolution (marked by dotted red boxes) (analysis by DNAnexus). The promoter region of *ICE1* contains three SPCH-binding motifs, CDCGTG, (red arrowheads) and two of the motifs lie within the two SPCH-binding peaks. The summit of the major SPCH-binding peak is marked by a blue arrowhead. (**B**) Nucleotide sequence view at the summit of the major SPCH-binding peak of *ICE1*. A SPCH-binding motif, CTCGTG, (marked by a red box) is 1 bp from the summit of the peak (blue arrowhead).

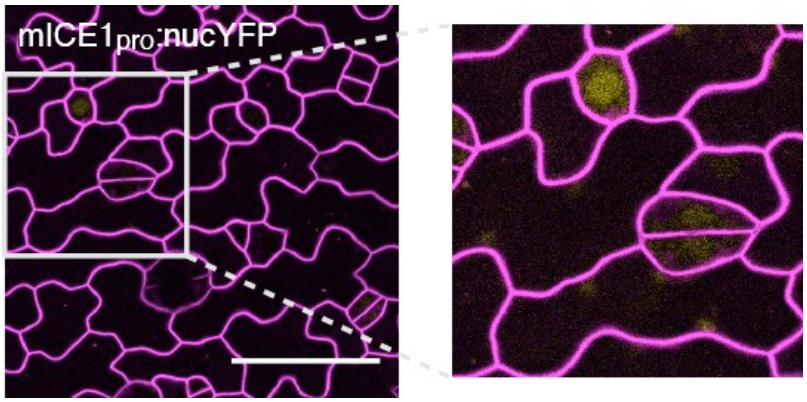


Fig. S12. Detection of expression of the mutated *ICE1* promoter reporter at higher gain.

Confocal images of mICE1pro:nucYFP as shown in Fig. 2H (left) and an indicated area of the same image with higher gain (right). Note the weak but still detectable expression of the mutated promoter reporter in stomatal lineage cells. Cell outlines (purple) were visualized with ML1pro:mCherry-RCI2A. Scale bar, 40 μm.

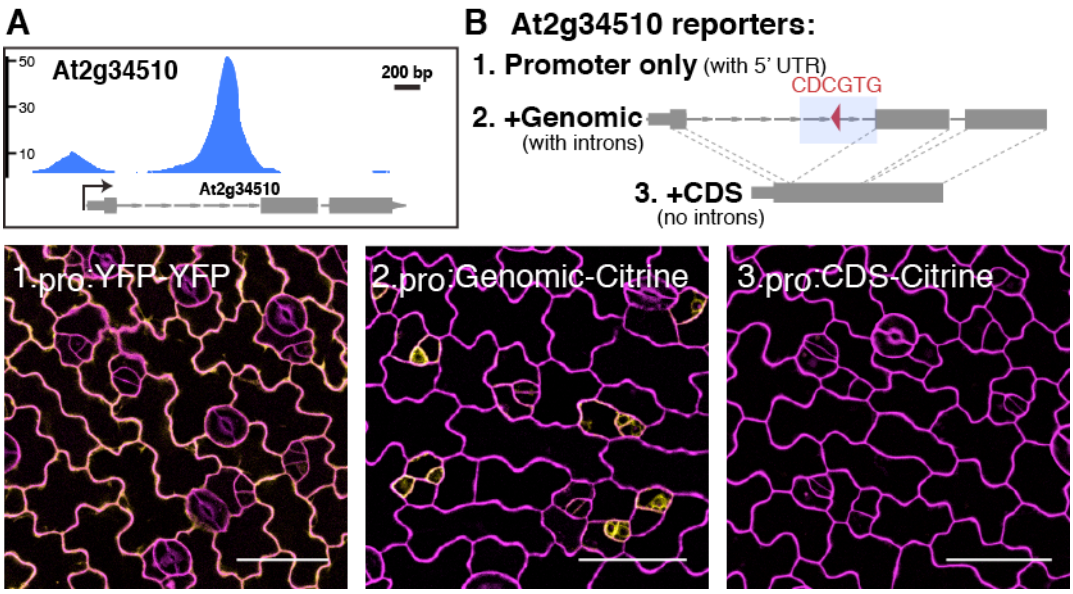


Fig. S13. A SPCH-binding intron confers meristemoid expression.

(**A**) SPCH ChIP-Seq profile of a gene of unknown biochemical function, At2g34510. Note the major SPCH-binding peak in its first intron. (**B**) Expression analysis of At2g34510 reporters. Only the intron-containing genomic fragment reporter (#2), but not the promoter alone (#1) or CDS (#3) reporters, displayed meristemoid-specific expression. Confocal images are of 4 dpg abaxial cotyledons. Genes of interest are in yellow and cell outlines (purple) were visualized with propidium iodide. Scale bar, 40 μm.

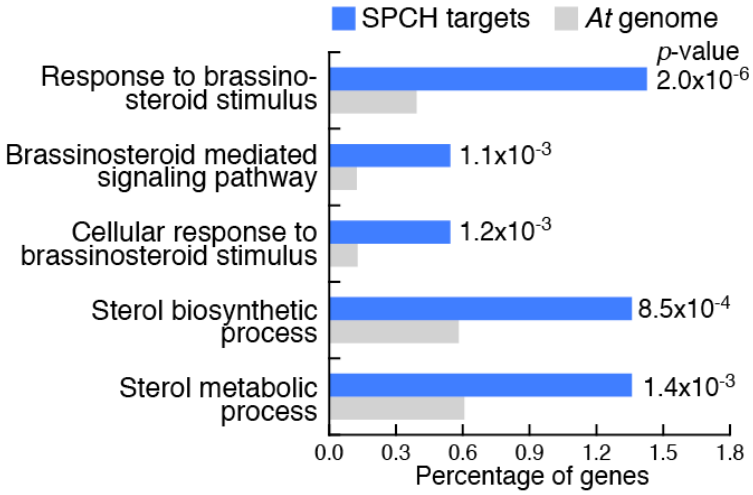


Fig. S14. Enrichment of Gene Ontology (GO) terms related to brassinosteroid (BR) biosynthesis and signaling in high-confidence SPCH target genes.

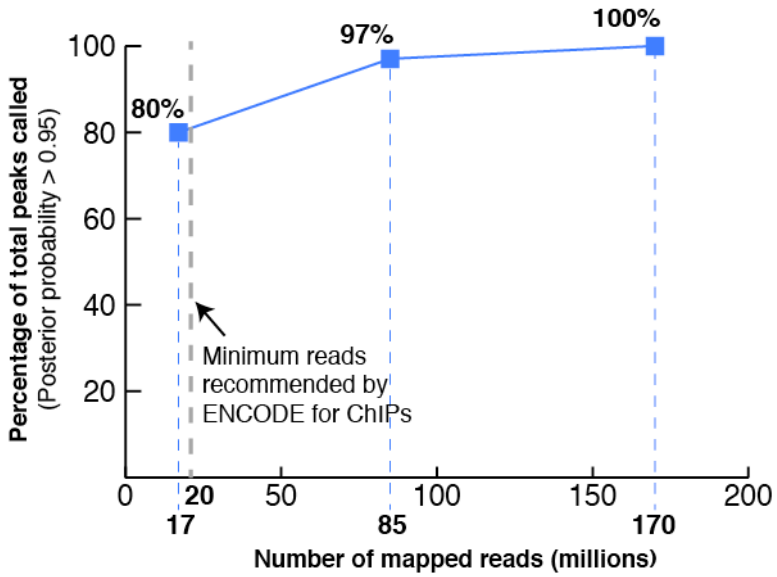


Fig. S15. SPCH-binding peak discovery by ChIP-Seq increases with sequencing depth. One of the advantages of the HiSeq sequencing platform is enhanced sequencing depth relative to the older GAII platform (>150 million reads per lane in a HiSeq's 50 bp run vs. ~20 million reads per lane in a GAII's 36 bp run). To estimate the benefit of greater coverage on peak discovery, random samples of 10% and 50% of the 170 million mapped reads produced by HiSeq were used to call SPCH-enriched regions via BayesPeak (posterior probability > 95%). Using 10% of the reads (17 million) mimics the output of GAII, and identifies 20% fewer peaks than the use of all HiSeq reads. Each data point represents the average of 10 random samples. The grey dashed line marks the minimum reads (20 million) recommended by the ENCODE consortium for ChIP-Seq experiments (*54*).

Table S1. False discovery rate (FDR) threshold used in select ChIP-Seq studies on plant developmental regulators

To assess the significance of the FDR threshold in our dataset, we chose three recent ChIP-Seq studies that employ the same peak caller, CSAR (*9*), for cross comparison. Among all these datasets, CSAR was able to detect the lowest peak score at the highest significance in our SPCH dataset, suggesting low background noise.

Gene name	FDR threshold	Corresponding peak score	Reference
APETALA1	< 0.001	79.6	(*11*)
AGAMOUS	< 0.001	38.14	(*55*)
JAGGED	< 0.01	1.85	(*10*)
SPEECHLESS	$< 1x10^{-6}$	1.62*	This study

*: Peak score ≥3 and ≥10 were used as a cut-off for further analysis (see table S2).

Table S2. Summary of SPCH ChIP-Seq results

Genotype	Total reads (millions)	Mapped reads (millions)		Genome-wide peaks (FDR <1x10^{-6} & posterior probability >0.95)		Associated genes	
		Bowtie2	**SOAP2**	**Score ≥3**	**Score ≥10**	**Score ≥3**	**Score ≥10**
SPCHpro:SPCH 2-4A-MYC; *spch3*	177	170 (95%)	166 (93%)	8,327	1,217*	8,958	1,517*
Col	172	132 (77%)	130 (75%)	-	-	-	-

*: The non-intergenic peaks with peak score ≥10 and their associated genes are defined as high-confidence SPCH-binding peaks and targets, respectively. The peak score cutoff was guided by known stomatal regulators.

Table S6. Overlaps between high-confidence SPCH targets and expression datasets

Inducible SPCH dataset		FACS-isolated stomatal lineage expression map			Stomatal mutant dataset (*13*)	
Overlap with up-regulated genes (6 or 8h post-induction)	**Overlap with down-regulated genes (6 or 8h post-induction)**	**Overlap with up-regulated genes in early meristemoid**	**Overlap with up-regulated genes in early and late meristemoid**	**Overlap with down-regulated genes in early meristemoid**	**Overlap with up-regulated genes in *scrmD mute***	**Overlap with up-regulated genes in *spch***
AT1G08200	AT1G01430	AT1G30690	AT1G15340	AT1G01480	AT1G01300	AT1G11260
AT1G12420	AT1G04240	AT1G33930	AT1G15570	AT1G03740	AT1G03590	AT1G22400
AT1G12845	AT1G05240	AT1G34245	AT1G18740	AT1G19180	AT1G03610	AT1G25400
AT1G12860	AT1G13245	AT2G21300	AT1G32640	AT1G19310	AT1G05630	AT1G25560
AT1G13360	AT1G15260	AT2G28070	AT1G64670	AT1G32230	AT1G07570	AT1G32920
AT1G13930	AT1G20650	AT2G28950	AT1G68400	AT1G68410	AT1G07630	AT1G54100
AT1G14440	AT1G21060	AT2G34510	AT1G69160	AT1G76180	AT1G08920	AT1G70090
AT1G16370	AT1G23480	AT2G41550	AT1G72970	AT2G28840	AT1G10850	AT1G70100
AT1G19570	AT1G30360	AT2G46710	AT1G80080	AT2G37040	AT1G11330	AT1G71030
AT1G26600	AT1G33240	AT2G46780	AT2G14890	AT2G38000	AT1G11850	AT1G72680
AT1G31935	AT1G49730	AT3G02120	AT2G39460	AT2G41420	AT1G12330	AT1G75750
AT1G34245	AT1G53830	AT3G04290	AT2G45590	AT2G46260	AT1G12430	AT1G76570
AT1G55450	AT1G67900	AT3G14000	AT3G06840	AT3G05220	AT1G12845	AT1G76590
AT1G55740	AT1G68530	AT3G15095	AT3G08670	AT3G07390	AT1G12860	AT1G80380
AT1G66160	AT1G69570	AT3G49900	AT3G12200	AT3G13520	AT1G13950	AT1G80440
AT1G80080	AT1G70940	AT3G53380	AT3G14760	AT3G51890	AT1G14440	AT2G16660
AT1G80640	AT1G71030	AT3G55280	AT3G18490	AT4G16520	AT1G15570	AT2G20670
AT2G18690	AT1G72180	AT3G61250	AT3G23730	AT4G18950	AT1G17200	AT2G30600
AT2G21660	AT1G72600	AT4G00480	AT3G46620	AT4G33030	AT1G18670	AT2G32150
AT2G25460	AT1G72610	AT4G14490	AT3G49220	AT4G36780	AT1G19330	AT2G40000
AT2G26530	AT2G16280	AT4G22560	AT3G49260	AT4G37980	AT1G21060	AT2G44080
AT2G27385	AT2G28470	AT4G22570	AT4G03210	AT5G01750	AT1G21090	AT2G46830
AT2G34510	AT2G28950	AT4G24570	AT4G11280	AT5G04220	AT1G21100	AT3G06500
AT2G39650	AT2G34620	AT4G28230	AT4G17245	AT5G05370	AT1G23710	AT3G07350
AT2G39870	AT2G35859	AT4G28250	AT4G25470	AT5G24470	AT1G26600	AT3G13450
AT2G40670	AT2G40000	AT5G01090	AT4G25620		AT1G28290	AT3G13750
AT2G43550	AT2G43010	AT5G05210	AT4G27280		AT1G29395	AT3G15620
AT2G44080	AT2G44130	AT5G41140	AT4G29020		AT1G32690	AT3G17510
AT2G44490	AT2G46830	AT5G47500	AT5G03040		AT1G33930	AT3G30775
AT3G09730	AT3G12320	AT5G50740	AT5G45800		AT1G34245	AT3G47340
AT3G14560	AT3G13310	AT5G53210	AT5G51750		AT1G44760	AT3G48360
AT3G14760	AT3G13750	AT5G60880	AT5G60210		AT1G55740	AT3G62860
AT3G22104	AT3G16370				AT1G61900	AT4G17245
AT3G22760	AT3G23030				AT1G64670	AT4G27450

AT3G26744
AT3G47295
AT3G47800
AT3G48580
AT3G52450
AT3G54400
AT3G54740
AT3G59900
AT3G62070
AT3G63450
AT4G01060
AT4G14040
AT4G14770
AT4G18360
AT4G18950
AT4G18970
AT4G23190
AT4G23920
AT4G25260
AT4G25620
AT4G25630
AT4G29020
AT4G31805
AT4G35930
AT4G37070
AT4G37310
AT4G37980
AT5G07180
AT5G07280
AT5G11550
AT5G13220
AT5G16350
AT5G19240
AT5G23860
AT5G53210
AT5G55120
AT5G55920
AT5G60880
AT5G62230
AT5G67260
AT3G45780
AT3G51950
AT3G53420
AT3G53720
AT3G57400
AT3G59060
AT3G59930
AT4G04330
AT4G18220
AT4G26700
AT4G34250
AT4G36220
AT4G38690
AT5G03120
AT5G04530
AT5G06980
AT5G08330
AT5G15310
AT5G19190
AT5G44680
AT5G45950
AT5G52900
AT5G54630
AT5G58140
AT5G61270
AT5G66580
AT1G66160
AT1G66400
AT1G68400
AT1G69160
AT1G72670
AT1G72970
AT1G74950
AT1G74960
AT1G75580
AT1G75640
AT1G76550
AT1G76560
AT1G76600
AT1G78260
AT1G78410
AT1G79630
AT1G80080
AT1G80290
AT2G02450
AT2G17470
AT2G22900
AT2G24580
AT2G26560
AT2G26700
AT2G34510
AT2G36040
AT2G36470
AT2G40670
AT2G41090
AT2G41820
AT2G42870
AT2G45400
AT2G45970
AT2G46420
AT2G46710
AT2G48030
AT3G02120
AT3G06120
AT3G06840
AT3G07390
AT3G09730
AT3G10570
AT3G12200
AT4G27657
AT4G35770
AT5G05410
AT5G05440
AT5G15410
AT5G19120
AT5G20250
AT5G43450
AT5G47610
AT5G52900
AT5G57560
AT5G59820

AT3G14000
AT3G14560
AT3G14760
AT3G15790
AT3G17640
AT3G18490
AT3G19540
AT3G19590
AT3G19850
AT3G21700
AT3G22104
AT3G22760
AT3G23730
AT3G24630
AT3G24670
AT3G26744
AT3G27060
AT3G44610
AT3G46110
AT3G47295
AT3G47800
AT3G49260
AT3G49900
AT3G50620
AT3G52450
AT3G54820
AT3G57070
AT3G60520
AT4G00480
AT4G14440
AT4G14770
AT4G18950
AT4G18970
AT4G23190
AT4G23920
AT4G25620
AT4G28230
AT4G28250
AT4G28260
AT4G29020
AT4G30230
AT4G31805
AT4G35930

AT4G37330
AT4G37740
AT4G37980
AT4G39390
AT5G01090
AT5G01990
AT5G02020
AT5G02260
AT5G04220
AT5G04770
AT5G04820
AT5G05190
AT5G07180
AT5G07280
AT5G07300
AT5G11550
AT5G13220
AT5G19730
AT5G20950
AT5G22940
AT5G23260
AT5G25210
AT5G46880
AT5G47050
AT5G47060
AT5G47500
AT5G47600
AT5G49170
AT5G49760
AT5G50740
AT5G51750
AT5G52120
AT5G52250
AT5G53210
AT5G54280
AT5G55120
AT5G60210
AT5G60880
AT5G62230
AT5G63810
AT5G66080
AT5G66650
AT5G67260

Table S7. List of SPCH target genes discussed in this study

Gene name	AGI code	Past report(s) on roles in stomatal development	Other refere-nce(s)
TOO MANY MOUTH (TMM)	AT1G80080	(*15*)	-
INDUCER OF CBF EXPRESSION 1 (ICE1)	AT3G26744	(*14*)	(*56*)
SCREAM 2 (SCRM2)	AT1G12860	(*14*)	-
MUTE	AT3G06120	(*7*, *8*)	-
ERECTA (ER)	AT2G26330	(*16*)	-
ERECTA-LIKE 1 (ERL1)	AT5G62230	(*16*)	-
ERECTA-LIKE 2 (ERL2)	AT5G07180	(*16*)	-
EPIDERMAL PATTERNING FACTOR 2 (EPF2)	AT1G34245	(*17*, *53*)	-
BREAKING OF ASYMMETRY IN THE STOMATAL LINEAGE (BASL)	AT5G60880	(*12*)	-
POLAR LOCALIZATION DURING ASYMMETRIC DIVISION AND REDISTRIBUTION (POLAR)	AT4G31805	Expression data (*13*)	-
POLAR-LIKE	AT5G10890	Expression data (*13*)	-
Unknown gene	AT2G34510	-	-
ARMADILLO REPEAT KINESIN 3 (ARK3)	AT1G12430	Expression data (*18*)	(*57*)
BRASSINOSTEROID-INSENSITIVE 2 (BIN2)	AT4G18710	(*21*, *22*, *58*)	(*59*)
CONSTITUTIVE PHOTOMORPHOGENIC DWARF (CPD)	AT5G05690	(*22*, *24*)	-
BES1-INTERACTING MYC-LIKE PROTEIN 2 (BIM2)	AT1G69010	-	(*27*)
BRASSINAZOLE-RESISTANT 1 (BZR1)	AT1G75080	-	(*25*, *60*, *61*)
BRI1-EMS-SUPPRESSOR 1 (BES1)	AT1G19350	-	(*26*, *62*)
BES1/BZR1 HOMOLOG 1 (BEH1)	AT3G50750	-	(*27*)
BES1/BZR1 HOMOLOG 2 (BEH2)	AT4G36780	-	(*27*)
BES1/BZR1 HOMOLOG 3 (BEH3)	AT4G18890	-	(*27*)
BES1/BZR1 HOMOLOG 4 (BEH4)	AT1G78700	-	(*27*)

Table S8. List of primers for ChIP-qPCR and RT-qPCR

Primers for ChIP-qPCR	(all written 5' to 3')			
Name in this study	**AGI code**	**Forward primer**	**Reverse primer**	**Amplicon size (bp)**
At2g34510 _intron	AT2G34510	TTTCGCCTTCACTACTTTTGC	TCACATTTATCAGCCACTACGG	106
BASLpro	AT5G60880	AATGCATTCATGTGTTGCAC	TGAGGGAGAAGAATGATTTGAA	105
BES1pro	AT1G19350	GGCCCCACATGTCTCTCTAT	GGTTTCGCGTTTTGAAGAAG	112
BIM2pro	AT1G69010	CGTGTCATCACAAGCTACGG	CTTGAGCTTTTGTCCGTCAGT	100
BIN2pro	AT4G18710	CACAAAGAGTCCGTGTCTGC	TCTTCGCCATCTTCTTTCGT	98
BZR1pro	AT1G75080	TGAACTCAATGTACTTGAACACACA	AAGAATCTCTCTTTCTTCTCCAACA	92
CPD_intron	AT5G05690	GGAAGAAGTAGTTGGGGGAAA	AGGCCAACATGCACATATCA	99
EPF1pro	AT2G20875	CCCAAGACAATGTCCCTCAT	GACCTTCATTAATGCACCCAAT	88
ERL2_3'UTR	AT5G07180	TGGGATAGGTGAAAGAGTAGGTG	GGATGTTTGCATTAAAGAGGAGA	100
ICE1pro	AT3G26744	TTCAATTCATGAAATCCTCGTG	GGAAGCTTGAAGACAATTCTGC	125
MUTEpro	AT3G06120	TTCTTTACGTTGTTCTAATACATATCG	TCGCTGTCTCGTGACATTCT	90
SPCHpro	AT5G53210	TTTGGCGTCTCTTTCTTTCC	TCTTCCTCGTTCGTTGTTTTG	91
TMMpro	AT1G80080	TCTAGGGCCCAACACAAGAC	TCACGGGATAACTCGAGGAG	108

Primers for RT-qPCR	(all written 5' to 3')			
Gene name	**AGI code**	**Forward primer**	**Reverse primer**	**Amplicon size (bp)**
ACTIN2	AT3G18780	TCTTCCGCTCTTTCTTTCCAAGC	ACCATTGTCACACACGATTGGTTG	77
BEH1	AT3G50750	TTTGTCTTGAAGCTGGTTGGATCG	TTCTGTTGGTCGAGAACCCTTTC	71
BEH2	AT4G36780	ATGGCACCACTTATCGCAAGGG	TGGTACGAAGGTGCAGGACTTG	136
BEH3	AT4G18890	TGCAATGAAGCTGGTTGGACTG	TCCATTGGTTTGCATCCCTTGC	68
BEH4	AT1G78700	TGATGGAACTACTTACCGCAAGGG	AGCACAGGGACTTGGCTGATAG	112
BES1	AT1G19350	TTTGTTCTGAAGCTGGTTGGGTTG	AGGTAGAGGCTTGTGTCCCTTG	71
BIM2	AT1G69010	AGTGCCTACTCGCATGAGTTATTG	TTGCTTGCGACAGATCAATGCC	76
BIN2	AT4G18710	TGCACCCGAGCTCATATTTGGTG	TCTCCGGGAAATAATGGCTGACC	111
BZR1	AT1G75080	TTGTGTTGAAGCTGGTTGGGTTG	GTAAAGGCTTGCATCCCTTGCG	68
CPD	AT5G05690	TTACCGCAAAGCCATCCAAGCG	TTTCATCACCACCACCGTCAACGC	64