

Cis and trans determinants of epigenetic silencing by Polycomb repressive complex 2 in *Arabidopsis*

Jun Xiao¹ , Run Jin^{1,12}, Xiang Yu^{1,12} , Max Shen¹, John D Wagner¹, Armaan Pai¹, Claire Song¹, Michael Zhuang¹, Samantha Klasfeld¹, Chongsheng He², Alexandre M Santos³, Chris Helliwell⁴ , Jose L Pruneda-Paz⁵, Steve A Kay⁶, Xiaowei Lin⁷, Sujuan Cui⁷, Meilin Fernandez Garcia⁸, Oliver Clarenz⁹, Justin Goodrich⁹, Xiaoyu Zhang³, Ryan S Austin^{10,11}, Roberto Bonasio²  & Doris Wagner¹ 

Disruption of gene silencing by Polycomb protein complexes leads to homeotic transformations and altered developmental-phase identity in plants^{1–5}. Here we define short genomic fragments, known as Polycomb response elements (PREs), that direct Polycomb repressive complex 2 (PRC2) placement at developmental genes regulated by silencing in *Arabidopsis thaliana*. We identify transcription factor families that bind to these PREs, colocalize with PRC2 on chromatin, physically interact with and recruit PRC2, and are required for PRC2-mediated gene silencing *in vivo*. Two of the *cis* sequence motifs enriched in the PREs are cognate binding sites for the identified transcription factors and are necessary and sufficient for PRE activity. Thus PRC2 recruitment in *Arabidopsis* relies in large part on binding of *trans*-acting factors to *cis*-localized DNA sequence motifs.

In both the plant and animal kingdoms, Polycomb-mediated gene repression is important for cell identity^{2,5–8}. The evolutionarily conserved PRC2 induces trimethylation of histone H3 at Lys27 (H3K27me3), an epigenetic mark that results in compaction of chromatin and silencing of gene expression at thousands of loci^{1–5,8,9}. After its establishment, the repressed chromatin state is mitotically heritable^{8,10}. Given that PRC2 has no inherent DNA binding specificity, a key question is how the Polycomb epigenetic machinery targets the loci it silences. In *Drosophila melanogaster*, multiple transcription factors (TFs) bind to *cis*-regulatory regions several hundreds of base pairs in length called PREs and recruit Polycomb complexes^{8,9}. Although a few such PREs were initially identified in mammals, recent studies instead implicate promoter-proximal unmethylated CpG islands in PRC recruitment⁷. Likewise, PREs with inherent silencing ability have been identified at a handful of loci in *Arabidopsis*^{10–14}, but it is unclear whether this mechanism broadly underpins PRC2 recruitment.

To elucidate the PRC2 targeting mechanism in *Arabidopsis*, we identified 132 high-confidence PRC2-regulated genes from our own and public genomic data sets^{15–17} (**Supplementary Fig. 1a** and **Supplementary Tables 1** and **2**) and computationally defined 170 candidate PREs (600 bp in length) associated with them. We selected five PREs from three loci—the PRC2 targets *AGAMOUS* (*AG*) and *SEPALLATA3* (*SEP3*)^{2,18} and a gene of unknown function, *At5g61120* (*At5g*)—to test their ability to recruit PRC2 and direct *de novo* H3K27me3 when randomly integrated into the genome. All five candidate plant PREs recruited PRC2 (represented by complex components FIE, EMF2 and MSI1; **Supplementary Fig. 1b**) and gained H3K27me3, as did a previously characterized PRE used as a positive control (PC_LEC2)¹² (**Fig. 1a,b** and **Supplementary Fig. 2a,b**; $n = 40$ transgenic lines). H3K27me3 is known to spread from the site of PRC2 recruitment to adjacent genomic locations^{8,9}; this was also observed at the PREs (**Supplementary Fig. 2a,b**). Random unlinked (NC_1 and NC_3) and linked (NC_2, linked with the *AG* locus) DNA fragments did not recruit PRC2 and/or gain H3K27me3.

In *Drosophila*, PREs not only recruit Polycomb complexes and become decorated with H3K27me3, but also repress linked genes in a Polycomb-dependent manner^{8,9}. Likewise, when placed between two constitutive promoters, the five candidate PREs significantly ($P < 0.05$, Mann–Whitney *U*-test) silenced three independent reporter genes (encoding GFP fluorescence, β -glucuronidase activity and herbicide resistance, respectively), as did PC_LEC2 (**Fig. 1c,d**, $n = 15$ independent transformants; **Supplementary Fig. 2c,d**, $n = 15$ independent transformants; and **Supplementary Fig. 2e**, $n = 60$ independent transformants). None of the negative-control DNA fragments had this effect. The ability of the PREs to silence active reporters was dependent on PRC2 in all cases (**Fig. 1d** and **Supplementary Fig. 2d,e**).

To determine which sequence-specific binding proteins associate with the five functional PREs, we performed high-throughput DNA

¹Department of Biology, University of Pennsylvania, Philadelphia, Pennsylvania, USA. ²Department of Cell and Developmental Biology, University of Pennsylvania, Philadelphia, Pennsylvania, USA. ³Department of Plant Biology, University of Georgia, Athens, Georgia, USA. ⁴CSIRO Agriculture and Food, Canberra, Australia. ⁵Division of Biology, University of California, San Diego, La Jolla, California, USA. ⁶Department of Neurology, Keck School of Medicine, University of Southern California, Los Angeles, California, USA. ⁷College of Life Science, Hebei Normal University, Hebei, People's Republic of China. ⁸Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, Pennsylvania, USA. ⁹Institute of Molecular Plant Science, School of Biological Sciences, University of Edinburgh, Edinburgh, UK. ¹⁰Agriculture & Agri-Foods Canada, London, Ontario, Canada. ¹¹Department of Biology, Western University, London, Ontario, Canada. ¹²These authors contributed equally to this work. Correspondence should be addressed to D.W. (wagnerdo@sas.upenn.edu).

Received 8 March; accepted 21 July; published online 21 August 2017; doi:10.1038/ng.3937

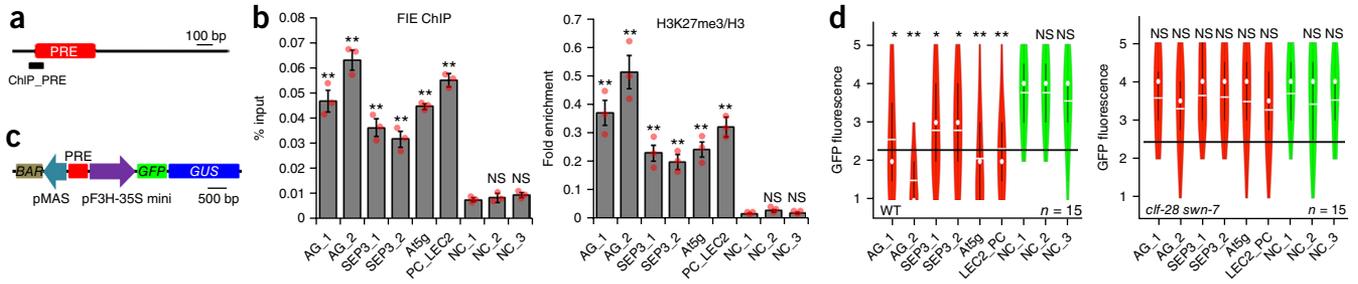


Figure 1 Identification of *Arabidopsis* DNA fragments with PRE activity. (a) A schematic of the construct used to test the ability of candidate PREs or negative control DNA fragments (NC) to recruit PRC2 and H3K27me3. The thick black line indicates the region tested by chromatin immunoprecipitation (ChIP) followed by qPCR. (b) Left, occupancy of the PRC2 component FIE on the tested fragments. Right, H3K27me3 accumulation relative to H3 on the tested fragments. Bars show mean \pm s.e.m. of three independent ChIP experiments (red dots). ** $P < 0.01$; NS, not significant ($P > 0.05$) relative to NC_1; one-tailed unpaired *t*-test. The *LEC2* PRE (ref. 12) served as a positive control (PC_LEC2). See also **Supplementary Figure 2a,b**. (c) A schematic of the construct used to test the ability of PREs to silence active reporters. Candidate PREs and control fragments were placed between two constitutive promoters, pF3H-35S mini and pMAS, with pF3H-35S mini driving expression of *GFP* and the β -glucuronidase gene *GUS*, and pMAS driving expression of the herbicide-resistance gene *BAR*. (d) Visual scores of GFP intensity (on a scale of 1 to 5, where 1 indicates no reporter expression and 5 indicates full reporter expression) in 15 independent transformants in the wild type (WT; Left) or a *prc2* mutant (*clf-28 swm-7*; Right). Shown are violin plots of GFP intensities of presumptive PREs and negative controls; red or green indicates range; white circle, median; white line, mean; vertical black line, lower to higher quartile. Black horizontal bars mark the median GFP fluorescence of the PRE populations in the wild-type background. * $P < 0.05$; ** $P < 0.01$; NS, not significant ($P > 0.16$) relative to NC_1; one-tailed Mann-Whitney *U*-test. Similar results were obtained for silencing of β -glucuronidase activity and herbicide resistance (**Supplementary Fig. 2c-e**).

binding assays using a library of 1,956 *Arabidopsis* TFs¹⁹. Our screen identified 233 PRE-binding TFs (**Supplementary Table 3**). We selected 55 TFs belonging to 20 families for further characterization on the basis of significant binding to multiple PREs (**Supplementary Table 3**).

The three families most enriched with PRE-associated TFs each contained a highly enriched subfamily (**Fig. 2a** and **Supplementary Table 3**): the C2H2 zinc-finger (ZnF) family²⁰ (C1-2iD subfamily; 4 of 6 members identified), the plant-specific APETALA2-like family²¹

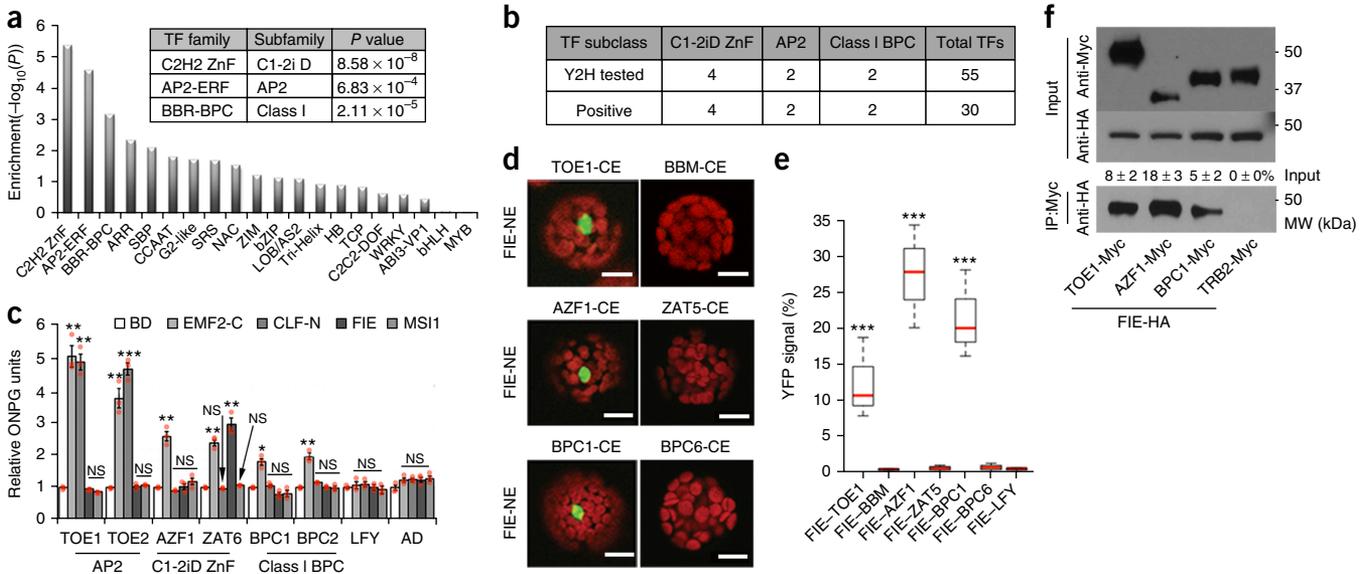


Figure 2 PRE-binding TFs physically interact with PRC2. (a) The TF families most enriched with PRE-binding TFs, as indicated by high-throughput yeast one-hybrid assays. *y*-axis, *P*-value, hypergeometric test. Inset, the most enriched TF subfamilies. (b,c) Interaction between PRE-binding TFs and PRC2 components as shown by medium-throughput yeast two-hybrid (Y2H) tests of all 55 PRE-binding TFs (b) and quantitative *o*-nitrophenyl- β -D-galactopyranoside (ONPG) assays of members of the most enriched TF subfamilies (c). The PRC2 components EMF2-C, CLF-N, FIE and MSI1 were fused to the GAL4 DNA binding domain (BD) and used in the Y2H growth and ONPG assays. BD alone served as a negative control. The transcription factors were fused to the GAL4 activation domain (AD). AD alone and the transcription factor LFY fused to AD served as negative controls. Bars show mean \pm s.e.m. of three independent yeast two-hybrid experiments (red dots). * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$ relative to BD; NS, not significant ($P > 0.05$) relative to BD; one-tailed unpaired *t*-test. (d,e) Interaction between PRE-binding TFs and PRC2 (represented by the single-copy PRC2 component FIE) as shown by BiFC in protoplasts. TFs tested were TOE1 (from AP2 subfamily), AZF1 (C1-2iD ZnF subfamily) and BPC1 (class I BPC subfamily). Controls were BBM (AP2/ANT subfamily), ZAT5 (C1-2iC ZnF subfamily), BPC6 (class II BPC subfamily) and LFY. Nuclear fluorescence is shown in d (scale bar, 10 μ m) and quantified in e. Box-and-whisker plots (e) show the median of three BiFC experiments comprising 150 cells each (red line), upper and lower quartiles (box), and minima and maxima (whiskers). *** $P < 0.001$ relative to controls; one-tailed Mann-Whitney *U*-test. (f) Coimmunoprecipitation of hemagglutinin (HA)-tagged PRC2 (FIE) after immunoprecipitation of Myc-tagged TFs in seedlings. TRB2 was a negative control. Values labeled “% input” represent mean \pm s.e.m. of relative immunoprecipitation (normalized to input FIE-HA and TF levels) from three experiments. **Supplementary Figure 3** shows coimmunoprecipitation in the presence of an endonuclease.

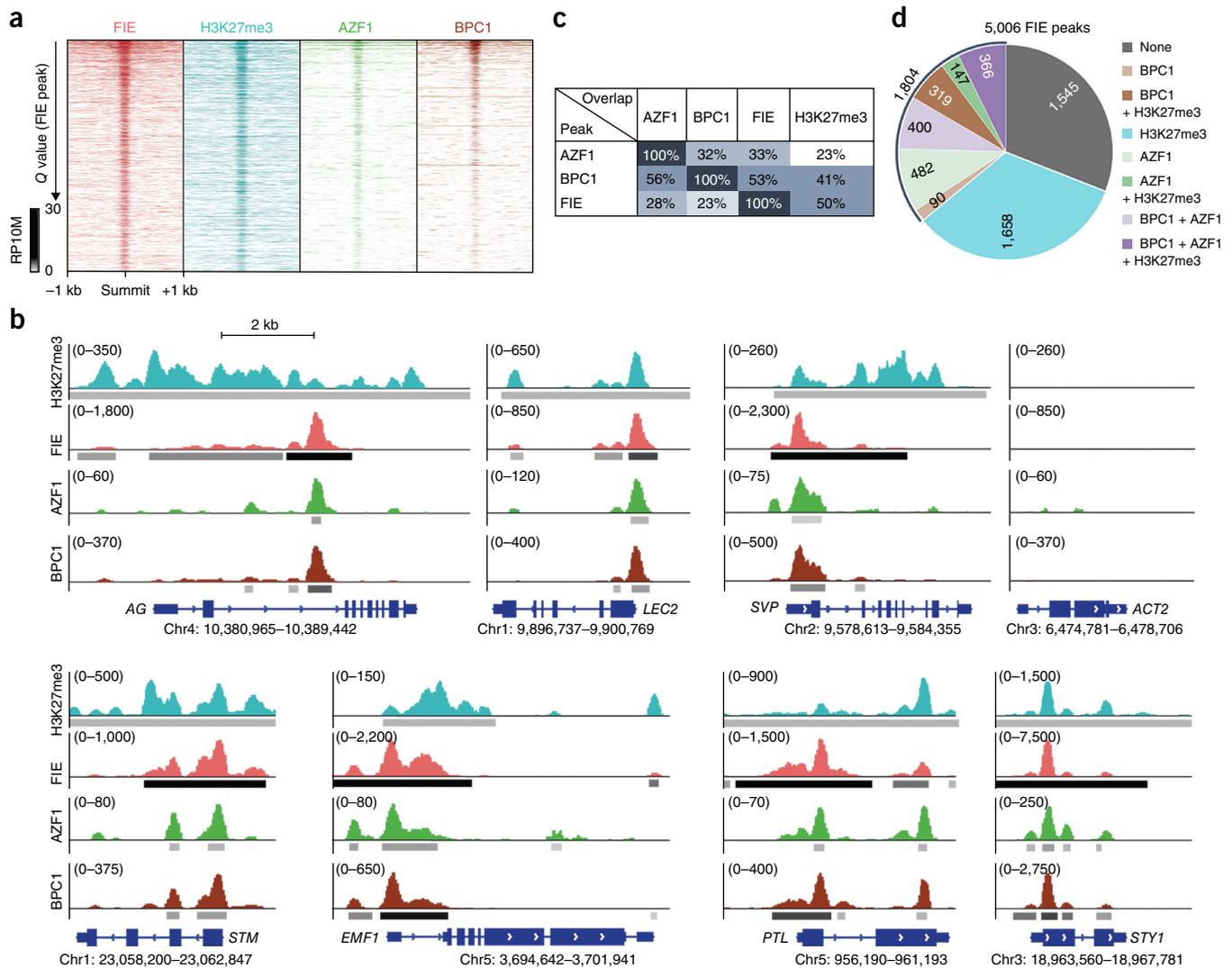


Figure 4 ChIP-seq analysis to test chromatin occupancy of FIE, H3K27me3, AZF1 and BPC1 in 30-h-old plants. **(a)** A heat map of background-corrected ChIP enrichment for PRC2 (FIE), H3K27me3, AZF1 and BPC1. Significant peaks ($Q < 10^{-10}$) were centered on FIE peak maxima and rank-ordered from highest (top) to lowest (bottom) FIE binding-peak significance. For each factor and mark, three independent ChIP-seq experiments were sequenced, as were three matched input controls. The highly consistent ChIP replicates were normalized by sequencing depth and averaged. RP10M, reads per 10 million mapped. **(b)** Browser view of input-subtracted ChIP-seq signals at *AG*, *LEC2* (positive control), *ACT2* (negative control NC_3 in Fig. 1) and additional FIE-bound loci. Significant peaks ($Q < 10^{-10}$) according to MACS2 are marked by horizontal bars, with the black saturation proportional to the Q value (as for the narrowPeak file format in ENCODE). As previously reported for seedlings¹⁵, about half of all FIE peaks overlapped with H3K27me3. **(c)** The percentage of ChIP-seq overlap (by row) of the significant ($Q < 10^{-10}$) FIE, AZF1 or BPC1 peaks with other significant peaks. Shading indicates the strength of overlap. **(d)** Numbers of significant ($Q < 10^{-10}$) FIE peaks that overlap with other significant peaks.

Cognate binding motifs (GA repeat and telobox) of the class I BPC and the C1-2iD ZnF TFs were also significantly ($P < 10^{-38}$ and $P < 10^{-219}$, respectively) enriched under the FIE binding peaks (**Supplementary Fig. 5d**). Gene Ontology term enrichment analysis linked the FIE, AZF1 and BPC1 targets to shoot development, flower patterning and organium development (**Supplementary Fig. 5e**).

Characteristic phenotypes associated with mutation of the PRC2 methyltransferase CLF are upward-curved leaves with partial floral identity as well as precocious flowering². Higher-order mutants from the class I BPC TFs or the C1-2iD ZnF TFs do not exhibit these phenotypes^{30,31} (**Supplementary Fig. 6**), suggesting combinatorial roles for the two TF families in Polycomb-mediated silencing. Two pieces of evidence support this idea. First, knockdown of either TF family significantly ($P < 0.05$) enhanced the leaf curling and the precocious flowering of the hypomorph *clf^R* mutant¹⁸ (**Supplementary**

Figs. 7 and 8). Second, simultaneous knockdown of both TF families (*BPC+ZnF^{KD}*) in the wild type triggered upward leaf curling and precocious flowering (**Fig. 5a,b**). The *BPC^{KD}clf^R*, *ZnF^{KD}clf^R* and *BPC+ZnF^{KD}* phenotypes were accompanied by a significant reduction in PRC2 (FIE) occupancy and in H3K27me3 at Polycomb target loci, and by significant derepression of the Polycomb targets (**Fig. 5c–e** and **Supplementary Figs. 7–9**). FIE occupancy (at peaks with $Q < 10^{-10}$ in the wild type) was also reduced genome-wide in 30-h-old *BPC+ZnF^{KD}* plants (**Fig. 5f,g** and **Supplementary Fig. 10**). By contrast, occupancy of BPC1 at a Polycomb target locus was not dependent on the presence of PRC2 (**Supplementary Figs. 11 and 12**).

Finally, we assessed the contribution of AZF1 and BPC1 to PRC2 recruitment by reciprocal gain-of-function tests. Tethering both the AZF1 and the BPC1 TF to an artificial promoter in isolated plant cells (protoplasts) triggered levels of FIE recruitment similar to

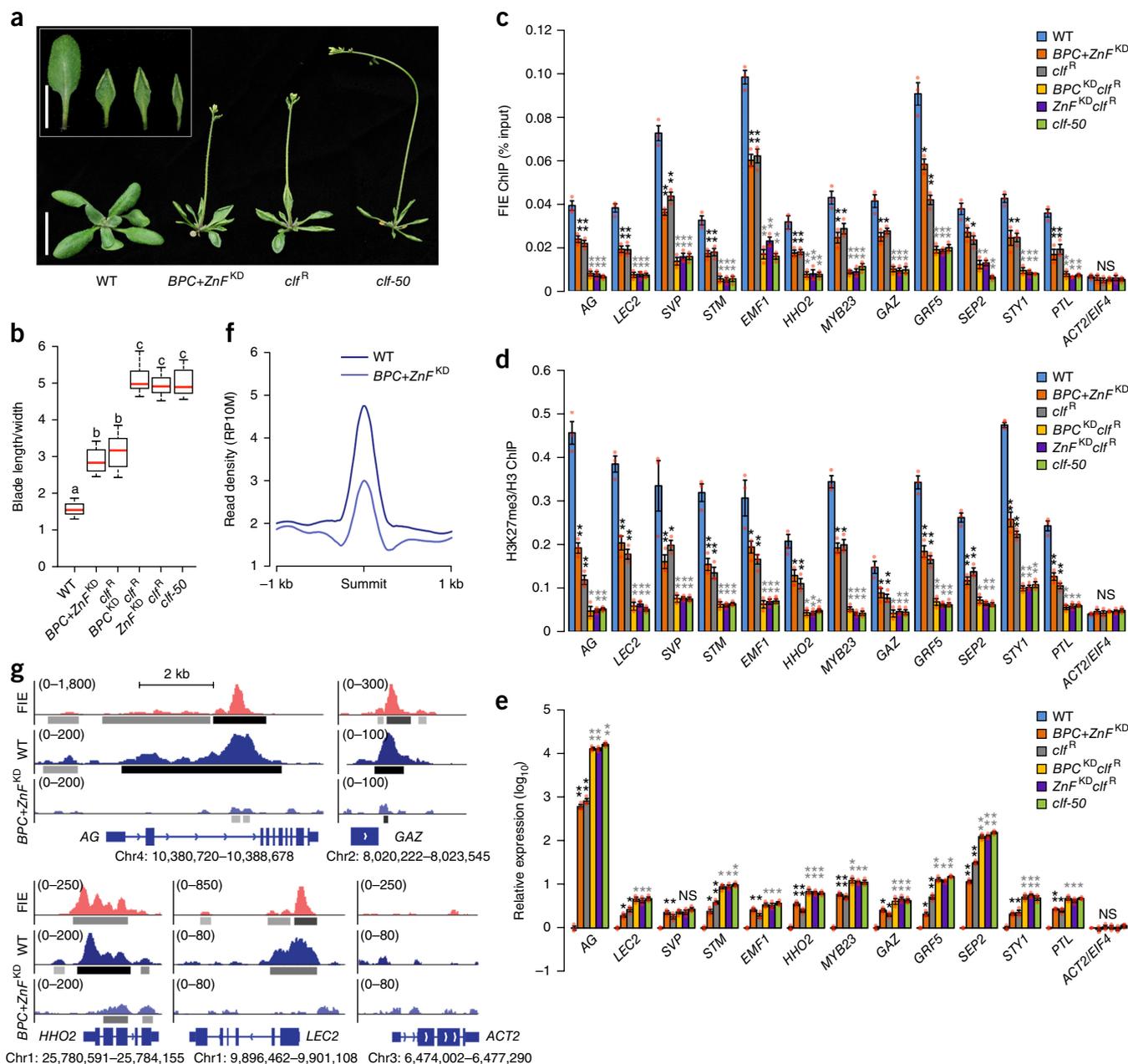


Figure 5 Class I BPC and C1-2iD ZnF TF families are required for Polycomb-mediated silencing and PRC2 recruitment *in planta*. **(a,b)** Precocious flowering and leaf curling in double-knockdown *BPC+ZnF^{KD}* plants compared to the wild type (WT) and to hypomorph *clf^R* and null *clf-50* mutants. Photographs in **a** show plants at 5 weeks of age; scale bar, 1 cm. Box-and-whisker plots in **b** quantify leaf curling, showing the median (red line; $n = 15$ plants), upper and lower quartiles (box edges), and minima and maxima (whiskers). Letters above boxes indicate significantly different groups ($P < 0.05$ based on Kruskal–Wallis test with Dunn’s *post hoc* test). **(c–e)** PRC2 (FIE) occupancy, H3K27me3 accumulation (normalized to H3 to control for nucleosome density) and gene expression in genotypes described in **a** and in *BPC^{KD}clf^R* or *ZnF^{KD}clf^R* plants. Negative-control loci were *ACT2* for ChIP and *EIF4* for gene expression. Expression in the mutant lines is shown relative to WT. Bars represent mean \pm s.e.m. from three independent experiments (red dots). Black asterisks, significantly different from WT; gray asterisks, significantly different from *clf^R*. (Asterisks are stacked vertically above the corresponding bar.) * $P < 0.05$; ** $P < 0.01$; NS, not significant ($P > 0.05$); one-tailed unpaired *t*-test. **(f,g)** ChIP-seq analysis of FIE occupancy in WT and *BPC+ZnF^{KD}* plants. For each factor and mark, three independent ChIP experiments were sequenced, as were three matched input controls. The highly consistent ChIP replicates were normalized by sequencing depth and averaged. FIE occupancy ($Q < 10^{-10}$) region metaplots are shown in **f**. RP10M, reads per 10 million mapped. Screenshots of input-subtracted FIE occupancy are shown in **g**. Red, previous FIE ChIP-seq (**Fig. 4**); blue, FIE ChIP-seq in WT (center) and in *BPC+ZnF^{KD}* (bottom).

those observed at an endogenous Polycomb target locus (**Fig. 6a**). In addition, overexpression of BPC1 or AZF1 in *clf^R* plants restored PRC2 occupancy and H3K27me3 at Polycomb target loci to nearly wild-type levels and largely rescued the leaf-curling defect of *clf^R* (**Supplementary Fig. 12**). These findings suggest combinatorial roles

for class I BPC and C1-2iD ZnF TFs in Polycomb-mediated silencing and PRC2 recruitment.

Here we uncover a PRC2-recruitment strategy in *Arabidopsis* that is similar to that in *Drosophila*, including roles for GA-repeat-binding and ZnF TFs in recruitment (**Fig. 6b**)^{7–9}. Further support for our

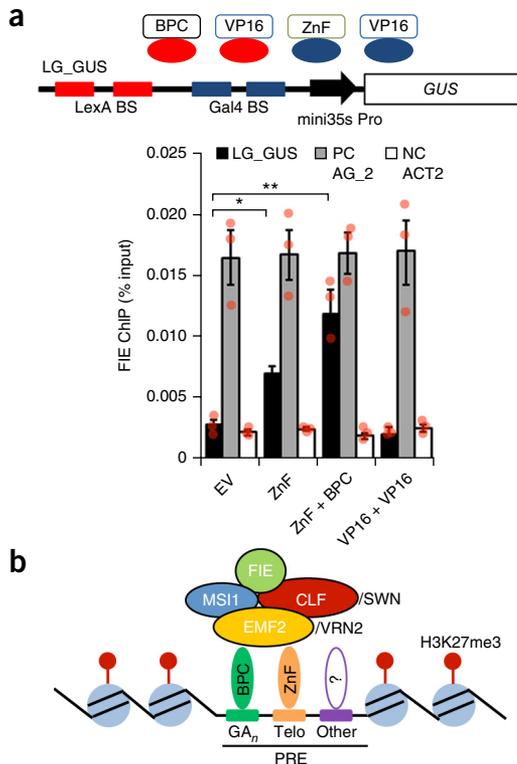


Figure 6 Tethering TFs from the class I BPC and C1-2iD ZnF subfamilies to the DNA to test PRC2 and FIE recruitment. **(a)** Top, a diagram showing tethering of BPC1 (BPC), AZF1 (ZnF) to an artificial promoter (LG_GUS). BPC1 and control construct VP16 (red ovals) were fused to the LexA DNA binding domain for recruitment to two LexA binding sites (BS) in LG_GUS. AZF1 and VP16 (blue ovals) were fused to the Gal4 DNA binding domain for recruitment to two Gal4 binding sites in LG_GUS. A 35S minimal promoter (mini35S Pro) drove β -glucuronidase expression (*GUS*). Bottom, results of tests of FIE recruitment using anti-FIE for ChIP-qPCR in protoplasts. Endogenous loci tested were *AGAMOUS* (*AG_2* region, positive control (PC)) and *ACT2* (negative control (NC)). Empty vector (EV) and VP16 served as control tethering vectors. Bars show mean \pm s.e.m. of three ChIP experiments (red dots). * $P < 0.05$; ** $P < 0.01$; relative to empty vector (EV); one-tailed unpaired *t*-test. **(b)** A model for PRC2 recruitment in *Arabidopsis*. The PRC2 complex is recruited to target loci by members of the class I BPC and C1-2iD ZnF transcription factor subfamilies, which recognize GA repeat (GA_n) and telobox (Telo) motifs in plant PREs, respectively. Additional *trans* factors (indicated by question mark) and *cis* elements ("Other") likely contribute to recruitment. See the text for details.

findings comes from a recent study of PRC2 recruitment to the *ABI4* locus³². Our data suggest that a similar logic underpins PRC2 recruitment in species from two kingdoms of life. The plant PREs we uncovered may recruit both PRC2 and PRC1, as GA repeats and telobox motifs are also linked to PRC1 occupancy in *Arabidopsis*^{26,27}. Additional PREs besides those we predicted probably exist and may act at different stages and in different tissues or conditions. Likewise, additional determinants of PRC2 recruitment remain unidentified. Their discovery, combined with the current data, should enable computational prediction of PREs for future epigenetic reprogramming of cell identity or function to enhance plant growth and yield.

URLs. R packages used: *vioplot*, <https://cran.r-project.org/web/packages/vioplot/index.html>; LOESS smoothing, <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/loess.html>; PCC analysis, <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/cor.html>; PCA analysis, <https://stat.ethz.ch/R-manual/R-devel/library/stats/>

<http://prcomp.html>. Motif analysis: Meme, <http://meme-suite.org/doc/download.html>; Weeder <http://159.149.160.51/modtools/>; MotifSampler, <http://bioinformatics.intec.ugent.be/MotifSuite/motifsampler.php>; AlignAce, <http://arep.med.harvard.edu/mrna-data/mrnasoft.html>; BioProspector, <http://ai.stanford.edu/~xslui/BioProspector/>; Cistome <http://bar.utoronto.ca/cistome>.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the [online version of the paper](#).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank U.-S. Lee for help with manuscript preparation, students of the BIOL425 lab course at the University of Pennsylvania (Spring 2015 and Fall 2015) for the telobox interactome screen, J. He for help with the fluorometric assay of *GUS* activity, and W. Li (Chinese Academy of Sciences) for statistical analyses. Support for the study comes from NSF MCB-1243757 and MCB-1614355 to D.W., NSF IOS-1238142 to X.Z., AAFRC GRDI-130 to R.S.A., BBSRC award G20266 to J.G., BBSRC award G20266 and DFG fellowship CL 393/1-1 to O.C., NIH NRSA 1F31GM112417-01 to M.F.G., NIH R01GM056006 to S.A.K. and J.L.P.-P. NIH R01GM067837 and RC2GM092412 to S.A.K., and NIH DP2MH107055, MoD 1-FY-15-344, and a Searle Scholar Program award to R.B.

AUTHOR CONTRIBUTIONS

D.W. and J.X. conceived of the study and J.X. conducted the majority of the experiments. O.C. and J.G. generated the CLF ChIP-chip data set. J.L.P.-P. and S.A.K. conducted the high-throughput DNA interactome screen. R.S.A. performed the motif analysis, C. Helliwell contributed to functional PRE analysis, and R.J. and M.S. contributed to TF-PRC2 interaction tests. A.P., C.S. and M.Z. identified telobox-binding TFs under the guidance of J.D.W. in the BIOL425 laboratory course held in the spring and fall semesters of 2015. C. He and R.B. performed library preps. R.B., X.Y., S.K., X.Z. and A.M.S. conducted bioinformatic analyses. M.F.G. developed the labeling protocol for EMSA. S.C. and X.L. raised anti-BPC1 antisera. D.W. wrote the paper with input from all authors.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the [online version of the paper](#).

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

- Förderer, A., Zhou, Y. & Turck, F. The age of multiplexity: recruitment and interactions of Polycomb complexes in plants. *Curr. Opin. Plant Biol.* **29**, 169–178 (2016).
- Goodrich, J. *et al.* A Polycomb-group gene regulates homeotic gene expression in *Arabidopsis*. *Nature* **386**, 44–51 (1997).
- Mozgova, I. & Hennig, L. The Polycomb group protein regulatory network. *Annu. Rev. Plant Biol.* **66**, 269–296 (2015).
- Pu, L. & Sung, Z.-R. PcG and trxG in plants—friends or foes. *Trends Genet.* **31**, 252–262 (2015).
- Xiao, J. & Wagner, D. Polycomb repression in the regulation of growth and development in *Arabidopsis*. *Curr. Opin. Plant Biol.* **23**, 15–24 (2015).
- Lewis, E.-B. A gene complex controlling segmentation in *Drosophila*. *Nature* **276**, 565–570 (1978).
- Blackledge, N.-P., Rose, N.-R. & Klose, R.-J. Targeting Polycomb systems to regulate gene expression: modifications to a complex story. *Nat. Rev. Mol. Cell Biol.* **16**, 643–649 (2015).
- Simon, J.-A. & Kingston, R.-E. Occupying chromatin: Polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put. *Mol. Cell* **49**, 808–824 (2013).
- Kassis, J.-A. & Brown, J.-L. Polycomb group response elements in *Drosophila* and vertebrates. *Adv. Genet.* **81**, 83–118 (2013).
- Sun, B. *et al.* Timing mechanism dependent on cell division is invoked by Polycomb inhibition in plant stem cells. *Science* **343**, 1248559 (2014).
- Lodha, M., Marco, C.-F. & Timmermans, M.-C. The ASYMMETRIC LEAVES complex maintains repression of KNOX homeobox genes via direct recruitment of Polycomb-repressive complex2. *Genes Dev.* **27**, 596–601 (2013).
- Berger, N., Dubreucq, B., Roudier, F., Dubos, C. & Lepiniec, L. Transcriptional regulation of *Arabidopsis* *LEAFY COTYLEDON2* involves *RLE*, a *cis*-element that regulates trimethylation of histone H3 at lysine-27. *Plant Cell* **23**, 4065–4078 (2011).

13. Yuan, W. *et al.* A *cis* cold memory element and a *trans* epigenome reader mediate Polycomb silencing of *FLC* by vernalization in *Arabidopsis*. *Nat. Genet.* **48**, 1527–1534 (2016).
14. Qüesta, J.-I., Song, J., Geraldo, N., An, H. & Dean, C. *Arabidopsis* transcriptional repressor VAL1 triggers Polycomb silencing at *FLC* during vernalization. *Science* **353**, 485–488 (2016).
15. Deng, W. *et al.* *Arabidopsis* Polycomb repressive complex 2 binding sites contain putative GAGA factor binding motifs within coding regions of genes. *BMC Genomics* **14**, 593 (2013).
16. Kim, S.-Y., Lee, J., Eshed-Williams, L., Zilberman, D. & Sung, Z.-R. EMF1 and PRC2 cooperate to repress key regulators of *Arabidopsis* development. *PLoS Genet.* **8**, e1002512 (2012).
17. Zhang, X. *et al.* Whole-genome analysis of histone H3 lysine 27 trimethylation in *Arabidopsis*. *PLoS Biol.* **5**, e129 (2007).
18. Lopez-Vernaza, M. *et al.* Antagonistic roles of *SEPALLATA3*, *FT* and *FLC* genes as targets of the Polycomb group gene *CURLY LEAF*. *PLoS One* **7**, e30715 (2012).
19. Pruneda-Paz, J.-L. *et al.* A genome-scale resource for the functional characterization of *Arabidopsis* transcription factors. *Cell Rep.* **8**, 622–632 (2014).
20. Kodaira, K.-S. *et al.* *Arabidopsis* Cys2/His2 zinc-finger proteins AZF1 and AZF2 negatively regulate abscisic acid-repressive and auxin-inducible genes under abiotic stress conditions. *Plant Physiol.* **157**, 742–756 (2011).
21. Kim, S., Soltis, P.-S., Wall, K. & Soltis, D.-E. Phylogeny and domain evolution in the *APETALA2*-like gene family. *Mol. Biol. Evol.* **23**, 107–120 (2006).
22. Monfared, M.-M. *et al.* Overlapping and antagonistic activities of *BASIC PENTACYSTEINE* genes affect a range of developmental processes in *Arabidopsis*. *Plant J.* **66**, 1020–1031 (2011).
23. Winter, C.-M. *et al.* *LEAFY* target genes reveal floral regulatory logic, *cis* motifs, and a link to biotic stimulus response. *Dev. Cell* **20**, 430–443 (2011).
24. Molitor, A.-M. *et al.* The *Arabidopsis* hnRNP-Q protein LIF2 and the PRC1 subunit LHP1 function in concert to regulate the transcription of stress-responsive genes. *Plant Cell* <https://dx.doi.org/10.1105/tpc.16.00244> (2016).
25. Wang, H. *et al.* *Arabidopsis* flower and embryo developmental genes are repressed in seedlings by different combinations of Polycomb group proteins in association with distinct sets of *cis*-regulatory elements. *PLoS Genet.* **12**, e1005771 (2016).
26. Zhou, Y., Hartwig, B., James, G.V., Schneeberger, K. & Turck, F. Complementary activities of TELOMERE REPEAT BINDING proteins and Polycomb group complexes in transcriptional regulation of target genes. *Plant Cell* **28**, 87–101 (2016).
27. Hecker, A. *et al.* The *Arabidopsis* GAGA-binding factor BASIC PENTACYSTEINE6 recruits the POLYCOMB-REPRESSIVE COMPLEX1 component LIKE HETEROCHROMATIN PROTEIN1 to GAGA DNA motifs. *Plant Physiol.* **168**, 1013–1024 (2015).
28. Kooiker, M. *et al.* BASIC PENTACYSTEINE1, a GA binding protein that induces conformational changes in the regulatory region of the homeotic *Arabidopsis* gene *SEEDSTICK*. *Plant Cell* **17**, 722–729 (2005).
29. Bouyer, D. *et al.* Polycomb repressive complex 2 controls the embryo-to-seedling phase transition. *PLoS Genet.* **7**, e1002014 (2011).
30. Ciftci-Yilmaz, S. & Mittler, R. The zinc finger network of plants. *Cell. Mol. Life Sci.* **65**, 1150–1160 (2008).
31. Meister, R.-J. *et al.* Definition and interactions of a positive regulatory element of the *Arabidopsis* *INNER NO OUTER* promoter. *Plant J.* **37**, 426–438 (2004).
32. Mu, Y. *et al.* BASIC PENTACYSTEINE proteins repress ABCISIC ACID INSENSITIVE4 expression via direct recruitment of the Polycomb-repressive complex 2 in *Arabidopsis* root development. *Plant Cell Physiol.* **58**, 607–621 (2017).

ONLINE METHODS

Plant material and treatment. Mutants and transgenic plants previously described: *clf-28*, *clf-28 swn-7* (ref. 33); *clf-50*, *clf^R* (pCLF:CLF-GR *clf-50*)¹⁸; *bpc123* (ref. 22); pFIE:FIE-HA *fie-11* (ref. 34); pEMF2:EMF2-3XFLAG *emf2* (ref. 35); pMSI1::GFP-MSI1 *msi1-1* (ref. 36); 35S::GFP-CLF *clf-50* (ref. 37). In *clf^R*, the *clf-50* RNA null mutant is partly rescued by 'leaky' nuclear translocation of pCLF:CLF-GR in the absence of steroid treatment. *clf-50* and *clf^R* are in the Ws accession, pFIE:FIE-HA *fie-11* is in the C24 accession, all other plants are in the Col-0 accession.

PRE tests. To test PRC2 recruitment and H3K27me3 enrichment at PREs or control fragments, we analyzed progeny pools of 40 random T1 plants. We adopted this strategy to minimize outliers caused, for example, by position effect. Independent pools tested gave similar results. Control fragments included a known PRE (PC_LEC2) and three random DNA fragments (NC_1, intron of At1g60200; NC_2, promoter of AG; NC_3, 3' UTR of Actin 2 (At3g18780)). GFP intensity was scored visually with a dissecting fluorescence microscope (Olympus, MVX10). A fluorometric (MUG) assay was used to quantify β -glucuronidase (GUS) activity as previously described^{38,39}, except that the 4-MU produced was normalized over the fresh weight of each plant. GFP and GUS reporter silencing was assayed in independent primary transformants (T1 plants). For the herbicide-resistance assay, primary transformants were transplanted into soil after selection and sprayed with the Basta herbicide (200 mg/l) (Bayer Crop Science) 2–3 d after transplanting. The survival rate was scored 5–7 d later.

Transgenic plants. Candidate PRE DNA fragments (~600 bp in length) were cloned into pFK205 (ref. 40) and transformed into wild type (Col-0), pMSI1:GFP-MSI1 *msi1-1* or pEMF2:EMF2-3XFLAG *emf2* for chromatin immunoprecipitation (ChIP) assays. Plants were selected on 1/2 Murashige and Skoog (MS) medium⁴¹ (Sigma) with 40 mg/l kanamycin.

A dual reporter system (pPRE-dual-rep) was generated containing the FLAVANONE 3-HYDROXYLASE regulatory region plus the 35S minimal promoter (pF3H-35S mini) driving expression of GFP and β -glucuronidase, and the mannopine synthase promoter (pMAS) driving the BAR gene. Candidate PRE or control fragments were cloned into pPRE-dual-rep and transformed into the wild type (Col-0) or *prc2* mutants (*clf-28*, *clf-28 swn-7/+*). PRE fragments for test of loss or gain of GA repeat and telobox motifs were synthesized (see **Supplementary Table 5**) (GenScript Inc. Company, Piscataway, NJ, US) and shuffled into pPRE-dual-rep. Plants were selected on 1/2 MS plates with 25 mg/l hygromycin.

For TF knockdown, 300-bp regions conserved in the PRC2 recruiting TF subfamilies (for example, C1-2iD ZnF) but not in the larger TF family (C1-2i ZnF) were PCR-amplified and inserted into vector pRNAi-GG. Plasmids were introduced into *Agrobacterium tumefaciens* strain GV3101 and transformed into *clf^R* and wild-type plants by floral dip⁴². T1 plants were selected on 1/2 MS plates with 40 mg/l kanamycin. *BPC^{KD}* plants were crossed with *ZnF^{KD}* plants to generate double-knockdown lines.

For overexpression of BPC1 or AZF1 in *clf^R*, the cDNA of BPC1 and AZF1 was cloned into vector pGWB12 (ref. 43). Plasmids were introduced into *A. tumefaciens* strain GV3101 and transformed into *clf^R* by floral dip⁴². T1 plants were selected on 1/2 MS plates with 40 mg/l kanamycin and 20 mg/l hygromycin.

For TF CoIP and ChIP, genomic fragments spanning the upstream intergenic region and the coding region for each TF were PCR amplified (BPC1 (–3,225 to +849 bp), AZF1 (–1,628 to +735 bp), TOE1 (–8,685 to +1,353 bp), TRB2 (–1,194 to +1,227 bp)), cloned into pEG303 (ref. 44) and transformed into pFIE:FIE-HA *fie-11* for CoIP and into the wild type (Col-0) for ChIP.

Phenotype quantification and qRT-PCR. Plants were grown at 22 °C in short-day conditions (8 h light/16 h dark, light intensity: ~140 μ mol/m²/s). The length and width of the blade of the fifth rosette leaf were measured for >15 plants at day 20. RNA was extracted from 3- to 7-d-old plants grown in short-day conditions, and quantitative real-time PCR was performed as previously described⁴⁵.

Identification of CLF-binding sites using ChIP-chip. ChIP was conducted with anti-GFP (Molecular Probes) in 35S::GFP-CLF *clf-50* plants³⁷ as

previously described¹⁷. Four independent replicates were performed. Amplification, labeling and microarray hybridization were as reported⁴⁶. We identified genomic regions enriched for CLF-binding sites by comparing CLF ChIP-chip to input DNA (four replicates), using the Tilemap program with the hidden Markov model (HMM) option⁴⁷. We merged adjacent probes with HMM posterior probabilities of $P > 0.5$ or higher into regions by requiring a minimal run of 50 bp and allowing a maximal gap of 200 bp, and we converted TAIR5 coordinates to TAIR10 coordinates with the 'update_coordinates.pl' script from TAIR. The 3,648 regions enriched for CLF binding are listed in **Supplementary Table 1**.

Identification of candidate PREs. Putative PRE-containing regions were identified on the basis of the following two conservative criteria. First, we compared previously published genome-wide distributions of H3K27me3, EMF1 and FIE^{15–17} with that of CLF generated in this study. EMF1 is a putative PRC1 component that frequently colocalizes with PRC2 (ref. 16). A candidate PRE-containing region was required to overlap with at least three of the four data sets (to account for potential false negative results in these data sets and the redundant contribution of other PRC components). A total of 1,504 regions were identified and assigned to 851 *Arabidopsis* genes on the basis of previously described criteria²³. Second, we required target genes associated with these putative PREs to be expressed in a highly tissue-specific manner or be derepressed in *prc2* mutants (*clf swn* and *fie*)^{15,33}. For the former, we analyzed previously published transcription profiles in different tissues⁴⁸, and defined tissue-specifically expressed genes as those with expression levels higher than 5 \times the baseline levels. Since all ChIP data (FIE, CLF, EMF1, H3K27me3) were from vegetative development, baseline was defined as the mean expression in samples ATGE_7, ATGE_87, ATGE_12, ATGE_26, ATGE_1, ATGE_19, ATGE_15, ATGE_13, ATGE_20, ATGE_21, ATGE_14, ATGE_17, ATGE_18, ATGE_91, ATGE_5, ATGE_16, ATGE_11 and ATGE_10 from AtGenExpress⁴⁸. PRE-linked genes derepressed in *clf swn* or *fie* were defined as in refs. 15,33. The combined expression filters resulted in the identification of 132 high-confidence PRC2-regulated genes. 170 candidate PREs were associated with the 132 PRC2-regulated genes. Five of the candidate PREs were selected for *in planta* PRE tests.

Motif prediction and mapping. *De novo* motif prediction was performed as previously published, with minor changes^{23,49–51}. Briefly, we applied the motif-prediction pipeline to a subset of the candidate PREs (those bound by CLF and FIE and marked by H3K27me3; 70 PREs) to enhance prediction performance and lower false positive rates⁵². Motifs were subsequently tested for enrichment within the entire PRE set. Motif width was set from 6 to 16 bp when applicable. The background set consisted of 600-bp genic terminal regions from the TAIR10 genome¹⁵. Motif enrichment was calculated using a non-parametric deterministic sampling as described previously ($Z \geq 3$; $P < 0.027$)²³. We filtered highly degenerate motifs from the results by using an information quality statistic (IQ > 20) defined as

$$IQ_L = \sum_L N_{L,A} \log_2 \left(\frac{F_{L,A}}{P_A} \right)$$

where $N_{L,A}$ is the count of residue type A at position L, $F_{L,A}$ is the frequency of occurrence of the residue of type A at position L in the PSSM, and P_A is the background frequency expected for residue A.

Predicted PSSMs were aligned by position of maximum average site-wise Euclidean distance and hierarchically clustered in the R statistical programming environment. We then chose a representative consensus PSSM from the PSSMs in each clade, along with a merged candidate, by maximizing for significance score and frequency within target genes. A functional-depth cutoff was used in mapping PSSMs to PREs and to genomic and ChIP data sequence sets (GAGA, 0.62; CTCC, 0.42; and CCG, 0.8 and 0 for telobox, CAA repeats and G-box). When mapping GAGA motif occurrences, we merged overlapping and nearby matches into a common region using the BEDTools merge function with a maximum distance between features of 8 bp (ref. 53).

Yeast two-hybrid tests. EMF2, CLF, MSI1, and FIE coding sequences were cloned into pDEST32 (Clontech) and introduced into yeast strain AH109

(MAT a). Full-length clones were used except for EMF2 and CLF. EMF2-C is a better interactor in yeast, and CLF-N overcomes the growth defects caused by full-length CLF⁵⁴. The 55 PRE-interacting TFs were cloned into pDEST22 (Clontech) and transfected into yeast strain Y187 (MAT α). Protein–protein interactions were tested after mating as described in the Matchmaker protocol (Clontech). We quantified interaction strength for a subset of TFs via *o*-nitrophenyl- β -D-galactopyranoside (ONPG) assays⁵⁵ for nine independent colonies for each interaction pair in three pools. In addition, different TF fragments (AZF1-N, 1–90 aa; AZF1-M, 91–194 aa; AZF1-C, 195–245 aa; BPC1-N, 1–140 aa; BPC1-C, 141–283 aa; TOE1-N, 1–151 aa; TOE1-M, 152–310 aa; TOE1-C, 311–464 aa) were cloned into pDEST22 and cotransformed with EMF2C or CLFN into yeast strain AH109, and yeast growth was scored.

Bimolecular fluorescence complementation (BiFC) assays. FIE and all TFs were cloned into pUC-SPV-NE^{GW} and pUC-SPV-CE^{GW} constructed by shuffling the split Venus-Gateway cassette from pDEST-VYNE/CE(R)^{GW} vectors⁵⁶ into pUC18. BiFC assays in *Arabidopsis* protoplasts were conducted and visualized as previously described^{57,58}. For each experiment, fluorescence was compared in protoplast populations prepared and transfected at the same time. Three independent BiFC experiments were performed for each combination of factors tested with at least 150 protoplasts scored per replicate. Representative images were taken with a confocal microscope with the same gain (Leica, LCS SL).

CoIP. CoIP was performed as described⁵⁹, with some modifications. Myc-gBPC1, Myc-gAZF1, Myc-gTOE1 or Myc-gTRB2 (TELOMER REPEAT BINDING 2 protein, negative control) was transformed into pFIE:FIE-HA *fie-11* plants. Three-day-old double-transgenic seedlings were harvested after growth in long-day conditions. Tissue was ground in protein extraction buffer (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1 mM EDTA, 10% glycerol, 0.2% Triton X-100, 1 mM PMSF, 1 \times protease inhibitor (Roche)), filtered, and centrifuged. The supernatant was incubated with anti-Myc (C3965 or 05-724, Sigma) coupled to protein A Dynabeads (Thermo Fisher) overnight at 4 °C. Beads were washed four times with wash buffer, and bound proteins were eluted with elution buffer containing 2% SDS for immunoblotting with HA-HRP-conjugated antibody (3F10, Roche). In some reactions Benzonase endonuclease (E1014, Sigma) was added into the protein sample (20 U) and incubated on ice for 1 h before immunoprecipitation. Full-length gel images for Western analyses after coIP and for protein abundance in BiFC experiments are shown in **Supplementary Figure 13**.

ChIP. ChIP was performed as previously described⁶⁰, with minor modifications. Extraction buffers I (0.4 M sucrose, 10 mM Tris-HCl, pH 8.0, 10 mM MgCl₂, 5 mM b-ME, 1 mM PMSF, 1 \times protease inhibitor (Roche)) and II (0.25 M sucrose, 10 mM Tris-HCl, pH 8.0, 10 mM MgCl₂, 1% Triton X-100, 5 mM b-ME, 1 mM PMSF, 1 \times protease inhibitor (Roche)) were used for protein extraction. The following antibodies that had previously been used for ChIP in *Arabidopsis* were used: anti-GFP (A6455, Thermo Fisher)⁶¹ for pMSI1:GFP-MSI1 *msi1-1* ChIP, anti-HA (12CA5, Roche)⁶¹ for pFIE:FIE-HA *fie-11* ChIP, anti-H3K27me3 (07-449, Millipore)^{33,62}, anti-H3 (07-690, Millipore)⁶², anti-FLAG (F3165, Sigma)⁶³ for pEMF2:EMF2-3XFLAG *emf2* ChIP, and anti-Myc (C3956, Sigma)⁶⁴ for gAZF1-Myc and gBPC1-Myc ChIP. Anti-FIE³⁴ antiserum was first used for ChIP here and showed much reduced occupancy in *prc2* mutants (**Fig. 5c**). The anti-BPC1 antiserum was generated in rabbits, using full-length recombinant BPC1 protein, and gave ChIP signal specifically in the wild type (**Supplementary Fig. 11**). Throughout, H3K27me3 was normalized over H3 to control for nucleosome density; all other ChIP reaction data are shown as % input. For most ChIP experiments, representative transgenic lines were used; alternatively, pools of 40 independent T1 progeny (or more) were used.

ChIP-seq in germinating embryos. For H3K27me3, FIE, AZF1 and BPC1 ChIP-seq, ChIP was performed on germinating embryos (30 h after imbibition) as described above, but eluted into a smaller volume (15 μ l in total), using the Qiagen MinElute PCR purification kit (Cat. No. 28004). Three independent ChIP and input reactions were sequenced. ChIP-DNA and input (after dilution to about 0.1–1 ng) were amplified with the SeqPlex DNA amplification kit (Sigma, SEQXE-10RXN) according to the manufacturer's instructions

with the following modifications: a first linear PCR was followed by a second round of amplification (less than ten cycles). After primer removal, qPCR was performed to test amplification of different genomic regions. We converted linearly amplified DNA from input chromatin and pull-downs to libraries for sequencing by performing end-repair followed by A-tailing and ligation of universal adaptors (all enzymes by Enzymatics, MA). Libraries were amplified to 50–100 nM with custom dual-indexing primers and sequenced with an Illumina NextSeq500 at a depth of >15 million reads per sample for pull-downs and >30 million reads per sample for inputs. We mapped reads to the TAIR10 genome release v31 using bowtie2 (ref. 65). Alignment files were converted to 1-bp resolution bigwig files and normalized by 10 million reads sequenced (RP10M) with custom scripts. All ChIP-seq replicates were highly similar to each other (see below). Bigwigs from the ChIP-seq replicates were averaged with WiggleTools⁶⁶. Significant peaks were identified with MACS2 version 2.1.1.20160309 with relevant input controls (C24 for FIE, Col for AZF1 and BPC1, total H3 pull-down for H3K27me3). Default MACS2 settings were used for FIE, AZF1, and BPC1, and the ‘-broad’ option was used for H3K27me3. Only peaks with $Q < 10^{-10}$ were considered.

Significant peaks were mapped to genes as previously described²³ with the HOMER⁶⁷ script annotatePeaks. Gene Ontology (GO) enrichment analysis was carried out with AgriGO⁶⁸ combined with manual curation to remove redundant terms.

For FIE ChIP-seq in wild-type and double TF-subfamily (class I BPC and C1-2iD ZnF) knockdown lines, anti-FIE³⁴ antiserum was used. ChIP DNA and input libraries (three replicates each) were generated with the ThruPLEX DNA-seq kit (RUBICON GENOMICS, R400406). Libraries were sequenced at a depth of >15 million reads per sample for pull-downs and >30 million reads per sample for inputs. Replicate comparison and significant peak identification were carried out as described above with the default MACS2 settings. Only peaks with $Q < 10^{-10}$ were considered.

To assess the change in FIE binding in *BPC+ZnF^{KD}* mutants relative to the wild type (WT), we extracted normalized reads mapping to WT FIE peaks (**Fig. 5f**) from WT and *BPC+ZnF^{KD}* FIE ChIP data sets using bedtools and plotted log₂(RPM) in a scatterplot. Reads increased or decreased in *BPC+ZnF^{KD}* were indicated ($P < 0.01$ and twofold change). For the region metaplot, detected FIE peaks were quantified from the normalized and input-subtracted number of reads in each 10-bp window (FIE peak signal) in WT and *BPC+ZnF^{KD}*. The average FIE peak signals in the window of $\pm 1,000$ bp around peak centers were plotted with the ‘loess’ smoothing function in R with a span of 0.2.

We assessed ChIP replicate concordance by computing pairwise Pearson correlation coefficients (PCCs) of the RP10M normalized read counts in each 10-kb window for all ChIP-seq data sets, using the ‘cor’ function with method ‘pearson’ in R. PCC was >0.95 for all ChIP-seq replicates. Principal component analysis was performed on the pairwise PCC matrix with the ‘prcomp’ function in R, and the results were converted to a scatterplot. Significant regions for all ChIP-seq experiments performed in 30-h-old plants (germinating embryos) are listed in **Supplementary Table 4**.

EMSA. A 60-bp fragment of the AG₂ PRE-containing two GA repeats or one telobox, as well as versions thereof with motif substitutions, were labeled with Cy5-dCTP (GE Healthcare Life Sciences) by end-repair with Klenow fragment (3'→5' exo-) (NEB). For EMSA and other oligonucleotide sequences, see **Supplementary Table 5**. Briefly, complementary single-strand DNA probes were synthesized (IDT) that, when annealed, gave rise to a two-nucleotide 3' overhang (with the last annealed nucleotide being a G). After Cy5 labeling, the probe was purified with Illustra MicroSpin G-25 columns (GE Healthcare Life Sciences).

Full-length *BPC1*, *AZF1* cDNAs were cloned into pET32a and transformed into *Escherichia coli* (BL21). Protein was isolated as previously described⁶⁹. For binding assays, a 20- μ l reaction containing 2 μ l of the protein extract, 3 μ l of 1 μ g/ μ l probe and 2 μ l of 10 \times binding buffer (0.1 M Tris-HCl, pH 7.5, 0.5 M NaCl, 10 mM DTT, 10 mM EDTA, 50% glycerol, 0.5 mg/ml poly(dI-dC), 1 mg/ml BSA) was used. For the AZF1 ZnF TF, the binding reaction was supplemented with 1 mM CaCl₂ and 0.1 mM ZnCl₂. Free and bound probes were separated on a 6% PAGE gel in 0.5 \times TBE at 100 V for 60 min. The gel was scanned by a Typhoon scanner (Typhoon 9410 variable mode imager, Amersham) at BP 670 with a gain of 600. Shifted and unshifted probes

were quantified with ImageJ. Full-length EMSA gel images are provided in **Supplementary Figure 13**.

Tethering assay. BPC1 and AZF1 coding sequences were cloned into LexA-DBD and Gal4-DBD vectors, respectively, as effectors⁷⁰. The reporter contained two repeats each of the LexA and Gal4 binding sites with a 35S minimal promoter driving β -glucuronidase (*GUS*)⁷⁰. 35S:LUC (firefly luciferase) was included to monitor transfection efficiency. Empty Gal4-DBD or LexA-DBD-VP16 and LexA-DBD-VP16 served as controls. 16 h after transfection, 1 × 10⁶ protoplasts were used for ChIP (anti-FIE antiserum)³⁴.

Y1H screen. Functionally defined PREs (AG, SEP3, At5g) were cloned as 300-bp DNA fragments into a modified pLacZi vector (Clontech) carrying *gLUC* instead of *LacZ* as a reporter gene and integrated into the genome of yeast strain YM4271. A robotic Y1H screen was carried out against 1,956 *Arabidopsis* TFs as previously described but with luciferase activity used as a readout¹⁹. The identified PRE-interacting TFs were further filtered on binding to multiple PRE fragments and binding strength.

For identification of telobox-motif-binding TFs, a telobox motif was inserted into a 30-bp region from a negative control fragment (NC_3; At3g18780 Actin2; Chr 3: 6,476,550–6,476,579). Three copies of the 30-bp NC_3 region with the telobox motif were inserted into bait vector pAbAi (Clontech), and the vector was then into the genome of yeast strain Y1HGold (Clontech). Y1H screening against a TF library with about 1,400 TFs⁷¹ was carried out according to the Matchmaker Gold Yeast One-Hybrid Library Screening System manual (Clontech), with aureobasidin A resistance used to select for binding.

Phylogenetic analyses. For phylogenetic trees, we carried out amino acid sequence alignment and generation of a neighbor-joining (NJ) phylogenetic tree using MEGA⁷² with default settings. Phylogenetic shadowing was performed essentially as previously described⁶¹, except that conserved regions were aligned using Clustal Omega (EMBL-EBI).

Statistical analysis. Statistical tests performed, sample size and *P* values are indicated in each figure legend. The investigator was not blinded to the group allocations during the experiment, and variation was not estimated within each group of data. Dependent variables were continuous, and all data points analyzed were independent. Throughout, the Kolmogorov–Smirnov⁷³ test was used to assess whether the data were normally distributed. For normally distributed data, an unpaired one-tailed *t*-test was used. In all other cases, non-parametric tests were used for two-group comparisons (Mann–Whitney *U*-test⁷⁴) and for multiple-group comparisons (Kruskal–Wallis test⁷⁵ combined with the Dunn's *post hoc* test⁷⁶). Variances in some of the groups compared differ for biological reasons. Rejecting the null hypothesis on the basis of these tests for this type of data implies that one group stochastically dominates a second group—that is, if a value *X* is randomly chosen from one group and a value *Y* is randomly chosen from another, then the probability of *X* > *Y* is greater than the probability of *Y* > *X*. The sample size was chosen on the basis of prior studies that showed significant effects with similar samples sizes (for an example, see ref. 57). *P* values for all tests performed, as well as additional statistical parameters, are listed in **Supplementary Table 6**.

Data availability. ChIP-chip (CLF) and ChIP-seq (all others) data were deposited in the NCBI's Gene Expression Omnibus (GEO) under series numbers GSE7065 (CLF), GSE7063 (input), GSE84483 (FIE, H3K27me3, AZF1 and BPC1) and GSE95562 (FIE in the wild type and in BPC+ZnF^{KD}).

Code availability. Motif-prediction programs are available from their individual websites: Meme v4.9.1, Weeder v1.4.2, AlignAce v4.0, MotifSampler v3.2 and Bioprospector v2004. For motif enrichment and downstream analysis we used the 'Cistome'⁷⁷ pipeline, which is freely available as a web-based application.

33. Lafos, M. *et al.* Dynamic regulation of H3K27 trimethylation during *Arabidopsis* differentiation. *PLoS Genet.* **7**, e1002040 (2011).
34. Wood, C.-C. *et al.* The *Arabidopsis thaliana* vernalization response requires a Polycomb-like protein complex that also includes VERNALIZATION INSENSITIVE 3. *Proc. Natl. Acad. Sci. USA* **103**, 14631–14636 (2006).
35. Kim, S.-Y., Zhu, T. & Sung, Z.-R. Epigenetic regulation of gene programs by EMF1 and EMF2 in *Arabidopsis*. *Plant Physiol.* **152**, 516–528 (2010).
36. Alexandre, C., Möller-Steinbach, Y., Schönrock, N., Gruissem, W. & Hennig, L. *Arabidopsis* MSI1 is required for negative regulation of the response to drought stress. *Mol. Plant* **2**, 675–687 (2009).
37. Schubert, D. *et al.* Silencing by plant Polycomb-group genes requires dispersed trimethylation of histone H3 at lysine 27. *EMBO J.* **25**, 4638–4649 (2006).
38. Maghuly, F. *et al.* Stress regulated expression of the *GUS*-marker gene (*uidA*) under the control of plant calmodulin and viral 35S promoters in a model fruit tree rootstock: *Prunus incisa* × *serrula*. *J. Biotechnol.* **135**, 105–116 (2008).
39. Blazquez, M. Quantitative *GUS* activity assay of plant extracts. *CSH Protoc.* **2007**, pdb.prot4690 (2007).
40. Mathieu, J., Yant, L.-J., Mürdter, F., Küttner, F. & Schmid, M. Repression of flowering by the miR172 target SMZ. *PLoS Biol.* **7**, e1000148 (2009).
41. Murashige, T. & Skoog, F. A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiol. Plant.* **15**, 473–497 (1962).
42. Clough, S.-J. & Bent, A.-F. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743 (1998).
43. Nakagawa, T. *et al.* Development of series of Gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. *J. Biosci. Bioeng.* **104**, 34–41 (2007).
44. Earley, K.-W. *et al.* Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J.* **45**, 616–629 (2006).
45. Yamaguchi, A. *et al.* The microRNA-regulated SBP-box transcription factor SPL3 is a direct upstream activator of *LEAFY*, *FRUITFULL*, and *APETALA1*. *Dev. Cell* **17**, 268–278 (2009).
46. Zhang, X. *et al.* Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell* **126**, 1189–1201 (2006).
47. Ji, H. & Wong, W.-H. TileMap: create chromosomal map of tiling array hybridizations. *Bioinformatics* **21**, 3629–3636 (2005).
48. Schmid, M. *et al.* A gene expression map of *Arabidopsis thaliana* development. *Nat. Genet.* **37**, 501–506 (2005).
49. Bailey, T.-L. & Elkan, C. The value of prior knowledge in discovering motifs with MEME. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **3**, 21–29 (1995).
50. Hughes, J.-D., Estep, P.-W., Tavazoie, S. & Church, G.-M. Computational identification of *cis*-regulatory elements associated with groups of functionally related genes in *Saccharomyces cerevisiae*. *J. Mol. Biol.* **296**, 1205–1214 (2000).
51. Pavesi, G., Mauri, G. & Pesole, G. An algorithm for finding signals of unknown length in DNA sequences. *Bioinformatics* **17**, S207–S214 (2001).
52. Hu, J., Li, B. & Kihara, D. Limitations and potentials of current motif discovery algorithms. *Nucleic Acids Res.* **33**, 4899–4913 (2005).
53. Quinlan, A.-R. & Hall, I.-M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
54. Chanvittana, Y. *et al.* Interaction of Polycomb-group proteins controlling flowering in *Arabidopsis*. *Development* **131**, 5263–5276 (2004).
55. Wang, H. *et al.* Yeast two-hybrid system demonstrates that estrogen receptor dimerization is ligand-dependent *in vivo*. *J. Biol. Chem.* **270**, 23322–23329 (1995).
56. Gehl, C., Waadt, R., Kudla, J., Mendel, R.-R. & Hänsch, R. New GATEWAY vectors for high throughput analyses of protein–protein interactions by bimolecular fluorescence complementation. *Mol. Plant* **2**, 1051–1058 (2009).
57. Wu, M.-F. *et al.* Auxin-regulated chromatin switch directs acquisition of flower primordium founder fate. *eLife* **4**, e09269 (2015).
58. Yoo, S.-D., Cho, Y.-H. & Sheen, J. *Arabidopsis* mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat. Protoc.* **2**, 1565–1572 (2007).
59. Xiao, J. *et al.* O-GlcNAc-mediated interaction between VER2 and TaGRP2 elicits *TaVRN1* mRNA accumulation during vernalization in winter wheat. *Nat. Commun.* **5**, 4572 (2014).
60. Yamaguchi, N. *et al.* PROTOCOLS: chromatin immunoprecipitation from *Arabidopsis* tissues. *Arabidopsis Book* **12**, e0170 (2014).
61. Yamaguchi, N. *et al.* A molecular framework for auxin-mediated initiation of flower primordia. *Dev. Cell* **24**, 271–282 (2013).
62. Aichinger, E. *et al.* CHD3 proteins and Polycomb group proteins antagonistically determine cell identity in *Arabidopsis*. *PLoS Genet.* **5**, e1000605 (2009).
63. Yamaguchi, N. *et al.* Gibberellin acts positively then negatively to control onset of flower formation in *Arabidopsis*. *Science* **344**, 638–641 (2014).
64. Li, K. *et al.* DELLA-mediated PIF degradation contributes to coordination of light and gibberellin signalling in *Arabidopsis*. *Nat. Commun.* **7**, 11868 (2016).
65. Langmead, B. & Salzberg, S.-L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
66. Zerbino, D.-R., Johnson, N., Juettemann, T., Wilder, S.-P. & Flicek, P. WiggleTools: parallel processing of large collections of genome-wide datasets for visualization and statistical analysis. *Bioinformatics* **30**, 1008–1009 (2014).
67. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
68. Du, Z., Zhou, X., Ling, Y., Zhang, Z. & Su, Z. agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res.* **38**, W64–W70 (2010).
69. Zhang, B., Wang, L., Zeng, L., Zhang, C. & Ma, H. *Arabidopsis* TOE proteins convey a photoperiodic signal to antagonize CONSTANS and regulate flowering time. *Genes Dev.* **29**, 975–987 (2015).
70. Tiwari, S.-B., Hagen, G. & Guilfoyle, T.-J. Aux/IAA proteins contain a potent transcriptional repression domain. *Plant Cell* **16**, 533–543 (2004).

71. Castrillo, G. *et al.* Speeding *cis-trans* regulation discovery by phylogenomic analyses coupled with screenings of an arrayed library of *Arabidopsis* transcription factors. *PLoS One* **6**, e21524 (2011).
72. Tamura, K. *et al.* MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739 (2011).
73. Massey, F.J. The Kolmogorov-Smirnov test for goodness of fit. *J. Am. Stat. Assoc.* **46**, 68–78 (1951).
74. Mann, H.-B. & Whitney, D.-R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.* **18**, 50–60 (1947).
75. Kruskal, W.-H. & Wallis, W.-A. Use of ranks in one-criterion variance analysis. *J. Am. Stat. Assoc.* **47**, 583–621 (1952).
76. Dunn, O.J. Multiple comparisons using rank sums. *Technometrics* **6**, 241–252 (1964).
77. Austin, R.-S. *et al.* New BAR tools for mining expression data and exploring *cis*-elements in *Arabidopsis thaliana*. *Plant J.* **88**, 490–504 (2016).