







# H2AK121ub in *Arabidopsis* associates with a less accessible chromatin state at transcriptional regulation hotspots

Xiaochang Yin<sup>1,8</sup>, Francisco J. Romero-Campero <sup>2,3,8</sup>✉, Pedro de Los Reyes <sup>2</sup>, Peng Yan<sup>4</sup>, Jing Yang <sup>5</sup>, Guangmei Tian<sup>6</sup>, XiaoZeng Yang <sup>5</sup>, Xiaorong Mo<sup>4</sup>, Shuangshuang Zhao<sup>7</sup>, Myriam Calonje <sup>2</sup>✉ & Yue Zhou <sup>1</sup>✉

Although it is well established that the Polycomb Group (PcG) complexes maintain gene repression through the incorporation of H2AK121ub and H3K27me3, little is known about the effect of these modifications on chromatin accessibility, which is fundamental to understand PcG function. Here, by integrating chromatin accessibility, histone marks and expression analyses in different *Arabidopsis* PcG mutants, we show that PcG function regulates chromatin accessibility. We find that H2AK121ub is associated with a less accessible but still permissive chromatin at transcriptional regulation hotspots. Accessibility is further reduced by EMF1 acting in collaboration with PRC2 activity. Consequently, H2AK121ub/H3K27me3 marks are linked to inaccessible although responsive chromatin. In contrast, only-H3K27me3-marked chromatin is less responsive, indicating that H2AK121ub-marked hotspots are required for transcriptional responses. Nevertheless, despite the loss of PcG activities leads to increased chromatin accessibility, this is not necessarily accompanied by transcriptional activation, indicating that accessible chromatin is not always predictive of gene expression.

<sup>1</sup>State Key Laboratory of Protein and Plant Gene Research, School of Advanced Agricultural Sciences, Peking-Tsinghua Center for Life Sciences, Peking University, 100871 Beijing, China. <sup>2</sup>Institute of Plant Biochemistry and Photosynthesis (IBVF-CSIC), Avenida Américo Vespucio 49, 41092 Seville, Spain. <sup>3</sup>Department of Computer Science and Artificial Intelligence (University of Sevilla), Avenida Reina Mercedes s/n, 41012 Seville, Spain. <sup>4</sup>State Key Laboratory of Plant Physiology and Biochemistry, College of Life Science, Zhejiang University, 310058 Hangzhou, China. <sup>5</sup>Beijing Agro-biotechnology Research Center, Beijing Key Laboratory of Agricultural Genetic Resources and Biotechnology, Beijing Academy of Agriculture and Forestry Sciences, 100097 Beijing, China. <sup>6</sup>Academy for Advanced Interdisciplinary Studies, Peking University, 100871 Beijing, China. <sup>7</sup>Key Laboratory of Plant Stress, Life Science College, Shandong Normal University, 250014 Jinan, China. <sup>8</sup>These authors contributed equally: Xiaochang Yin, Francisco J. Romero-Campero. ✉email: [fran@us.es](mailto:fran@us.es); [myriam.calonje@ibvf.csic.es](mailto:myriam.calonje@ibvf.csic.es); [yue\\_zhou@pku.edu.cn](mailto:yue_zhou@pku.edu.cn)

PcG complexes maintain gene repression by incorporating histone modifications within chromatin<sup>1–3</sup>. While PcG repressive complex 1 (PRC1) has histone 2A E3 ubiquitin ligase activity<sup>4–8</sup>, PRC2 has histone 3 lysine 27 (H3K27) trimethyltransferase activity<sup>9–12</sup>. Nevertheless, despite the important role of these chromatin marks in regulating transcription in eukaryotes, their effect on chromatin is not yet clear. Several pieces of evidence suggested that animal PcG proteins mediate chromatin compaction<sup>13–15</sup>. However, a recent report proposed that although PcG occupied promoters exhibit reduced accessibility, this does not rely on PcG proteins. Instead, PRC1 plays a role in increasing nucleosome occupancy and decreasing nucleosomal spacing<sup>16</sup>.

In *Arabidopsis thaliana* (*Arabidopsis*), the composition of PcG complexes is still a matter of debate. While PRC2 core components are well conserved to their animal counterparts and the catalytic activity is mainly carried out by the EZ homologs CURLY LEAF (CLF) and SWINGER (SWN) during sporophyte development<sup>3,17</sup>, the identity of PRC1 is less clear. Despite the H2A monoubiquitination module is formed by the Polycomb Group ring finger (PCGF) and RING1 homologs BMI1 (A, B, or C) and RING1 (A or B)<sup>2,18</sup>, association of the plant-specific proteins EMBRYONIC FLOWER 1 (EMF1) and LIKE-HETEROCHROMATIN PROTEIN 1 (LHP1) to this complex has turned to be controversial. EMF1 was proposed to be a PRC1 component due to its ability to in vitro mediate chromatin compaction as *Drosophila melanogaster* (*Drosophila*) PRC1 component Posterior Sex Combs (Psc) does<sup>19,20</sup>. However, recent data showed that EMF1 co-purifies with PRC2 (refs. 21–23) and is required for H3K27me3 marking<sup>19,24,25</sup>. Likewise, LHP1 was proposed to perform similar function to *Drosophila* PRC1 component Polycomb (Pc)<sup>26,27</sup>; however, LHP1 also co-purifies with PRC2 (refs. 21–23) and is involved in maintenance/spreading of H3K27me3 (ref. 21) but dispensable for H2AK121ub marking<sup>28</sup>. Thus, these two proteins actually seem to be PRC2-associated component.

Interestingly, although H2AK121ub and H3K27me3 marks often co-localize at genes (H2AK121ub/H3K27me3-marked genes), H3K27me3-marked peaks are generally much longer than that of H2AK121ub<sup>28</sup>. Accordingly, H3K27me3 marks usually cover the complete gene region, whereas H2AK121ub marks are enriched at the region surrounding the transcriptional start site (TSS)<sup>28</sup>. In addition, there are also genes marked with only-H2AK121ub or H3K27me3 (ref. 28) (only-H2AK121ub and only-H3K27me3, respectively), indicating that these marks may play independent roles.

H3K27me3 marks in *Arabidopsis* have been linked to regions displaying locally reduced DNA accessibility<sup>29</sup>. According to this, the promoters of expressed genes display high chromatin accessibility and are depleted of H3K27me3 marks<sup>30,31</sup>. However, it is not clear whether H3K27me3 has a direct role in regulating chromatin accessibility. On the other hand, the role of PRC1-mediated H2AK121ub is far from clear. Although several data support a repressive role of this modification<sup>6,8,28</sup>, it has been proposed to be associated with gene responsiveness and its repressive function seems to require PRC2 recruitment<sup>32</sup>. In any case, nothing is known about a possible role of H2AK121ub in regulating chromatin accessibility in plants.

In this work, by integrating chromatin accessibility, histone marks, and expression analyses in different *Arabidopsis* PcG mutants, we show that H2AK121ub marks associate with a less accessible but permissive chromatin at transcriptional regulation hotspots, which are sites enriched for the binding of transcription factors (TFs). Chromatin accessibility can be further reduced by EMF1 and PRC2 activity. However, while H2AK121ub/H3K27me3-mediated inaccessible chromatin is still

transcriptionally responsive, only-H3K27me3 marked chromatin is less responsive.

## Results

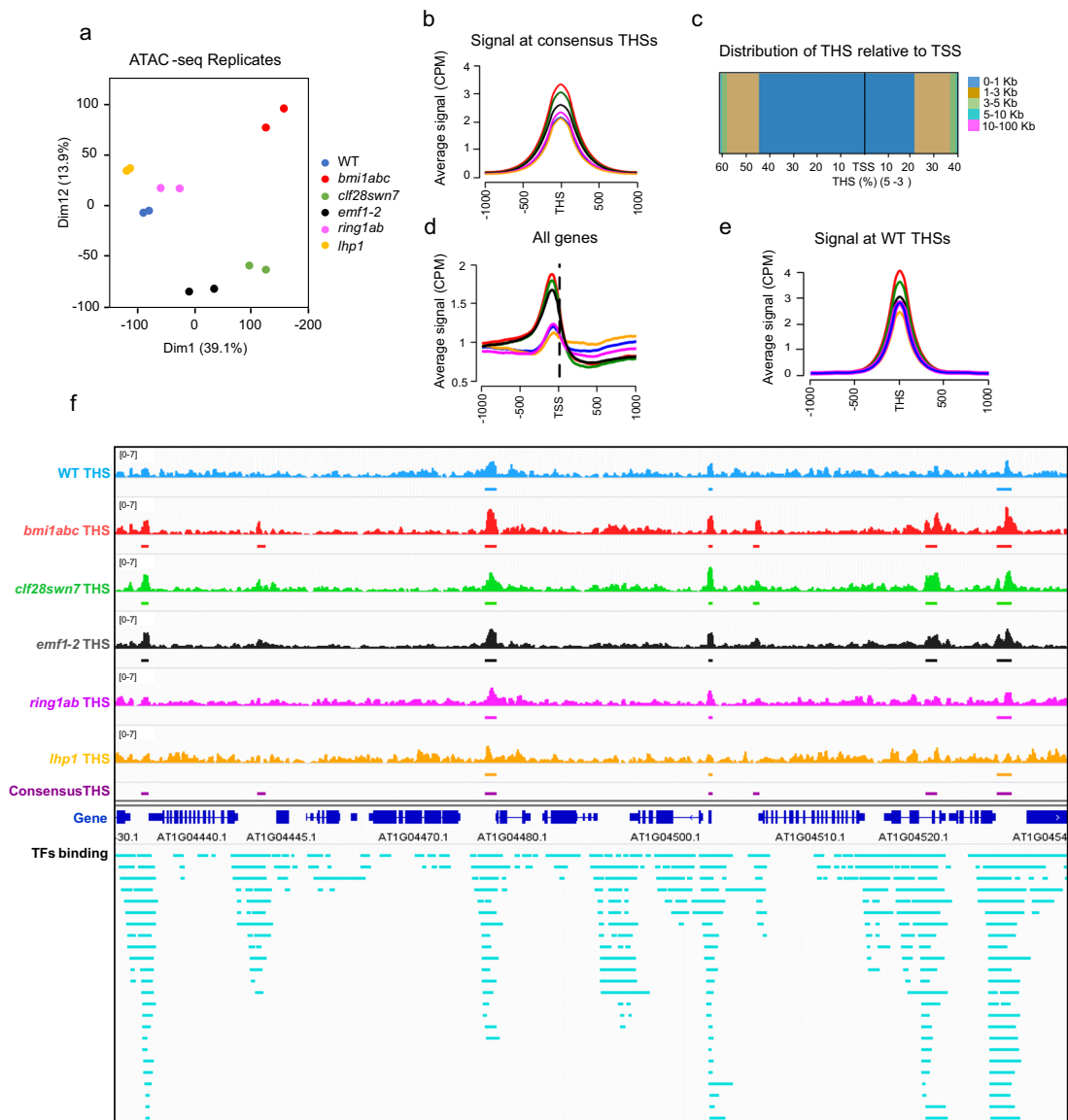
### H2AK121ub hallmarks hotspots for transcriptional regulation.

To investigate a possible role of H3K27me3 and H2AK121ub in regulating chromatin accessibility, we first performed Assay for Transposase-Accessible Chromatin sequencing (ATAC-seq) in wild-type Col-0 (WT) and different PcG mutants at 10 days after germination (DAG). We selected *bmi1abc*, *clf28sw7*, and *emf1-2* severe mutants<sup>8,19,33</sup> that cannot undergo the transition from embryonic to vegetative development after germination in the aerial part of the seedling. However, unlike *bmi1abc* and *clf28sw7*, which in addition have a stunted root displaying embryonic traits, *emf1-2* is able to develop a WT-like root (Supplementary Fig. 1). We in addition included *ring1ab* weak and *lhp1* mutants that are able to develop vegetative and floral organs in spite of several developmental alterations<sup>34,35</sup> (Supplementary Fig. 1).

Principal component analysis (PCA) of ATAC-seq data revealed that WT, *ring1ab*, and *lhp1* clustered together, whereas *clf28sw7* and *emf1-2* on one side, and *bmi1abc* on the other, constituted two distant and distinct clusters, indicating differences between them (Fig. 1a). We next identified Tn5 hypersensitive sites (THSs) for each genotype as those genomic regions exhibiting a significant accumulation of signal from the corresponding ATAC-seq data in the two different replicates. We found that the number of THSs in WT was lower than that in the different mutants (11,351 in WT, 12,803 in *lhp1*, 13,829 in *ring1ab*, 13,472 in *emf1-2*, 17,162 in *clf28sw7*, and 17,420 in *bmi1abc*; Supplementary Fig. 2). In addition, most of WT THSs were also accessible regions in mutants, indicating that loss-of-function of these PcG proteins leads to emergence of new THSs (Supplementary Fig. 2). To compare the contribution of the different mutants in increasing chromatin accessibility, THSs from the different genotypes were merged together into a list of consensus THSs. The list consisted in highly confident consensus THSs exhibiting an accessibility signal greater than 3 CPM with a *q*-value <0.05 in at least one of the genotypes (17,372; Supplementary Data 1). Average accessibility signal at consensus THSs was much higher in severe mutants than in *ring1ab* and *lhp1*, in which the signal was similar to WT (Fig. 1b). Consensus THSs were mostly localized within the 1 kb region upstream of the TSS (Fig. 1c). Accordingly, we found an accessibility signal peak at the region prior to the TSS in all genotypes (Fig. 1d). We then analyzed the average accessibility signal at WT THSs in the different genotypes (Fig. 1e). Interestingly, the embryonic mutants *bmi1abc*, *clf28sw7*, and *emf1-2* showed higher accessibility signals at this subset of THSs than the genotypes that develop vegetative tissues, as is the case of WT, *ring1ab*, and *lhp1* mutants. This is consistent with the fact that undifferentiated embryonic cells tend to have a more accessible chromatin state than differentiated cells<sup>36–38</sup>.

We found that consensus THSs co-localized with regions highly enriched in transcription factor (TF)-binding sites (enrichment = 7.5, *p* value <10<sup>−4</sup>; Fig. 1f) as determined by genome-wide identification of the binding sites of 100 TFs (Supplementary Data 2), supporting that THSs indicate the presence of transcriptional regulation hotspots<sup>30,36,39</sup>. According to this, a high number of transcriptional regulation hotspots that were inaccessible in WT became accessible in one or more mutants (Fig. 1f).

Nevertheless, the emergence of new THSs in mutants could be the result of a direct or indirect consequence of loss of PcG regulation. Therefore, we compared the genome-wide

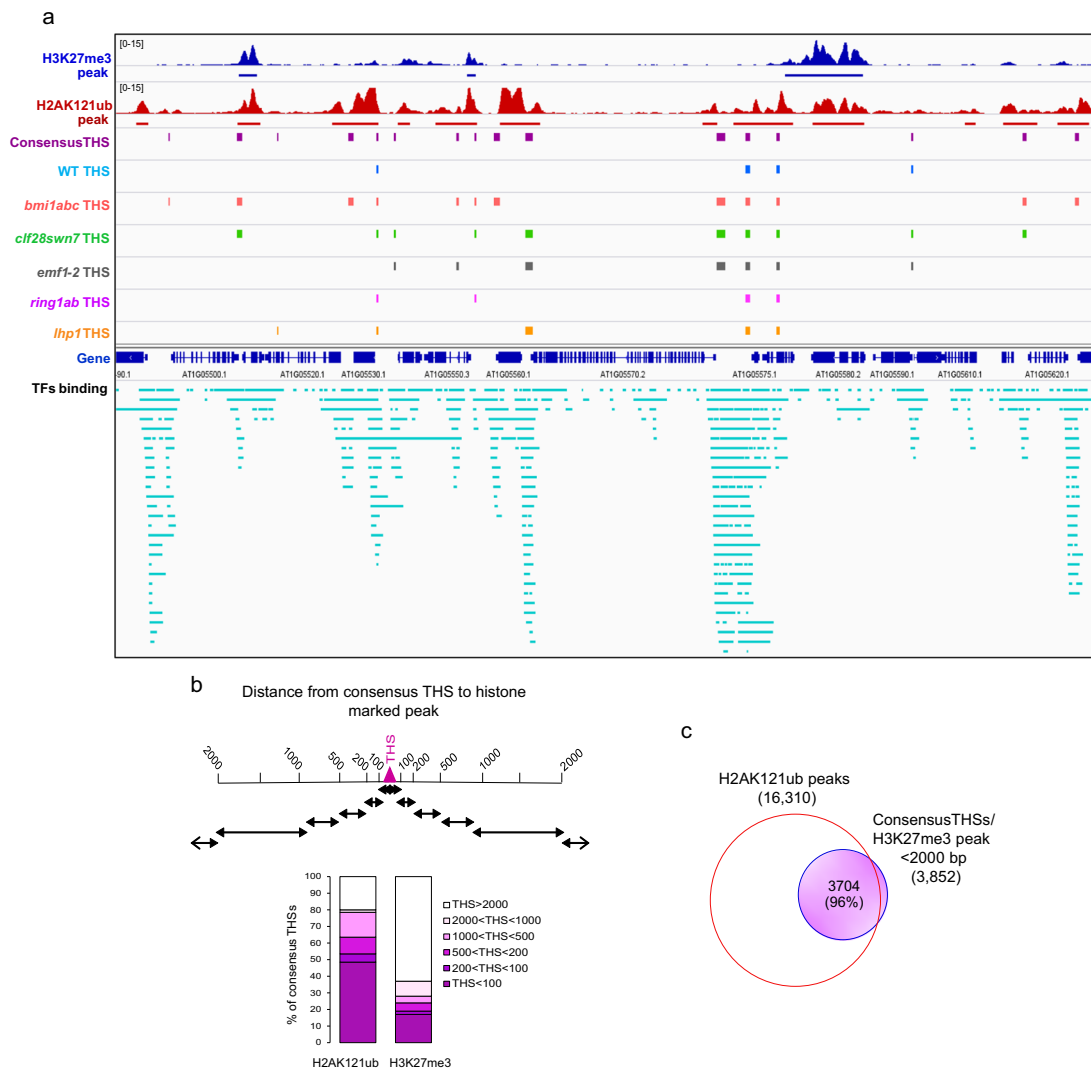


**Fig. 1 Consensus THSs co-localize with transcriptional regulation hotspots.** **a** PCA analysis of WT, *bmi1abc*, *clf28sw7*, *emf1-2*, *lhp1*, and *ring1ab* ATAC-seq replicates at 10 DAG. **b** Average signal levels at the center of consensus THSs in WT, *bmi1abc*, *clf28sw7*, *emf1-2*, *lhp1*, and *ring1ab* mutants. **c** Distribution of consensus THSs relative to TSS. The percentage of THSs found in different intervals is indicated in different colors. **d** Average accessibility profile of coding genes at the region surrounding the TSS in WT and mutants. **e** Average signal levels at the center of WT THSs in WT, *bmi1abc*, *clf28sw7*, *emf1-2*, *lhp1*, and *ring1ab* mutants. **f** Screenshot showing distribution of genotype-specific THSs, consensus THSs, and TFs binding sites (see also Supplementary Data 2). Each genotype is indicated in different color (WT in blue, *bmi1abc* in red, *clf28sw7* in green, *emf1-2* in black, *ring1ab* in pink, and *lhp1* in orange). Consensus THSs are indicated in magenta. Gene regions are indicated in dark blue and the binding sites of TFs in greenish blue.

distribution of H2AK121ub and H3K27me3 marks in WT to that of consensus THSs, genotype-specific THSs, and TFs binding sites (Fig. 2a). We found that consensus THSs tend to associate with PcG-marked regions (Fig. 2a and Supplementary Fig. 3). Intriguingly, around 80% of consensus THSs showed an H2AK121ub peak in its vicinity ( $\leq 2$  kb distance from THS) (Fig. 2a, b); furthermore, 50% of consensus THSs overlapped or had an H2AK121ub peak at less than 100 bp away (Fig. 2b). In contrast, this was not so evident for H3K27me3, as only around 40% of consensus THSs showed an H3K27me3 peak within the next 2 kb (Fig. 2b) and 96% of these regions co-localized with H2AK121ub (Fig. 2c). All together, these results strongly suggest that the loss of PcG function plays a role in increasing accessibility at transcriptional regulation hotspots; moreover, that H2AK121 monoubiquitination hallmarks these hotspots.

**H2AK121ub and H3K27me3 are differentially affected in PcG mutants.** To investigate the role of H3K27me3 and H2AK121ub marks in regulating chromatin accessibility, we next analyzed the signal levels of these modifications at H2AK121ub/H3K27me3, only-H3K27me3, and only-H2AK121ub marked genes in the different mutants (Fig. 3a–d). For this, we used previously published ChIP-seq data for some of the mutants and WT<sup>28</sup>, and generated new datasets for the remaining mutants together with a new WT dataset (see “Methods” section). We redefined H2AK121ub/H3K27me3, only-H3K27me3, and only-H2AK121ub marked genes as those genes displaying the corresponding modification/s in one of the two WT datasets, resulting in similar gene lists than the ones previously published<sup>28</sup> (see “Methods” section).

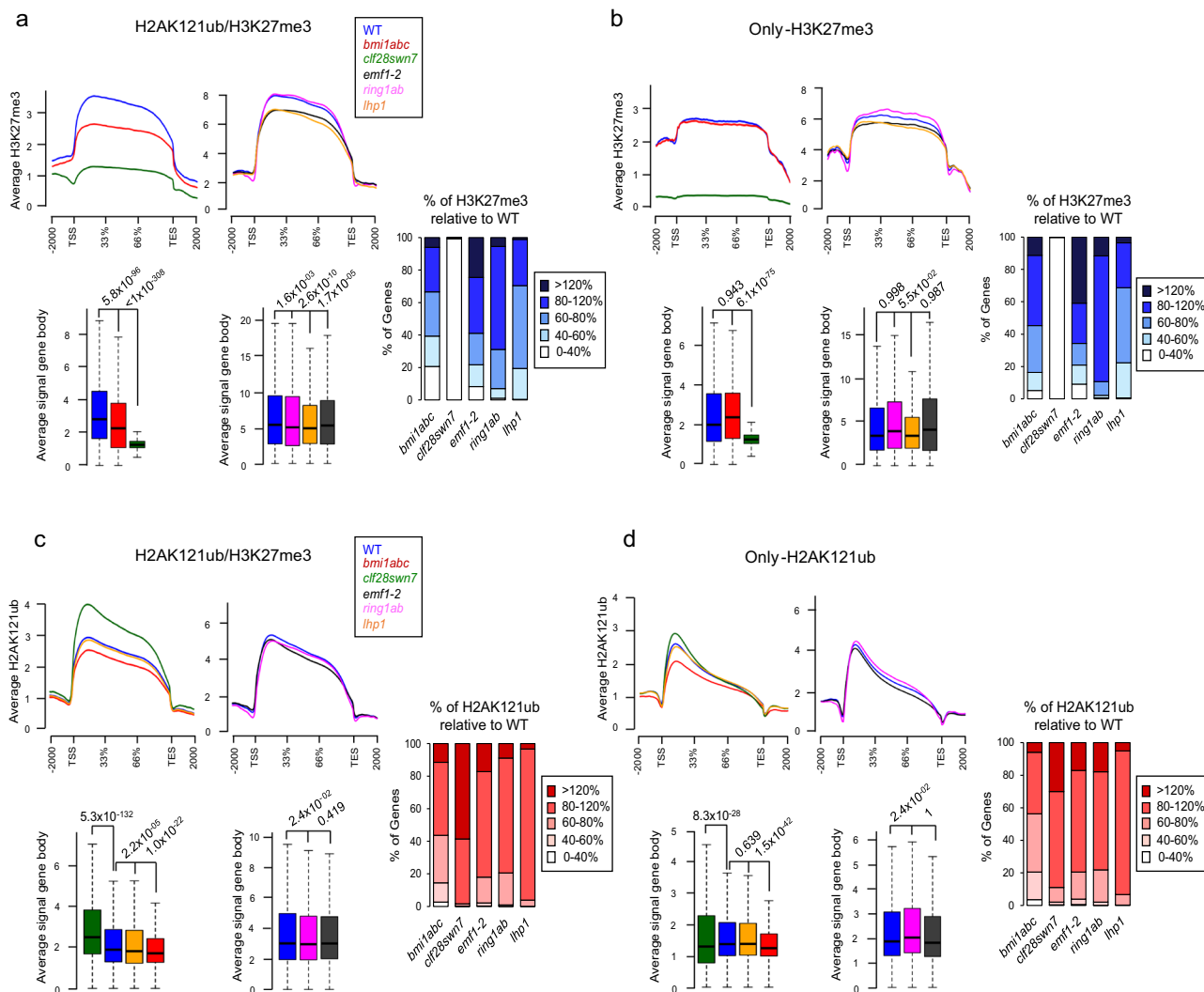
Average H3K27me3 levels at H2AK121ub/H3K27me3-marked genes were almost undetectable in *clf28sw7* and significantly decreased in *bmi1abc* compared to WT as previously described<sup>28</sup>



**Fig. 2 H2AK121ub monoubiquitination hallmarks hotspots for transcriptional regulation.** **a** Screenshots showing distribution of H3K27me3 and H2AK121ub peaks in WT, consensus THSs, genotype-specific THSs, and TFs binding sites. H3K27me3 and H2AK121ub peaks are indicated in blue and red, respectively. Consensus THSs, genotype-specific THSs, gene regions, and TFs binding sites color code is as in Fig. 1f. **b** Percentage of consensus THSs showing an H2AK121ub or H3K27me3 region at different non-overlapping distance intervals as indicated. **c** Venn diagram showing the overlap between consensus THSs having an H3K27me3 peak within the next 2 kb and H2AK121ub peaks.

(Fig. 3a), which is consistent with a requirement of PRC1 activity for H3K27me3 marking. Accordingly, ChIP-reChIP experiments using anti-H3K27me3 and anti-H2AK121ub antibodies verified that these marks co-localize at some nucleosomes of H2AK121ub/H3K27me3-marked genes (Supplementary Fig. 4). We also found reduced levels of H3K27me3 in *lhp1* and *emf1-2* at these genes, although the differences with respect to WT were less significant than in *bmi1abc*. Conversely, *ring1ab* weak mutant showed WT-like levels (Fig. 3a). Nevertheless, analyzing the levels of H3K27me3 at individual genes in *ring1ab*, we found specific genes displaying less than 60% of WT levels (Fig. 3a and Supplementary Fig. 5), which is consistent with previous results<sup>40</sup>. The low percentage of genes showing strongly reduced levels of H3K27me3 in *ring1ab* (around 7%) is most probably due to the presence of RING1 activity in this knockdown mutant. All together, these data indicate that all these proteins are required for appropriate H3K27me3 marking at H2AK121ub/H3K27me3 genes. On the contrary, average global H3K27me3 levels at only-H3K27me3 genes were significantly affected in *clf28swn7* and *lhp1* but not in strong *bmi1abc*, weak *ring1ab*, and unexpectedly *emf1-2* (Fig. 3b).

Interestingly, despite we found decreased levels of H3K27me3 at many genes in *emf1-2*, which supports that EMF1 plays an important role in H3K27me3 deposition<sup>19,24,25</sup>, we also found a high percentage of both H2AK121ub/H3K27me3 and only-H3K27me3 genes with increased levels of H3K27me3 in *emf1-2* compared to WT (Fig. 3a, b and Supplementary Fig. 6). Phenotypic analysis of the different mutants indicated that unlike BMI1 and CLF/SWN, EMF1 might not be decisive for the incorporation of H3K27me3 marks in root tissue as *emf1-2* can develop a WT-like root (Supplementary Fig. 1). Since we used whole seedlings for ChIP-seq analyses, the fact that *emf1-2* displays a stunted shoot development but a WT-like root implies a higher proportion of root cells in *emf1-2* samples compared to WT, which develops shoot and root tissues. Therefore, the apparently increased levels of H3K27me3 at some genes in *emf1-2* might be a consequence of the different ratio of root cells between *emf1-2* and WT samples, which lead to increased average H3K27me3 levels in *emf1-2* at root repressed genes. Supporting this, Gene Ontology (GO) analysis of genes with increased H3K27me3 levels in *emf1-2* showed an enrichment in processes involved in photosynthesis and response to light, corresponding



**Fig. 3 H2AK121ub and H3K27me3 levels in mutant and WT seedlings.** **a** Metagenes plots of H3K27me3 coverage at H2AK121ub/H3K27me3-marked genes in the different mutants. Left panel shows results using previously published data<sup>28</sup> and right panel using data generated in this work (see “Methods”). Bar plots indicate the percentage of genes displaying different levels of H3K27me3 marks relative to WT in each mutant. Box plots below metagenes plots show average signal at gene body in the different genotypes. The median (middle line), upper and lower quartiles (boxes), and minimum and maximum values (whiskers) are indicated. *P* values of differences between WT and mutants according to one-sided Mann–Whitney–Wilcoxon test are also indicated. **b** Same as in **a** but for only-H3K27me3-marked genes. Box plots below metagenes plots show average signal at gene body in the different genotypes. The median (middle line), upper and lower quartiles (boxes), and minimum and maximum values (whiskers) are indicated. *P* values of differences between WT and mutants according to one-sided Mann–Whitney–Wilcoxon test are indicated. **c** Metagenes plots of H2AK121ub coverage at H2AK121ub/H3K27me3-marked genes in the different mutants. Left panel results using previously published data<sup>28</sup> and right panel using data generated in this work (see “Methods”). Bar plots indicate the percentage of genes displaying different levels of H2AK121ub marks relative to WT in each mutant. Box plots below metagenes plots show average signal at gene body in the different genotypes. The median (middle line), upper and lower quartiles (boxes), and minimum and maximum values (whiskers) are indicated. *P* values of differences between WT and mutants according to one-sided Mann–Whitney–Wilcoxon test are indicated. **d** Same as in **c** but for only-H2AK121ub-marked genes. Box plots below metagenes plots show average signal at gene body in the different genotypes. The median (middle line), upper and lower quartiles (boxes), and minimum and maximum values (whiskers) are indicated. *P* values of differences between WT and mutants according to one-sided Mann–Whitney–Wilcoxon test are indicated. In all cases (**a–d**), data were generated from *n* = 2 biological independent ChIP-seq samples. The list of only-H2AK121ub, H2AK121ub/H3K27me3, and only-H3K27me3 marked genes are provided as a Source Data file.

to genes marked with H3K27me3 in both WT and *emf1-2* roots (Supplementary Fig. 6).

Regarding H2AK121ub marks, average H2AK121ub signal levels at H2AK121ub/H3K27me3 and only-H2AK121ub marked genes were significantly reduced in *bmi1abc*, increased in *clf28swm7*, and unaltered in *lhp1*, as previously reported<sup>28</sup> (Fig. 3c, d). The fact that *clf28swm7* displayed increased H2AK121ub levels at both H2AK121ub/H3K27me3 and only-H2AK121ub marked genes (Fig. 3c, d) suggests an indirect

consequence, since this effect was observed at only-H2AK121ub genes that are not CLF or SWN targets<sup>41</sup>. On the other hand, except for a few genes, we did not find significantly altered levels of H2AK121ub in *ring1ab* and *emf1-2* (Fig. 3c, d), which is consistent with a reduced but not eliminated RING1 activity and a low impact of EMF1 in H2AK121ub marking, respectively. All together, these data show that the levels of H2AK121ub are differentially affected in all these PcG mutants.

**H2AK121ub and EMF1 reduce chromatin accessibility at PcG targets.** Since we observed a generalized increase in chromatin accessibility in severe mutants associated with the embryonic phenotype, we investigated whether the loss of PcG marks actually plays a role in regulating chromatin accessibility. For this, we analyzed the accessibility profile around the TSS of only-H2AK121ub, H2AK121ub/H3K27me3 and only-H3K27me3 marked genes dividing the genes in groups according to their levels of H2AK121ub or H3K27me3 in the different mutants compared to WT (Fig. 4 and Supplementary Data 3). We first focused on only-H2AK121ub-marked genes (Fig. 4a). We found that decreasing levels of H2AK121ub led to a significant progressive increase in accessibility (Fig. 4a) whereas increasing levels led to reduced accessibility, supporting the association of H2AK121ub marks with a less accessible chromatin. However, we also found that expression levels of only-H2AK121ub-marked genes in WT, although not as high as in active genes lacking PcG marks (non-PcG targets), were higher than in H2AK121ub/H3K27me3, and only-H3K27me3-marked genes (Supplementary Fig. 7). Thus, as gene expression has been correlated with accessible chromatin<sup>30,42</sup>, these results suggest that H2AK121ub marks at only-H2AK121ub-marked genes associate with a less accessible although still permissive chromatin. Then, we analyzed H2AK121ub/H3K27me3-marked genes, which are generally repressed<sup>28</sup>. Since the incorporation of H3K27me3 at these genes is dependent on H2AK121ub<sup>28</sup>, their lower expression levels (Supplementary Fig. 7) suggests a role of H3K27me3 marks in further decreasing accessibility. Accordingly, we found that reduced levels of H2AK121ub in *bmi1abc* induced a significant loss of H3K27me3 and an increase in accessibility (Fig. 4b). Surprisingly, in *clf28swm7*, in which H3K27me3 is fully eliminated and H2AK121ub increases at many genes, accessibility was not increased as much as in *bmi1abc*, suggesting that either H2AK121ub and/or some other factor is preventing a gain of chromatin accessibility in this mutant. Importantly, although the levels of H2AK121ub were unaltered at H2AK121ub/H3K27me3-marked genes in *emf1-2*, genes with reduced levels of H3K27me3 in this mutant showed higher accessibility than in *clf28swm7*, indicating that EMF1, acting downstream H2AK121ub, contributes in reducing chromatin accessibility. Increased chromatin accessibility in *emf1-2* was also observed at only-H3K27me3-marked genes, which are not regulated by *bmi1abc* (Fig. 4c), supporting that EMF1 plays a role of in establishing an inaccessible chromatin.

We noted that the accessibility peak at only-H3K27me3-marked genes in *emf1-2* was smaller than that at H2AK121ub/H3K27me3 genes (Fig. 4b, c), suggesting that chromatin accessibility at only-H3K27me3 genes is less susceptible to a decrease in the levels of H3K27me3. On the other hand, although H3K27me3 levels were affected in *lhp1* and *ring1ab*, we did not find significant accessibility changes at any of the gene subsets.

**Accessibility increases in mutants independently of transcription.** Gene expression has been correlated with increased chromatin accessibility<sup>31,37,42</sup>. In line with this, we found that in WT the genes displaying the highest expression levels were the ones with highest chromatin accessibility (Supplementary Fig. 8). Hence, as loss of PcG function causes upregulation of a considerable number of PcG target and non-target genes<sup>26,28,33,40,43</sup> (Supplementary Fig. 9a), PCA analysis of the transcriptome in the different mutants (Supplementary Fig. 9b) showed a similar clustering to the one observed with ATAC-seq data (Fig. 1a). However, it is also known that many PcG target genes do not become upregulated in PcG mutants despite losing PcG marks<sup>26,28,33,40,43</sup>. Therefore, we wondered whether the increased

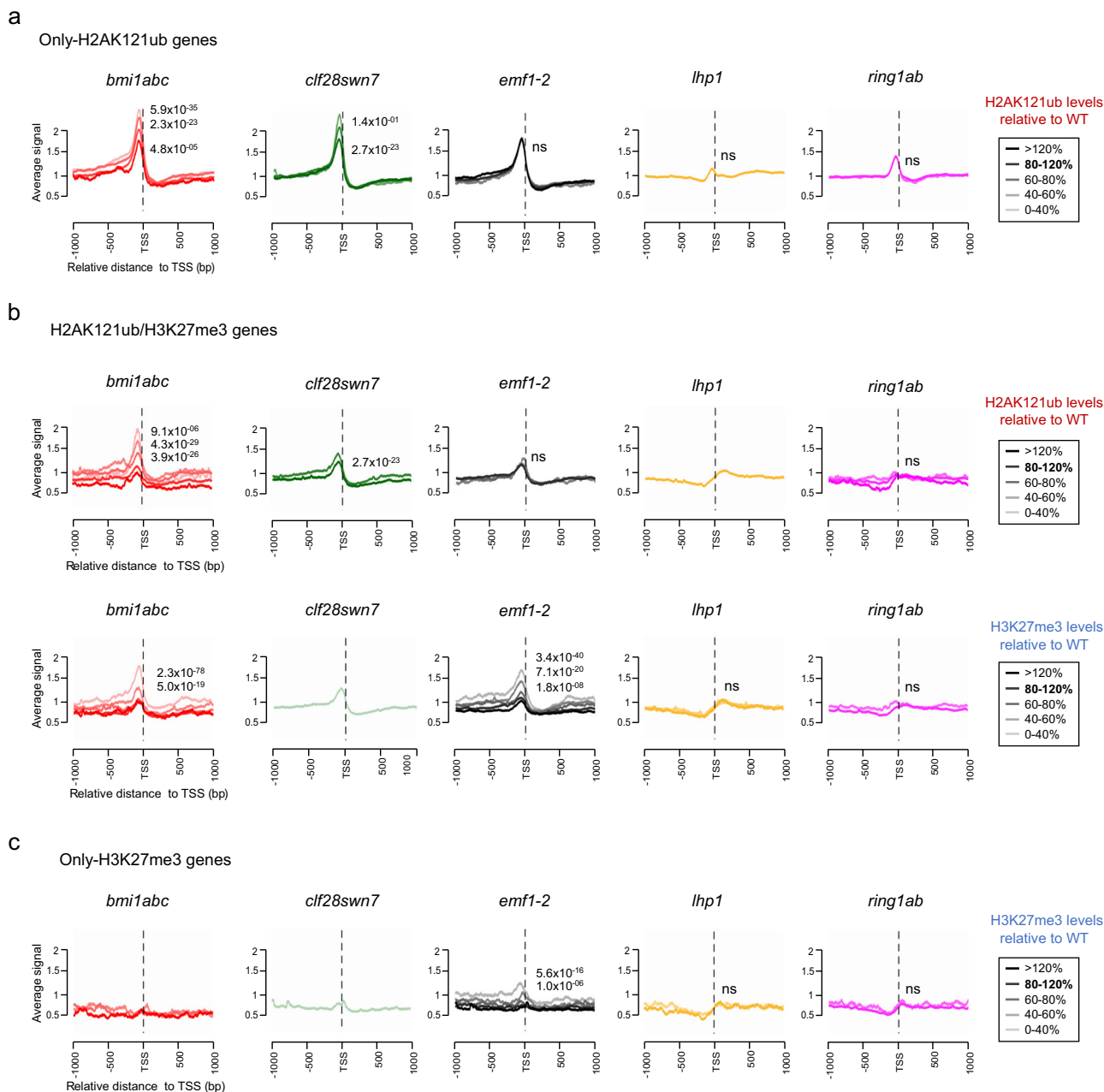
chromatin accessibility found at PcG targets in *bmi1abc*, *clf28swm7*, and *emf1-2* was a consequence of transcriptional upregulation of target genes or if it was caused by the loss of PcG function independently of the transcriptional state. To investigate this, we integrated accessibility, histone mark levels, and gene expression data in the different mutants. Since accessibility changes at only-H2AK121ub, H2AK121ub/H3K27me3, or only-H3K27me3 marked genes in mutants were evident at genes displaying less than 60% of WT H2AK121ub and/or H3K27me3 levels (Fig. 4), we analyzed whether or not the expression of the genes with less than 60% of WT marks (cutoff  $\log_2FC \leq -0.74$ ) was altered in *bmi1abc*, *clf28swm7*, and *emf1-2* (Fig. 5). Scatter plots in mutants representing accessibility versus histone mark levels confirmed that reduced levels of H2AK121ub and/or H3K27me3 led to increased chromatin accessibility (Fig. 5a–c and Supplementary Fig. 10). Interestingly, despite a high percentage of these genes were upregulated in mutants, we found a similar or even higher percentage of genes displaying unaltered expression levels. Remarkably, the percentage of genes with unaltered expression levels was particularly high in the subset of only-H3K27me3-marked genes (Fig. 5a–c). All together, these results indicate that the increased chromatin accessibility found at PcG targets is not caused by gene expression.

## Discussion

PcG complexes play central roles in eukaryotic gene regulation. However, despite these complexes have been proposed to define a closed chromatin conformation<sup>44</sup>, little is known about their role in regulating chromatin accessibility, especially in plants. Instead, PcG function has been extensively correlated to gene repression<sup>44</sup>.

While the association of PRC2 marking and gene repression is widely accepted in plants, PRC1-mediated H2AK121ub has been recently proposed to be associated with gene responsiveness<sup>32</sup>. Despite PRC1 activity is required to recruit PRC2 for H3K27me3 marking at many genes, there are also genes only marked with H2AK121ub or H3K27me3 (refs. 28,32). Interestingly, average expression levels of only-H2AK121ub genes are higher than H2AK121ub/H3K27me3 or only-H3K27me3 (ref. 28), which may argue against a repressive role of this modification. In line with this, a recent report proposed that H2AK121ub at H2AK121ub/H3K27me3-marked genes does not play a repressive role by itself, as despite it allows recruitment of PRC2, and thus gene repression, the removal of H2AK121ub at these genes seems to prevent gene reactivation by interfering with active H3K27me3 demethylation<sup>32</sup>. However, we found that the expression levels of only-H2AK121ub-marked genes were significantly lower than non-PcG active genes, supporting that this modification may play a role in modulating gene expression.

Nevertheless, although it is true that inaccessible chromatin is refractory to gene expression, chromatin accessibility by itself does not impose a transcriptional state<sup>45</sup>. Therefore, to understand PcG function it is fundamental to determine whether PRC1, PRC2, or their respective hallmarks play a role in regulating chromatin accessibility. Our data revealed that indeed PcG activities regulate chromatin accessibility. We found that reduced levels of H2AK121ub led to increased chromatin accessibility, whereas increased levels of this modification caused the opposite effect, supporting a role of this modification in regulating chromatin accessibility. It might be argued that the BMI1 proteins rather than H2AK121ub marks are the ones responsible for this effect. However, this role is supported by the fact that accessibility changes associated with the levels of H2AK121ub are observed in *clf28swm7*. In this mutant, the expression of *BMI1s* is not altered and the recruitment of the BMI1s should not be affected, as PRC1 recruitment is independent of PRC2. In any case, our data cannot



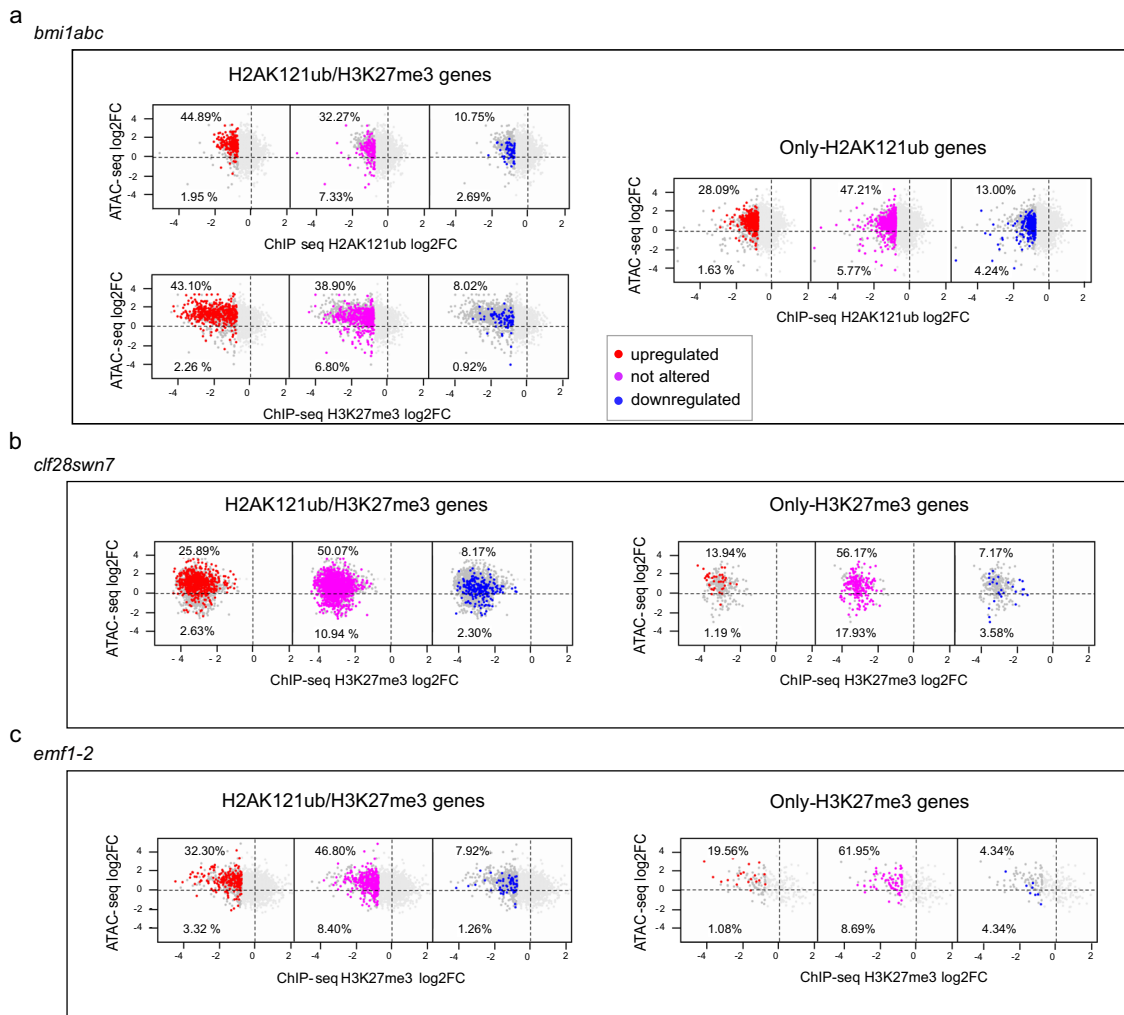
**Fig. 4 PRC1-mediated H2AK121ub and EMF1 in collaboration with PRC2 activity are required to reduce chromatin accessibility.** **a** Accessibility profiles around the TSS of only-H2AK121ub-marked genes in the different mutants. Genes were classified into different groups depending on their levels of H2AK121ub marks relative to WT. The profile of the different groups is indicated in a color increasing palette from lowest to highest H2AK121ub levels. **b** Accessibility profiles around the TSS of H2AK121ub/H3K27me3-marked genes in the different mutants, grouping the genes according to their levels of H2AK121ub marks (upper panels) or H3K27me3 (bottom panels) relative to WT. The profile of the different gene groups is indicated in a color increasing palette from lowest to highest relative H2AK121ub or H3K27me3 levels. **c** Accessibility profiles around the TSS of only-H3K27me3-marked genes in the different mutants, grouping the genes according to their levels of H3K27me3 marks relative to WT. The profile of the different gene groups is indicated in a color increasing palette from lowest to highest relative H3K27me3 levels. Gene groups with a very small number of genes were excluded as they produced noisy profiles (see Supplementary Data 3). In all cases (**a-c**), *p* values of significant differences between the profile of genes with WT-like levels (80-120%) and those of genes with different levels are indicated according to one-sided Mann-Whitney-Wilcoxon test.

rule out the possibility of a direct implication of BMI1 proteins in regulating chromatin accessibility instead of H2AK121ub.

We also found that H2AK121ub associates with transcriptional regulation hotspots, which are sites enriched for the binding of a wide variety of TFs, and thus, required for transcriptional responses. Hence, we propose that H2AK121ub marks help to create a less accessible although still permissive chromatin at these regulatory hotspots (Fig. 6). This is consistent with the

proposed hypothesis that H2AK121ub could provide a standby mode between activation and repression<sup>32</sup>.

In addition, our results indicate that chromatin accessibility can be further reduced by EMF1. This plant-specific protein has been shown to associate with PRC2 and to be required for H3K27me3 marking<sup>19,21-25</sup>. Accordingly, loss of EMF1 function affected H3K27me3 levels but not H2AK121ub. However, *emf1-2* ChIP-seq results showed that not all H3K27me3-marked genes



**Fig. 5 Increased chromatin accessibility in PcG mutants is not necessarily accompanied by gene expression.** **a** Scatter plots showing the relationship between THS accessibility and marks levels (H2K121ub or H3K27me3) at H2AK121ub/H3K27me3 (left) and only-H2K121ub marked genes (right) in *bmi1abc* mutant. **b** Scatter plots showing the relationship between THS accessibility and H3K27me3 levels at H2AK121ub/H3K27me3 (left) and only-H3K27me3 marked genes (right) in *clf28swn7* mutant. **c** Scatter plots showing the relationship between THS accessibility and H3K27me3 levels at H2AK121ub/H3K27me3 (left) and only-H3K27me3 marked genes (right) in *emf1-2* mutant. In each panel, top numbers indicate the percentage of genes displaying altered or unaltered expression levels within the subset of genes showing less than 60% of WT H2K121ub or H3K27me3 (cutoff  $\log_2FC \leq -0.74$ ) and increased accessibility compared to WT. While bottom numbers indicate the same for the genes with less than 60% of WT H2K121ub or H3K27me3 and decreased accessibility compared to WT. The sum of top and bottom numbers from each panel represent 100% of the genes with less than 60% of WT H2K121ub or H3K27me3 and altered accessibility compared to WT. Upregulated expression is indicated in red, not altered expression in pink, and downregulated expression in blue. Source data are provided as a Source Data file.

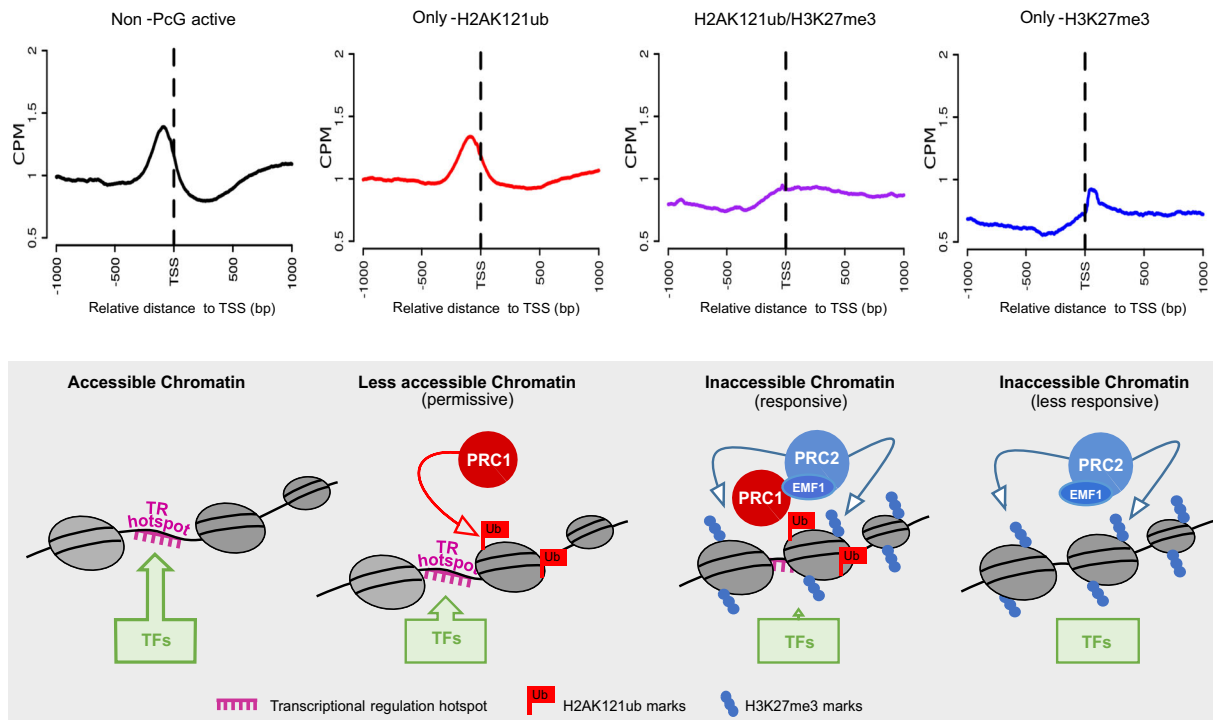
display reduced levels. Furthermore, *emf1-2* phenotype suggests that EMF1 is dispensable for H3K27me3 marking at root tissue. Interestingly, a recent phylogenetic analysis of EMF1 showed that ancestral EMF1 gene duplicated twice resulting in three orthologous groups that, next to EMF1, contain three novel out-paralogs<sup>46</sup>. This finding rises the possibility that these paralogs display EMF1-like functions in different tissues or in the regulation of different subsets of genes.

In any case, we found that the H3K27me3-marked genes that lose H3K27me3 in *emf1-2* displayed higher chromatin accessibility than *clf28swn7*, indicating that EMF1 acting upstream H3K27me3 marking plays an important role in regulating chromatin accessibility. Interestingly, EMF1 has the ability to in vitro mediate chromatin compaction<sup>20</sup>. Local and/or higher order chromatin compaction affects chromatin accessibility<sup>45</sup>. Furthermore, local chromatin compaction has been shown to precede the establishment of H3K27me3 in mouse embryonic stem

cells<sup>47</sup>, which supports the connection between EMF1 and PRC2 activity. EMF1 ability to in vitro compact chromatin is shared with *Drosophila* PRC1 component Psc<sup>20</sup>, which combines BMI1 and EMF1 functions in a single polypeptide<sup>48</sup>, and with vertebrate Cbx2 (ref. 13), which is one of the Pc homologs<sup>49</sup>, indicating that, unlike in animals, EMF1 function in plants is associated to PRC2.

Based on our results, we also propose that chromatin at H2AK121ub/H3K27me3-marked genes associates with inaccessible but responsive chromatin, most likely due to the presence of H2AK121ub-marked transcriptional regulation hotspots, which may be important to allow gene reprogramming (Fig. 6). Conversely, only-H3K27me3-marked genes, which are generally not linked with these hotspots, are less responsive (Fig. 6), suggesting that the lack of H2AK121ub-marked transcriptional hotspots may prevent the ectopic expression of these genes.





**Fig. 6 H2AK121ub, H3K27me3, and the combination of both marks define distinct chromatin accessibility states.** Accessibility profiles around the TSS of non-PcG-marked active genes, only-H2AK121ub, H2AK121ub/H3K27me3, and only-H3K27me3 marked genes in WT (upper panels). Model proposed for the role of H2AK121ub and/or H3K27me3 marks in defining different chromatin states that control transcriptional responses to different degrees (bottom panels).

In vertebrate, a recent report showed that despite PcG target promoters exhibit reduced chromatin accessibility, removal of PRC1, but not PRC2, causes reduced nucleosome occupancy and increase nucleosome spacing without affecting chromatin accessibility<sup>16</sup>. We found this result surprising, as one would have predicted that these effects would yield at least locally increased chromatin accessibility. In any case, under our experimental conditions and integrating chromatin accessibility, histone marks levels, and expression data in different PcG mutants, we were able to show that the loss of PcG activities in *Arabidopsis* leads to increased chromatin accessibility at PcG targets. Furthermore, our findings indicate that this increased chromatin accessibility is not caused by gene expression, which explains the lack of transcriptional activation of many PcG target genes that lose PcG marks in mutant plants<sup>26,28,33,40,43</sup>.

## Methods

**Plant materials and growth conditions.** *Arabidopsis thaliana* WT Col-0, *bmi1abc*<sup>8</sup>, *clf28sw7* (ref. 33), *emf1-2* (ref. 50), *lhp1* (ref. 51), and *ring1ab*<sup>34</sup> mutants were grown under long-day condition (16 h of light/8 h of dark) at 22 °C on MS agar plates containing 1% sucrose and 0.7% agar.

**ATAC-seq and data analysis.** Nuclei isolation of ATAC-seq experiments were performed as previously described following the sucrose sedimentation protocol<sup>52</sup>. Briefly, nuclei from 10-day-old whole seedlings were separated from other cellular debris by passing the nuclei suspension through a sucrose cushion during centrifugation. After centrifugation, the contaminating organelles and debris are visible at the top of the tube and nuclei at the bottom. Next, 30,000 nuclei were counted and resuspended in 50  $\mu$ L transposition reaction mix using the True-PrepTM DNA Library Prep Kit V2 for Illumina® (Vazyme, TD501). Reactions were incubated for 30 min at 37 °C before isolation DNA using the VAHTSTM DNA Clean Beads (Vazyme, N411). Two independent biological replicates were processed for next-generation sequencing library preparation. Libraries were prepared by PCR amplification using two index (i5 and i7) Illumina barcodes (Vazyme, TD202) with 13 cycles. Libraries were purified with a VAHTSTM DNA Clean Beads. ATAC libraries were sequenced on Illumina HiSeq-Xten PE150 by generating 2  $\times$  150-bp paired-end reads.

We obtained around 45 million paired-end reads for each independent biological replicates of the different genotypes. Quality trimming and adapters removal were performed using Trimmomatic<sup>53</sup>. The organelle genomes were excluded from the TAIR10 *Arabidopsis* reference genome before reads were mapped using bowtie. PCR duplicates were discarded from the mapped reads using Picard (Picard Toolkit, 2019, Broad Institute, GitHub Repository).

<http://broadinstitute.github.io/picard/>; Broad Institute). Regions with an artifactual massive amount of unique mapped reads were identified. Specifically, 91 regions were collected into a list of blacklisted genomic regions (Supplementary Data 4). In order to prevent interference with subsequent analysis steps, reads mapped to the blacklisted genomic regions were eliminated using samtools. Next, THSs were identified by performing a peak calling analysis using MACS2 with parameters—nomodel—shift -100—extsize 200—q 0.05. For each genotype, genotype-specific THSs were determined by intersecting the THSs found on each individual replicate using bedtools. A final set of consensus THSs was generated by merging the THSs identified in each genotype. Using different deepTools functions we generated bigwig files storing the genome-wide accessibility signal measured as CPM for the different genotype replicates. Average accessibility profiles for the different genotypes were generated using the bioconductor R package ChIPpeakAnno. An accessibility signal matrix was generated where rows represented THSs and columns genotype replicates. For each THS, its average accessibility signal was computed for each sample. Rows corresponding to THSs exhibiting a maximum signal level below the 5% lower percentile of the signal distribution, approximately 3 CPM were filtered out in order to reduce noise in our analysis. After log<sub>2</sub> transformation of the filtered matrix, the bioconductor R package limma was used to identify significantly reduced or increased accessibility levels in the different mutants when compared to WT-like levels (80–120% levels). The statistical significance of differences in accessibility around TSS were assessed using the non-parametric Mann–Whitney–Wilcoxon test implemented in the Wilcox.test function in R. Regression analysis was performed using the *lm* (linear model) function in R. The overall significance of the regression analysis was carried out using the *F*-test.

The bioconductor R package Chipseeker<sup>54</sup> was used to study the location of THS around the TSS. The percentages of consensus THSs located nearby histone marks were determined using different functions from deepTools.

Transcription factor-binding sites (TFBS) have been determined in this study by reanalyzing in a uniform manner 100 independent ChIP-seq datasets previously published covering most TF families in *Arabidopsis thaliana* (see Supplementary Data 2). The TFBS enrichment at the THS was estimated using Monte Carlo simulations. First, we computed the median number of TFBS overlapping with our consensus THSs using deepTools. Then, we generated 10<sup>4</sup> sets of random genomic regions using a custom R script. These random sets were composed of as many genomic regions as consensus THSs. Moreover, the random regions have the same

length as our THSs. For each one of the randomly generated sets we computed the median number of TFBS overlapping with the corresponding random genomic regions. By comparing these random values to the actual number of TFBS in our consensus THSs we estimated the corresponding *p* value and enrichment.

**ChIP-seq and data analysis.** H2AK121ub and H3K27me3 ChIP-seq data of *bmi1abc*, *clf28swm7*, and WT at 7 DAG, and H2AK121ub data of *lhp1* at 7 DAG were previously generated<sup>28</sup>. H2AK121ub and H3K27me3 ChIP-seq data of WT, *ring1ab*, *emf1-2*, and H3K27me3 data of *lhp1* at 10 DAG were generated for this work.

ChIP-seq experiments were performed as previously described<sup>55</sup>. In brief, whole seedlings were fixed in 1% formaldehyde. Chromatin was extracted from fixed tissue and fragmented using a Bioruptor® Pico (Diagenode) in fragments of 200–500 bp. The sheared chromatin was immunoprecipitated overnight using the following antibodies: Anti-H2Aub (Cell Signaling Technology, 8240S, dilution 1:100) and anti-H3K27me3 (Millipore, 07-449, dilution 1:300). Immunocomplexes were captured using Protein A Sepharose beads CL-4B (GE Healthcare). After washing the Protein-A beads, chromatin was eluted and the crosslinking was reversed overnight at 65 °C. The DNA from the immunoprecipitated chromatin was treated with RNase and proteinase K and purified by phenol–chloroform extraction followed by ethanol precipitation. For ChIP-seq, two immunoprecipitations from independent biological replicates were processed for next-generation sequencing library preparation. All libraries were prepared with end repair, A-tailing and ligation of Illumina-compatible adapters using the Ovation® Ultralow Library Systems (NuGEN). The ligated product was amplified with 14 cycles of PCR. DNA of a size range between 200 and 600 bp was purified from an agarose gel. Amplification was confirmed by testing an aliquot of the library before and after amplification by qPCR. Libraries were sequenced on Illumina HiSeq-Xten PE150 by generating 2 × 150-bp paired-end reads.

Each sample for the different genotypes generated approximately 12 million paired-end reads. Bowtie<sup>56</sup> was used to map reads to the TAIR10 *Arabidopsis* reference genome. The tool boxes for the analysis of high-throughput sequencing data deepTools<sup>57</sup> and samtools<sup>58</sup> were used to generate bigwig files storing H3K27me3 and H2AK121ub genome-wide signal levels measured as CPM (counts per million). Specifically, the function `bamCoverage` was used to compute normalized signal levels using the parameter `-normalizeUsing CPM`. Mark peaks were detected using MACS2 (ref. 59) with default parameters for each individual replicate of the WT genotype. Specifically, the function `bamCoverage` was used to compute normalized signal levels using the parameter `-normalizeUsing CPM`. Mark peaks were detected using MACS2 (ref. 59) with default parameters for each individual replicate of the WT genotype. A specific input for each of the two datasets was considered, namely an input for the previously published data corresponding to 7 DAG and another specific input for the data generated in this study from 10 DAG seedlings. The final set of mark peaks were defined as the intersection of the ones detected for each replicate and the combination obtained from the 7 DAG and 10 DAG WT data using bedtools<sup>60</sup>. Target genes were associated to mark peaks using the bioconductor R package `ChIPpeakAnno`<sup>61</sup> when an overlap was detected with either the gene body or promoter defined as the region 750 bp upstream from the TSS. A matrix was constructed where rows represented marked genes and columns the different genotype samples. For each marked gene a signal level in each one of the genotype samples under study was computed as the average signal across the corresponding histone mark peak. In order to remove noisy peaks, we filtered out genes with signals below 3 CPM, approximately corresponding to the 5% lower percentile of the signal distribution. After log2 transformation of the filtered matrix the bioconductor R package `limma`<sup>62</sup> was used to identify significantly reduced or increased histone mark signal levels in the different mutants when compared to the WT.

Metagene plots were generated with a custom R script available from our github repository (<https://doi.org/10.5281/zenodo.4304639>)<sup>69</sup>. The statistical significance of differences in metagene plots were assessed using the non-parametric Mann–Whitney–Wilcoxon test implemented in the `Wilcox.test` function in R. Similarly, the identification of binding sites for the TFs under analysis was performed using bowtie for read mapping and macs2 for peak calling.

**ChIP-reChIP-qPCR analysis.** ChIP-reChIP experiments were performed as previously described<sup>63</sup>. In brief, 10-day-old WT seedlings were fixed in 1% formaldehyde. Chromatin was extracted from fixed tissue and fragmented using a Bioruptor® Pico (Diagenode) in fragments of 500–1000 bp. ChIP was first performed with anti-H3K27me3 (Millipore, 07-449, dilution 1:300). Immunocomplexes were captured using Protein A Sepharose beads CL-4B (GE Healthcare). After washing the Protein-A beads, one fraction was processed as in a conventional ChIP assay in order to revert the crosslinking and purify the DNA, and another fraction was used to elute the immunoprecipitated protein–DNA complexes in Re-ChIP elution buffer (2 mM EDTA, 500 mM NaCl, 0.1% SDS, 1% NP40) by incubation for 30 min at 37 °C. The eluate was diluted 20 times with ChIP dilution buffer supplemented with 50 µg of BSA and protease inhibitor. Then, the second ChIP was performed with anti-H2Aub (Cell Signaling Technology, 8240S, dilution 1:100). Immunocomplexes were captured using Protein A Sepharose beads CL-4B (GE Healthcare). After washing the Protein-A beads, chromatin was eluted, de-

crosslinked, and the DNA was extracted. The DNA obtained in the two sequential ChIPs were used as PCR template to amplify *PLT5*, *SOC1*, and *AG* using specific primers (Supplementary Table 1).

**RNA-seq and data analysis.** WT at 7 DAG RNA-seq data of *bmi1abc* have been previously published<sup>28</sup>. RNA-seq data of WT and the remaining mutants at 10 DAG were generated for this work.

Total RNA was extracted using E.Z.N.A.® Plant RNA Kit (Omega, R6827-01) from 10-day-old whole seedlings following the manufacturer’s instructions. Three independent biological replicates were processed for next-generation sequencing library preparation. Libraries were prepared using 1 µg of total RNA with the VAHTSTM Total RNA-seq (H/M/R) Library Prep Kit for Illumina® (Vazyme, NR603) according to the manufacturer’s instructions. Libraries were sequenced on Illumina HiSeq-Xten PE150 by generating 2 × 150-bp paired-end reads.

Each independent biological replicate of the different genotypes generated approximately 20 million paired-end reads. Quality control was performed using FASTQC. Read mapping to the TAIR10 *Arabidopsis* reference genome was carried out using HISAT2 (ref. 64). Transcript assembly and gene expression estimation measured as FPKM (fragments per kilobase of exon and million of mapped reads) were computed using Stringtie<sup>65</sup> and the bioconductor R package `ballgown`<sup>66</sup>. Differential gene expression analysis was performed using the bioconductor R package `limma`<sup>62</sup>. Specifically, a log2 fold-change of ±1 and a *p* value of 0.05 was used to determine activated, repressed, and unaltered genes. GO enrichment analysis was performed over the different gene sets of interest using the bioconductor R package `clusterProfiler`<sup>67</sup>. PCA was carried out using the R package `FactoMineR` v2.3 (ref. 68).

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

Data supporting the findings of this work are available within the paper and its Supplementary Information files. A reporting summary for this article is available as a Supplementary Information file. The datasets and plant materials generated and analyzed during the current study are available from the corresponding author upon request. ATAC-seq, ChIP-seq, and RNA-seq datasets generated in this study have been deposited in the Gene Expression Omnibus (GEO) under accession [GSE155378](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE155378). Previously generated ChIP-seq and RNA-seq data are under accession [GSE89358](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE89358). Source data are provided with this paper.

## Code availability

Several custom R scripts that were used to generate signal profiles have been deposited in the GitHub repository (<https://doi.org/10.5281/zenodo.4304639>)<sup>69</sup>.

Received: 2 July 2020; Accepted: 7 December 2020;

Published online: 12 January 2021

## References

- Ringrose, L. & Paro, R. Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. *Annu. Rev. Genet.* **38**, 413–443 (2004).
- Merini, W. & Calonje, M. PRC1 is taking the lead in PcG repression. *Plant J. Cell Mol. Biol.* **83**, 110–120 (2015).
- 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).
- Wang, H. et al. Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **431**, 873–878 (2004).
- Cao, R., Tsukada, Y.-I. & Zhang, Y. Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing. *Mol. Cell* **20**, 845–854 (2005).
- Bratzel, F., López-Torrejón, G., Koch, M., Del Pozo, J. C. & Calonje, M. Keeping cell identity in *Arabidopsis* requires PRC1 RING-finger homologs that catalyze H2A monoubiquitination. *Curr. Biol.* **20**, 1853–1859 (2010).
- Bratzel, F. et al. Regulation of the new *Arabidopsis* imprinted gene *AtBMI1C* requires the interplay of different epigenetic mechanisms. *Mol. Plant* **5**, 260–269 (2012).
- Yang, C. et al. VAL- and AtBMI1-mediated H2Aub initiate the switch from embryonic to postgerminative growth in *Arabidopsis*. *Curr. Biol.* **23**, 1324–1329 (2013).
- Müller, J. et al. Histone methyltransferase activity of a *Drosophila* Polycomb group repressor complex. *Cell* **111**, 197–208 (2002).
- Cao, R. et al. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**, 1039–1043 (2002).

11. Makarevich, G. et al. Different Polycomb group complexes regulate common target genes in *Arabidopsis*. *EMBO Rep.* **7**, 947–952 (2006).
12. Mozgova, I. & Hennig, L. The Polycomb group protein regulatory network. *Annu. Rev. Plant Biol.* **66**, 269–296 (2015).
13. Grau, D. J. et al. Compaction of chromatin by diverse Polycomb group proteins requires localized regions of high charge. *Genes Dev.* **25**, 2210–2221 (2011).
14. Trojer, P. et al. L3MBTL2 protein acts in concert with PcG protein-mediated monoubiquitination of H2A to establish a repressive chromatin structure. *Mol. Cell* **42**, 438–450 (2011).
15. Deaton, A. M. et al. Enhancer regions show high histone H3.3 turnover that changes during differentiation. *eLife* **5**, e15316 (2016).
16. King, H. W., Fursova, N. A., Blackledge, N. P. & Klose, R. J. Polycomb repressive complex 1 shapes the nucleosome landscape but not accessibility at target genes. *Genome Res.* **28**, 1494–1507 (2018).
17. Mozgova, I., Köhler, C. & Hennig, L. Keeping the gate closed: functions of the polycomb repressive complex PRC2 in development. *Plant J. Cell. Mol. Biol.* **83**, 121–132 (2015).
18. Calonje, M. PRC1 marks the difference in plant PcG repression. *Mol. Plant* **7**, 459–471 (2014).
19. Calonje, M., Sanchez, R., Chen, L. & Sung, Z. R. EMBRYONIC FLOWER1 participates in polycomb group-mediated AG gene silencing in *Arabidopsis*. *Plant Cell* **20**, 277–291 (2008).
20. Beh, L. Y., Colwell, L. J. & Francis, N. J. A core subunit of Polycomb repressive complex 1 is broadly conserved in function but not primary sequence. *Proc. Natl Acad. Sci. USA* **109**, E1063–E1071 (2012).
21. Derkacheva, M. et al. *Arabidopsis* MSI1 connects LHP1 to PRC2 complexes. *EMBO J.* **32**, 2073–2085 (2013).
22. Liang, S. C. et al. Kicking against the PRCs—a domesticated transposase antagonises silencing mediated by polycomb group proteins and is an accessory component of polycomb repressive complex 2. *PLoS Genet.* **11**, e1005660 (2015).
23. Bloomer, R. H. et al. The *Arabidopsis* epigenetic regulator ICU11 as an accessory protein of Polycomb Repressive Complex 2. *Proc. Natl Acad. Sci. USA* **117**, 16660–16666 (2020).
24. 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).
25. Li, Z., Fu, X., Wang, Y., Liu, R. & He, Y. Polycomb-mediated gene silencing by the BAH-EMF1 complex in plants. *Nat. Genet.* **50**, 1254–1261 (2018).
26. Turck, F. et al. *Arabidopsis* TFL2/LHP1 specifically associates with genes marked by trimethylation of histone H3 lysine 27. *PLoS Genet.* **3**, e86 (2007).
27. Zhang, X. et al. The *Arabidopsis* LHP1 protein colocalizes with histone H3 Lys27 trimethylation. *Nat. Struct. Mol. Biol.* **14**, 869–871 (2007).
28. Zhou, Y., Romero-Campero, F. J., Gómez-Zambrano, Á., Turck, F. & Calonje, M. H2A monoubiquitination in *Arabidopsis thaliana* is generally independent of LHP1 and PRC2 activity. *Genome Biol.* **18**, 69 (2017).
29. Shu, H., Wildhaber, T., Siretskiy, A., Gruijsem, W. & Hennig, L. Distinct modes of DNA accessibility in plant chromatin. *Nat. Commun.* **3**, 1281 (2012).
30. Tannenbaum, M. et al. Regulatory chromatin landscape in *Arabidopsis thaliana* roots uncovered by coupling INTACT and ATAC-seq. *Plant Methods* **14**, 113 (2018).
31. Frerichs, A., Engelhorn, J., Altmüller, J., Gutierrez-Marcos, J. & Werr, W. Specific chromatin changes mark lateral organ founder cells in the *Arabidopsis* inflorescence meristem. *J. Exp. Bot.* **70**, 3867–3879 (2019).
32. Kralemann, L. E. M. et al. Removal of H2Aub1 by ubiquitin-specific proteases 12 and 13 is required for stable Polycomb-mediated gene repression in *Arabidopsis*. *Genome Biol.* **21**, 144 (2020).
33. Lafos, M. et al. Dynamic regulation of H3K27 trimethylation during *Arabidopsis* differentiation. *PLoS Genet.* **7**, e1002040 (2011).
34. Xu, L. & Shen, W.-H. Polycomb silencing of KNOX genes confines shoot stem cell niches in *Arabidopsis*. *Curr. Biol.* **18**, 1966–1971 (2008).
35. Gaudin, V. et al. Mutations in LIKE HETEROCHROMATIN PROTEIN 1 affect flowering time and plant architecture in *Arabidopsis*. *Dev. Camb. Engl.* **128**, 4847–4858 (2001).
36. Sijacic, P., Bajic, M., McKinney, E. C., Meagher, R. B. & Deal, R. B. Changes in chromatin accessibility between *Arabidopsis* stem cells and mesophyll cells illuminate cell type-specific transcription factor networks. *Plant J. Cell Mol. Biol.* **94**, 215–231 (2018).
37. Gaspar-Maia, A., Alajem, A., Meshorer, E. & Ramalho-Santos, M. Open chromatin in pluripotency and reprogramming. *Nat. Rev. Mol. Cell Biol.* **12**, 36–47 (2011).
38. Schlesinger, S. & Meshorer, E. Open chromatin, epigenetic plasticity, and nuclear organization in pluripotency. *Dev. Cell* **48**, 135–150 (2019).
39. Maher, K. A. et al. Profiling of accessible chromatin regions across multiple plant species and cell types reveals common gene regulatory principles and new control modules. *Plant Cell* **30**, 15–36 (2018).
40. 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).
41. Shu, J. et al. Genome-wide occupancy of histone H3K27 methyltransferases CURLY LEAF and SWINGER in *Arabidopsis* seedlings. *Plant Direct* **3**, e00100 (2019).
42. Alvarez, J. M. et al. Local changes in chromatin accessibility and transcriptional networks underlying the nitrate response in *Arabidopsis* roots. *Mol. Plant* **12**, 1545–1560 (2019).
43. Bouyer, D. et al. Polycomb repressive complex 2 controls the embryo-to-seedling phase transition. *PLoS Genet.* **7**, e1002014 (2011).
44. Schuettengruber, B., Bourbon, H.-M., Di Croce, L. & Cavalli, G. Genome regulation by Polycomb and Trithorax: 70 years and counting. *Cell* **171**, 34–57 (2017).
45. Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* **20**, 207–220 (2019).
46. Berke, L. & Snel, B. The plant Polycomb repressive complex 1 (PRC1) existed in the ancestor of seed plants and has a complex duplication history. *BMC Evol. Biol.* **15**, 44 (2015).
47. Yuan, W. et al. Dense chromatin activates Polycomb repressive complex 2 to regulate H3 lysine 27 methylation. *Science* **337**, 971–975 (2012).
48. Brunk, B. P., Martin, E. C. & Adler, P. N. *Drosophila* genes posterior sex combs and suppressor two of zeste encode proteins with homology to the murine bmi-1 oncogene. *Nature* **353**, 351–353 (1991).
49. Völkel, P., Le Faou, P., Vandamme, J., Pira, D. & Angrand, P.-O. A human Polycomb isoform lacking the Pc box does not participate to PRC1 complexes but forms protein assemblies and represses transcription. *Epigenetics* **7**, 482–491 (2012).
50. Moon, Y.-H. et al. EMF genes maintain vegetative development by repressing the flower program in *Arabidopsis*. *Plant Cell* **15**, 681–693 (2003).
51. Kotake, T., Takada, S., Nakahigashi, K., Ohto, M. & Goto, K. *Arabidopsis* TERMINAL FLOWER 2 gene encodes a heterochromatin protein 1 homolog and represses both FLOWERING LOCUS T to regulate flowering time and several floral homeotic genes. *Plant Cell Physiol.* **44**, 555–564 (2003).
52. Bajic, M., Maher, K. A. & Deal, R. B. Identification of open chromatin regions in plant genomes using ATAC-Seq. *Methods Mol. Biol.* **1675**, 183–201 (2018).
53. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
54. Yu, G., Wang, L.-G. & He, Q.-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015).
55. Reimer, J. J. & Turck, F. Genome-wide mapping of protein-DNA interaction by chromatin immunoprecipitation and DNA microarray hybridization (ChIP-chip). Part A: ChIP-chip molecular methods. *Methods Mol. Biol.* **631**, 139–160 (2010).
56. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
57. Ramirez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
58. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
59. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
60. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
61. Zhu, L. J. et al. ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics* **11**, 237 (2010).
62. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
63. Truax, A. D. & Greer, S. F. ChIP and Re-ChIP assays: investigating interactions between regulatory proteins, histone modifications, and the DNA sequences to which they bind. *Methods Mol. Biol.* **809**, 175–188 (2012).
64. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
65. Kovaka, S. et al. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* **20**, 278 (2019).
66. Frazee, A. C. et al. Ballgown bridges the gap between transcriptome assembly and expression analysis. *Nat. Biotechnol.* **33**, 243–246 (2015).
67. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omic J. Integr. Biol.* **16**, 284–287 (2012).
68. Lê, S., Josse, J. & Husson, F. FactoMineR: an R package for multivariate analysis. *J. Stat. Softw.* **25**, 1–18 (2008).

69. Yin, X. et al. H2AK121ub in *Arabidopsis* associates with a less accessible chromatin state at transcriptional regulation hotspots. Available from GitHub repository. <https://doi.org/10.5281/zenodo.4304639> (2020).

### Acknowledgements

This work is supported by startup funds from State Key Laboratory for Protein and Plant Gene Research, School of Advanced Agricultural Sciences, Peking-Tsinghua Center for Life Sciences, Peking University, grant 31970532 from the National Natural Science Foundation of China and grants BIO2016-76457-P, PID2019-106664GB-I00, and BIO2017-84066-R from Spanish Ministry of Science and innovation. P.d.l.R. is supported by BIO2017-83629-R, granted to Dr. F. Valverde.

### Author contributions

X. Yin, P.Y., and X.M. performed ATAC-seq experiment. X. Yin, G.T., and S.Z. performed ChIP-seq experiment. J.Y., Y.Z., X.M., S.Z., X. Yin, J.Y. and X. Yang performed RNA-seq experiments. F.J.R.-C. and P.d.l.R. analyzed high-throughput sequencing data (ATAC-seq, ChIP-seq and RNA-seq, and TF-binding sites, respectively); M.C., Y.Z., and F.J.R.-C. interpreted the data and M.C. wrote the manuscript. All authors read and approved the final manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41467-020-20614-1>.

**Correspondence** and requests for materials should be addressed to F.J.R.-C., M.C. or Y.Z.

**Peer review information** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021