Edinburgh Research Explorer

Cis- and trans-determinants of epigenetic silencing by Polycomb Repressive Complex 2 in Arabidopsis

Cis- and trans-determinants of epigenetic silencing by Polycomb Repressive Complex 2 in Arabidopsis

Jun Xiao[1], Run Jin[1#], Xiang Yu[1#], Max Shen[1], John Wagner[2], Armaan Pai[2], Claire Song[2], Michael Zhuang[2], Samantha Klasfeld[1], Chongsheng He[3], Alexandre M. Santos[4], Chris Helliwell[5], Jose L Pruneda-Paz[6], Steve A Kay[7], Xiaowei Lin[8], Sujuan Cui[8], Meilin Fernandez Garcia[9], Oliver Clarenz[10], Justin Goodrich[10], Xiaoyu Zhang[4], Ryan S. Austin[11,12], Roberto Bonasio[3] & Doris Wagner[1]*

[1]Department of Biology, University of Pennsylvania, Philadelphia, PA, USA.
[2]Lab course BIOL425 Fall and Spring 2015, Department of Biology, University of Pennsylvania, Philadelphia, PA, USA
[3]Department of Cell and Developmental Biology, University of Pennsylvania, Philadelphia, PA, USA
[4]Department of Plant Biology, University of Georgia, Athens, GA, USA.
[5]CSIRO Agriculture and Food, Canberra, Australia.
[6]Division of Biology, University of California San Diego, La Jolla, CA, USA.
[7] Department of Neurology, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA.
[8]College of Life Science in Hebei Normal University, Hebei, P.R. China
[9]Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, PA, USA.
[10]University of Edinburgh, Edinburgh, UK.
[11]Agriculture & Agri-Foods Canada, London, ON, Canada.
[12]Department of Biology, Western University, London, ON, Canada.

[#]: equal contribution

*Correspondence should be addressed to D.W. (wagnerdo@sas.upenn.edu)

28   **Disruption of gene silencing by Polycomb Complexes leads to homeotic transformations**

29   **and altered developmental phase identity in plants[1-5]. Here we define short genomic**

30   **fragments, Polycomb Response Elements (PREs), that direct Polycomb Repressive**

31   **Complex 2 (PRC2) placement at developmental genes regulated by silencing in**

32   **Arabidopsis. We identify transcription factor families that bind to these PREs, co-localize**

33   **with PRC2 on chromatin, physically interact with and recruit PRC2, and are required for**

34   **Polycomb silencing in vivo. Two of the cis sequence motifs enriched in the PREs are**

35   **cognate binding sites for the identified transcription factors and are necessary and**

36   **sufficient for PRE activity. Thus PRC2 recruitment in plants relies in large part on binding**

37   **of trans-acting factors to cis-localized DNA sequence motifs.**

38   In both the plant and animal kingdoms, Polycomb repression is important for cell identity[2,5-8]. The

39   evolutionarily conserved PRC2 complex trimethylates lysine 27 of histone H3 (H3K27me3), an

40   epigenetic mark that results in compaction of chromatin and silencing of gene expression at thousands of

41   loci[1-5,8,9]. After its establishment, the repressed chromatin state is mitotically heritable[8,10]. Given that

42   PRC2 has no inherent DNA binding specificity, a key question is how the Polycomb epigenetic

43   machinery targets loci it silences. In Drosophila, multiple transcription factors (TFs) bind to *cis* regulatory

44   regions several hundreds of base pairs in length called PREs and recruit Polycomb complexes[8,9]. Despite

45   initial identification of a few such PREs in mammals, recent studies instead implicate promoter proximal

46   unmethylated CpG islands in PRC recruitment[7]. Likewise, PREs with inherent silencing ability were

47   identified at a handful of loci in Arabidopsis[10-14], but it is unclear whether this mechanism broadly

48   underpins PRC2 recruitment.

49

50   To elucidate the PRC2 targeting mechanism in Arabidopsis, we identified 132 high-confidence PRC2-

51   regulated genes from our own and public genomic datasets[15-17] (Supplementary Fig. 1a and

52   Supplementary Tables 1 and 2) and computationally defined 170 candidate PREs (600 bp in length)

53   associated with them. We selected five PREs from three loci (the PRC2 targets *AGAMOUS* (*AG*) and

54   *SEPALLATA3* (*SEP3*)[2,18] and a gene of unknown function (*At5g61120* (*At5g*)) to test their ability to

55   recruit PRC2 and direct *de novo* H3K27 trimethylation when randomly integrated into the genome. All

56   five candidate plant PREs recruited PRC2 (represented by complex components FIE, EMF2 and MSI1,

57   Supplementary Fig. 1b) and gained H3K27me3 as did a previously characterized control PRE

58   (PC_LEC2)[12] (Fig. 1 a, b and Supplementary Fig. 2a,b; n=40 transgenic lines). H3K27me3 is known to

59   spread from the site of PRC2 recruitment to adjacent genomic locations[8,9], this was also observed at the

60   PREs (Supplementary Fig. 2a, b). Random unlinked (NC_1 and NC_3) and linked (*AG* locus, NC_2)

61   DNA fragments did not recruit PRC2 and or gain H3K27me3.

62

63   In Drosophila, PREs not only recruit Polycomb complexes and become decorated with H3K27me3, they

64   also repress linked genes in a Polycomb dependent manner[8,9]. Likewise, when placed between two

65   constitutive promoters, the five candidate PREs significantly (P <0.05, Mann Whitney *U*-test) silenced

66   three independent reporter genes (GFP fluorescence, beta-glucuronidase activity and herbicide

67   resistance), as did PC_LEC2 (Fig. 1c,d, n=15 and Supplementary Fig. 2c,d, n=15 and Supplementary Fig.

68   2e, n=60 independent transformants). None of the control DNA fragments had this effect. The ability of

69   the PREs to silence active reporters was dependent on PRC2 in all cases (Fig. 1d, and Supplementary Figs.

70   2d-e).

71

72   To determine which sequence-specific binding proteins associate with the five functional PREs, we

73   performed high-throughput DNA binding assays using a library of 1956 Arabidopsis TFs[19]. Our screen

74   identified 233 PRE-binding TFs (Supplementary Table 3). Fifty-five TFs belonging to 20 families were

75   selected for further characterization on the basis of significant binding to multiple PREs (Supplementary

76   Table 3). Among the 20 PRE-associated TF families, three (sub) families were most significantly

77   enriched (Fig. 2a and Supplementary Table 3): the C2H2 Zinc finger family[20] (C1-2iD ZnF subfamily, 4

78   of 6 members identified), the plant specific APETALA2-like family[21] (AP2 subfamily, 2 of 6 members

79   identified) and the plant specific BASIC PENTACYSTEINE (BPC) family[22] (class I subfamily, 2 of 3

80   members identified).

81

82   In yeast-two-hybrid tests, more than 50 percent of the 55 PRE-interacting transcription factors - including

83   all of the identified C1-2iD Zn-finger, AP2 and class I BPC subfamily transcription factors - physically

84   interacted with at least one PRC2 complex component (Fig. 2b,c and Supplementary Fig. 3a,

85   Supplementary Table 3). We confirmed contact between select members of the transcription factor

86   subfamilies (TOE1, AZF1and BPC1) and the PRC2 complex (represented by FIE) using bimolecular

87   fluorescence complementation (BiFC) in plant cells (Fig. 2d,e and Supplementary Fig. 3b). No BiFC

88   signal was observed for members of closely related TF subfamilies (Fig. 2d,e). Using plants stably

89   expressing tagged versions of TOE1, AZF1 and BPC1 from their endogenous promoters, we showed that

90   the transcription factors co-immunoprecipitated with the PRC2 complex (FIE) in intact plants (Fig. 2f).

91   Co-immunoprecipitation was also observed in the presence of an endonuclease (Supplementary Fig. 3c).

92

93   In a parallel approach, we identified six motifs enriched in the 170 computationally defined PREs with a

94   *de novo* motif analysis pipeline[23] (Fig. 3a,b and Supplementary Table 2). Many of these motifs, in

95   particular the GA repeat and the telobox, had previously been correlated with Polycomb occupancy in

96   Arabidopsis[15,24-27]. To test the biological roles of the GA repeat and telobox motifs in Polycomb silencing,

97   we mutated both motifs in the AG_2 and the At5g PRE (Fig.1). Mutation of GA repeats and telobox

98   motifs significantly reduced the ability of both PREs to silence an active reporter (Fig. 3c). The residual

99   activity of the mutated PREs suggests the presence of additional cis motifs with a role in PRC2

100   recruitment. The GA repeat and telobox motifs of the AG_2 and At5g PRE are conserved in species of

101   the *Brassicaceae* (Fig.3d). Addition of two GA repeats and teloboxes to a DNA fragment that does not

102   recruit PRC2 (NC_1; Fig. 1b) resulted in a synthetic PRE (NC+), which significantly silenced an active

103   reporter in a PRC2-dependent manner (Fig. 3c). The data suggest that the GA repeat and telobox motifs

104   are necessary and sufficient for PRE activity.

105

106   Prior studies had identified GA repeats as the cognate binding sites of class I BPC transcription factors[28]

107   and we confirmed binding of the PRC2 interacting class I BPC TF BPC1 to the GA repeat by

108   electrophoretic mobility shift assay (EMSA) (Supplementary Fig. 4a). We conducted a motif-based DNA

109   interaction screen for TFs that bind the telobox. The screen identified members of the PRE-binding and

110   PRC2-interacting C1-2iD ZnF TF subfamily (Supplementary Fig. 4b,c). Association of the C1-2iD ZnF

111   TF AZF1 with the telobox was verified by EMSA (Fig. 3e). Our findings link sequence motifs important

112   for PRE function to recruitment of TFs that physically interact with PRC2.

113

114   We next assessed the genome-wide overlap between chromatin occupancy of PRC2 (FIE) and the two TF

115   families (AZF1 and BPC1) in 30-hour-old plants, a stage when PRC2 function becomes essential for

116   plant development[29]. Significant ($Q<10^{-10}$) binding peaks of FIE, AZF1 and BPC1 co-localized with each

117   other and with H3K27me3 peaks both globally and at individual loci (Fig. 4a,b and Supplementary Fig.

118   5). 23% of all FIE-bound regions overlapped with BPC1 peaks, while 28% overlapped with AZF1 peaks

119   (Fig. 4c). This overlap was significantly larger than expected by chance for the peak associated genes

120   ($P<10^{-307}$ and $P<10^{-105}$ (hypergeometric test), respectively). In total, 1804 FIE peaks (35%) overlapped

121   with an AZF1 or a BPC1 peak and at 42% of these peaks both TFs were present (Fig. 4d). Cognate

122   binding motifs (GA repeat and telobox) of the class I BPC and the C1-2iD TFs were also significantly

123   enriched under the FIE binding peaks (Supplementary Fig. 5d). Gene Ontology term enrichment analysis

124   links the FIE, AZF1and BPC1 targets to shoot development, flower patterning and gynoecium

125   development (Supplementary Fig. 5e).

126

127   Characteristic phenotypes of mutants in the PRC2 methyltransferase CLF are upwards curled leaves with

128   partial floral identity as well as precocious flowering[2]. Higher order mutants in the class I BPC TFs or the

129   C1-2iD C2H2 ZnF TFs do not exhibit these phenotypes[30,31] (Supplementary Fig. 6), suggesting

130   combinatorial roles for the two TF families in Polycomb silencing. Two pieces of evidence support this

131   idea. Firstly, knockdown of either TF family significantly enhanced the leaf curling and the precocious

132   flowering of the hypomorph *clf*[R] mutant[18] (Supplementary Figs. 7, 8). Secondly, simultaneous knockdown

133   of both TF families (*BPC+ZnF*[KD]) in the wild type triggered upwards leaf curling and precocious

134   flowering (Fig. 5a,b). The *BPC*[KD]*clf*[R], *ZnF*[KD]*clf*[R] and *BPC+ZnF*[KD] phenotypes were accompanied by a

135   significant reduction in PRC2 (FIE) occupancy and in H3K27 trimethylation at Polycomb target loci and

136   by significant de-repression of the Polycomb targets (Fig. 5c-e and Supplementary Figs. 7-9). FIE

137   occupancy -at peaks with $Q<10^{-10}$ in the wild type- was also reduced genome-wide in 30 hr old

138   *BPC+ZnF*[KD] plants (Fig. 5f,g and Supplementary Fig. 10). By contrast, occupancy of BPC1 at a

139   Polycomb target locus was not dependent on presence of PRC2 (Supplementary Figs. 11, 12).

140

141   Finally, we assessed the contribution of AZF1 and BPC1 to PRC2 recruitment by reciprocal gain of

142   function tests. Tethering both the AZF1 and the BPC1 TF to an artificial promoter in isolated plant cells

143   (protoplasts) triggered levels of FIE recruitment similar to those observed at an endogenous Polycomb

144   target locus (Fig. 6a). In addition, overexpression of BPC1 or AZF1 in clf^R plants restored PRC2

145   occupancy and H3K27me3 at Polycomb target loci to near wild-type levels and largely rescued the leaf

146   curling defect of clf^R (Supplementary Fig. 12). These findings suggest combinatorial roles for class I BPC

147   and C1-2iD ZnF TFs in Polycomb silencing and PRC2 recruitment.

148

149   Here we uncover a PRC2 recruitment strategy in Arabidopsis that is strikingly similar to that in

150   Drosophila – including roles for GA repeat-binding and Zn-finger TFs in recruitment (Fig. 6b)[7-9]. Further

151   support for our findings comes from a recent study of PRC2 recruitment to the *ABI4* locus[32]. Our data

152   suggest that a similar logic underpins PRC2 recruitment in species from two kingdoms of life. The plant

153   PREs we uncovered may recruit both PRC2 and PRC1, since GA repeat and telobox motifs also link to

154   PRC1 occupancy in Arabidopsis[26,27]. Additional PREs besides those we predicted likely exist and may

155   act at different stages and in different tissues or conditions. Likewise, additional determinants of PRC2

156   recruitment remain unidentified. Their discovery combined with the current data should enable

157   computational prediction of PREs for future epigenetic reprograming of cell identity or function to

158   enhance plant growth and yield.

159

160   **URLs.**

161   R packages used:"vioplot" https://cran.r-project.org/web/packages/vioplot/index.html; leoss

162   smoothing https://stat.ethz.ch/R-manual/R-devel/library/stats/html/loess.html , PCC analysis

163   https://stat.ethz.ch/R-manual/R-devel/library/stats/html/cor.html, PCA analysis https://stat.ethz.ch/R-

164   manual/R-devel/library/stats/html/prcomp.html.

165   .

181

AUTHOR CONTRIBUTIONS

182

183   D.W. and J.X. conceived of the study and J.X. conducted the majority of the experiments. O.C. and J.G.

184   generated the CLF ChIP-chip dataset. J.P.P. and S.A.K. conducted the high throughput DNA interactome

185   screen. R.A. performed the motif analysis, C.H. contributed to functional PRE analysis and R.J. and M.S.

186   to transcription factor/PRC2 interaction tests. A.P., C.S. and M.Z. identified telobox binding TFs under

187   the guidance of J.W. Cho.H. and R.B. performed library preps. R.B., X.Y., S.K. X.Z. and A.S. conducted

188   bioinformatic analysis. M.F. developed the labeling protocol for EMSA. S.C. and X.L. raised anti-BPC1

189   antisera. D.W. wrote the paper with input from all authors.

190

COMPETING FINANCIAL INTERESTS

191

192   The authors declare no competing financial interests.

193

194   **References**
195
196   1.   Förderer, A., Zhou, Y. & Turck, F. The age of multiplexity: Recruitment and interactions of
197   Polycomb complexes in plants. in *Current Opinion in Plant Biology* Vol. 29 169-178 (2016).
198   2.   Goodrich, J. *et al.* A Polycomb-group gene regulates homeotic gene expression in Arabidopsis.
199   *Nature* **386**, 44-51 (1997).
200   3.   Mozgova, I. & Hennig, L. The polycomb group protein regulatory network. *Annu Rev Plant Biol*
201   **66**, 269-296 (2015).
202   4.   Pu, L. & Sung, Z.-R. PcG and trxG in plants - friends or foes. *Trends in genetics : TIG* **31**, 252-62
203   (2015).
204   5.   Xiao, J. & Wagner, D. Polycomb repression in the regulation of growth and development in
205   Arabidopsis. *Current opinion in plant biology* **23**, 15-24 (2015).
206   6.   Lewis, E.-B. A gene complex controlling segmentation in Drosophila. *Nature* **276**, 565-70
207   (1978).
208   7.   Blackledge, N.-P., Rose, N.-R. & Klose, R.-J. Targeting Polycomb systems to regulate gene
209   expression: modifications to a complex story. *Nat Rev Mol Cell Biol* **16**, 643-9 (2015).
210   8.   Simon, J.-A. & Kingston, R.-E. Occupying Chromatin: Polycomb Mechanisms for Getting to
211   Genomic Targets, Stopping Transcriptional Traffic, and Staying Put. *Molecular Cell* **49**, 808-824
212   (2013).
213   9.   Kassis, J.-A. & Brown, J.-L. Polycomb group response elements in Drosophila and vertebrates.
214   *Adv Genet* **81**, 83-118 (2013).
215   10.   Sun, B. *et al.* Timing mechanism dependent on cell division is invoked by Polycomb eviction in
216   plant stem cells. *Science* **343**, 1248559 (2014).
217   11.   Lodha, M., Marco, C.-F. & Timmermans, M.-C. The ASYMMETRIC LEAVES complex
218   maintains repression of KNOX homeobox genes via direct recruitment of Polycomb-repressive
219   complex2. *Genes Dev* **27**, 596-601 (2013).
220   12.   Berger, N., Dubreucq, B., Roudier, F., Dubos, C. & Lepiniec, L. Transcriptional regulation of
221   Arabidopsis LEAFY COTYLEDON2 involves RLE, a cis-element that regulates trimethylation
222   of histone H3 at lysine-27. *Plant Cell* **23**, 4065-4078 (2011).
223   13.   Yuan, W. *et al.* A cis cold memory element and a trans epigenome reader mediate Polycomb
224   silencing of FLC by vernalization in Arabidopsis. *Nat Genet* **48**, 1527-1534 (2016).
225   14.   Qüesta, J.-I., Song, J., Geraldo, N., An, H. & Dean, C. Arabidopsis transcriptional repressor
226   VAL1 triggers Polycomb silencing at FLC during vernalization. *Science (New York, N.Y.)* **353**,
227   485-8 (2016).
228   15.   Deng, W. *et al.* Arabidopsis Polycomb Repressive Complex 2 binding sites contain putative
229   GAGA factor binding motifs within coding regions of genes. *BMC genomics* **14**, 593 (2013).
230   16.   Kim, S.-Y., Lee, J., Eshed-Williams, L., Zilberman, D. & Sung, Z.-R. EMF1 and PRC2 cooperate
231   to repress key regulators of Arabidopsis development. *PLoS Genet* **8**, e1002512 (2012).
232   17.   Zhang, X. *et al.* Whole-genome analysis of histone H3 lysine 27 trimethylation in Arabidopsis.
233   *PLoS Biology* **5**, 1026-1035 (2007).
234   18.   Lopez-Vernaza, M. *et al.* Antagonistic roles of SEPALLATA3, FT and FLC genes as targets of
235   the polycomb group gene CURLY LEAF. *PLoS One* **7**, e30715 (2012).
236   19.   Pruneda-Paz, J.-L. *et al.* A genome-scale resource for the functional characterization of
237   Arabidopsis transcription factors. *Cell Rep* **8**, 622-632 (2014).
238   20.   Kodaira, K.-S. *et al.* Arabidopsis Cys2/His2 zinc-finger proteins AZF1 and AZF2 negatively
239   regulate abscisic acid-repressive and auxin-inducible genes under abiotic stress conditions. *Plant*
240   *Physiol* **157**, 742-756 (2011).
241   21.   Kim, S., Soltis, P.-S., Wall, K. & Soltis, D.-E. Phylogeny and domain evolution in the
242   APETALA2-like gene family. *Mol Biol Evol* **23**, 107-120 (2006).
243   22.   Monfared, M.-M. *et al.* Overlapping and antagonistic activities of BASIC PENTACYSTEINE
244   genes affect a range of developmental processes in Arabidopsis. *Plant J* **66**, 1020-1031 (2011).

245   23.   Winter, C.-M. *et al*. LEAFY target genes reveal floral regulatory logic, cis motifs, and a link to
246   biotic stimulus response. *Developmental cell* **20**, 430-43 (2011).
247   24.   Molitor, A.-M. *et al*. The Arabidopsis hnRNP-Q Protein LIF2 and the PRC1 subunit LHP1
248   function in concert to regulate the transcription of stress-responsive genes. *The Plant Cell*,
249   tpc.00244.2016 (2016).
250   25.   Wang, H. *et al*. Arabidopsis Flower and Embryo Developmental Genes are Repressed in
251   Seedlings by Different Combinations of Polycomb Group Proteins in Association with Distinct
252   Sets of Cis-regulatory Elements. *PLoS Genetics* **12**, 1-25 (2016).
253   26.   Zhou, Y., Hartwig, B., Velikkakam James, G., Schneeberger, K. & Turck, F. Complementary
254   activities of TELOMERE REPEAT BINDING proteins and Polycomb Group complexes in
255   transcriptional regulation of target genes. *The Plant Cell* **28**, TPC2015-00787-RA (2015).
256   27.   Hecker, A. *et al*. The Arabidopsis GAGA-Binding Factor BASIC PENTACYSTEINE6 Recruits
257   the POLYCOMB-REPRESSIVE COMPLEX1 Component LIKE HETEROCHROMATIN
258   PROTEIN1 to GAGA DNA Motifs. *Plant physiology* **168**, 1013-24 (2015).
259   28.   Kooiker, M. *et al*. BASIC PENTACYSTEINE1, a GA binding protein that induces
260   conformational changes in the regulatory region of the homeotic Arabidopsis gene SEEDSTICK.
261   *Plant Cell* **17**, 722-729 (2005).
262   29.   Bouyer, D. *et al*. Polycomb repressive complex 2 controls the embryo-to-seedling phase
263   transition. *PLoS Genetics* **7**(2011).
264   30.   Ciftci-Yilmaz, S. & Mittler, R. The zinc finger network of plants. *Cell Mol Life Sci* **65**, 1150-
265   1160 (2008).
266   31.   Meister, R.-J. *et al*. Definition and interactions of a positive regulatory element of the Arabidopsis
267   INNER NO OUTER promoter. *Plant J* **37**, 426-438 (2004).
268   32.   Mu, Y. *et al*. BASIC PENTACYSTEINE Proteins Repress ABSCISIC ACID INSENSITIVE4
269   Expression via Direct Recruitment of the Polycomb-Repressive Complex 2 in Arabidopsis Root
270   Development. *Plant Cell Physiol* **58**, 607-621 (2017).
271

Figure legends
273
274 **Figure 1** Identification of *Arabidopsis* DNA fragments with PRE activity.
275 (**a**) Construct to test ability of candidate PREs or control DNA fragments (NC) to recruit PRC2 and
276 H3K27me3. Below: region tested by ChIP-qPCR.
277 (**b**) Top: Occupancy of the PRC2 component FIE at fragments tested. Bottom: H3K27me3 accumulation
278 relative to H3 at fragments tested. Shown are mean ± SEM of three independent ChIP experiments (red
279 dots). **, P< 0.01; ns, not significant P>0.05 one-tailed unpaired *t* test relative to NC_1. The *LEC2* PRE[12],
280 serves as positive control (PC_LEC2). See also Supplementary Fig. 2a,b.
281 (**c**) Construct to test ability of PREs to silence active reporters. Candidate PREs and control fragments
282 were placed between two constitutive promoters (pF3H and pMAS) driving expression of GFP, GUS or
283 an herbicide resistance gene (Bar), respectively.
284 (**d**) GFP intensity visually scored (1 = no reporter expression to 5 = full reporter expression) in 15
285 independent transformants in the wild type (WT) (top) or a *prc2* mutant (*clf-28 swn-7*) (bottom). Violin
286 plot of GFP intensities of presumptive PREs (left) and negative controls (right): range, median= white
287 circle, mean= white line, lower to higher quartile= vertical black line. Black bar: median GFP
288 fluorescence of the PRE populations in the wild type background. *, P<0.05; **, P<0.01; ns, not
289 significant (P>0.16) relative to NC_1, one-tailed Mann-Whitney *U* test. Similar results were obtained for
290 silencing of beta-glucuronidase activity and herbicide resistance (Supplementary Fig. 2c-e).
291
292 **Figure 2** PRE-binding TFs physically interact with PRC2.
293 (**a**) TF families most enriched as PRE-binding based on high-throughput yeast-one-hybrid assays. Y-axis:
294 P-value hypergeometric test. Inset: Most enriched TF subfamilies.
295 (**b, c**) Interaction between PRE-binding TFs and PRC2 by medium throughput yeast-two-hybrid test for
296 all 55 PRE-binding TFs (b) and quantitative ortho-Nitrophenyl-β-galactoside (*ONPG*) assays for
297 members of the most enriched TF subfamilies (c). Negative controls: LFY TF and AD alone. Mean ±
298 SEM of three independent yeast-two-hybrid experiments (red dots). *, P<0.05; **, P< 0.01; ns, not
299 significant P> 0.05 relative to BD, one-tailed unpaired *t*-test.
300 (**d, e**) Interaction between PRE-binding TFs and the PRC2 complex (represented by the single copy PRC2
301 component FIE) by bimolecular fluorescence complementation in protoplasts. TFs tested (subfamily):
302 TOE1 (AP2), AZF1 (C1-2iD Zn Finger) and BPC1 (class I BPC) and controls: BBM (AP2/ANT), ZAT5
303 (C1-2iC Zn Finger), BPC6 (class II BPC) plus LFY. Nuclear fluorescence (d) and quantification (e). Bar
304 =10 µm. Box and whisker plot with median of three BiFC experiments comprising 150 cells each (red
305 line), upper and lower quartile (box) and minima, maxima (whiskers). P-value: one tailed Mann-Whitney
306 *U* test, ***, P<0.001 relative to controls.
307 (**f**) Co-immunoprecipitation of HA-tagged PRC2 (FIE) after IP of MYC tagged TFs in seedlings. TRB2:
308 negative control. % input values represent mean ± SEM of relative IP (normalized to input FIE-HA and
309 TF levels) from three experiments. Supplementary Fig. 3 shows co-IP in the presence of an endonuclease.
310
311 **Figure 3** Cis motifs enriched in plant PREs required for PRE activity.
312 (**a, b**) Sequence logos for position-specific scoring matrices (PSSMs) of motifs identified by *de novo*
313 motif analysis (a) and their enrichment (p-values; converted from the Z-scores of the motif enrichment
314 calculations using a normal distribution) and frequency (%) in the 170 candidate PREs (b).
315 (**c**) Test of GA repeat and telobox function. Top: PRE Diagram indicating GA repeats and telobox motifs
316 mutated in PREs or added to NC_1. Below: Effect of mutation of GA repeat and telobox motifs in the
317 AG_2 and At5g PREs (AG_2_mu or At5g_mu) or insertion of GA repeat and telobox motifs into the
318 negative control fragment NC_1 (NC+) on PRE activity in populations of independent primary
319 transformants (n=15) in the wild type (WT) or the *prc2* mutant (*clf-28*). Violin plot: range, median= white
320 circle, mean= white line, lower to higher quartile= vertical black line. *, p<0.05; **, p<0.01; one-tailed
321 Mann-Whitney *U* test.

322   (**d**) Evolutionarily conservation of GA repeats and telobox motifs in the AG_2 PRE and At5g_PRE in
323   *Brassicaceae* species. Vertical lines: additional PRE sequences omitted, *fully conserved nucleotide; :
324   and . partially conserved nucleotide.
325   (**e**) Electrophoretic mobility shift assay (EMSA) to test association of the AZF1 C1-2iD ZnF TF with the
326   telobox but not mutant versions thereof. % complex: fraction of shifted DNA (mean ± SEM) from three
327   independent EMSA experiments. HIS-TF, unrelated TF control.
328
329   **Figure 4** ChIPseq analysis to test chromatin occupancy of FIE, H3K27me3, AZF1 and BPC1 in 30-hour-
330   old plants.
331   (**a**) Heatmap of background-corrected ChIP enrichment for PRC2 (FIE), H3K27, AZF1 and BPC1.
332   Significant peaks ($Q<10^{-10}$) were centered on FIE peak maxima and rank ordered from highest (top) to
333   lowest (bottom) FIE binding peak significance. For each factor and mark, three independent ChIPseq
334   experiments were performed and sequenced as were three matched input controls. The highly consistent
335   ChIP replicates were normalized by sequencing depth and averaged.
336   (**b**) Browser view of input subtracted ChIPseq signals at *AG*, *LEC2* (PC) and *ACT2* (NC_3) and additional
337   FIE-bound loci. Significant peaks ($Q < 10^{-10}$) according to MACS2 are marked by horizontal bars, with
338   the black saturation proportional to the Q value (as for the narrowPeak file format by ENCODE). As
339   previously reported for seedlings[15] about half of all FIE peaks overlapped with H3K27me3.
340   (**c**) Percent ChIPseq overlap (by row) for the significant ($Q < 10^{-10}$) FIE, AZF1 or BPC1 peaks with other
341   significant peaks. Shading indicates strength of overlap.
342   (**d**) Fraction of significant ($Q < 10^{-10}$) FIE peaks that overlap with other significant peaks.
343
344   **Figure 5** Class I BPC and C1-2iD ZnF TF families are required for Polycomb silencing and PRC2
345   recruitment *in planta*.
346   (**a, b**) Flowering time and leaf curling in double BPC and ZnF family knockdown plants ($BPC+ZnF^{KD}$)
347   compared to wild type (WT), hypomorph clf^{R} and null *clf-50* mutants. Bar = 1cm. (b) Quantification of
348   leaf curling. Box and whisker plot with median (red line; n=15 plants), upper and lower quartile (box) and
349   minima, maxima (whiskers). Different letters above bars indicate significantly different groups, p<0.05
350   based on Kruskal-Wallis test with Dunn's posthoc test.
351   (**c-e**) PRC2 (FIE) occupancy, H3K27me3/H3 accumulation and gene expression in genotypes described
352   in (a) and in BPC^{KD} clf^{R} or ZnF^{KD} clf^{R} plants. Negative control loci: *ACT2* for ChIP, *EIF4* for gene
353   expression. Expression in the mutant lines is shown relative to the wild type. Mean ± SEM from three
354   independent experiments (red dots). Black asterisks- significantly different from WT; grey asterisks –
355   significantly different from clf^{R}. * P<0.05; ** P< 0.01, ns P>0.05; one-tailed unpaired student *t* test.
356   (**f, g**) ChIPseq analysis of FIE occupancy in wild type (WT) and $BPC+ZnF^{KD}$ plants. For each factor and
357   mark, three independent ChIP experiments were sequenced as were three matched input controls. The
358   highly consistent ChIP replicates were normalized by sequencing depth and averaged. (f) FIE occupancy
359   ($Q < 10^{-10}$) region metaplot in wild type (WT) and $BPC+ZnF^{KD}$ samples, (g) screenshots of input
360   subtracted FIE occupancy. red: previous FIE ChIPseq (Fig. 4), blue: FIE ChIPseq data in WT (center) and
361   in $BPC+ZnF^{KD}$ (bottom).
362
363   **Figure 6** Tethering class I BPC and C1-2iD ZnF TF family to the DNA and test of PRC2/FIE recruitment.
364   (**a**) Tethering of BPC1 (BPC-LexA DBD) and AZF1 (ZnF-Gal4 DBD) to an artificial promoter
365   (LG_GUS) and test of FIE recruitment using anti –FIE antibody for ChIP qPCR in protoplasts.
366   Endogenous loci tested: *AGAMOUS* (AG_2, positive control (PC)) and *ACT2* (negative control (NC)).
367   Empty vector (EV) and VP16 serve as control tethering vectors. Shown are mean ± SEM of three ChIP
368   experiments (red dots). * P<0.05; ** P< 0.01; relative to empty vector (EV), one-tailed unpaired *t* test. (**b**)
369   Model for PRC2 recruitment. See text for details.

370   **METHODS:**

371   **Plant material and treatment**

372   Mutants and transgenic plants previously described: *clf-28, clf-28 swn-7*[33]; *clf-50, clf*[R] (pCLF:CLF-GR

373   *clf-50*)[18]; *bpc123*[22] pFIE:FIE-HA *fie-11*[34]; pEMF2:EMF2-3XFLAG *emf2*[35]; pMSI1::GFP-MSI1 *msi1-1*[36];

374   35S::GFP-CLF *clf-50*[37]. In *clf*[R], the *clf-50* RNA null mutant is partly rescued by 'leaky' nuclear

375   translocation of pCLF:CLF-GR in the absence of steroid treatment. *clf*-50 and *clf*[R] are in the Ws

376   accession, pFIE:FIE-HA *fie-11* is in the C24 accession, all other plants are in the Col-0 accession.

377

378   **PRE tests**

379   To test PRC2 recruitment and H3K27me3 by PREs or control fragments, progeny pools of 40 random T1

380   plants were analyzed. This strategy was adopted to minimize outliers caused –for example- by position

381   effect. Independent pools tested gave similar results. Control fragments included a known PRE

382   (PC_LEC2) and three random DNA fragments (NC_1, intron of At1g60200; NC_2, promoter of AG;

383   NC_3, 3'UTR of Actin 2 (At3g18780)). GFP intensity was scored visually using a dissecting

384   fluorescence microscope (Olympus, MVX10). A fluorometric (MUG) assay was used to quantify beta-

385   glucuronidase (GUS) activity as previously described[38,39]except that the 4-MU produced was normalized

386   over the fresh weight of each plant. GFP and GUS reporter silencing was assayed in independent primary

387   transformants (T1 plants). For the herbicide resistance assay, primary transformants were transplanted

388   into soil after selection and sprayed with the Basta herbicide (200mg/l) (Bayer Crop Science) 2-3 days

389   after transplanting. Survival rate was scored 5-7 days later.

390

391   **Transgenic plants**

392   Candidate PRE DNA fragments (~600bp in length) were cloned into pFK205[40] and transformed into wild

393   type (Col-0), pMSI1:GFP-MSI1 *msi1-1* or pEMF2:EMF2-3XFLAG *emf2* for chromatin

394   immunoprecipitation (ChIP) assays. Plants were selected on 1/2 Murashige and Skoog (MS) medium[41]

395   (Sigma) with 40mg/l kanamycin.

396

397   A dual reporter system (pPRE-dual-rep) was generated containing the FLAVANONE 3-

398   HYDROXYLASE regulatory region plus the 35S minimal promoter (pF3H-35S mini) driving expression

399   of GFP and beta-glucuronidase and the mannopine synthase promoter (pMAS) driving the BAR gene.

400   Candidate PRE or control fragments were cloned into pPRE-dual-rep and transformed into wild type

401   (Col-0) or *prc2* mutants (*clf*-28, *clf-28 swn-7/+*). PRE fragments for test of loss- or gain- of GA repeat

402   and telobox motifs were synthesized (see Supplementary Table 5) (GenScript Inc. Company, Piscataway,

403   NJ, US) and shuffled into pPRE-dual-rep. Plants were selected on 1/2 MS plates with 25mg/l

404   hygromycin.

405

406   For TF knockdown, 300 bp regions conserved in the PRC2 recruiting TF subfamilies (e.g. C1-2iD Zn

407   finger) but not in the larger TF family (C1-2i Zn finger) were PCR amplified and inserted into vector

408   pRNAi-GG. Plasmids were introduced into *A. tumefaciens* strain GV3101 and transformed into *clf^R* and

409   wild-type plants using floral dip[42]. T1 plants were selected on 1/2 MS plates with 40mg/l kanamycin.

410   *BPC^{KD}* plants were crossed with *ZnF^{KD}* plants to generate double knockdown lines.

411

412   For overexpression of BPC1 or AZF1 in *clf^R*, the cDNA of BPC1 and AZF1 was cloned into vector

413   pGWB12[43]. Plasmids were introduced into *A. tumefaciens* strain GV3101 and transformed into *clf^R* using

414   floral dip[42]. T1 plants were selected on 1/2 MS plates with 40mg/l kanamycin and 20mg/l hygromycin.

415

416   For TF CoIP and ChIP, genomic fragments spanning the upstream intergenic region and the coding

417   region for each TF were PCR amplified [BPC1 (-3225 to +849 bp), AZF1 (-1628 to +735 bp), TOE1 (-

418   8685 to +1353 bp), TRB2 (-1194 to +1227 bp)], cloned into pEG303[44] and transformed into pFIE:FIE-

419   HA *fie-11* for CoIP and into wild type (Col-0) for ChIP.

420

Phenotype quantification and qRT-PCR

422   Plants were grown at 22 in short day conditions (8 hr light/ 16hr dark, light intensity: ~140 μMol/m^{2}s^1).

423   The length and width of the blade of the fifth rosette leaf was measured for > 15 plants at day 20. RNA

424   was extracted from 3- to 7-day-old plants grown in short day condition and quantitative real-time PCR

425   was performed as previously described[45].

426

Identification of CLF-binding sites using ChIP-chip

428   ChIP was conducted with anti-GFP antibodies (Molecular Probes) in 35S::GFP-CLF *clf-50* plants[37] as

429   previously described[17]. Four independent replicates were performed. Amplification, labeling and

430   microarray hybridization was as reported[46]. Genomic regions enriched for CLF binding sites were

431   identified by comparing CLF ChIP-chip to input DNA (4 replicates) using the Tilemap program with the

432   Hidden Markov model (HMM) option[47]. Adjacent probes with HMM posterior probabilities of p>0.5 or

433   higher were merged into regions by requiring a minimal run of 50 bp and allowing a maximal gap of 200

434   bp, and TAIR5 coordinates were converted to TAIR10 coordinates using the "update_coordinates.pl"

435   script from TAIR. The 3,648 regions enriched for CLF binding are listed in Supplementary Table 1.

436

Identification of candidate PREs

437 Putative PRE-containing regions were identified based on the following two conservative criteria. First,

439 we compared previously published genome-wide distribution of H3K27me3, EMF1 and FIE[15-17] with that

440 of CLF generated in this study. EMF1 is a putative PRC1 component that frequently co-localizes with

441 PRC2[16]. A candidate PRE-containing region was required to overlap with at least 3 of the 4 datasets (to

442 account for potential false negative results in these dataset and the redundant contribution of other PRC

443 components). A total of 1,504 regions were identified and assigned to 851 Arabidopsis genes based on

444 previously described criteria[23]. Second, we required target genes associated with these putative PREs be

445 expressed in a highly tissue-specific manner or be de-repressed in *prc2* mutants (*clf swn* and *fie*)[15,33]. For

446 the former, we analyzed previously published transcription profiles in different tissues[48], and defined

447 tissue-specifically expressed genes as those with expression levels higher than 5x the baseline levels.

448 Since all ChIP data (FIE, CLF, EMF1, H3K27me3) was from vegetative development, baseline was

449 defined as the mean expression in samples ATGE_7; ATGE_87; ATGE_12; ATGE_26; ATGE_1;

450 ATGE_19; ATGE_15; ATGE_13; ATGE_20; ATGE_21; ATGE_14; ATGE_17; ATGE_18; ATGE_91;

451 ATGE_5; ATGE_16; ATGE_11; ATGE_10 from AtGen Express[48]. PRE linked genes de-repressed in *clf*

452 *swn* or *fie* were defined as in[15,33]. The combined expression filters resulted in identification of 132 high-

453 confidence PRC2-regulated genes. 170 candidate PREs were associated with the 132 PRC2-regulated

454 genes. Five of the candidate PREs were selected for *in planta* PRE tests.

Motif prediction and mapping

457 *De novo* motif prediction was performed as previously published with minor changes[23,49-51]. Briefly, we

458 applied the motif prediction pipeline to a subset of the candidate PREs (those bound by CLF, EMF2 and

459 marked by H3K27me3, 70 PREs) to enhance prediction performance and lower false positive rates[52].

460 Motifs were subsequently tested for enrichment within the entire PRE set. Motif width was set from 6-

461 16bp when applicable. The background set consisted of 600bp genic terminal regions from the TAIR10

462 genome[15]. Motif enrichment was calculated using a non-parametric deterministic sampling as described

463 previously ($Z \geq 3$; $p < 0.027$)[23]. Highly degenerate motifs were filtered from the results using an

464 information quality statistic ($IQ > 20$) defined as:

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

465 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

466 of type A at position L in the PSSM, and P_A is the background frequency expected for residue A.

468   Predicted PSSMs were aligned by position of maximum average site-wise Euclidean distance and

469   hierarchically clustered in the R statistical programming environment. A representative consensus PSSM

470   was then chosen from the PSSMs in each clade, along with a merged candidate, by maximizing for

471   significance score and frequency within target genes. A functional-depth cutoff was used in mapping

472   PSSMs to PREs and to genomic and ChIP data sequence sets (GAGA: 0.62, CTCC: 0.42 and CCG: 0.8

473   and 0 for telobox, CAA repeats and G-box). When mapping GAGA motif occurrences, overlapping and

474   nearby matches were merged into a common region using the BEDTools merge function with a maximum

475   distance between features of 8bp[53].

476

477   **Yeast-two-hybrid tests**

478   EMF2, CLF, MSI1, and FIE were cloned into pDEST32 (Clontech) and introduced into yeast strain

479   AH109 (MAT a). Full-length clones were used except for EMF2 and CLF. EMF2C is a better interactor

480   in yeast, while CLFN overcomes the growth defects caused by full-length CLF[54]. The 55 PRE-interacting

481   TFs were cloned into pDEST22 (Clontech) and transfected into yeast strain Y187 (MAT α).

482   Protein−protein interactions were tested after mating as described in the Matchmaker protocol (Clontech).

483   Interaction strength was quantified for a subset of TFs using ortho-nitrophenyl-β-D-galactopyranoside

484   (ONPG) assays[55] for nine independent colonies for each interaction pair in three pools. In addition,

485   different TF fragments (AZF1-N: 1-90 aa; AZF1-M: 91-194 aa; AZF1-C: 195-245 aa; BPC1-N: 1-140 aa;

486   BPC1-C: 141-283 aa; TOE1-N: 1-151 aa; TOE1-M: 152-310 aa; TOE1-C: 311-464 aa) were cloned into

487   pDEST22 and co-transformed with EMF2C or CLFN into yeast strain AH109 followed by scoring of

488   yeast growth.

489

490   **Bimolecular Fluorescence Complementation (BiFC) assays**

491   FIE and all TFs were cloned into pUC-SPV-NE[GW] and pUC-SPV-CE[GW] constructed by shuffling the split

492   Venus-Gateway cassette from pDEST-VYNE/CE(R)[GW] vectors[56] into pUC18. BiFC assays in

493   Arabidopsis protoplasts were conducted and visualized as previously described[57,58]. For each experiment,

494   fluorescence was compared in protoplast populations prepared and transfected at the same time. Three

495   independent BiFC experiments were performed for each combination of factors tested with at least150

496   protoplasts scored per replicate. Representative images were taken with a confocal microscope with the

497   same gain (Leica, LCS SL).

498

499   **CoIP**

500   Co-IP was performed as described[59] with some modifications. Myc-gBPC1, Myc-gAZF1, Myc-gTOE1 or

501   Myc-gTRB2 (TELOMER REPEAT BINDING 2 protein, negative control) were transformed into

502   pFIE:FIE-HA *fie-11* plants. Three-day-old double transgenic seedlings were harvested after growth in

503   long day conditions. Tissue was ground in protein extraction buffer (20 mM Tris-HCl pH8.0, 150 mM

504   NaCl, 1 mM EDTA, 10% Glycerol, 0.2% Triton X-100, 1 mM PMSF, 1×Protease inhibitor (Roche)),

505   filtered, and centrifuged. The supernatant was incubated with anti-Myc antibody (C3965 or 05-724,

506   Sigma) coupled to protein A Dynabeads (ThermoFisher) overnight at 4 degrees. Beads were washed four

507   times with wash buffer, and bound proteins eluted with elution buffer containing 2% SDS for

508   immunoblotting with HA-HRP conjugated antibody (3F10, Roche). In some reactions Benzonase

509   endonuclease (E1014, Sigma) was added into the protein sample (20 U) and incubated on ice for 1hr prior

510   to immunoprecipitation. Full length gen images for Western analyses after co-IP and for protein

511   abundance in BiFC experiments are sown in Supplementary Figure S13.

512

513   **ChIP**

514   ChIP was performed as previously described[60], with minor modifications. Extraction buffer I (0.4M

515   Sucrose, 10mM Tris-HCl pH8.0, 10mM MgCl2, 5mM b-ME, 1mM PMSF, 1X Protease inhibitor

516   (Roche)) and II (0.25M Sucrose, 10mM Tris-HCl pH8.0, 10mM MgCl2, 1% Triton X-100, 5mM b-ME,

517   1mM PMSF, 1X Protease inhibitor (Roche)) were used for protein extraction. The following antibodies

518   that had previously been used for ChIP in Arabidopsis were employed: anti-GFP (A6455, Thermo

519   Fisher)[61] for pMSI1:GFP-MSI1 *msi1-1* ChIP, anti-HA (12CA5, Roche)[61] for pFIE:FIE-HA *fie-11* ChIP,

520   anti- H3K27me3 (07-449, Millipore)[33,62], anti-H3 (07-690, Millipore)[62], anti-FLAG (F3165, Sigma)[63] for

521   pEMF2:EMF2-3XFLAG *emf2* ChIP and anti-Myc (C3956, Sigma)[64] for gAZF1-Myc and gBPC1-Myc

522   ChIP. Anti-FIE[34] antiserum was first used for ChIP here and showed much reduced occupancy in *prc2*

523   mutants (Fig. 5c). The anti-BPC1 antiserum was generated in rabbits using full length recombinant BPC1

524   protein and gave ChIP signal specifically in the wild type (Supplementary Fig. 11). Throughout

525   H3K27me3 was normalized over H3 to control for nucleosome density, all other ChIP reactions are

526   shown as % input. For most ChIP experiments, representative transgenic lines were used, alternatively

527   pools of 40 independent T1 progeny (or more) were used.

528

529   **ChIP-seq in germinating embryos**

530   For H3K27me3, FIE, AZF1 and BPC1 ChIPseq, ChIP was performed on germinating embryos (30 hrs

531   after imbibition) as described above, but eluted into a smaller volume (15ul in total), using the Qiagen

532   MinElute PCR purification kit (Cat. No. 28004). Three independent ChIP and input reactions were

533   sequenced. ChIP-DNA and input (after dilution to 0.1~1 ng) were amplified using the SeqPlex DNA

534   Amplification kit (Sigma, SEQXE-10RXN) according to the manufacturer's instructions with the

535   following modifications: a first linear PCR was followed by a second round of amplification (<10 cycles).

536 After primer removal, qPCR was performed to test amplification of different genomic regions. Linearly

537 amplified DNA from input chromatin and pull-downs was converted to libraries for sequencing by

538 performing end-repair followed by A-tailing and ligation of universal adapters (all enzymes by

539 Enzymatics, MA). Libraries were amplified to 50–100 nM using custom dual indexing primers and

540 sequenced with an Illumina NextSeq500 at a depth of > 15 million reads per sample for pull-downs and >

541 30 million reads per sample for inputs. We mapped reads to the TAIR10 genome release v31 using

542 bowtie2[65]. Alignment files were converted to 1-bp resolution bigwig files and normalized by 10 million

543 reads sequenced (RP10M) using custom scripts. All ChIPseq replicates were highly similar to each other

544 (see below). Bigwigs from the ChIPseq replicates were averaged using WiggleTools[66]. Significant peaks

545 were identified with MACS2 version 2.1.1.20160309 using relevant input controls (C24 for FIE, Col for

546 AZF1 and BPC1, total H3 pull-down for H3K27me3). Default MACS2 settings were used for FIE, AZF1,

547 and BPC1, and the "--broad" option was used for H3K27me3. Only peaks with Q-value $< 10^{-10}$ were

548 considered.

549

550 Significant peaks were mapped to genes as previously described[23] using the HOMER[67] script

551 annotatePeaks. Gene Ontology (GO) enrichment analysis was performed using AgriGO[68] combined with

552 manual curation to remove redundant terms.

553

554 For FIE ChIPseq in WT and double TF subfamily (class I BPC and C1-2iD ZnF) knock-down lines, anti-

555 FIE[34] antiserum was used. ChIP-DNA and input libraries (three replicates each) were generated using the

556 ThruPLEX DNA-seq kit (RUBICON GENOMICS, cat. R400406). Libraries were sequenced at a depth of

557 > 15 million reads per sample for pull-downs and > 30 million reads per sample for inputs. Replicate

558 comparison and significant peak identification was as described above using default MACS2 settings.

559 Only peaks with $Q < 10^{-10}$ were considered.

560

561 To assess the change in FIE binding in $BPC + ZnF^{KD}$ mutants relative to the wild type (WT), normalized

562 reads mapping to WT FIE peaks (Fig. 5f) were extracted from WT and $BPC + ZnF^{KD}$ FIE ChIP datasets

563 using bedtools and plotted [$\log_2$(RPM)] in a scatterplot. Reads increased or decreased in $BPC + ZnF^{KD}$

564 were indicated (P-value <0.01 and 2-fold change). For the region metaplot, detected FIE peaks were

565 quantified using the normalized and input-subtracted number of reads in each 10 bp window (FIE

566 peak signal) in WT and $BPC + ZnF^{KD}$. The average FIE peak signals in the window of +/- 1000 bp

567 around peak centers were plotted with the "loess" smoothing function in R with span = 0.2.

568

569   ChIP replicate concordance was assessed by computing pairwise Pearson Correlation Coefficients (PCC)

570   of the RP10M normalized read counts in each 10 kb window for all ChIPseq datasets using the "cor"

571   function with method "pearson" in R. PCC was > 0.95 for all ChIPseq replicates. Principal component

572   analysis (PCA) was performed on the pairwise PCC matrix using the "prcomp" function in R and

573   converted to a scatterplot. Significant regions for all ChIPseq experiments performed in 30-hr-old plants

574   (germinating embryos) are listed in Supplementary Table 4.

575

576   **EMSA**

577   A 60 bp fragment of the AG_2 PRE containing two GA repeats or one telobox, as well as versions thereof

578   with motif substitutions were labeled with Cy5-dCTP (GE Healthcare Life Sciences) by end-repair with

579   Klenow Fragment (3´→ 5´ exo-) (NEB). For EMSA and other oligonucleotide sequences see

580   Supplementary Table 5. Briefly, complementary single strand DNA probes were synthesized (IDT) that,

581   when annealed, gave rise to a two nucleotide 3' overhang (with the last annealed nucleotide being a G).

582   After Cy5 labeling, the probe was purified using Illustra MicroSpin G-25 columns (GE Healthcare Life

583   Sciences).

584

585   Full-length *BPC1*, *AZF1* cDNAs were cloned into pET32a and transformed into in *E. coli* (BL21). Protein

586   was isolated as previous described[69]. For binding assays, a 20 µL reaction containing 2 µL of the protein

587   extract, 3 µL of 1 pg/uL probe and 2 µL of 10× binding buffer (0.1 M Tris-HCl at pH 7.5, 0.5 M NaCl, 10

588   mM DTT, 10 mM EDTA, 50% glycerol, 0.5 mg/mL poly(dI-dC),1 mg/mL BSA) was used. For the AZF1

589   Zn-finger TF, the binding reaction was supplemented with 1mM $CaCl_2$ and 0.1 mM $ZnCl_2$. Free and

590   bound probes were separated on a 6% PAGE gel in 0.5× TBE at 100 V for 60 min. The gel was scanned

591   by a Typhoon scanner (Typhoon 9410 variable mode imager, Amersham) at BP 670 with a gain of 600.

592   Shifted and unshifted probes were quantified using Image J. Full length EMSA gel images are provided in

593   Supplementary Figure S13.

594

595   **Tethering assay**

596   BPC1 and AZF1 were cloned into LexA_DBD and Gal4_DBD vectors, respectively, as effectors[70]. The

597   reporter contained 2 repeats each of the LexA and Gal4 binding sites with a 35S minimal promoter

598   driving β-glucuronidase (GUS)[70]. 35S: LUC (firefly luciferase) was included to monitor transfection

599   efficiency. Empty Gal4_DBD or Gal4_DBD_VP16 and LexA_DBD_VP16 served as controls. 16 hrs

600   after transfection, $2x10^5$ protoplasts were used for ChIP (anti-FIE antiserum)[34].

601

Y1H screen

Functionally defined PREs (AG, SEP3, At5g) were cloned as 300bp DNA fragments into a modified pLacZi vector (Clontech) carrying *gLUC* instead of *LacZ* as reporter gene and integrated into the genome of yeast strain YM4271. A robotic Y1H screen was carried out against 1956 Arabidopsis TFs as previously described but using luciferase activity as readout[19]. The identified PRE-interacting TFs were further filtered on binding to multiple PRE fragments and binding strength.

For identification of telobox motif binding transcription factors, a telobox motif was inserted into a 30 bp region from a negative control fragment (NC_3; At3g18780 Actin2; Chr3: 6,476,550-6,476,579). Three copies of the 30 bp NC_3 region with telobox motif were inserted into bait vector pAbAi (Clontech), followed by integration into the genome of yeast strain Y1HGold (Clontech). Y1H screening against a TF library with about 1400 transcription factors[71] was carried out according to the Matchmaker® Gold Yeast One-Hybrid Library Screening System manual (Clontech), using Aureobasidin A resistance to select for binding.

Phylogenetic analyses

For phylogenetic trees, amino acid sequence alignment and generation of a neighbor-Joining (NJ) phylogenetic tree was performed using MEGA[72] with default settings. Phylogenetic shadowing was performed essentially as previously described[61], 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–Smirnov[73] 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 employed for two group comparisons (Mann-Whitney *U*-test[74]) and for multiple group comparison (Kruskal–Wallis test[75] combined with the Dunn's posthoc[76] test). Variances in some of the groups compared differ for biological reasons. Rejecting the null hypothesis based on these tests for this type of data implies that one group stochastically dominates a second group, that is to say 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 based on prior studies that

635 showed significant effects using similar samples sizes, for example see[57]. P-values for all tests

636 performed, as well as additional statistical parameters are listed in Supplementary Table 6.

637

638 **Data availability**

639 ChIPchip (CLF) and ChIPseq (all others) data were deposited in the Gene Expression Omnibus (GEO)

640 under series numbers GSE7065 (CLF) and GSE7063 (input), GSE84483 (FIE, H3K27me3, AZF1 and

641 BPC1) and GSE95562 (FIE in wild type and in $BPC+ZnF^{KD}$).

642

643 **Code availability**

644 Motif prediction programs are available from their individual websites: Meme v4.9.1, Weeder v1.4.2,

645 AlignAce v4.0, MotifSampler v3.2 and Bioprospector v2004. Motif enrichment and downstream analysis

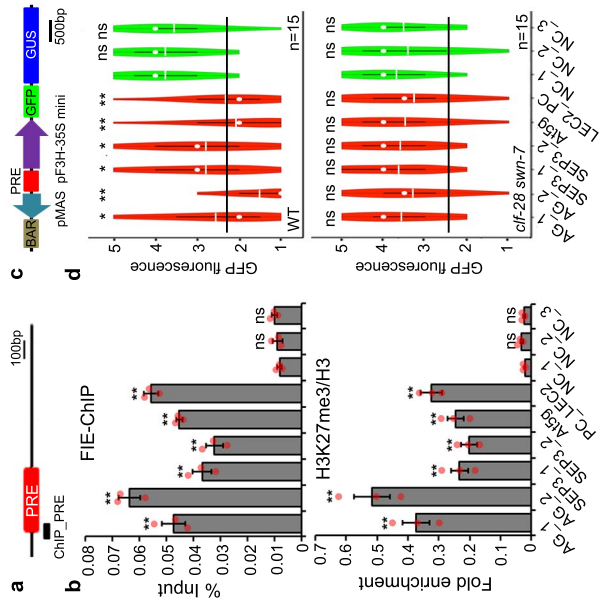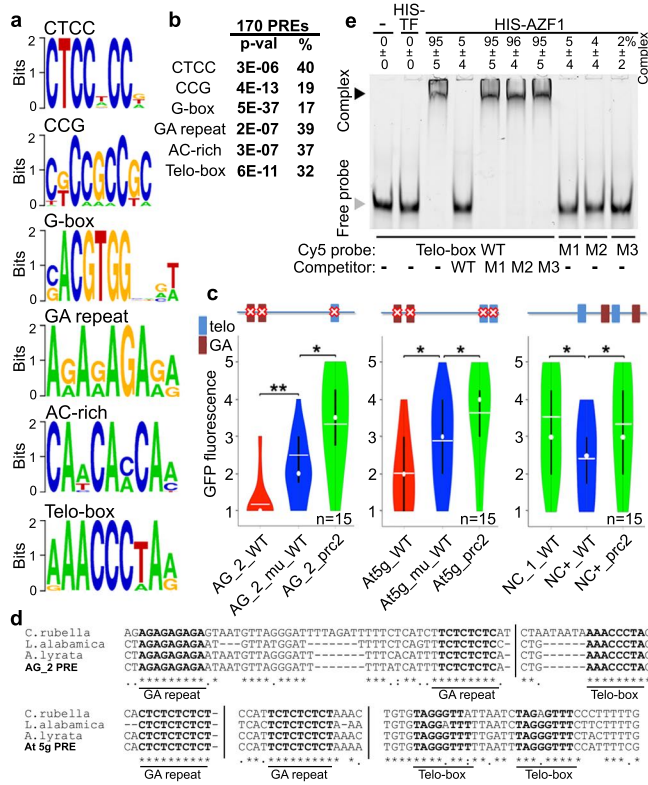646 used the "Cistome"[77] pipeline, which is freely available as a web-based.
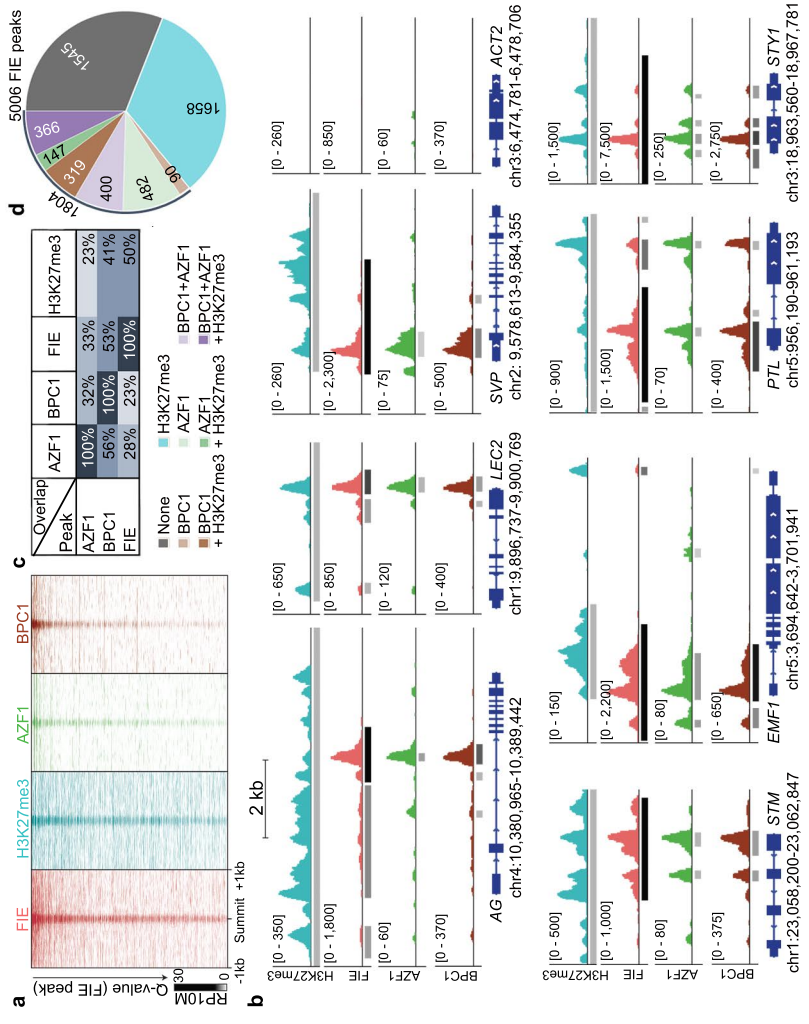
647

648 **Methods only references**

649 33. Lafos, M. *et al*. Dynamic regulation of H3K27 trimethylation during Arabidopsis differentiation.
650   *PLoS genetics* **7**, e1002040 (2011).
651 34. Wood, C.-C. *et al*. The Arabidopsis thaliana vernalization response requires a polycomb-like
652   protein complex that also includes VERNALIZATION INSENSITIVE 3. *Proc Natl Acad Sci U S*
653   *A* **103**, 14631-14636 (2006).
654 35. Kim, S.-Y., Zhu, T. & Sung, Z.-R. Epigenetic regulation of gene programs by EMF1 and EMF2
655   in Arabidopsis. *Plant Physiol* **152**, 516-528 (2010).
656 36. Alexandre, C., Moller-Steinbach, Y., Schonrock, N., Gruissem, W. & Hennig, L. Arabidopsis
657   MSI1 is required for negative regulation of the response to drought stress. *Mol Plant* **2**, 675-687
658   (2009).
659 37. Schubert, D. *et al*. Silencing by plant Polycomb-group genes requires dispersed trimethylation of
660   histone H3 at lysine 27. *EMBO J* **25**, 4638-4649 (2006).
661 38. Maghuly, F. *et al*. Stress regulated expression of the GUS-marker gene (uidA) under the control
662   of plant calmodulin and viral 35S promoters in a model fruit tree rootstock: Prunus incisa x
663   serrula. *J Biotechnol* **135**, 105-16 (2008).
664 39. Blazquez, M. Quantitative GUS Activity Assay of Plant Extracts. *CSH Protoc* **2007**, pdb
665   prot4690 (2007).
666 40. Mathieu, J., Yant, L.-J., Murdter, F., Kuttner, F. & Schmid, M. Repression of flowering by the
667   miR172 target SMZ. *PLoS Biol* **7**, e1000148 (2009).
668 41. Murashige, T. & Skoog, F. A Revised Medium for Rapid Growth and Bio Assays with Tobacco
669   Tissue Cultures. *Physiologia Plantarum* **15**, 473-497 (1962).
670 42. Clough, S.-J. & Bent, A.-F. Floral dip: a simplified method for Agrobacterium-mediated
671   transformation of Arabidopsis thaliana. *Plant J* **16**, 735-743 (1998).
672 43. Nakagawa, T. *et al*. Development of series of gateway binary vectors, pGWBs, for realizing
673   efficient construction of fusion genes for plant transformation. *J Biosci Bioeng* **104**, 34-41 (2007).
674 44. Earley, K.-W. *et al*. Gateway-compatible vectors for plant functional genomics and proteomics.
675   *Plant J* **45**, 616-629 (2006).
676 45. Yamaguchi, A. *et al*. The microRNA-regulated SBP-Box transcription factor SPL3 is a direct
677   upstream activator of LEAFY, FRUITFULL, and APETALA1. *Dev Cell* **17**, 268-278 (2009).
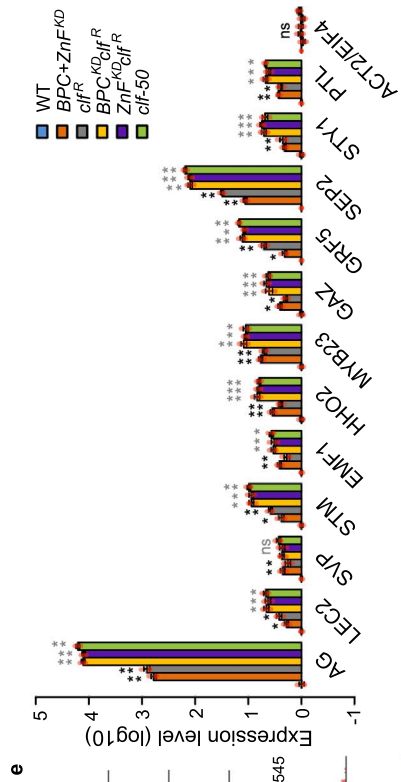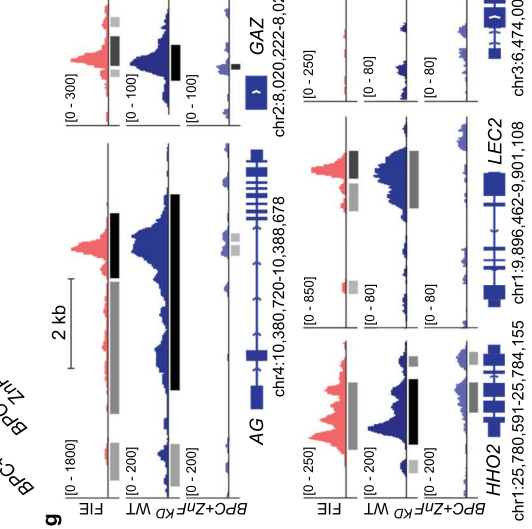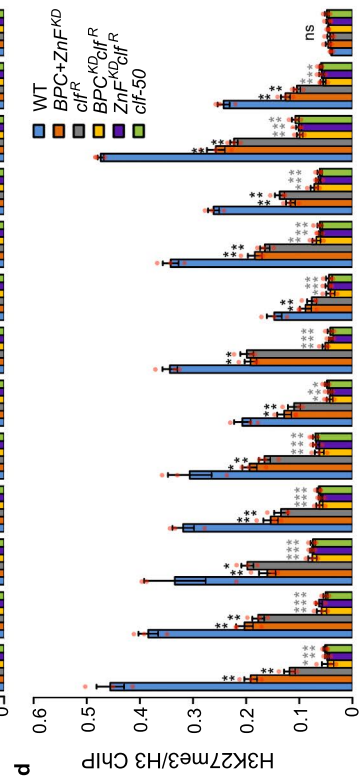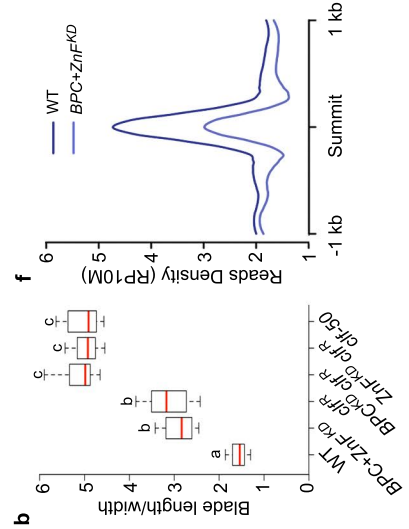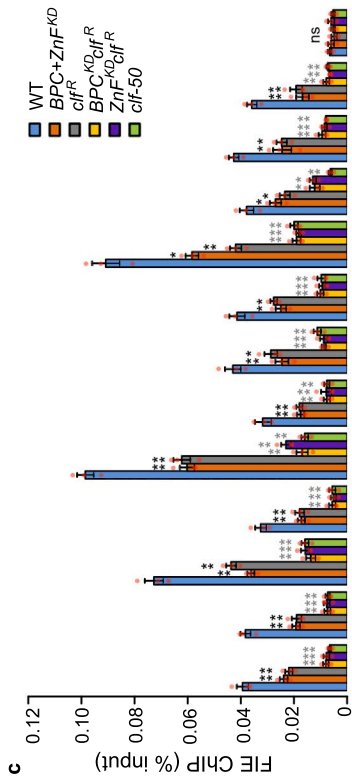
678   46.   Zhang, X. *et al*. Genome-wide high-resolution mapping and functional analysis of DNA
679   methylation in arabidopsis. *Cell* **126**, 1189-1201 (2006).
680   47.   Ji, H. & Wong, W.-H. TileMap: create chromosomal map of tiling array hybridizations.
681   *Bioinformatics* **21**, 3629-3636 (2005).
682   48.   Schmid, M. *et al*. A gene expression map of Arabidopsis thaliana development. *Nat Genet* **37**,
683   501-506 (2005).
684   49.   Bailey, T.-L. & Elkan, C. The value of prior knowledge in discovering motifs with MEME. *Proc*
685   *Int Conf Intell Syst Mol Biol* **3**, 21-29 (1995).
686   50.   Hughes, J.-D., Estep, P.-W., Tavazoie, S. & Church, G.-M. Computational identification of cis-
687   regulatory elements associated with groups of functionally related genes in Saccharomyces
688   cerevisiae. *J Mol Biol* **296**, 1205-1214 (2000).
689   51.   Pavesi, G., Mauri, G. & Pesole, G. An algorithm for finding signals of unknown length in DNA
690   sequences. *Bioinformatics* **17 Suppl 1**, S207-14 (2001).
691   52.   Hu, J., Li, B. & Kihara, D. Limitations and potentials of current motif discovery algorithms.
692   *Nucleic Acids Res* **33**, 4899-913 (2005).
693   53.   Quinlan, A.-R. & Hall, I.-M. BEDTools: A flexible suite of utilities for comparing genomic
694   features. *Bioinformatics* **26**, 841-842 (2010).
695   54.   Chanvivattana, Y. *et al*. Interaction of Polycomb-group proteins controlling flowering in
696   Arabidopsis. *Development* **131**, 5263-5276 (2004).
697   55.   Wang, H. *et al*. Yeast two-hybrid system demonstrates that estrogen receptor dimerization is
698   ligand-dependent in vivo. *J Biol Chem* **270**, 23322-23329 (1995).
699   56.   Gehl, C., Waadt, R., Kudla, J., Mendel, R.-R. & Hansch, R. New GATEWAY vectors for high
700   throughput analyses of protein-protein interactions by bimolecular fluorescence
701   complementation. *Mol Plant* **2**, 1051-1058 (2009).
702   57.   Wu, M.-F. *et al*. Auxin-regulated chromatin switch directs acquisition of flower primordium
703   founder fate. *Elife* **4**, e09269 (2015).
704   58.   Yoo, S.-D., Cho, Y.-H. & Sheen, J. Arabidopsis mesophyll protoplasts: a versatile cell system for
705   transient gene expression analysis. *Nat Protoc* **2**, 1565-1572 (2007).
706   59.   Xiao, J. *et al*. O-GlcNAc-mediated interaction between VER2 and TaGRP2 elicits TaVRN1
707   mRNA accumulation during vernalization in winter wheat. *Nat Commun* **5**, 4572 (2014).
708   60.   Yamaguchi, N. *et al*. PROTOCOLS: Chromatin Immunoprecipitation from Arabidopsis Tissues.
709   *Arabidopsis Book* **12**, e0170 (2014).
710   61.   Yamaguchi, N. *et al*. A molecular framework for auxin-mediated initiation of flower primordia.
711   *Dev Cell* **24**, 271-282 (2013).
712   62.   Aichinger, E. *et al*. CHD3 proteins and polycomb group proteins antagonistically determine cell
713   identity in Arabidopsis. *PLoS Genet* **5**, e1000605 (2009).
714   63.   Yamaguchi, N. *et al*. Gibberellin acts positively then negatively to control onset of flower
715   formation in Arabidopsis. *Science* **344**, 638-41 (2014).
716   64.   Li, K. *et al*. DELLA-mediated PIF degradation contributes to coordination of light and gibberellin
717   signalling in Arabidopsis. *Nat Commun* **7**, 11868 (2016).
718   65.   Langmead, B. & Salzberg, S.-L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-
719   359 (2012).
720   66.   Zerbino, D.-R., Johnson, N., Juettemann, T., Wilder, S.-P. & Flicek, P. WiggleTools: Parallel
721   processing of large collections of genome-wide datasets for visualization and statistical analysis.
722   *Bioinformatics* **30**, 1008-1009 (2014).
723   67.   Heinz, S. *et al*. Simple combinations of lineage-determining transcription factors prime cis-
724   regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-89 (2010).
725   68.   Du, Z., Zhou, X., Ling, Y., Zhang, Z. & Su, Z. agriGO: a GO analysis toolkit for the agricultural
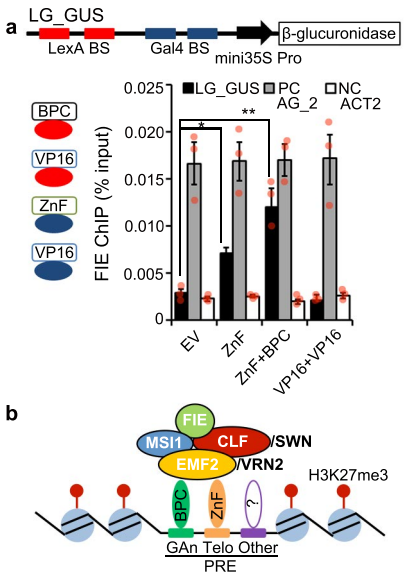726   community. *Nucleic Acids Res* **38**, W64-70 (2010).

727   69.   Zhang, B., Wang, L., Zeng, L., Zhang, C. & Ma, H. Arabidopsis TOE proteins convey a
728   photoperiodic signal to antagonize CONSTANS and regulate flowering time. *Genes Dev* **29**, 975-
729   987 (2015).
730   70.   Tiwari, S.-B., Hagen, G. & Guilfoyle, T.-J. Aux/IAA proteins contain a potent transcriptional
731   repression domain. *The Plant cell* **16**, 533-43 (2004).
732   71.   Castrillo, G. *et al*. Speeding cis-trans regulation discovery by phylogenomic analyses coupled
733   with screenings of an arrayed library of Arabidopsis transcription factors. *PLoS One* **6**, e21524
734   (2011).
735   72.   Tamura, K. *et al*. MEGA5: molecular evolutionary genetics analysis using maximum likelihood,
736   evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**, 2731-9 (2011).
737   73.   Massey, F.J. The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American
738   Statistical Association* **46**, 68–78 (1951).
739   74.   Mann, H.-B. & Whitney, D.-R. On a Test of Whether one of Two Random Variables is
740   Stochastically Larger than the Other. *Annals of Mathematical Statistics* **18**, 50-60 (1947).
741   75.   Kruskal, W.-H. & Wallis, W.-A. Use of Ranks in One-Criterion Variance Analysis. *Journal of
742   the American Statistical Association* **47**, 583-621 (1952).
743   76.   Dunn, O.J. Multiple comparisons using rank sums. *Technometrics* **6**, 241–252 (1964).
744   77.   Austin, R.-S. *et al*. New BAR Tools for Mining Expression Data and Exploring Cis-Elements in
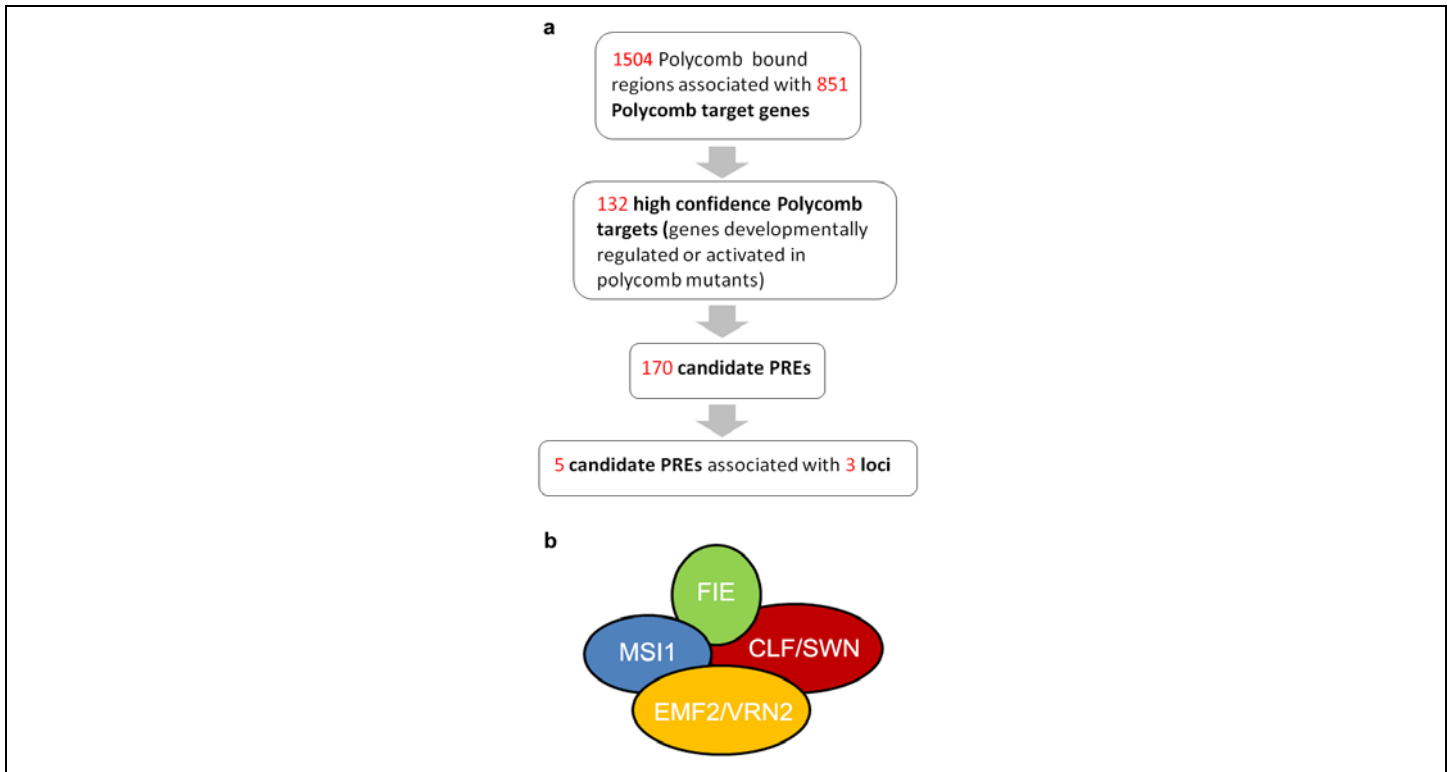745   Arabidopsis thaliana. *Plant J* (2016).
746

a

CTCC

CCG

G-box

GA repeat

AC-rich

Telo-box

b

<u>170 PREs</u>

	p-val	%
CTCC	3E-06	40
CCG	4E-13	19
G-box	5E-37	17
GA repeat	2E-07	39
AC-rich	3E-07	37
Telo-box	6E-11	32

e

	−	HIS-TF		HIS-AZF1						
	0 ± 0	0 ± 0	95 ± 5	5 ± 4	95 ± 5	96 ± 4	95 ± 5	5 ± 4	4 ± 4	2% ± 2

Complex

Free probe

Cy5 probe: Telo-box WT | M1 M2 M3
Competitor: − − − WT M1 M2 M3 − − −

c

telo
GA

GFP fluorescence

** *

AG_2_WT AG_2_mu_WT AG_2_prc2   n=15

* *

At5g_WT At5g_mu_WT At5g_prc2   n=15

* *

NC_1_WT NC+_WT NC+_prc2   n=15

d

```
C.rubella    AGAGAGAGAGAGTAATGTTAGGGATTTTAGATTTTTCTCATCTTCTCTCTCAT   CTAATAATAAAACCCTACAT
L.alabamica  CTAGAGAGAGAAT---GTTATGGAT-------TTTTCTCAGTTTCTCTCTCA-   CTG-----AAACCCTAGA
A.lyrata     CTAGAGAGAGAATAATGTTAGGGATT------TTTCACATTTTCTCTCTCA-   CTG-----AAACCCTAGA
AG_2 PRE     CTAGAGAGAGAATAATGTTAGGGATT------TTTATCATTTTCTCTCTCA-   CTGAGAGAAAACCCTAGA
             ..*********.*   ****.****   ***.:**.**********           **..   **********
             └GA repeat┘                                    └Telo-box┘        └Telo-box┘

C.rubella    CACTCTCTCTCT-   CCATTCTCTCTAAAC   TGTGTAGGGTTATTAATCTAGAGTTTCCCTTTTTG
L.alabamica  --CTCTCTCTCT-   TCACTCTCTCTA-AA   TGTGTAGGATTTTAATCTAGGGTTTCTTCTTTTTG
A.lyrata     CACTCTCTCTCT-   CCATTCTCTCTAAAC   TGTGTAGGGTTTTTAATTTAGGGTTTCTACTTTTG
At5g PRE     CACTCTCTCTCT-   CCATTCTCTCTAAAA   TGTGTAGGGTTTTAATTTAGGGTTTCCATTTTCG
             .**********    .*.*.*********.*    *******.**.*.***.**..*****        ***.*
             └GA repeat┘    └GA repeat┘     └Telo-box┘        └Telo-box┘
```

a

BPC1 AZF1 H3K27me3 FIE

RP10M

Q-value (FIE peak) 0 30

−1kb Summit +1kb   −1kb +1kb

b

H3K27me3 FIE AZF1 BPC1

AG chr4:10,380,965-10,389,442 — [0 - 350], [0 - 1,800], [0 - 60], [0 - 370]

LEC2 chr1:9,896,737-9,900,769 — [0 - 650], [0 - 850], [0 - 120], [0 - 400]

SVP chr2: 9,578,613-9,584,355 — [0 - 260], [0 - 2,300], [0 - 75], [0 - 500]

ACT2 chr3:6,474,781-6,478,706 — [0 - 260], [0 - 850], [0 - 60], [0 - 370]

STM chr1:23,058,200-23,062,847 — [0 - 500], [0 - 1,000], [0 - 80], [0 - 375]

EMF1 chr5:3,694,642-3,701,941 — [0 - 150], [0 - 2,200], [0 - 80], [0 - 650]

PTL chr5:956,190-961,193 — [0 - 900], [0 - 1,500], [0 - 70], [0 - 400]

STY1 chr3:18,963,560-18,967,781 — [0 - 1,500], [0 - 7,500], [0 - 250], [0 - 2,750]

2 kb

c

Peak \ Overlap	AZF1	BPC1	FIE	H3K27me3
AZF1	100%	32%	33%	23%
BPC1	56%	100%	53%	41%
FIE	28%	23%	100%	50%

None; BPC1; BPC1 + H3K27me3; H3K27me3; AZF1; AZF1 + H3K27me3; BPC1+AZF1; BPC1+AZF1 + H3K27me3

d

5006 FIE peaks

1658; 1545; 366; 147; 319; 400; 482; 90; 1804

a

LG_GUS

LexA BS | Gal4 BS | mini35S Pro | β-glucuronidase

BPC

VP16

ZnF

VP16

■ LG_GUS ■ PC AG_2 □ NC ACT2

FIE ChIP (% input)

0.025
0.020
0.015
0.010
0.005
0

*   **

EV   ZnF   ZnF+BPC   VP16+VP16

b

FIE
MSI1   CLF /SWN
EMF2 /VRN2

BPC ZnF ?

H3K27me3

GAn Telo Other
PRE

Supplementary Figure 1

High confidence PRC2 targets and candidate PREs.

(a) Flowchart for identification of candidate Arabidopsis PREs. We identified 1504 genomic regions marked by at least 3 of the following: H3K27me3, FIE, CLF or EMF1[15-17] (CLF ChIP-chip data: GSE7065) and linked these to 851 genes as previously described[23]. 132 of the 851 genes were significantly upregulated in *prc2* mutants[15,33] or strongly developmentally regulated[48] and thus considered high confidence PRC2 regulated genes. 170 candidate PREs were associated with the 132 genes. From the 170 PREs we selected 5 associated with 3 genes for test of PRE activity. See methods for additional details.

(b) Composition of the PRC2 complexes in the Arabidopsis sporophyte (diploid generation). Complex components: one of two SET domain methyltransferases (CURLY LEAF (CLF) or SWINGER (SWN)), one of two VEFS domain proteins (EMBRYONIC FLOWER2 (EMF2) or VERNALIZATION2 (VRN2)), a WD40 domain protein that can recognize H3K27me3 (FERTILIZATION INPENDPENT ENDOSPERM (FIE)) and a histone binding protein (MSI1)[3].
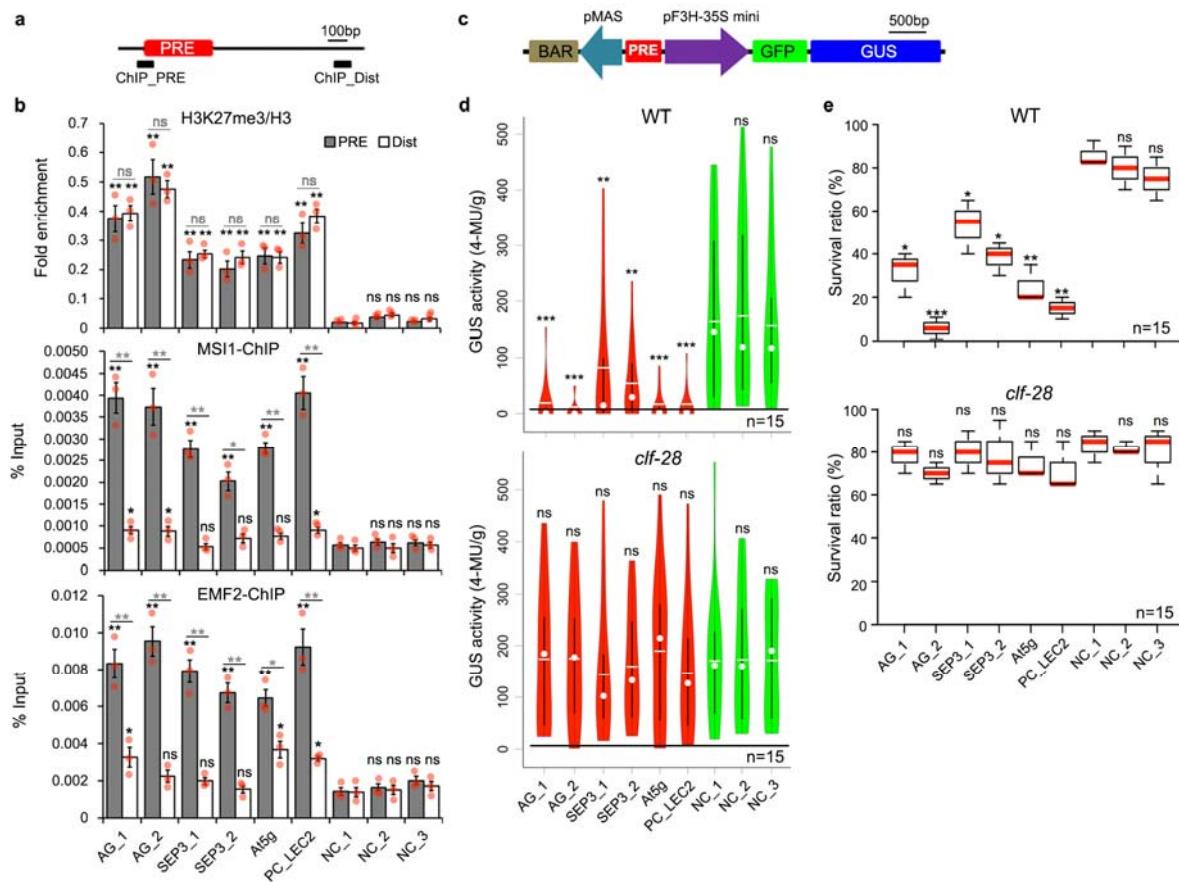
Supplementary Figure 2

Test of PRE activity.

(a) Construct to test ability of candidate PREs or control DNA fragments (NC) to recruit PRC2 and H3K27me3. Below: regions tested: PRE or a distal site (Dist).

(b) H3K27me3/H3 abundance (top) and occupancy of PRC2 components MSI1 (middle) and EMF2 (bottom) assessed by ChIP-qPCR. Mean ± SEM from three experiments (red dots). Black asterisks- significantly different relative to NC_1; grey asterisks – significantly different occupancy at the PREs relative to the distal site (Dist). * P<0.05; ** P< 0.01, ns P>0.05; one-tailed unpaired student *t* test.

(c) Construct to test PRE-mediated reporter silencing.

(d) Fluorometric assay of beta-glucuronidase (GUS) activity of 15 independent transformants in the wild type (WT) (top) or a *PRC2* mutant (*clf-28*) (bottom). Violin plot of GUS activity in PREs (left) or negative controls (right): range, median= white circle, mean= white line, lower to higher quartile= vertical black line. Black bar: median GUS activity of all PRE populations in the wild-type background. * P<0.05; ** P<0.01; *** P<0.001; ns, not significant (P>0.25) relative to NC_1, one-tailed Mann-Whitney *U* test.

(e) Herbicide resistance (survival rate) conferred by the *BAR* gene product in n=60 independent T1 plants in the wild-type (top) or in the *prc2* (*clf-28*) mutant (bottom) background. Box and whisker plot with median (red line), upper and lower quartile (box) and minima, maxima (whiskers). P -value (one-tailed Mann–Whitney *U* test): ns, not significant; P> 0.07; * P<0.05; ** P<0.01; *** P<0.001 relative to NC_1.
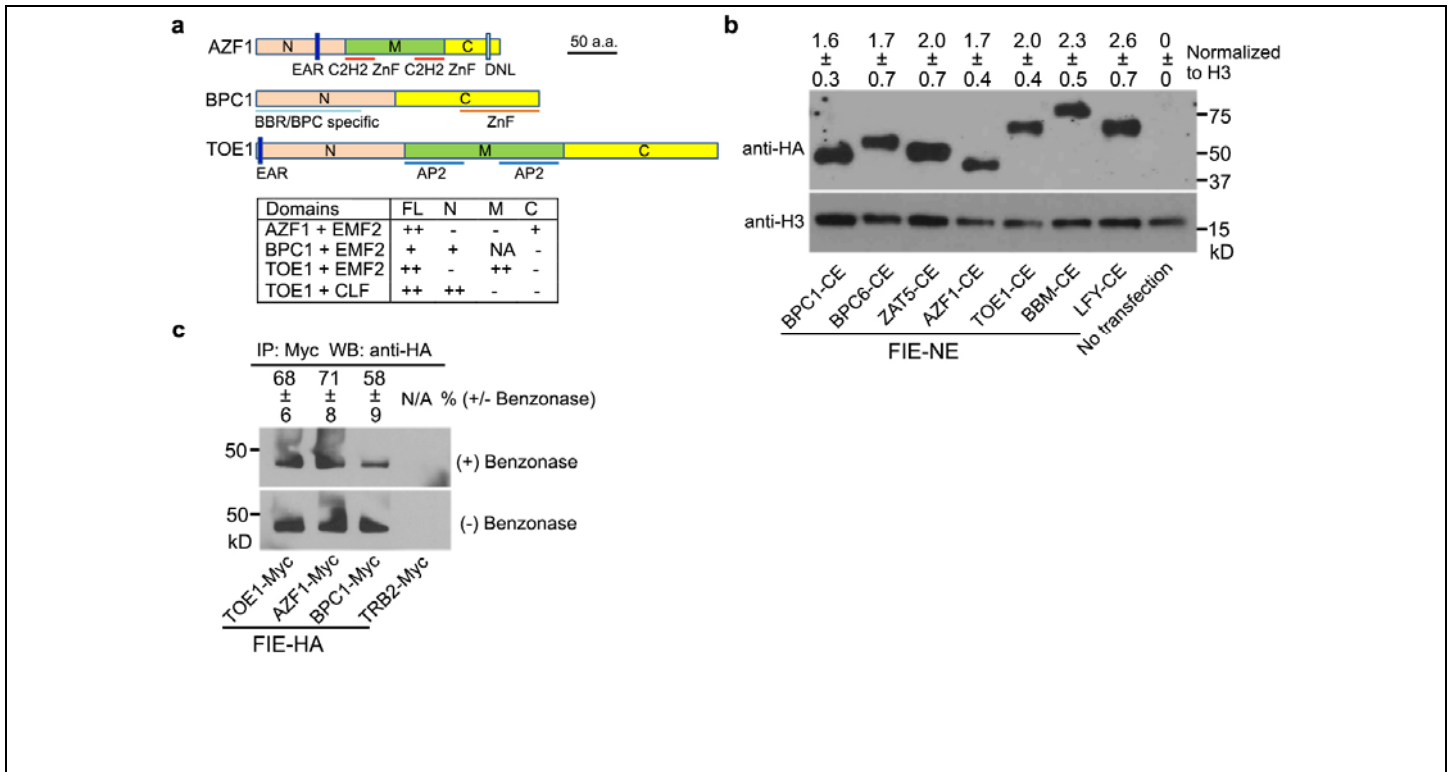
Supplementary Figure 3

**Physical interaction of PRE-binding transcription factors with PRC2.**

(a) Domains of TFs important for interaction with PRC2. Yeast-two-hybrid interaction tests using EMF2 and CLF as bait and AZF1, BPC1 and TOE1 full-length proteins (FL), N-terminal domains (N), middle regions (M), or C-terminal domains (C) as prey. Known protein motifs[20,21,78] are shown below. NA = Not applicable. No single known protein motif is responsible for the TF/PRC2 interactions. a.a.: amino acid.

(b) Detection of TF abundance in plant cells co-transfected with FIE-NE and used for the BiFC analyses in Fig. 2d, e. Above: mean ± SEM of TF levels relative to those of histone H3 from three independent BiFC experiments.

(c) Co-IP of Myc-tagged TFs and PRC2 in the presence or absence of the benzonase endonuclease[79]. TOE1, AZF1 and BPC1 co-immunoprecipitate with PRC2 (HA-tagged FIE) in the absence and presence of the nuclease. Above: FIE co-IP in the presence relative to absence of benzonase, mean ± SEM from two independent co-IP experiments. The TRB2 sequence-specific binding protein serves as negative control. For expression levels of all 4 Myc-tagged proteins see Fig. 2f.

78.   Wanke, D. *et al.* Alanine zipper-like coiled-coil domains are necessary for homotypic dimerization of plant GAGA-factors in the nucleus and nucleolus. *PLoS One* **6**, e16070 (2011).
79.   Fiil, B.-K., Qiu, J.-L., Petersen, K., Petersen, M. & Mundy, J. Coimmunoprecipitation (co-IP) of Nuclear Proteins and Chromatin Immunoprecipitation (ChIP) from Arabidopsis. *CSH Protoc* **2008**, pdb prot5049 (2008).
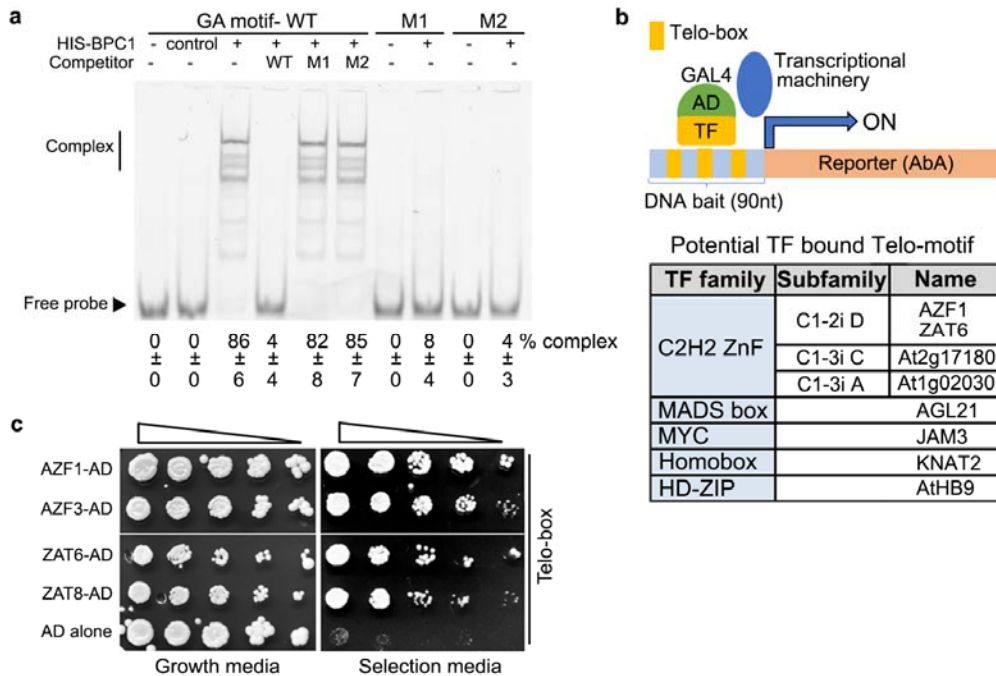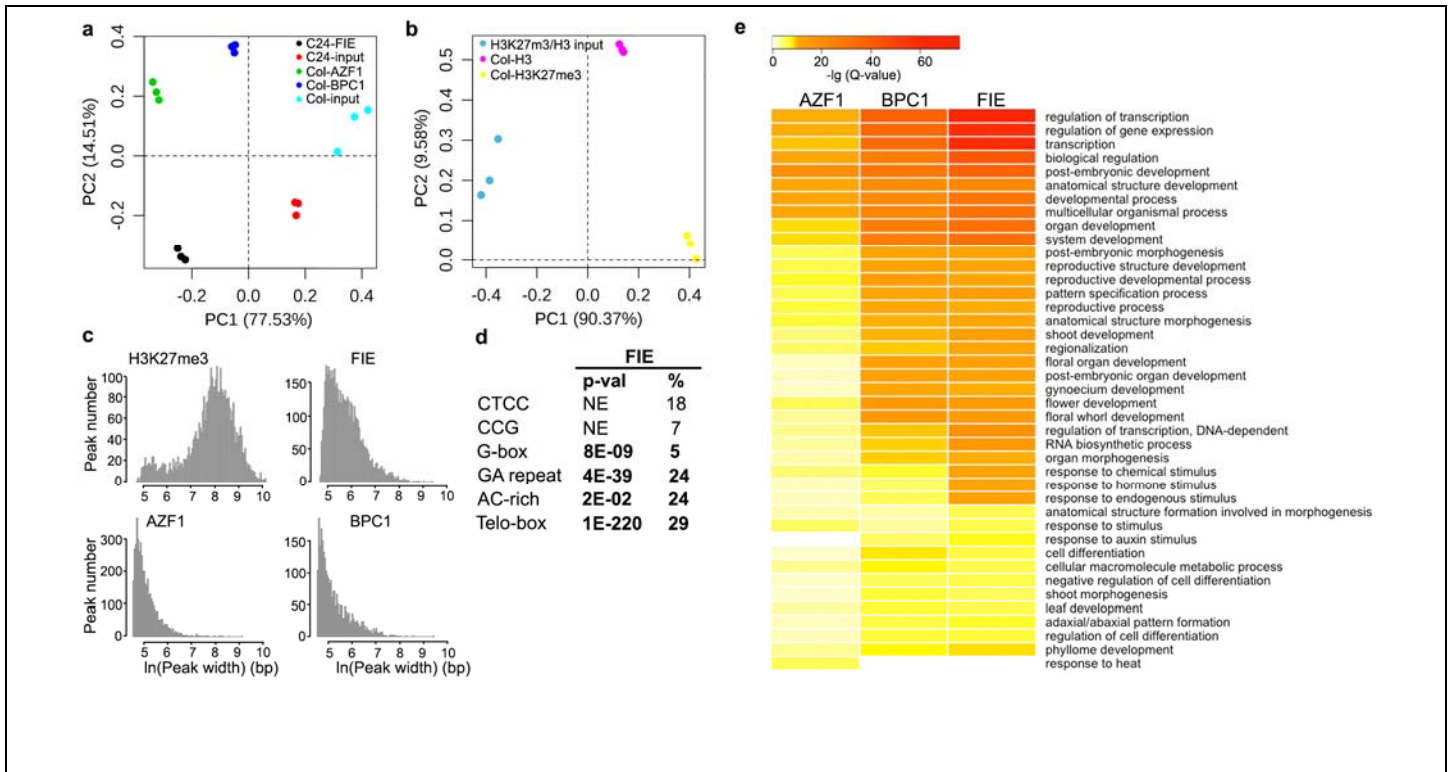
Supplementary Figure 4

GA repeat and telobox motifs are bound by class I BPC and C1-2iD Zn Finger transcription factors.

(a) Electrophoretic mobility shift assay to test association of BPC1 with the GA repeats or with mutated versions thereof. WT, M1 and M2 were also used as cold competitors. Class I BPC TFs are known to oligomerize[22,28,78]. This, combined with the presence of two GA repeats in the tested DNA fragment, may explain why multiple shifted bands are observed when the protein is complexed with the DNA. Below: fraction of shifted DNA (% complex). Mean ± SEM of three EMSA experiments.

(b) Motif-based yeast one hybrid screen identified members of the C1-2iD family of Zn-finger TFs as telobox (**AAACCCTA**) binding transcription factors. The screen preferentially identified C2H2 ZnF proteins, in particular those of the C1-2iD subfamily (AZF1 and ZAT6).

(c) Confirmation of telobox interactome screen. Ten-fold serial dilutions of the yeast strains containing the bait DNA (b) integrated into the genome and plasmids containing the C1-2iD Zn-finger TFs indicated were plated on growth media (left) or on selection media (containing 600ng/ml aureobasidin A fungicide; right). See Supplementary Fig. 7 for a phylogenetic tree of ZnF TFs. The thin white line indicates where the plate image was cut.

78.   Wanke, D. *et al.* Alanine zipper-like coiled-coil domains are necessary for homotypic dimerization of plant GAGA-factors in the nucleus and nucleolus. *PLoS One* **6**, e16070 (2011).
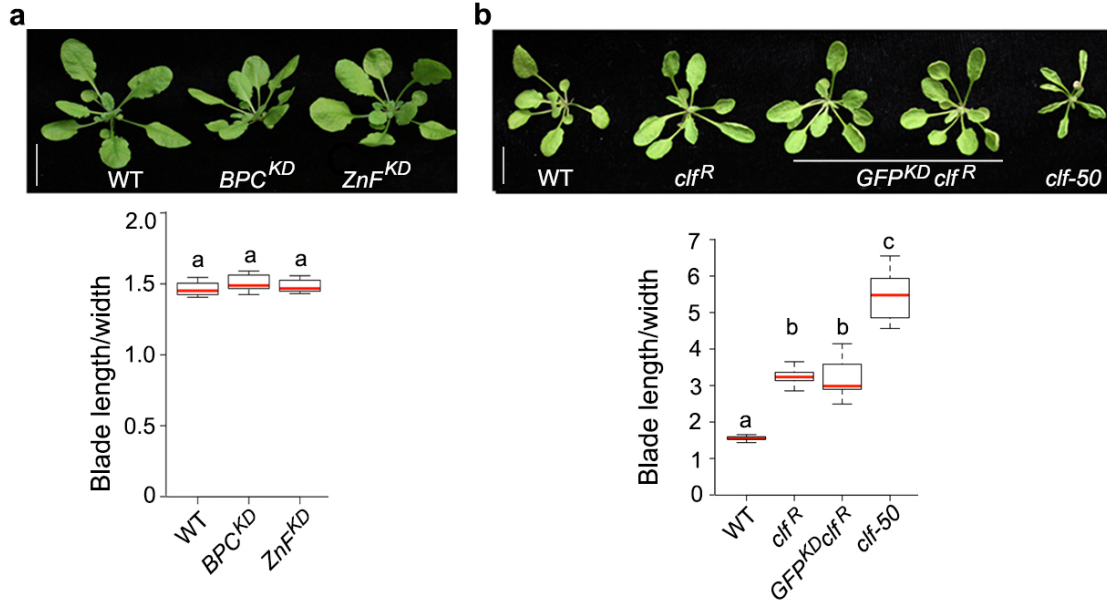
Supplementary Figure 5

FIE, AZF1, BPC1 and H3K27me3 ChIPseq analysis.

(a, b) Principal component analysis (PCA) of RPM-normalized ChIP and input DNA reads for narrow (a) and broad (b) peak calling.

(c) Peak size distribution.

(d) Enrichment (p-values, converted from the Z-scores of the motif enrichment calculations using a normal distribution) and frequency (%) of PRE cis motifs under FIE-bound peaks. NE: not enriched (p>0.5).

(e) Functional classification of PRC2 (FIE), C1-2iD ZnF (AZF1) and class I BPC (BPC1) peak associated genes. Enrichment of Gene Ontology terms (FDR<10^{-5} in at least one of the datasets) for the genes associated with FIE, BPC1 and AZF1 peaks (Q<10^{-10}). The majority of the significant BPC1 and AZF1 target gene Gene Ontology terms are also significant Gene Ontology terms of FIE targets. Enriched GO terms include regulation of transcription, postembryonic development (reproductive development, shoot development, gynoecium development) and hormone response.
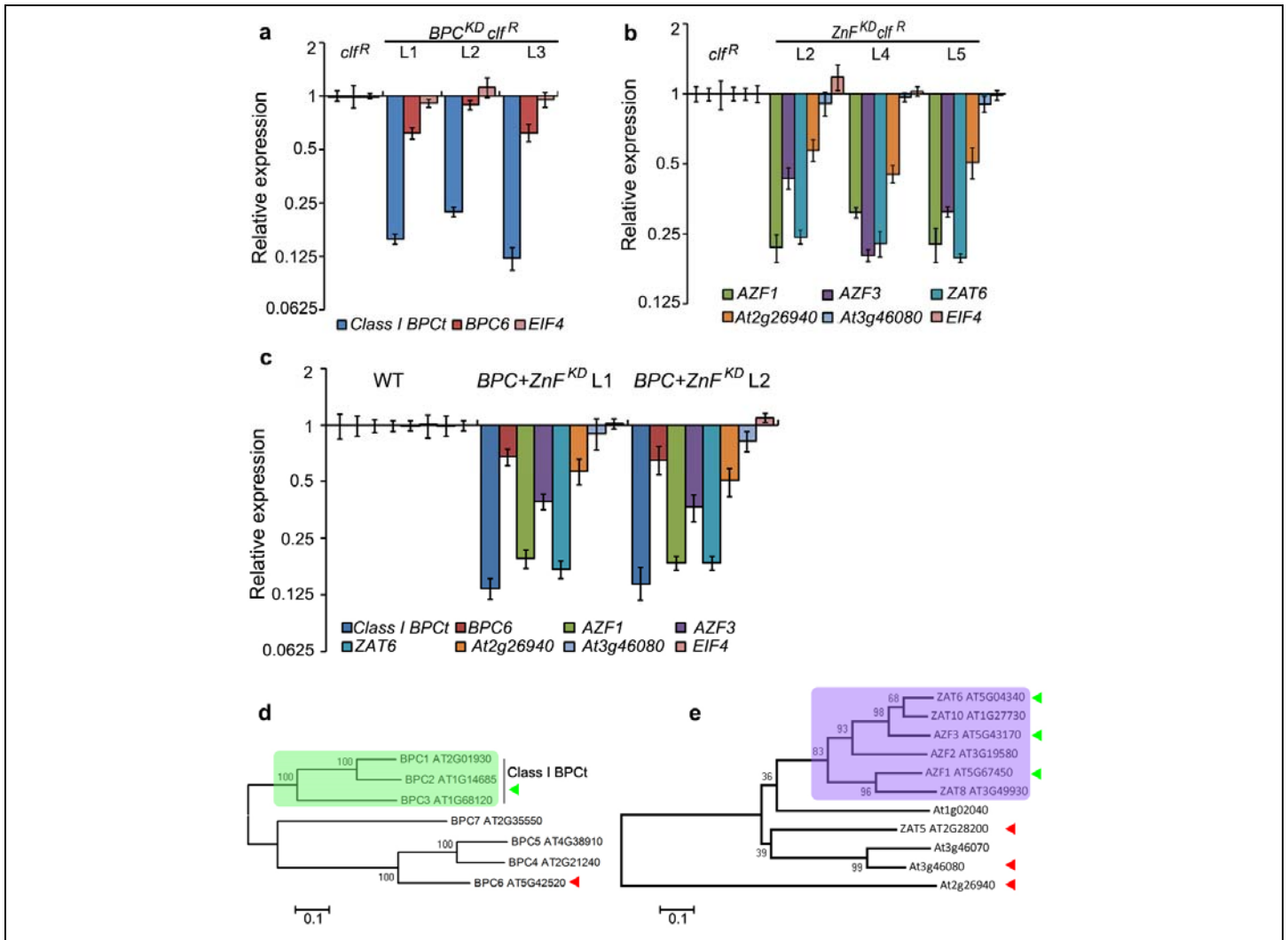
Supplementary Figure 6

Transcription factor family knockdown in the wild type and control knockdown in *clf^R*.

(**a**) class I BPC knockdown in the wild type (WT) does not cause leaf curling. Top: Representative images, bar = 1cm. Bottom: quantification of phenotypes. Box and whisker plot with median (red line, n=15 independent lines), upper and lower quartile (box) and minima, maxima (whiskers). Different letters above bars indicate significantly different groups, P<0.05 based on Kruskal-Wallis test with Dunn's posthoc test.

(**b**) Control knock-down (*GFP^{KD}*) in the hypomorph *clf^R* mutant does not enhance *clf^R* leaf curling. Top: Representative images, bar = 1cm. Bottom: quantification of phenotypes. Box and whisker plot with median (red line), upper and lower quartile (box) and minima, maxima (whiskers). Different letters above bars indicate significantly different groups, P<0.05 based on Kruskal-Wallis test with Dunn's posthoc test.

Supplementary Figure 7

RNAi-mediated knockdown of class I BPC and C1-2iD ZnF TF families.

(**a, b**) Class I BPC TF family knockdown by RNAi (*BPC^KD*; a) or C1-2iD C2H2 Zn-finger TF knockdown (*ZnF^KD*; b) was assayed in independent transgenic lines in the hypomorph *clf^R* mutant background. Expression of all genes tested is shown relative to that in *clf^R* (mean ± SEM) of one experiment for 3 independent lines. In each case, strong knockdown of the targeted TF family is observed (class I BPC and AZF1, AZF3, ZAT6), with more minor effects on distantly related BPC or ZnF TFs (BPC6, At2g26940, At3g46080). The translation initiation factor EIF4 served as qRT-PCR control.

(**c**) Simultaneous knockdown of Class I and C1-2iD C2H2 Zn TFs by RNAi (*BPC + ZnF^KD*) in the wild type tested as described in (a, b) in 2 independent lines.

(**d, e**) Phylogenetic tree of the BBR-BPC family of TFs (see also[22]) (d) and of select C2H2 Zinc finger TFs (see also[20]) plus a subset of additional C1-2i C2H2 ZnF proteins in Arabidopsis (e). Branch length is indicated. Light green shading highlights the class I BPC TFs and light purple shading the C1-2iD Zn-finger proteins targeted for knockdown. Triangles point to the genes tested by qRT-PCR in (a-c) or used as controls in BiFC (Fig. 2d,e; BPC6 and ZAT5). Triangle color indicates RNAi targets (green) or distantly related genes (red). Note that the primer pair employed for class I BPCt qRT-PCR amplifies all 3 genes in that clade.
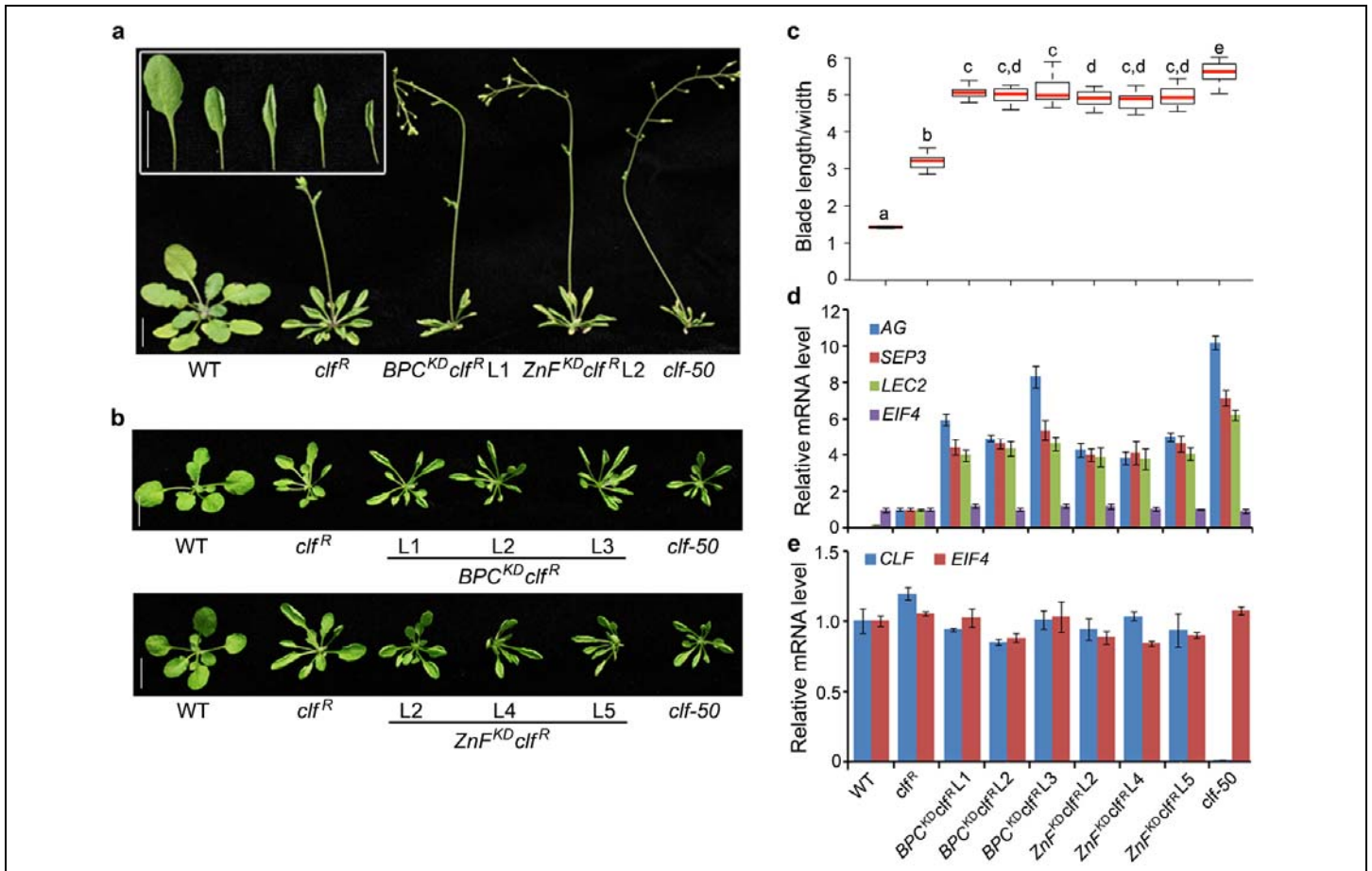
Supplementary Figure 8

**The effect of knockdown of class I BPC or C1-2iD Zn-finger TF family in the hypomorph *clf^R* mutant.**

(**a**) Leaf curling (inset) and flowering time in wild type (WT), the weak *clf^R* mutant, class I BPC knockdown in *clf^R*, C1-2iD ZnF knock down in *clf^R* and in *clf-50* RNA null mutant, bar = 1cm.

(**b**) Leaf curling phenotype of independent *BPC^KD clf^R* (top) and *ZnF^KD clf^R* (bottom) transgenic lines, bar = 1cm.

(**c**) Phenotype quantification of genotypes in (b). Box and whisker plot with median (red line, n=15 independent lines), upper and lower quartile (box) and minima, maxima (whiskers). Different letters above bars indicate significantly different groups (p<0.05) based on Kruskal-Wallis test with Dunn's posthoc test.

(**d**) De-repression of Polycomb target gene expression (*AG*, *SEP3*, *LEC2*) in the genotypes in (b). Expression relative to the parental line (*clf^R*) is shown. The housekeeping gene *EIF4* served as control.

(**e**) The abundance of the *CLF* mRNA relative to the wild type as measured by qRT-PCR. CLF levels are unchanged in the *BPC^KDclf^R* and the *ZnF^KDclf^R* plants shown in (b) compared to the parental line (WT). Shown are mean ± SEM of one experiment for three independent transgenic lines (d,e).

1
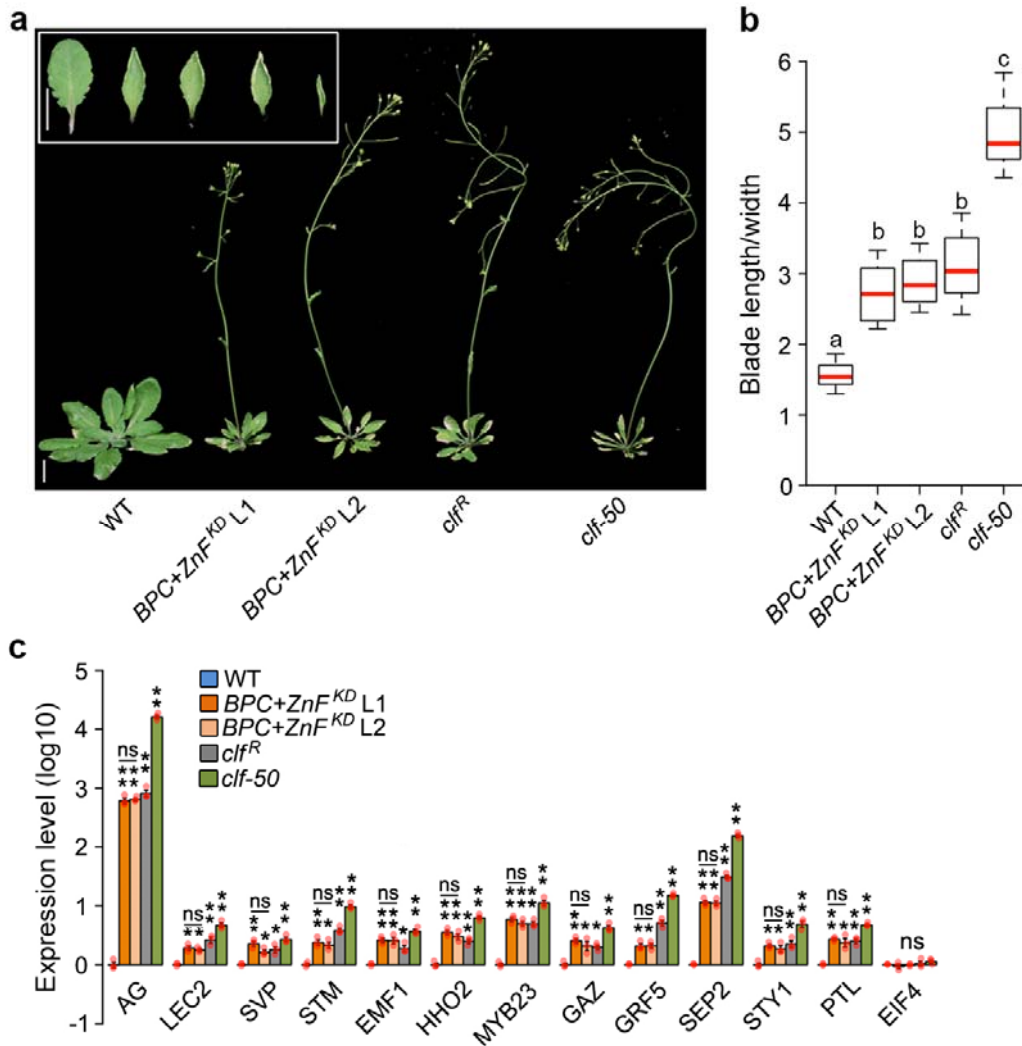
Supplementary Figure 9

Phenotype of simultaneous knockdown of C1-2iD ZnF and class I BPC transcription factors.

(a) Leaf curling (inset) and early flowering in two independent BPC^{KD} ZnF^{KD} TF family knockdown lines, the hypomorph clf^R mutant and the null clf-50 mutant relative to the wild type (WT), bar = 1cm.

(b) Quantification of the phenotypes in (a). Box and whisker plot with median (red line, n=15 independent lines), upper and lower quartile (box) and minima, maxima (whiskers). Different letters above bars indicate significantly different groups (p<0.05) based on Kruskal-Wallis test with Dunn's posthoc test.

(c) Misexpression of polycomb target genes in the genotypes listed in (a). Expression in the mutant lines is show relative to that of the wild type. Mean ± SEM from three experiments (red dots). * P<0.05; ** P< 0.01, ns P>0.05; relative to the wild type, one-tailed unpaired *t* test. The two independent BPC^{KD} ZnF^{KD} TF family knockdown lines are not significant different from each other, ns P>0.05; one-tailed unpaired *t* test.

2
3
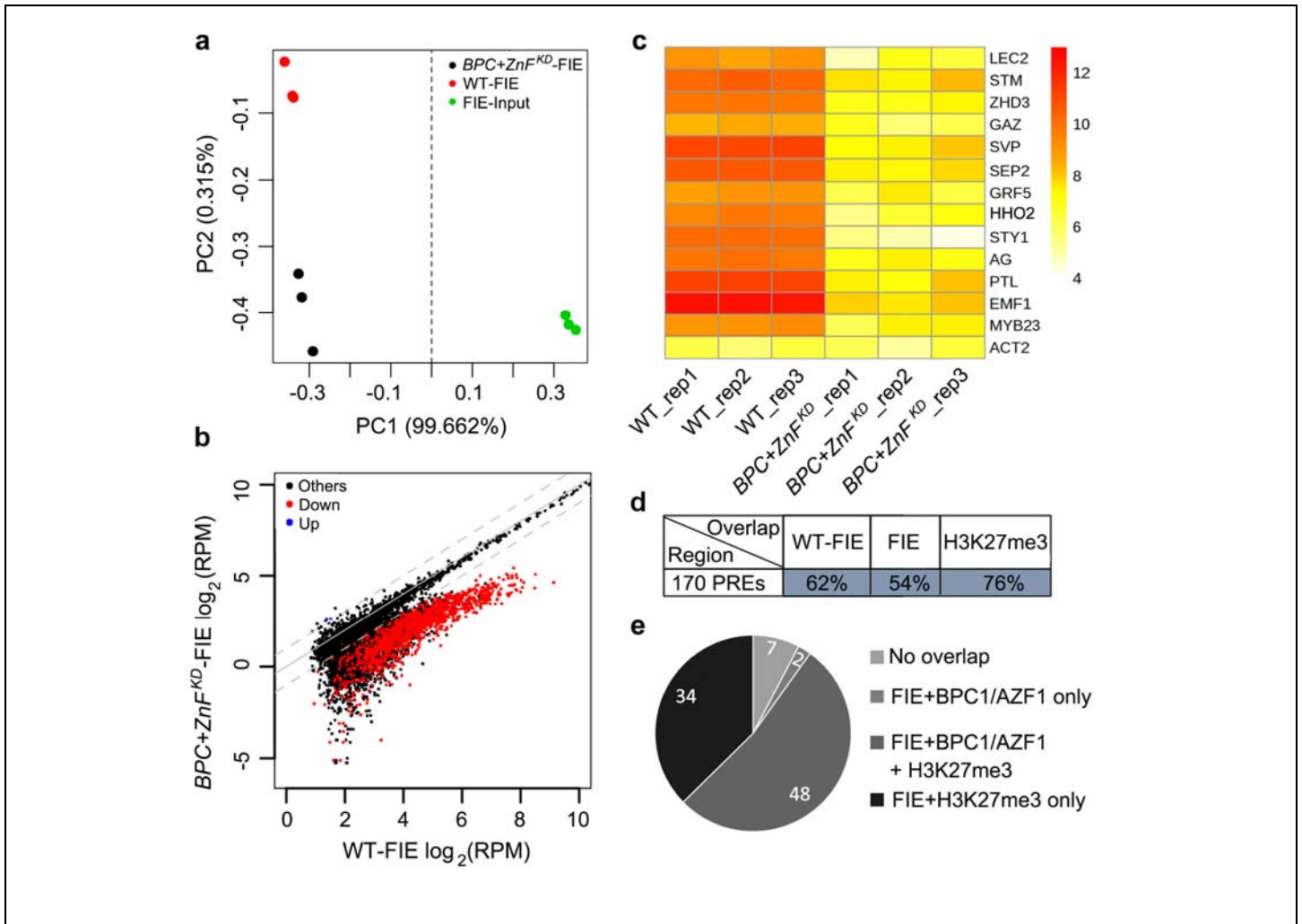
Supplementary Figure 10

**FIE ChIP in wild type or *BPC+ZnF* ^{KD} plants.**

(**a**) Principal component analysis (PCA) of RPM-normalized ChIP and input DNA.

(**b**) Comparison of FIE ChIPseq reads from wild-type (WT) and *BPC+ZnF* ^{KD} plants. Shown are ChIPseq reads mapping to the WT-FIE peak regions. Regions with significantly (P-value <0.01 and 2-fold change) increased and decreased read density in *BPC+ ZnF* ^{KD} are indicated by blue and red color, respectively.

(**c**) Heatmap of RPM normalized read counts of FIE ChIPseq at the targets tested by qRT-PCR in Figure 5.

(**d**) Percent overlap table (by row) of 170 candidate PREs with significant (Q < 10^{-10}) ChIPseq peaks for FIE in the wild type background (from Figure 5), and FIE and H3K27me3 (from Figure 4). Shading highlights strength of overlap.

(**e**) Piechart for the overlap between significant (Q<10^{-10}) FIE, TF (BPC1/AZF1) and H3K27me3 peaks identified by ChIPseq in 30-hr-old plants and the 170 computationally defined candidate PREs from the vegetative phase of development.
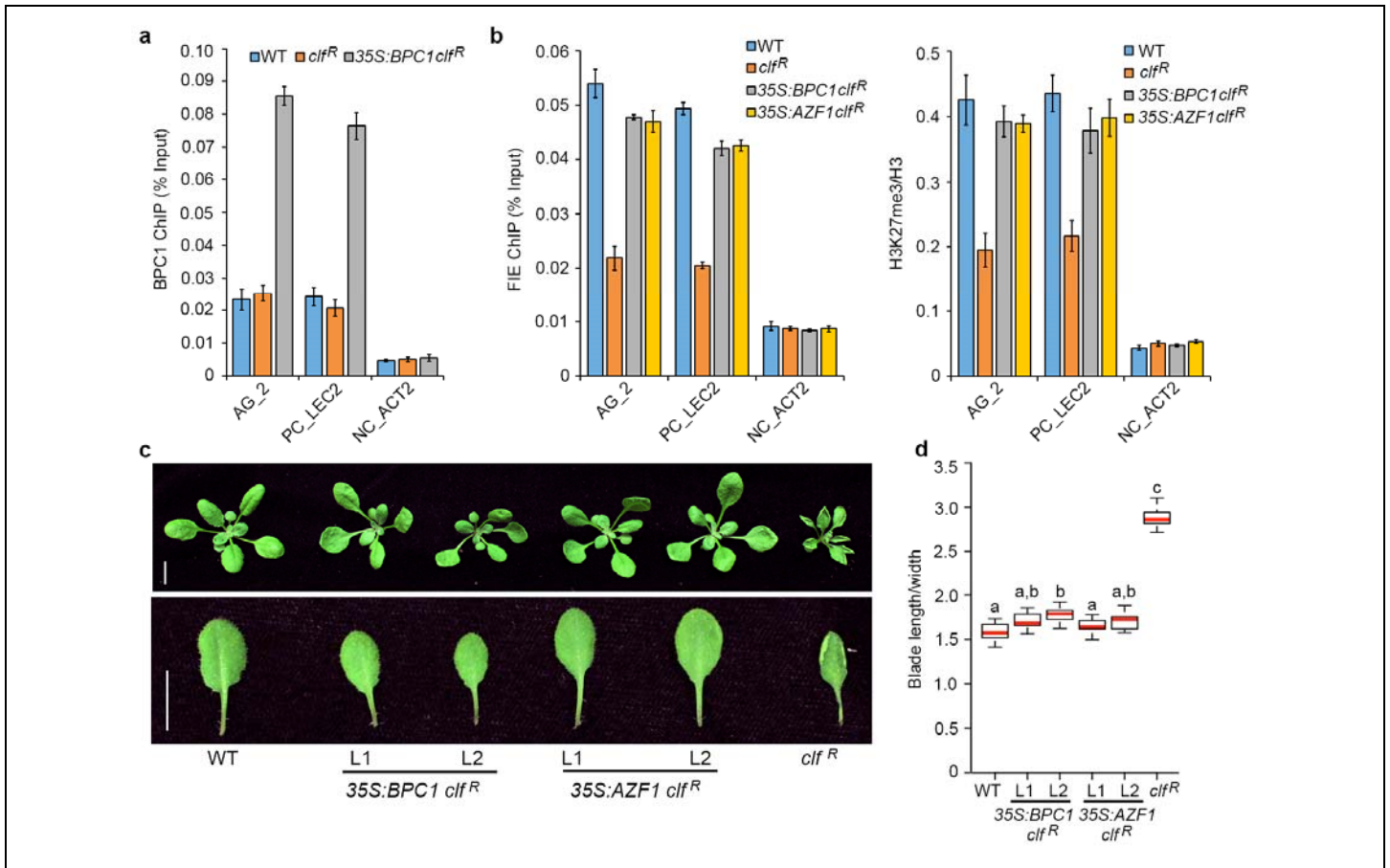
4
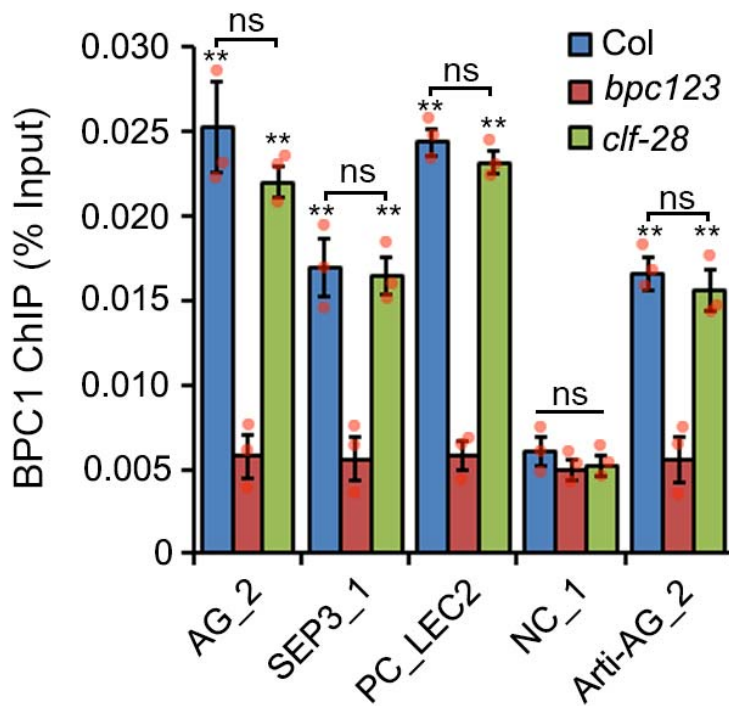
Supplementary Figure 11

BPC1 TF chromatin occupancy in *prc2* mutant.

ChIP using antiserum against class I BPC proteins to test BPC1 occupancy in the wild type (WT), the *bpc123* triple mutant and the *clf-28* (*prc2*) mutant. Binding was assayed at endogenous PREs (AG_2, SEP3_1, PC_LEC2), a control locus (NC_1) or a PRE reporter (Arti-AG_2). Shown are mean ± SEM of three experiments (red dots). ** P< 0.01, relative to *bpc123* mutant, one-tailed unpaired *t* test. ns, no significant difference (P>0.08) between BPC1 binding in Col and *clf-28*, one-tailed unpaired *t* test.

5

6

Supplementary Figure 12

Gain-of BPC1 or AZF1 function rescues a hypomorph *prc2* mutant.

(a) ChIP using antiserum against class I BPC proteins (see Supplementary Fig. 11) to test BPC1 occupancy in the wild type (WT), the hypomorph *clf*[R] (*prc2*) mutant and 35S:BPC1 *clf*[R]. Binding was assayed at endogenous PREs (AG_2, PC_LEC2) or an endogenous control region (NC_ ACT2). Shown is the mean ± SEM of two independent TF overexpression lines.

(b) ChIP using antiserum against FIE (PRC2; left) or H3K27me3/H3 (right) in the wild type (WT), *clf*[R], 35S:BPC1 *clf*[R] and 35S:AZF1 *clf*[R]. Shown is the mean ± SEM of two independent TF overexpression lines.

(c, d) Leaf curling in wild type (WT), two independent lines (L1 and L2) of 35S:BPC1 *clf*[R] and of 35S:AZF1 *clf*[R] and the hypomorph *clf*[R] mutant. (c) Representative images. Bar = 1cm. (d) Quantification of leaf curling in the genotypes shown in (c). Box and whisker plot with median (red line, n=15 independent lines), upper and lower quartile (box) and minima, maxima (whiskers). Different letters above bars indicate significantly different groups, $p < 0.05$ based on Kruskal-Wallis test with Dunn's posthoc test.

7
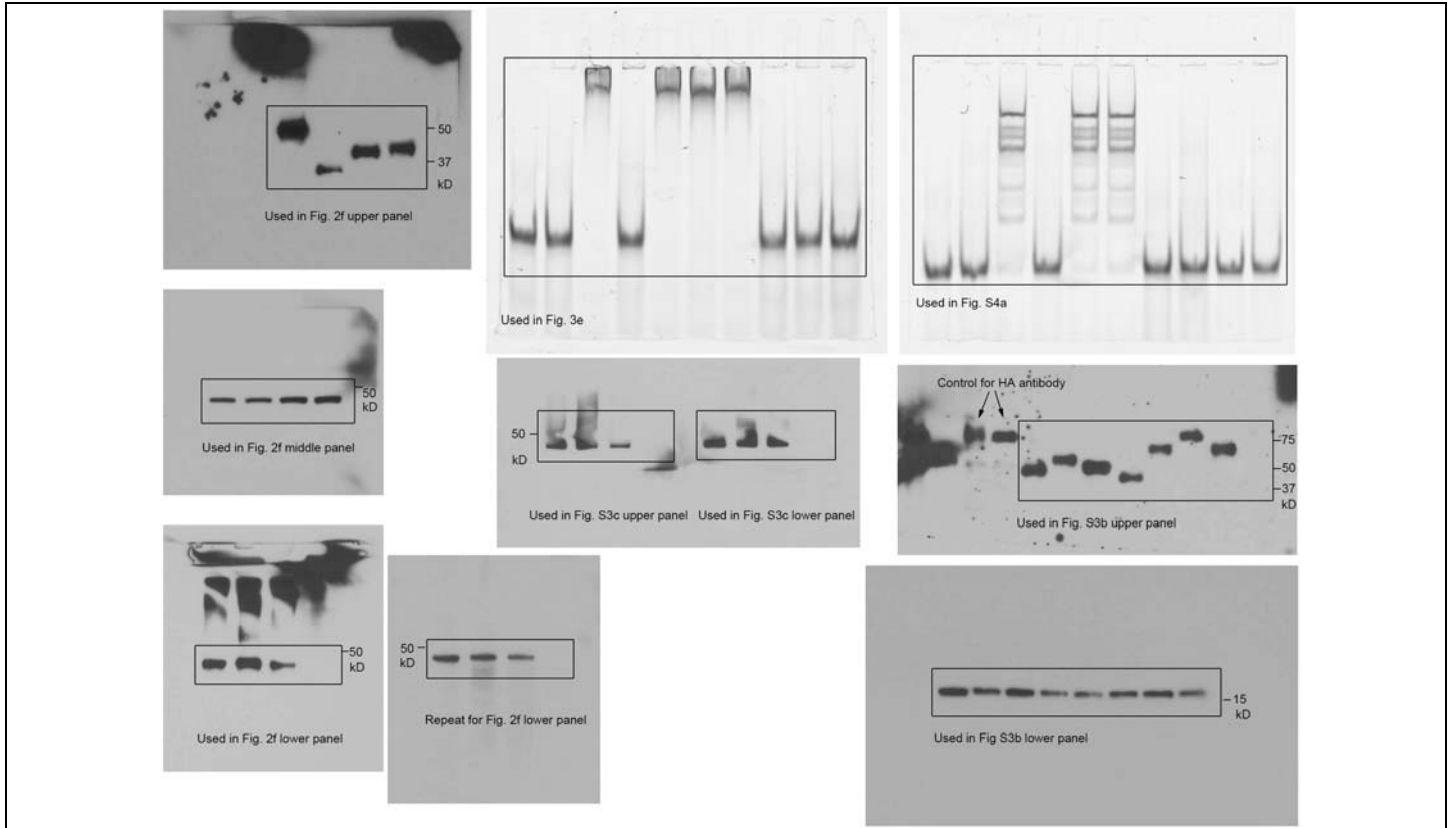
Supplementary Figure 13

Full length gel images for all figures.

Left: Images for Co-IP. Two images are provided for the lower panel, the left image has non-specific signal at the top.
Center (top): Image for telobox EMSA; (below): Co-IP in the absence and presence of Benzonase.
Right (top): Image for GA repeat EMSA; (below): Quantification of proteins in protoplasts used for BiFC.

8

9

10

11

12