

RESEARCH ARTICLE

Chromatin state architecture governs transcription factor accessibility across plant genomes

Vikas Shukla^{1,2}, Elin Axelsson¹, Tetsuya Hisanaga^{1,3}, Jim Haseloff⁴, Frédéric Berger^{1*}, Facundo Romani^{4*}

1 Gregor Mendel Institute, Austrian Academy of Sciences, BioCenter, Vienna, Austria, **2** Vienna BioCenter PhD Program, Doctoral School of the University of Vienna and Medical University of Vienna, Vienna, Austria, **3** Current address: Biological Sciences, Nara Institute of Science and Technology, Ikoma, Nara, Japan, **4** Department of Plant Sciences, University of Cambridge, Cambridge, United Kingdom

* frederic.berger@gmi.oeaw.ac.at (FB); fr391@cam.ac.uk (FR)



OPEN ACCESS

Citation: Shukla V, Axelsson E, Hisanaga T, Haseloff J, Berger F, Romani F (2026) Chromatin state architecture governs transcription factor accessibility across plant genomes. *PLoS Genet* 22(1): e1012015. <https://doi.org/10.1371/journal.pgen.1012015>

Editor: John M. Grealley, Albert Einstein College of Medicine, UNITED STATES OF AMERICA

Received: December 4, 2025

Accepted: December 25, 2025

Published: January 22, 2026

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pgen.1012015>

Copyright: © 2026 Shukla et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution,

Abstract

The complexity of varied modifications of chromatin composition is integrated in archetypal combinations called chromatin states that predict the local potential for transcription. The degree of conservation of chromatin states has not been established amongst plants, and how they interact with transcription factors is unknown. Here we identify and characterize chromatin states in the flowering plant *Arabidopsis thaliana* and the bryophyte *Marchantia polymorpha*, showing a large degree of functional conservation over more than 450 million years of land plant evolution. We used this new resource of conserved plant chromatin states to understand the influence of chromatin states on gene regulation. We established the preferential association of chromatin states with binding sites and activity of transcription factors. These associations define three main groups of transcription factors that bind upstream of the transcription start site, at the + 1 nucleosome or further downstream of the transcription start site and broadly associate with distinct biological functions including a list of potential candidate pioneer factors we know little about in plants, compared to their important roles in animal stem cells and early development.

Author summary

In eukaryotes, DNA is tightly associated with histone proteins. Histone covalent modifications and histones isoforms, also called histone variants provide most of the complexity of chromatin and associate in specific combinations called chromatin states. Here we establish the broad conservation of chromatin states and their association with transcription in land plants. We design an index that predicts transcription factor binding and identify distinct groups of transcription factors based on their occupation of chromatin states. Our findings suggest that chromatin defines a specific environment that reflects transcriptional activity.

and reproduction in any medium, provided the original author and source are credited.

Data availability statement: Datasets from the analyses are provided as supplementary data files. Pipeline and scripts used for data analysis are available on the following GitHub link: https://github.com/Gregor-Mendel-Institute/vikas_states_tf_study_2025.git. NGS data are deposited in GEO under accession numbers: GSE302232.

Funding: This project was funded by the Austrian academy of sciences with the following grants from the Austrian Research Fund to FB: FWF P32054, P36231, PAT1104523 and PAT6138924, by the European Union's Framework Programme for Research and Innovation Horizon 2020 (2014–2020) to TH: the Marie Curie Skłodowska Grant Agreement no. 847548 (VIP2), by the Leverhulme Trust: Early Career Fellowship (ECF-2023-534 to FR) and the Isaac Newton Trust (23.08(f) to FR). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Eukaryotic genomes are organized into chromatin, a dynamic polymer of DNA and histone proteins that control access to the underlying genetic information. Chromatin structure, shaped by nucleosome positioning, histone modifications, histone variants, DNA methylation, and higher-order interactions, plays a fundamental role in gene regulation, genome stability, and chromosome organization. Advances in genome-wide profiling techniques have enabled the classification of chromatin states, defined by combinatorial patterns of histone marks and other epigenetic features, which reflect functionally distinct regulatory environments [1–3]. Chromatin states provide a powerful framework to annotate genome function, delineate regulatory elements, and uncover specialised domains of chromatin within constitutive heterochromatin occupied by transposable elements (TEs), facultative heterochromatin occupied by genes transcriptionally repressed and euchromatin occupied by expressed genes.

A major function of chromatin is to regulate transcription by controlling the accessibility of DNA to transcription factors (TFs). In turn, TFs bind to specific cis-regulatory elements and influence gene expression by recruiting transcriptional machinery and chromatin-modifying complexes. In metazoans, models of chromatin states have proven especially useful in distinguishing functional modes of TF activity [4–6]. Some TFs preferentially bind to open, transcriptionally active chromatin, while others engage nucleosomal DNA in repressive contexts, functioning as pioneer TFs that remodel chromatin to allow binding of additional TFs [6–9]. These distinct modes of chromatin engagement are often predictive of TF mechanisms, such as the recruitment of the Polycomb repressive complex (PRC) [10–12] or histone acetyltransferases (HATs) [13,14].

In plants, chromatin state analysis has primarily focused on *Arabidopsis thaliana*, where recent efforts have defined high-resolution maps covering developmental stages and environmental conditions [15,16]. This architecture shares analogous features with chromatin states in animals. These analyses constitute a valuable resource for functional genomics, but outside *Arabidopsis thaliana* reports of chromatin states are scarce [17,18] and the extent to which these combinatorial patterns of histone modifications are maintained across plant evolution remains an open question.

In contrast to animals, the interplay between TFs and chromatin has remained largely unexplored in plants. Early studies in *Arabidopsis thaliana* reported that most TFs bind open chromatin regions, with little variation in chromatin state preference across TF families [19], and the functional association between chromatin and TFs has remained unresolved. Notably, the best characterized pioneer TF so far is LEAFY (LFY), but other candidates among the MADS-box transcription factors have also been proposed as potential pioneer factors [20–22]. Although hundreds of TF binding profiles have been generated in *Arabidopsis*, the field lacks a unified framework to relate these profiles to chromatin context and transcriptional activity. Genome-wide TF binding assays such as ChIP-seq or DAP-seq often reveal that many binding events are non-productive, i.e., they do not lead to measurable changes in gene expression [23,24]. Unlike in animals, there is no unified chromatin–TF framework akin to ENCODE or modENCODE [25], complicating efforts to classify TFs based on binding context or regulatory output.

Here, by comparing *Arabidopsis thaliana* and *Marchantia polymorpha*, we showed the overall conservation of the role of histone H2A variants and histone post-translational modifications in defining conserved chromatin states. To characterize TF association with chromatin states, we defined a TF activity score aggregating both genome-wide TF binding data and co-regulation of gene expression. Using this strategy, we showed that TFs can be grouped into distinct categories based on chromatin state preferences. These are involved in distinct biological functions, including a subset of potential candidate pioneer factors. We propose that TF–chromatin relationships follow similar principles in both species, supporting the conservation of regulatory strategies across land plants.

Results

Chromatin states exhibit positional preference for gene regulatory regions of *Arabidopsis thaliana*

Chromatin states integrate the specific combinatorial occupancy of histone modifications and variants [2] at a near nucleosome resolution (bin size = 200 bp; see Methods). These states help delineate major functional domains of chromatin and provide insights into transcriptional regulation. We independently used ChromHMM [2] to recalculate the chromatin states of the *Arabidopsis thaliana* genome (TAIR 10) using the series of genomic profiles of histone variants and most abundant histone PTMs used by [15]. We fine-tuned the chromatin state annotation pipeline and used stricter quality control steps (see Methods), to obtain a refined and more robust definition of chromatin states [15,16]. Each group of states were defined by a specific combination of histone variants and histone PTMs (Fig 1A) and were associated with diverse levels of transcripts (S1A Fig), DNA methylation (S1B–S1D Fig), and were differentially associated with transposable elements, genes or intergenic regions (Fig 1B).

The new model contained 26 states including seven states of constitutive heterochromatin (H1 – H7) on transposable elements, four intergenic states (I1 – I4), five states of facultative heterochromatin (F1 – F5) on lowly expressed protein coding genes, and ten states in euchromatin on expressed genes (E1 – E10) (Fig 1A). The enrichment in specific classes of H2A variants marked the identity of each group of states, with H2A.W in constitutive heterochromatin, H2A.Z in facultative heterochromatin, and H2A (and H2A.X) on expressed genes. Only minor differences separated this set of chromatin states from those previously described [15]. The previous states H4 becomes annotated as H2, and the previous states H2 and H3 split in states H3 – H5, and the reannotation of the facultative chromatin states including the state F1 split into I1 and I2 (S1E Fig).

We investigated the positional preference of chromatin states along chromosomes and observed that the states of constitutive heterochromatin (H1–H7) were primarily enriched on transposable elements and showed a positional preference with respect to the centromere of chromosomes (S1F Fig). The state H1 showed an exclusive preference for centromeric chromatin while states H2 to H7 were successively enriched from pericentromeres (H2–H5) to chromosome arms (H6 and H7) (S1F Fig). This constitutive heterochromatin represents circa 20% of the genome (S1G Fig).

The other states were primarily enriched at genes, and we performed neighbourhood enrichment analysis with the Transcription Start Site (TSS) or the Transcription Termination Site (TTS) as anchors. (Fig 1C; see Methods). In euchromatin, expressed genes were occupied by a remarkable sequential enrichment pattern of chromatin states. The state I4 was most specifically enriched in the promoter and terminator regions of most protein-coding genes while the states E1 – E8 were organized from the promoter towards the gene body. The states E1, E2, and E3 showed enrichment in positions corresponding approximately to the 1st, 2nd, and 3rd nucleosomes, respectively, for most of the expressed protein-coding genes. The states E4 and E5 occupied gene bodies and E5 was also marked by high CG methylation and enrichment in H3.3 compared with state E4 (Figs 1A, 1C and S1B), states E6, E7, and E8 occupied the end of the gene body (Fig 1C). Plotting the enrichment of major histone H2A variants and chromatin modifications over genes occupied by states E1–E8 supported the predominance of H2A.Z associated with H3K4me3 in the first nucleosomes while the gene body is occupied by H2A and H2A.X with H3K36me3 and H2Bub (Fig 1D). A succession of facultative heterochromatin states F3 – F5 occupied gene bodies (Fig 1C) although their organization was less complex compared with euchromatin states. Contrasting

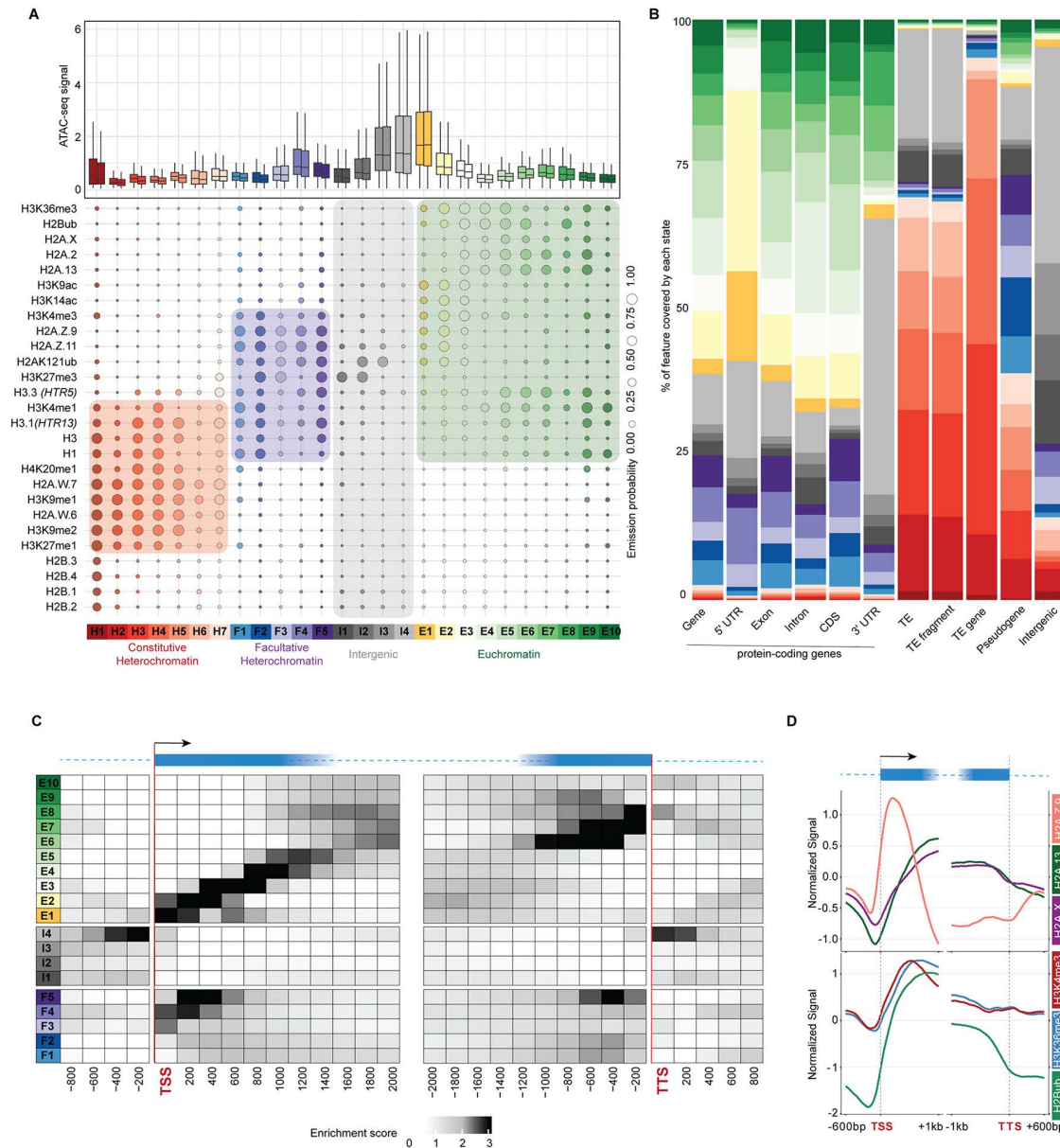


Fig 1. Chromatin states of *Arabidopsis thaliana*. (A) Bubble plot showing the emission probabilities for histone modifications/variants across the 26 chromatin states of *Arabidopsis thaliana*. The size of the bubble represents the emission probability ranging from 0 to 1. Coloured rectangles demarcate the classification of states into major domains of chromatin. The box plot on top shows the average ATAC-seq signal for each state representing chromatin accessibility. The two boxes per state are the two replicates of the ATAC-seq experiment. (B) Stacked bar plot showing the overlap between annotated genomic features and chromatin states. (C) Neighbourhood Enrichment analysis of chromatin states representing the fold enrichment for each state at fixed positions relative to the anchor position (TSS and TTS). (D) Metaplot showing the enrichment of prominent histone variants and modifications associated with active transcription. The enrichments here are plotted on genes overlapping states E1-E8. (n = 15234/ 27416 protein coding genes).

<https://doi.org/10.1371/journal.pgen.1012015.g001>

with the intergenic state I4, the other intergenic states I1, I2 and I3 were enriched in H3K27me3 and H2AK121ub, and showed lower gene expression (Figs 1A and S1A). These features suggest that the intergenic states I1, I2, and I3 are likely related to facultative heterochromatin, together occupying ca 30% of the genome (S1G Fig). Contrasting with the graded patterns of other chromatin states, the states E9 and E10 (ca. 5% of euchromatin) and the states F1 and F2 did

not show a strong positional enrichment relative to TSS or TTS (Fig 1C), but shared enrichment in linker histone H1 and were associated with genes expressed at very low levels (Figs 1A and S1A), suggesting a specific role of H1 in defining these chromatin states in addition to the major role of histone PTMs and H2A variants.

Conservation of chromatin states in land plants

To investigate the conservation of chromatin states in land plants, we used the same approach as above to define chromatin states in *Marchantia polymorpha*. This model liverwort shares a similar genome organization with model species of hornworts and mosses that, together, form the monophyletic group of bryophytes, which diverged from vascular plants about 450 mya [26]. To the set of publicly available histone-modification enrichment profiles [27,28], we added profiles of histone H2A variants detected with specific antibodies for H2A.X.1, H2A.X.2, H2A.Z, and H2A.M.2, which are expressed in vegetative tissues of *Marchantia* [29]. Enrichment of chromatin features was mapped onto the newest telomere-to-telomere chromosome-level genome assembly v7.1 [30]. Using ChromHMM we described fifteen chromatin states (Fig 2A; see Methods) which were annotated based on their specific enrichment in different genomic features (Fig 2B).

Intergenic states covered approximately a quarter of the genome (S2A Fig). Chromatin states associated with specific levels of gene expression (S2B Fig), chromatin accessibility (Fig 2A), and DNA methylation (S2C–S2E Fig). Based on the enrichment of known constitutive heterochromatin marks, including H3K9me1/2 and H3K27me1, we ascribed six distinct states to constitutive heterochromatin (MpH1–MpH6) (Fig 2A) that covered 39% of the genome (S2A Fig). These states were also distinguished by the lowest accessibility, enrichment in the H2A variant H2A.M.2 (Fig 2A), the highest degree of DNA methylation (S2C–S2E Fig) and were primarily present on transposable elements, with a minor representation on protein-coding genes (Fig 2B). The hallmark of heterochromatin states is their enrichment in H2A.M.2 and H3K9me2, but they are divided into three groups (Fig 2A). MpH5 and MpH6 do not carry additional marks, except for methylated cytosines in MpH5, while states MpH3 and MpH4 are enriched in H3K9me1 and H3K27me1. Unlike heterochromatin states in *Arabidopsis*, *Marchantia* heterochromatin states MpH1 and MpH2, which represent more than a third of constitutive heterochromatin are also enriched in H3K27me3 with less DNA methylation than MpH3 and MpH4.

Facultative heterochromatin states MpF1 and MpF2, both enriched in H2A.Z and H2A ubiquitination (H2Aub), were associated with the lowest range of gene expression and covered ca 16% of the genome (Figs 2A, 2B, S2A, and S2B) and the state MpF1 was also enriched in H3K27me3. As in *Arabidopsis*, facultative heterochromatin states were also enriched in H3K4me1 and H3K4me3 in *Marchantia*. Neighborhood enrichment analysis anchored on TSS/TTS from MpTak-1v7.1 [30] revealed that intergenic state Mpl1 was positioned upstream of MpF1. These two H3K27me3-enriched states together represented one third of facultative heterochromatin, exhibited low accessibility, and primarily marked repressed genes (Figs 2A–C, S2A, and S2B). States Mpl2 and MpF2 occupied the remaining two thirds of facultative heterochromatin in *Marchantia*. Unlike MpF1, these states lacked enrichment of H3K27me3, associated with more accessible chromatin, and marked less repressed genes (Figs 2A–C and S2B), consistent with previous findings [31]. Mpl2 occupied 19% of the genome and altogether facultative heterochromatin in *Marchantia* occupied 37% of the genome which was much higher than in *Arabidopsis*.

Regions with most open chromatin and the highest gene expression represented euchromatin (Figs 2A and S2B) and comprised the intergenic state Mpl3 and four genic states E1–E4 covering altogether circa 24% of *Marchantia* genome (Figs 2A, 2B and S2A). The state Mpl3 showed enrichment in the intergenic regions and the neighbourhood enrichment analyses anchored on TSS/TTS showed that the intergenic state Mpl3 corresponds to a region located in intergenic regions upstream of the TSS (Fig 2B and 2C). Contrary to *Arabidopsis*, the promoters of *Marchantia* defined by the region just upstream of the TSS showed enrichment of H2Aub and the elongation mark H3K36me3, along with other euchromatic marks. We considered whether this could result from imprecise TSS definitions in the *Marchantia* genome. The current annotations are based on Iso-seq data [27], a long-read sequencing approach that accurately determines TSS positions, particularly for highly expressed genes. We further validated these annotations by examining CAGE-seq profiles, which

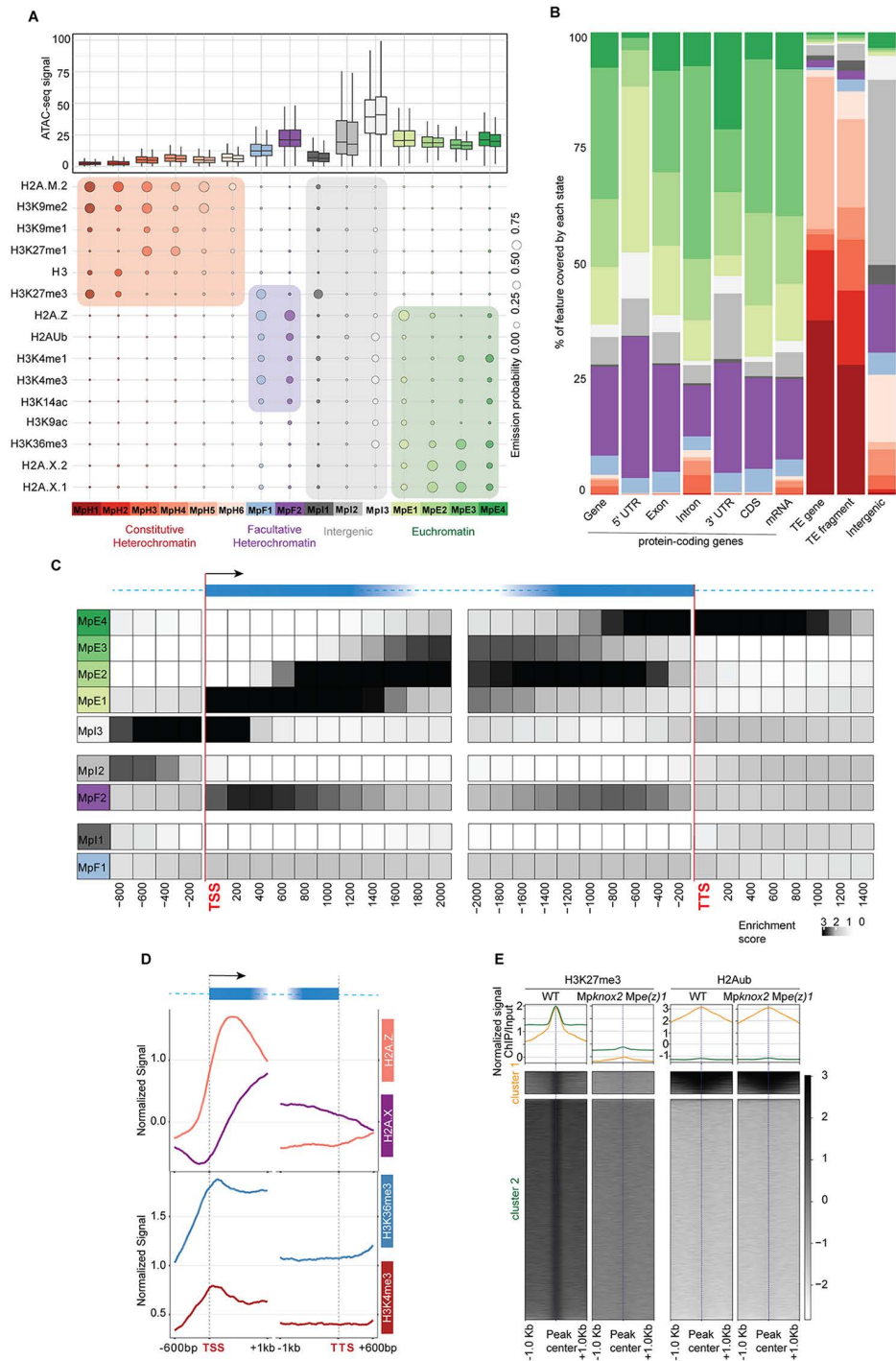


Fig 2. Chromatin states of *Marchantia polymorpha*. (A) Bubble plot showing the emission probabilities for histone modifications/variants across the 15 chromatin states of *Marchantia polymorpha*. The size of the bubble represents the emission probability ranging from 0 to 1. Coloured rectangles demarcate the classification of states into major domains of chromatin. The box plot at top shows the average ATAC-seq signal for each state representing chromatin accessibility. The two boxes per state are the two replicates of the ATAC-seq experiment. (B) Stacked bar plot showing the overlap between annotated genomic features and chromatin states. (C) Neighbourhood Enrichment analysis of chromatin states representing the fold enrichment for each state at fixed positions relative to the anchor position (TSS and TTS). (D) Metaplot showing the enrichment of prominent histone variants and modifications associated with active transcription. The enrichments here are plotted on genes overlapping states MpE1-MpE4. ($n = 10715/ 17944$ protein coding genes). (E) Enrichment heatmap and metaplots showing the apparent disassociation between H3K27me3 and H2Aub marks in wild

type and *Mpknox2 Mpe(z)1* double mutants. The enrichment of both marks is plotted on wild type H3K27me3 peaks (n=25159). The peaks have been clustered using k-means clustering into two clusters where the cluster 1 represents peaks of H3K27me3 shared with H2AKub while cluster 2 represents majority of H3K27me3 peaks that do not overlap with H2AKub.

<https://doi.org/10.1371/journal.pgen.1012015.g002>

revealed that most CAGE-seq peaks [27,32] aligned with the annotated TSS positions (S2F Fig). While this supports the general accuracy of current TSS annotations, refinements in TSS definitions may help reconcile the observed enrichment patterns. It should also be noted that our chromatin state analysis uses 200bp bins hence we cannot exclude the possibility that mapping of states on some genes may be offset by one bin in the upstream direction, an issue that future improvements in genome annotations may resolve. The most accessible, non-intergenic state is MpE1, which marked the +1 nucleosome of most protein-coding genes of *Marchantia* (Fig 2A and 2C). This state, with strong enrichment of H2A.Z and H3K36me3, is present in the first few nucleosomes of active genes while the states MpE2–MpE4 successively occupy the gene body and are enriched in H2A.X (Fig 2A, 2C, and 2D). To evaluate the relative role of histone H2A variants and histone PTMs in defining chromatin states, we calculated a new matrix of chromatin states in the absence of data on either H2A variants or histone PTMs and used the Jaccard index to compare both resulting matrices to the original 15 chromatin states (S2G Fig). A high Jaccard index reports high similarity. We observed that *in silico* removal of H2A variants perturbed the definition of constitutive heterochromatin states less strongly than removal of histone PTMs while the influence of H2A variants and histone PTMs was comparable on the chromatin states of facultative heterochromatin and euchromatin. Accordingly, the pattern of enrichment of H2A.Z and H2A.X over expressed genes were similar between *Marchantia* and *Arabidopsis* (Figs 1D and 2D).

To directly compare the chromatin states of *Arabidopsis* and *Marchantia*, we first calculated a 15-states model in *Arabidopsis* based on analogous set of histone modifications and H2A variants used to identify chromatin states of *Marchantia*. The 15 and the 26 states models were highly comparable with a notable reduction of the complexity of constitutive heterochromatin to only four states and of euchromatin to only five states (Fig 1A and S3A). To compare the 15 chromatin states of *Arabidopsis* and *Marchantia*, we performed hierarchical clustering of emission probabilities of the two models (S2H Fig). The resulting tree supported the overall similarities of the constitutive heterochromatin and euchromatin states. In contrast, while the typical H3K27me3 enriched in facultative heterochromatin states of the two species clustered, the *Marchantia* facultative heterochromatin MpF2 clustered with the *Arabidopsis* euchromatic states E2, which was likely due to the prominent enrichment in H2A.Z in their definition. We noted that chromatin accessibility of Mpl2 and MpF2 were similar to accessibility of euchromatin states than in *Arabidopsis* (Fig 2A and S3A) suggesting an ambivalent role for this subtype of facultative heterochromatin in *Marchantia*. To further compare the impact of PRC2 on histone modifications that define in a major manner the chromatin states, we analysed enrichment of H3K27me3, H2Aub, H3K4me3, H3K9me1, and H3K27me1 in the *Marchantia Mpknox2 Mpe(z)1* double mutant, which exhibited reduced PRC2 activity [28,33] (Fig 2E and S2I). As shown for PRC2 deficient mutants in *Arabidopsis* [34,35] peaks enriched in H3K27me3 in the wild type were depleted in the mutant deprived of PRC2 in *Marchantia*. However, the loss of PRC2 activity in *Marchantia* did not affect the deposition of H2Aub (Fig 2E) presumably by PRC1 activity [34]. Conversely, the loss of PRC1 affected the deposition of both H3K27me3 and H2Aub [34] suggesting a partial independence of the two Polycomb repressive pathways in *Marchantia*.

While our data supported an overall conservation of the chromatin states between the two species, we uncovered unique features of *Marchantia* chromatin organization. H3K27me3 was predominantly associated with transposable elements, with only a minor fraction marking protein-coding genes in facultative heterochromatin (Fig 2B). Instead, H2A.Z and H2AKub were the dominant marks of facultative heterochromatin, which is consistent with the mild effect of the loss of H3K27me3 on vegetative development [28] compared with *Arabidopsis* [22,36]. The promoter state MpE1 enriched in the transcription elongation mark H3K36me3, and the long persistence of E4 after the TTS creates an asymmetry between

the intergenic regions upstream and downstream of each gene (Fig 2C). This is in contrast with *Arabidopsis*, where the intergenic state I4 symmetrically occupies these two regions (Fig 2C). Notwithstanding potential inaccuracies in TSS/TTS coordinates, this asymmetry suggests that in *Marchantia*, gene orientation may define distinct chromatin environments in their vicinity, through mechanisms yet to be uncovered.

The distinct positional preferences of chromatin states in *Arabidopsis* versus *Marchantia*, particularly their sequential organization around gene regulatory regions, suggested a functional relationship with transcriptional regulation. To examine this in further detail, we ranked genes by their expression levels in *Arabidopsis* seedlings and *Marchantia* thalli. Highly expressed genes were depleted of facultative chromatin states in both species (Fig 3A and 3D).

In the case of *Arabidopsis*, I3 was also depleted in highly expressed genes and enriched in genes with low expression (Fig 3A and 3B). This observation was validated using experimentally derived TSS positions for *Arabidopsis* to examine the role of potential inaccuracies in TSS definitions (S4A–S4F Fig). The state E1, which was close to the TSS, was enriched in highly expressed genes in both species but more prominently in *Marchantia*, where it included the proximal promoter (Fig 3A and 3D).

Genes with low expression in vegetative tissues may be transcriptionally active at specific developmental stages or under environmental stimuli. To explore this, we assessed the coefficient of variation (CV) of gene expression across a panel of RNA-seq datasets spanning *Marchantia* and *Arabidopsis* development. The CV, defined as the standard deviation divided by the mean expression level, is independent of absolute expression and distinguishes ubiquitously expressed genes from those with variable expression in different conditions. Genes with high CVs exhibited greater coverage of facultative chromatin states, whereas ubiquitously expressed genes were largely depleted of these states (Fig 3B and 3E). In contrast, euchromatin states showed the opposite trend. A similar pattern was observed when analysing chromatin states within coding regions, with a much more prominent participation of euchromatin states (S4G–S4J Fig).

To determine whether facultative chromatin states are associated with specific tissues, we examined genes with tissue-specific expression in both species. Contrasting with ubiquitously expressed genes, promoters of these genes exhibited a strong association with facultative chromatin states (Fig 3C and 3F), supporting the hypothesis that facultative heterochromatin states are associated with cell differentiation and developmental transitions, with a prevalence of H3K27me3 enriched states. The loss of this marks resulted in strong ectopic expression of the genes covered primarily by both types of facultative heterochromatin (Fig 3G and 3H).

Besides differences related to the organisation of chromatin states on expressed genes and the relative extension of facultative heterochromatin marked by H3K27me3 versus H2Aub, our analyses conclude a broad conservation of chromatin states composition, and their association with transcriptional activity of genes and silencing of transposable elements in *Marchantia* and *Arabidopsis*.

Transcription factors associate with specific chromatin states of *Arabidopsis*

The specific patterns of enrichment of histone modifications around the TSS both in expressed and repressed genes raised the possibility that specific transcription factors may preferentially associate with specific chromatin states. To investigate this hypothesis, we examined the genome-wide association patterns between transcription factors and chromatin states. To investigate TF binding preferences, we compiled a comprehensive dataset of ChIP-seq, DAP-seq, and position weight matrix (PWM)-based datasets from *Arabidopsis* and assigned TF-bound regions to chromatin states. TF occupancy combining data from ChIP-seq and DAP-seq was evaluated using an approach similar to that applied in human TF studies [5]. ChIP-seq peaks showed a strong preference for intergenic states (especially I3 and I4) and E1 (associated with the TSS) and were depleted in constitutive heterochromatin and euchromatin states present on bodies of expressed genes but interestingly some TF enrichment was found in states of facultative heterochromatin (Fig 4A), to a large extent consistent with the accessibility constraints (Fig 1A).

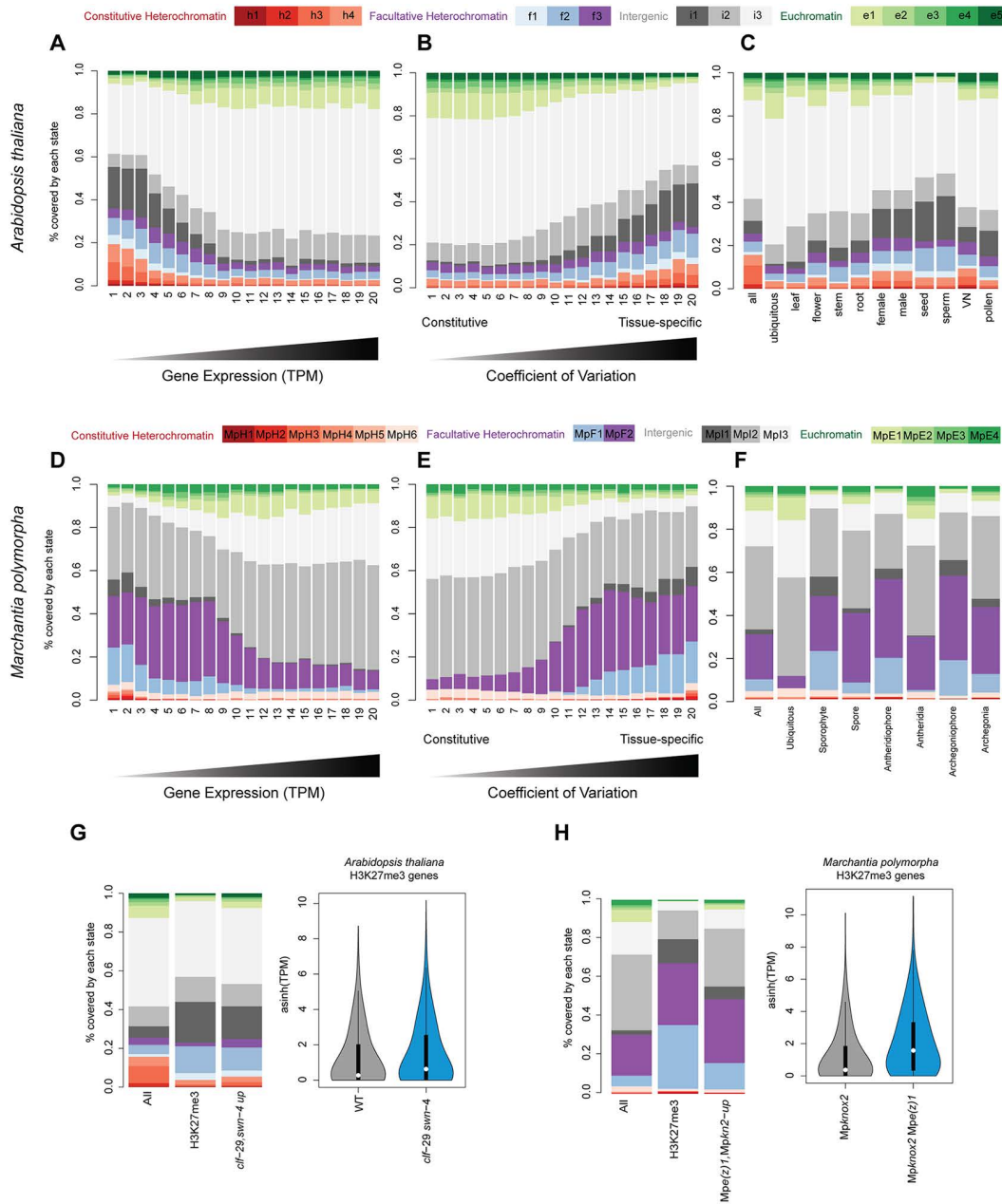


Fig 3. Association between chromatin states and transcription across promoters in *Arabidopsis* and *Marchantia*. (A, D) Bar plot of the proportion of chromatin states for 20 bins across promoters of genes ordered by transcripts per million (TPM) *Arabidopsis* leaves (A) and *Marchantia* thallus (D). (B, E) Bar plot of the proportion of chromatin states across promoters for 20 bins of genes ordered by coefficient of variation (CV) across different tissues of *Arabidopsis* (B) or *Marchantia* (E). (C, F) Bar plot the proportion of chromatin states across promoters for genes all genes or differentially expressed in specific tissues in *Arabidopsis* (C) and *Marchantia* (F). (G, H) Bar plot of the proportion of chromatin states across promoters for genes covered by H3K27me3 or differentially expressed in PRC2 mutant in *Arabidopsis* (G) and *Marchantia* (H). *clf/swr* double mutant for *Arabidopsis* and *Mpknox2 Mpe(z)1* double mutant for *Marchantia*, all genes are shown as a reference in both cases. In the right, $\text{asinh}(\text{TPM})$ of expression values for the same group of genes are shown for both mutant and the WT. The color code for the diagram represents the chromatin states in the 15 states models.

<https://doi.org/10.1371/journal.pgen.1012015.g003>

Prediction of TF binding sites using DAP-seq, which does not depend on chromatin accessibility [37], exhibited a broader distribution across chromatin states with some preference for intergenic states (Fig 4B). In contrast, PWM-based TF binding predictions, which only rely on DNA binding motifs, displayed a nearly uniform distribution (Fig 4C).

The observed differences between experimental approaches highlight the need for integrative analyses to fully resolve TF-chromatin state relationships in plants [5]. TF binding does not always lead to transcriptional regulation, as ChIP-seq or DAP-seq frequently detects peaks at sites that are not relevant for transcriptional activation or repression (Alvarez et al., 2021). To extend our analysis, we developed a method to estimate TF activity, incorporating co-expression data as a proxy for functional interactions [38]. For this, we classified the TF DNA-binding data between chromatin states and call a list of the proximal targets for each TF and chromatin states. We then extracted a list of all positively and negatively co-expressed genes for the TF that was used for that experiment. Finally, we calculated the enrichment between the two lists, for each state, against a random set of gene of equal size. This allowed us to distinguish between passive and active binding and reduce technical bias, particularly from *in vitro* binding. Ultimately, we hypothesize this score with highlight interactions between the chromatin and the TF that are functionally relevant. We adapted a method to calculate chromatin state preference scores by evaluating the enrichment of TF binding to their putative targets, positively and negatively co-expressed genes (see Methods). Overall, low scores of TF occupancy and TF activity were broadly correlated with in heterochromatin states, and the highest scores were observed in chromatin accessible regions such as the intergenic states (Fig 4D and 4E), which is the main factor that differentiated ChIP-seq from DAP-seq and PWM (Fig 4A–4E).

With this method to estimate TF activity, the scores of TF occupancy and activity converged between the different methods. To look at different patterns of chromatin preferences among TFs, we kept only the data from ChIP-seq and DAP-seq experiments and eliminated experiment with low scores across all chromatin states (score < 8). In total, we collected binding profiles corresponding to ~300 TFs in *Arabidopsis*. Using hierarchical clustering, we identified distinct associations between TFs and specific chromatin states (Fig 4F). Interestingly, the different TF cluster also display distinct binding preferences relative to the TSS (Fig 4G). This observation was validated using experimentally derived TSS positions for *Arabidopsis* to examine the role of potential inaccuracies in TSS definitions (S5A–S5C Fig).

Unlike *Arabidopsis*, *Marchantia* lacks a comprehensive dataset of ChIP-seq or DAP-seq experiments and unfortunately our efforts to map the linker histone H1 and the variants of histone H3 and H2B have not been successful. However, previous studies have shown that TF DNA binding preferences are highly conserved across land plants [39], allowing to derive PWMs in *Marchantia* from PWMs from orthologous *Arabidopsis* proteins. Here, we combined recently published 24 DAP-seq experiments for *Marchantia* TFs with PWM from protein-binding microarrays [40] to serve as an alternative approach to infer TF binding. As with *Arabidopsis*, PWM-derived occupancy was evenly distributed across chromatin states (S2J Fig). However, when TF activity was estimated using co-expression, a striking shift emerged (S2K Fig). The most accessible chromatin states Mp12 and MpF2 were the most enriched, followed by MpE2, mirroring the patterns observed in *Arabidopsis*. A clustering analysis revealed further preferential association of clusters of TFs with states enriched either before the TSS (Mp11 and Mp12) or after the TSS (MpF2 and MpE2) (S2L Fig). However, the scarcity of the data and the fact that we compared TF binding in diploid tissues of *Arabidopsis* with haploid tissues in *Marchantia* limited our analysis and further investigation is required to determine whether similar regulatory strategies are employed in the interplay between TFs and chromatin states to establish the degree of conservation of these interactions across land plants.

Remarkably in *Arabidopsis*, many TF families showed preferential association with one or two chromatin states (Fig 4H) that may be associated with different mechanisms of action. The largest group (Cluster 3, 32% of all TFs) was associated primarily with I4 and bound upstream of the TSS (Fig 4F and 4G). This cluster comprised TFs from the WRKY, bHLH, bZIP, and HB families that bound to open chromatin of expressed genes (state I4) to regulate gene expression (Fig 4H and S3 Table), likely following the classic mode of RNA Polymerase II recruitment and stabilization of the pre-initiation complex [41]. A GO term analysis indicated that TFs of cluster 3 primarily participated in biosynthetic processes, photosynthesis, circadian rhythm, and other housekeeping functions. Other TFs associated with euchromatic chromatin

states formed clusters 4 and 5. Notably, TF families present only in cluster 5 were involved in hormone signalling and flower development, including ARF [42]. Most TFs from cluster 4 clearly bound after the TSS and several of these were associated with the control of the cell cycle, including E2FA, which is a typical marker of cell division. Clusters 4 and 5 also included the recently described GATC binding TFs [43], a large fraction of the AP2/EREBP family involved in stress response [44,45] and a large fraction of C2C2-GATA and G2-like TFs.

On the other hand, the TFs of Cluster 1 and 2 preferentially bound sites associated with facultative chromatin states (Figs 1A and 4F). Cluster 2 contained TFs primarily controlling development, and more particularly root and vascular development. Hence these TFs presumably control genes, which are not strongly active in leafy tissues of seedlings, in agreement with cluster 2 TFs occupying facultative heterochromatin states I1-I3 (Fig 4F). Many of these TFs binding sites were located upstream of the TSS (Fig 4G) and are thus expected to bind to typical promoters. In contrast, cluster 1 TFs primarily associated with the position of the first nucleosome downstream of the TSS (Fig 4G). Cluster 1 TFs are not expressed at the developmental stage analyzed (S5D Fig) and their loci are occupied by facultative heterochromatin. As a result, predicting the first nucleosome position on their target genes relies entirely on TSS annotation rather than chromatin state information, making these predictions less certain compared to the expressed TFs in clusters 3–5. Most of these TFs controlled meristem and flower development and were significantly enriched in MADS-domain TFs, including APETALA1, PISTILLATA, APETALA3, SEPALATA3, and the pioneer factor LEAFY (LFY), which regulates flower development. Together with their target genes, these TFs are not expressed in seedlings where the chromatin states were described [22,46] (Fig 4F and S5 Table), which aligns with their binding to facultative heterochromatin. Overall, our results demonstrate that chromatin states can serve as a framework to classify TF-chromatin interactions. This analysis revealed distinct regulatory strategies among plant TFs involving proximal promoters in intergenic regions, the +1 nucleosomes, or cis elements downstream of the TSS most likely acting as enhancers or silencers.

Discussion

Here, we identified chromatin states in vegetative tissues of the liverwort *Marchantia* and found that their composition and association with transcription are strikingly conserved with the chromatin states of *Arabidopsis*. Overall, orthologous H2A variants and histone PTMs in *Marchantia* show the same associations and are equally important in defining chromatin states as has been shown in *Arabidopsis* [15]. The variants H2A and H2A.X and the PTM H3K36me3 mark euchromatin; H3K27me3 and H2A.Z mark facultative heterochromatin; in constitutive heterochromatin, H3K9me1/2 associates with the variants H2A.W or H2A.M, which share C-terminal motifs enriched in lysine residues [20]. Over protein-coding genes, chromatin states assemble in staggered arrays from the TSS to the TTS. Based on the integration of several datasets to map TF binding, we identified the preference of certain families and functional groups of TFs for specific chromatin states present either in the 5'UTR, the +1 nucleosome, or downstream of the TSS, suggesting a high complexity of cis elements controlling gene expression and the involvement of chromatin states in TF recruitment, either through direct recognition of chromatin marks or through co-regulatory mechanisms.

The evolutionary distance between the two species and their distinct genome organization makes *Marchantia* an ideal model for exploring the conservation of chromatin states and highlighting the evolution of fundamental chromatin regulatory mechanisms. Overall, the organization of chromatin states in *Marchantia* confirms our previous evaluation of profiles of chromatin modifications and their conservation amongst bryophytes [28]. Similarly, chromatin states described in *Arabidopsis* are representative of flowering plants [47,48]. We observed species-specific heterochromatin organization, including the dispersed constitutive heterochromatin, the less complex and less abundant facultative heterochromatin in *Marchantia* marked by H3K27me3, a predominant share of regions with this mark also associated with TE in bryophytes [49]. In *Marchantia*, euchromatin organization differs from *Arabidopsis*. The marks H3K4me1 and H3K36me3 reflecting transcriptional elongation and confined to the gene bodies in *Arabidopsis*, extend beyond the TTS in *Marchantia*, suggesting that signals for transcriptional termination differ between flowering plants and bryophytes. However, we cannot

exclude that this pattern may result from a combination of technical limitations including missing transcripts in the current MpTak1v7.1 genome annotation (e.g., lncRNAs) or potential annotation errors (e.g., incorrect gene ends or read-through transcripts). Overall, together with the similarities in the organization of histone modifications among bryophytes [28] and flowering plants [15,48] our data support the conservation of chromatin states and their relationship with gene expression in land plants. It is remarkable that this conservation is seen when comparing haploid gametophytic tissues of bryophytes and diploid sporophytic tissues of angiosperms, suggesting that chromatin organization in extant land plants reflects the organization of chromatin states in the haploid ancestors of land plants. The broad conservation of active chromatin states is also supported by a recent survey of chromatin states across eukaryotes while plants have evolved specific forms of repressive chromatin [50]. Specifically, we show a certain degree of independence between chromatin states governed by PRC1 and PRC2 in *Marchantia*, which contrasts with animals that connect the activities of the two modifiers through hierarchical recruitment models of “writing” and “reading” histone modifications H2AK121ub and H3K27me3 [51,52]. These differences support conclusions from phylogenetic analyses, which place the origins of the two independent types of Polycomb activities in the LECA, followed by divergence of their associations during evolution of different eukaryotic groups [53]. Our results and the differences in the phenotypes of the mutants in PRC1 and PRC2 in *Arabidopsis* [54] further support the conclusion that PRC1 and PRC2 have remained largely independent from each other in plants.

In animals, distinct forms of association with chromatin distinguish functional types of TFs, including those associated with facultative heterochromatin [4,6]. Comparable systematic studies were missing in plants. Here we integrate stable TF-chromatin interactions identified with ChIP-seq with transient, context-dependent interactions and show that plant TFs associate with preferred chromatin states. Many TFs appear to follow a “canonical” mode of action, binding to open chromatin and regulating transcription by facilitating the assembly or stabilization of RNA polymerase II and the initiation of transcription (S6A Fig). In *Arabidopsis*, this mechanism is predominant among TFs that regulate constitutive genes or modulate expression quantitatively. However, TF binding in *Arabidopsis* is not always associated with measurable transcriptional changes [24]. Notably, open chromatin is potentially occupied by multiple TFs, suggesting extensive competition for binding. This could be due to open chromatin, which is associated with highly expressed genes and permissive for TF binding. Such regions crowded in TF binding are also referred to as highly occupied target (HOT) regions, which have been identified before both in plants (19) and animals [55]. These binding events are yet functional and likely important for gene regulation (19) but they can also be associated with redundant or passive activity [19].

Genes that require tight spatiotemporal regulation—remaining off in most tissues and activating only under specific developmental or environmental conditions, are more likely controlled by TFs capable of acting within facultative chromatin states. These TFs may interact with chromatin remodelers to induce local chromatin accessibility or directly access DNA protected by nucleosomes, resembling the pioneer TF paradigm described in animals [56,57]. LFY functions as a pioneer TF by binding nucleosome-occupied sites at its targets APETALA1 and AGAMOUS, altering chromatin accessibility via SWI/SNF chromatin remodelers [20,58]. MADS-domain TFs, bZIP TFs, LEAFY COTYLEDON1 (LEC1), TERMINAL FLOWER 1, FLOWERING LOCUS T, and MONOPTEROUS, have been proposed as potential candidate pioneer factors [59]. Further experiments including careful analysis of TSS in relevant cell types and chromatin profiling will be required to investigate the potential mechanism of action of TFs of clusters 1 and 2 to qualify as pioneer factors. Thus, TF preferences for specific chromatin states provide a sequence-independent axis for classifying plant TFs and inferring their regulatory modes. Generalizing this observation, our work captures the chromatin preference signature of pioneer TFs and expands the list of potential candidate pioneer TFs in plants.

Materials and methods

Reagents and tools

Reagents and Tools are listed in [Table 1](#).

Table 1. Reagents and tools.

Reagent/Resource	Reference or Source	Identifier or Catalog Number
Experimental Models		
<i>Marchantia polymorpha</i> Tak-1	[60]	NA
<i>Arabidopsis thaliana</i> Col-0	Nottingham Arabidopsis Stock Centre	NA
Antibodies		
Rabbit polyclonal Anti-H2A.X. antibody	This paper	NA
Rabbit polyclonal Anti-H2A.X.2 antibody	This paper	NA
Rabbit polyclonal Anti-H2A.Z antibody	This paper	NA
Rabbit polyclonal Anti-H2A.M.2 antibody	This paper	NA
Chemicals, Enzymes and other reagents		
cOmplete Protease Inhibitor Cocktail	Roche	Cat#11697498001
Protein A/G beads	Thermo Fisher Scientific	NA
Proteinase K	Thermo Fisher Scientific	NA
RNase A	Thermo Fisher Scientific	NA
MinElute PCR purification kit	Qiagen	NA
Ovation Ultralow Library Systems V2	Tecan, Männedorf, Switzerland	NA
Software		
BEDTools suite		v2.27.1
FastQC		v0.11.5
Trim Galore		v0.6.5
Bowtie2		v2.4.1
SAMtools		v1.9
Picard		v2.22.8
deepTools		v3.1.2
ChromHMM		v1.23
valr (R package)		v0.6.6
Tidyverse (R package)		NA
GenomicRanges (R package)		v1.56
circize (R package)		NA
ComplexHeatmap (R package)		NA
ChIPseeker (R package)		NA
pheatmap (R package)		NA
MEME Suite (FIMO)		NA
PANTHER		NA
Other		
Covaris E220 High-Performance Focused Ultrasonicator	Covaris	E220
Covaris milliTUBE	Covaris	NA
Illumina HiSeq v4 sequencer	Illumina	NA

<https://doi.org/10.1371/journal.pgen.1012015.t001>

Methods and protocols

Profiling of *Marchantia polymorpha* H2A variants using ChIP-seq. ChIP experiments were performed using a previously described protocol with some modifications [29]. Two-week-old thalli started from gemmae of *Marchantia polymorpha* Tak-1 wild type were collected and cross-linked using 1% formaldehyde in 1 × PBS under vacuum on ice

for 10 min. The cross-linking reaction was quenched by adding 2 M glycine to achieve a final concentration of 0.125 M under vacuum on ice for 10 min. Excess solution was removed from cross-linked tissue by blotting with paper towels. Cross-linked tissue was then snap frozen in liquid nitrogen and ground to a fine powder using mortar and pestle. The powder was transferred into a 50-ml plastic tube and suspended in 40 ml of MP1 buffer (10 mM MES-KOH pH 5.3, 10 mM NaCl, 10 mM KCl, 0.4 M sucrose, 2% (w/v) Polyvinyl pyrrolidone (PVP 10), 10 mM MgCl₂, 10 mM 2-mercaptoethanol, 6 mM EGTA, 1 × cOmplete protease inhibitor cocktail). Suspended samples were then filtered twice through one layer of Miracloth, once through a 40 μm nylon mesh, and twice through a 10 μm nylon mesh. Filtered samples were centrifuged at 3000 × g at 4°C for 10 min, and the supernatant was discarded. The pellet was washed using 15 ml of MP2 buffer (10 mM MES-KOH buffer pH 5.3, 10 mM NaCl, 10 mM KCl, 0.25 M sucrose, 10 mM MgCl₂, 10 mM 2-mercaptoethanol, 0.2% Triton-X 100, 1 × cOmplete protease inhibitor cocktail) three times. The final pellet was then resuspended in 5 ml of MP3 buffer (10 mM MES-KOH pH 5.3, 10 mM NaCl, 10 mM KCl, 1.7 M sucrose, 2 mM MgCl₂, 10 mM 2-mercaptoethanol, 1 × cOmplete protease inhibitor cocktail) and centrifuged at 16 000 × g at 4°C for 1 h. After centrifugation, the supernatant was discarded, and the pellet was resuspended in 900 μl of covaris buffer (0.1% SDS, 1 mM EDTA, 10 mM Tris-HCl pH 8.0, 1 × cOmplete protease inhibitor cocktail). The resuspended pellet containing the chromatin fraction was fragmented using a Covaris E220 High-Performance Focused Ultrasonicator for 15 min at 4°C (duty factor, 5.0; peak incident power, 140.0; cycles per burst, 200) in a 1-ml Covaris milliTUBE. Sheared chromatin was centrifuged at 20 000 × g at 4°C for 10 min, and the supernatant was transferred into a new 5-ml tube and diluted by adding 2.7 ml of ChIP dilution buffer (0.01% SDS, 1.1% Triton-X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl pH 8.0, 167 mM NaCl). Diluted chromatin was cleared by incubating with protein A/G beads (Thermo Fisher Scientific, Waltham, MA, USA) at 20 rpm rotating at 4°C for 1 h. Beads were removed by magnetic racks and precleared chromatin was separated into five tubes and incubated with 1 μg of specific antibodies for histone H2A variants at 20 rpm spinning at 4°C overnight. Chromatin bound by antibodies was collected by incubating with protein A/G beads for 3 h. The beads were collected by magnetic racks and washed twice with a low salt wash buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100 and 0.1% SDS), once with a high salt wash buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 2 mM EDTA, 1% Triton X-100 and 0.1% SDS), once with a LiCl wash buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.25 M LiCl, 1% IGEPAL CA-630 and 0.1% sodium deoxycholate), and twice with a TE buffer (10 mM Tris-HCl pH 8.0 and 1 mM EDTA). Immunoprecipitated DNA was eluted using 500 μl elution buffer (1% SDS and 0.1 M NaHCO₃) at 65°C for 15 min. To reverse cross-link, eluted DNA was mixed with 51 μl of reverse cross-link buffer (40 mM Tris-HCl pH 8.0, 0.2 M NaCl, 10 mM EDTA, 0.04 mg ml⁻¹ proteinase K; Thermo Fisher Scientific) and incubated at 45°C for 3 h and then at 65°C for 16 h. After cross-link reversal, DNA was treated with 10 μg of RNase A (Thermo Fisher Scientific), incubated at room temperature for 30 min and purified using the MinElute PCR purification kit (Qiagen). ChIP-seq library was generated from ChIPed DNAs using Ovation Ultralow Library Systems V2 (Tecan, Männedorf, Switzerland). The ChIP-seq libraries were sequenced on illumina Hiseq v4 to generate 50 bp single end reads.

ChIP-seq data collection and processing. The publicly available ChIP-seq datasets for *Arabidopsis* were downloaded from our previous study [15]. A detailed description of the sources of raw files can be found in the (S3 Table). The description of publicly available ChIP-seq datasets and the datasets generated for this study to calculate the chromatin states of *Marchantia* is also available in the (S4 Table).

Raw BAM files were converted to FASTQ files using the “bamtofastq” sub-command of the “BEDTools suite” v2.27.1 [61]. Sequencing quality of the raw files was evaluated using quality reports generated by FastQC v0.11.5 [62]. Reads were trimmed using “Trim Galore” v0.6.5 (DOI: 10.5281/zenodo.5127898). The trimmed reads were aligned to the reference genome: TAIR10 in case of *Arabidopsis* and MpTak-1 v7.1 in case of *Marchantia* using Bowtie2 v2.4.1 [63] and further processed using SAMtools v1.9 [64] and BEDTools v2.27.1 [61]. Duplicates were removed using Picard v2.22.8 (<https://broadinstitute.github.io/picard/>) to generate the aligned BAM files. The correlation between ChIP samples was evaluated using the multiBamSummary and plotCorrelation functions of deepTools v3.1.2 [65]. The bamCompare function

of deepTools v3.1.2 was used to normalize the ChIP signal with Input/H3 and generate BigWig files with normalized ChIP signal. The code for the above-mentioned analysis can be found here: https://github.com/Gregor-Mendel-Institute/vikas_states_tf_study_2025.git.

Chromatin state calculations. The aligned BAM files generated as explained above were used to calculate the chromatin states using the BinarizeBAM and LearnModel commands of ChromHMM v1.23 [2] with default parameters and models ranging from 2–50 states were learned. BAM files were binarized in 200bp bins for model learning to characterize chromatin states using the multivariate Hidden Markov Model (HMM) to identify the combinatorial spatial patterns in the ChIP-seq signal of multiple chromatin marks. The models with 2–50 chromatin states were generated in each case. A 26-state model was chosen for the full *Arabidopsis* datasets based on two reasons: (1) As the raw data used to calculate states was the same as the one used in [15] with the main difference being the employment of stricter alignment parameters and filtering of previously included low confidence reads. These changes did not lead to extensive differences from the 26-state model in [15] (S1E Fig). A follow up on all models between 23–29 states revealed that models with less than 26 states showed loss of complexity (merging of states associated with distinct genomic features) and states with more than 26 states showed appearance of states that can no longer be associated with distinct genomic features. Hence, we decided to pick the 26-state model to be optimal for our analysis (S1 Data).

The chromatin states of *Marchantia* were calculated with the same pipeline. Given the comparatively smaller set of chromatin marks available in *Marchantia*, a smaller number of chromatin states was to be expected. A 15-state model was chosen for *Marchantia* (S2 Data) using the similar strategy as in [15]. To compare the chromatin state of *Arabidopsis* and *Marchantia*, we used only the chromatin marks that were available in both species and generated a less complex, comparable 15-state model for *Arabidopsis* (S3 Data).

Annotation of chromatin states. The BED files of the chromatin states generated by ChromHMM were imported into R v4.3.2 (<https://www.R-project.org/>). The chromatin state regions were overlapped with the genomic features defined in *Arabidopsis* genome TAIR10 (GFF3 file downloaded from [here](#)) and *Marchantia* genome MpTak-1 v7.1 (GFF3 file downloaded from [here](#)) using the “bed_intersect” function of R package valr v0.6.6 [66]. The results were plotted as stacked bar plots using the Tidyverse [67] package of R (Figs 1B, 2B, and S3B).

The NeighbourhoodEnrichment command of ChromHMM v1.23 [2] was used to calculate the enrichment of each state relative to the Transcription Start Site (TSS) or Transcription Termination Site (TTS) as anchors that were extracted from aforementioned GFF files. The resultant matrix was imported into R v4.3.2 (<https://www.R-project.org/>) and the heatmap was plotted using the Tidyverse [67] package of R (Figs 1C, 2C, and S3H).

To evaluate the genomic localization of states, especially the heterochromatic states (H1:H7) in *Arabidopsis*, histograms with binwidth of 50,000 bp were plotted on chromosome 1 using the Tidyverse [67] package of R showing the preferential localization of heterochromatic states near the centromere region (S1F Fig).

We identified the states of constitutive heterochromatin by their strong enrichment in the regions containing Transposable Elements (TE), TE genes and TE fragments (Figs 1B, 2B, and S3B) along with high emission probabilities of known heterochromatic marks including H2A.W, H3K9me1/2 among others (Figs 1A, 2A, and S3A). The distinction between the states of euchromatin and facultative heterochromatin was made by the high emission probabilities of marks associated with each type of chromatin state. The states of intergenic region showed little to no enrichment of any specific mark, representing Nucleosome Free Regions (NFR) and constituted about 20% of the genomes in both *Arabidopsis* and *Marchantia*. (S1G, S2A, and S3C Figs). We decided to annotate the states with the following nomenclature strategy:

- (1) Euchromatin: The states of euchromatin were named with letter “E” in *Arabidopsis* 26-state model, with letter “e” in *Arabidopsis* 15-state model, and with letters “MpE” followed by a number in *Marchantia* 15-state model. The order of the numbers was assigned based on their enrichment relative to the TSS going from TSS to gene body (Figs 1C, 2C, and S3H).

- (2) Constitutive Heterochromatin: The states of constitutive heterochromatin were named with letter “H” in *Arabidopsis* 26-state model, with letter “h” in *Arabidopsis* 15-state model, and with letters “MpH” in *Marchantia* 15-state model. The order of the numbers was assigned based on their localization on chromosomes going from centromere to pericentromere in case of *Arabidopsis* states ([S1F Fig](#)).
- (3) Facultative Heterochromatin: The states of facultative heterochromatin were named with letter “F” in *Arabidopsis* 26-state model, with letter “f” in *Arabidopsis* 15-state model, and with letters “MpF” in *Marchantia* 15-state model.
- (4) Intergenic states: The states of intergenic regions were named with letter “I” in *Arabidopsis* 26-state model, with letter “i” in *Arabidopsis* 15-state model, and with letters “Mpl” in *Marchantia* 15-state model.

Analysis of chromatin states. The emission matrices generated by ChromHMM were imported into R v4.3.2 (<https://www.R-project.org/>). The state names and colours were reassigned using the strategy aforementioned, followed by plotting the emission matrices as bubble plots using the Tidyverse [\[67\]](#) package of R ([Figs 1A, 2A, and S3A](#)).

The percentage of genome covered by each state was calculated as follows: (Total number of base pairs covered by each state/ Genome size) * 100. The results were plotted as a pie chart using the Tidyverse [\[67\]](#) package of R ([S1G, S2A, and S3C Figs](#)).

We compared the chromatin states published in [\[15\]](#) with our 26-state model by importing the BED files for both models into R followed by overlapping the two matrices to create the transition matrix containing the occurrences of state transition between the two models. The data was plotted as an alluvial plot using the “circlize” [\[68\]](#) package of R. Similar analysis also led to the comparison of the *Arabidopsis* 26 and 15-state models ([S1E and S3I Figs](#)).

To show the similarities between the states of *Arabidopsis* 15-state model and the states from extensive 26-state model, after removing the additional marks from the 26-state model that were missing in the 15-state model.

To show the similar association of chromatin marks between the states of *Arabidopsis* and *Marchantia* 15-states models, hierarchical clustering of the two emission matrices was performed. H2A.W of *Arabidopsis* and H2A.M.2 of *Marchantia* were used interchangeably for comparison purposes. The two matrices were imported into R and merged for the common marks. The heatmap and clustering was performed using ComplexHeatmap [\[69\]](#) package of R ([S2H Fig](#)).

Validation of Neighbourhood Enrichment analysis using the experimentally determined TSS and TTS of *Arabidopsis*. We used TAIR10 rather than Araport11 as the reference genome because several studies [\[70,71\]](#) have demonstrated that TAIR10 gene borders are better defined for protein-coding genes, which are the focus of our analysis. To investigate the influence of potential inaccuracies in TAIR10-derived TSS/TTS positions, we incorporated experimentally determined gene borders obtained from full-length RNA-seq and high-throughput 5’ and 3’ tag sequencing data [\[71\]](#). We mapped the experimentally determined TSS/TTS peaks from [\[71\]](#) to their nearest TAIR10-annotated protein-coding genes, successfully updating the TSS/TTS positions for 19,102 genes. The distribution of the updated TSS/TTS as compared to TAIR10 based TSS/TTS positions was plotted as density plot using the Tidyverse (33) package of R ([S7A Fig](#)).

To assess the impact of potential inaccuracies in the TAIR10 derived TSS/TTS positions on our neighbourhood enrichment analysis, we used experimentally validated TSS/TTS positions derived from [\[71\]](#) to evaluate the neighbourhood enrichment using three different TSS/TTS definition sets: (1) 19,102 genes with experimentally updated TSS/TTS positions based on [\[71\]](#); (2) all 27,206 protein-coding genes, comprising the 19,102 genes with experimentally updated TSS/TTS and the remaining 8,104 genes with TAIR10-based TSS/TTS; and (3) all 27,206 protein-coding genes using only TAIR10-based TSS/TTS positions ([S7B–S7D Figs](#)). The Jaccard Index of Similarities between the three sets of matrices was calculated and plotted using the Tidyverse (33) package of R ([S7E Fig](#)).

RNA-seq, ATAC-seq, and Bisulfite-seq data analysis. The Bisulfite-seq and ATAC-seq data for wild type *Arabidopsis* seedlings was downloaded from GEO accession: [GSE146948](#) [\[72\]](#) while the RNA-seq data was downloaded from [\[73\]](#). All data was aligned to TAIR10 using bowtie2 v2.4.1 [\[63\]](#) and further processed using SAMtools v1.9 [\[64\]](#) and BEDTools v2.27.1 [\[61\]](#). BigWig files for ATAC-seq and Bisulfite-seq and counts file for RNA-seq was imported into R. The data was

overlapped with states regions file using the `valr` v0.6.6 [66]. The results were plotted as box plots using the Tidyverse [67] package of R (Figs 1A, S1A–S1D, and S3D–S3G).

The Bisulfite-seq, ATAC-seq, and RNA-seq data for *Marchantia* was downloaded from [27] accession numbers SRA: PRJNA553138 and PRJDB8530. All data was aligned to MpTak-1 v7.1 genome and the boxplots were generated with the pipeline mentioned above (Figs 2A and S2B–S2E).

Association of chromatin states with expression. To assess the relationship between chromatin states and gene expression, we extracted promoter sequences from *Arabidopsis thaliana* (TAIR10) and *Marchantia polymorpha* (Tak1 v7.1), defining promoters as the 1,000 bp region upstream of the transcription start site (TSS) using the promoter function from the GenomicRanges package (v1.56) in R. For each gene in *Marchantia* genome, we determined the overlap of these promoter regions with annotated chromatin states using the `findOverlaps` and `pintersect` functions.

For gene expression data, we compiled a diverse set of publicly available RNA-seq experiments from MarpolBase expression [74] for *Marchantia* and EvoRepro for *Arabidopsis* [75]. To examine chromatin state association with transcriptional output, we ranked genes by their expression levels—measured in transcripts per million (TPM)—in the primary vegetative tissues where chromatin state data were derived (*Arabidopsis* leaf and *Marchantia* thallus). Genes were grouped into 20 equal-sized bins based on TPM values, and for each bin, we calculated the average proportion of each chromatin state per bin and visualized the distribution.

Similarly, to assess chromatin state dynamics across broader expression datasets, we analysed the same set of RNA-seq samples and estimated transcriptional variability by calculating the coefficient of variation (CV) for each gene, defined as the ratio of the standard deviation to the mean TPM across samples. Genes were ranked by CV, divided into bins, and plotted as described above.

For tissue-specific expression, we identified genes with at least a two-fold change in expression relative to their average expression across all tissues and assessed chromatin state distributions accordingly. A similar approach was applied to analyse chromatin state occupancy over gene bodies. To identify genes marked by H3K27me3, we applied the method described in [27] with data from [27] and [33].

Transcription factor occupancy and association with chromatin states. To analyse TF occupancy, we compiled publicly available *Arabidopsis thaliana* ChIP-seq and DAP-seq BED files from multiple studies [19,37,76]. For *Marchantia polymorpha*, we inferred TF binding sites by mapping position weight matrices (PWMs) from [37] to the *Marchantia* v7.1 genome using FIMO (MEME Suite) [77]. In both cases, each significant peak region in the genome was classified according to chromatin state using the `OverlapsAny` function from GenomicRanges package in R. For that purpose, only the centre of the peak was used to assign the chromatin state.

TF binding enrichment for each chromatin state was calculated following [6]. Specifically, for each TF and chromatin state (*s*), enrichment was determined using: $(a_s/b)/(c_s/d)$, where a_s is the total number of bases in a peak call in *s*; *b* is the total number of bases in a peak call; c_s is the total number of bases in *s*; *d* is the total number of bases of the chromatin state.

To calculate TF activity, we incorporated gene co-expression data. First, genes associated with TF peaks were annotated using the ChIPseeker package [78] with default parameters. Then, co-expression data from ATTEDII [79] (version Ath-r.c3-1) was used to filter TF target genes. Both positively (score < 2000) and negatively (score > max(score) - 1000) co-expressed genes were retained. The same enrichment formula described above was applied to this filtered subset. After this filter, we used the same formula than before.

To calculate the score of TF activity, we assessed statistical Enrichment of co-Expressed TF targets in chromatin states, we performed Fisher's exact tests (alternative hypothesis = greater). For each chromatin state, we compared TF targets with positively or negatively co-expressed genes as described above. Peaks were annotated as described above, and gene IDs were extracted for each chromatin state. As null distribution, we considered a random sample of genes from the whole genome using Table 2.

Table 2. Data to score TF activity.

Count (genes with peaks + co expressed)	count (genes with peaks)
Count (co expressed)	27205 for <i>Arabidopsis</i> or 18328 for <i>Marchantia</i>

<https://doi.org/10.1371/journal.pgen.1012015.t002>

The TF-chromatin state score was computed as: $-\log_{10}(\text{p-value}) + \log_2(\text{odds ratio} + 1)$ and normalized using the maximum score for each TF.

For each method we generated an $N_{\text{TF}} \times N_{\text{state}}$ matrix containing all computed scores and visualized chromatin state occupancy patterns using the pheatmap R package. For aggregate analyses across all TFs, scores were summed and normalized to the maximum score for each state.

In subsequent analyses, TFs with maximum scores below 8 were filtered out to exclude low-confidence results. TF family enrichment was assessed using Fisher's exact tests, employing TF classifications from [37] and PlantTFDB [80]. GO term enrichment analysis of biological processes in *Arabidopsis* was performed using PANTHER [81] with default parameters.

Supporting information

S1 Fig. Genome-wide chromatin state profiling and functional characterization in *Arabidopsis thaliana*. (A) Box plot showing the expression of protein-coding genes overlapping with each chromatin state in Transcripts per Million (TPM). (B-D) Box plot showing levels of CpG (B), CHH (C), and CHG (D) methylation for all chromatin states. (E) Flow diagram showing the overlap (in bp) between the chromatin states defined in this study with states from Jamge et al., 2023. The color code for the flow diagram represents the chromatin states in the current model. (F) Genomic distribution of heterochromatic states (H1–H7) from bottom to top on Chr1 of *Arabidopsis thaliana*. (G) Pie chart showing the percentage of the genome covered by each state. (TIF)

S2 Fig. Genome-wide chromatin state profiling and functional characterization in *Marchantia polymorpha*. (A) Pie chart showing the percentage of the genome covered by each state of *Marchantia polymorpha*. (B) Box plot showing the expression of protein-coding genes overlapping with each chromatin state of *Marchantia polymorpha* in Transcripts per Million (TPM). (C-E) Box plot showing levels of CpG (C), CHH (D), and CHG (E) methylation for all chromatin states of *Marchantia polymorpha*. (F) Enrichment heatmap of CAGE-seq signal around the TSS defined in *Marchantia* genome MpTak1v7.1. The green rectangle represents the bin size reference for the chromatin state as calculated by chromHMM. (G) Heatmap showing the Jaccard similarity index between the states generated using the whole model and states using a subset of marks, i.e., excluding a set of marks and variants as indicated on the x-axis. (H) Hierarchical clustering of the emission probabilities of chromatin states of *Marchantia* and a down-sampled 15-states model of *Arabidopsis* representing the conservation of chromatin state definition in the two distantly related species. (I) Heatmap and metaplots showing the enrichment of H3K27me1, H3K9me1, and H3K4me3 marks in wild type and *Mpknox2 Mpe(z)1* double mutants. The enrichment of both marks is plotted on wild type H3K27me3 peaks ($n=25159$). The two clusters are the same as obtained in Fig 2E. (J, K) Boxplot of TF occupancy (J) and TF activity (K) scores across different chromatin states. The later scores are calculated as the statistical enrichment of TF binding over putative targets (co-expressed genes). (L) Heatmap and hierarchical clustering of TF activity scores for PWM across chromatin states. Colour scale is shown at the bottom. Scores are normalized to the maximum value of each TF. Low unnormalized scores were discarded. The colour code for the diagrams represents the chromatin states in the current 15-states models. (TIF)

S3 Fig. Functional characterization of a 15-state chromatin model in *Arabidopsis thaliana*. (A) Bubble plot showing the emission probabilities for histone modifications/variants for a down-sampled 15-state model of *Arabidopsis thaliana*.

The size of the bubble represents the emission probability ranging from 0 to 1. Coloured rectangles demarcate the classification of states into major domains of chromatin. The box plot on top shows the average ATAC-seq signal for each state representing chromatin accessibility. The two boxes per state are two replicates of the ATAC-seq experiment. (B) Stacked bar plot showing the overlap between annotated genomic features and the chromatin states in the 15-state model of *Arabidopsis thaliana*. (C) Pie chart showing the percentage of the genome covered by each state of the 15-state model of *Arabidopsis thaliana*. (D) Box plot showing the expression of protein-coding genes overlapping with each chromatin state of the 15-state model of *Arabidopsis thaliana* in Transcripts per Million (TPM). (E-G) Box plot showing levels of CpG, CHH, and CHG methylation respectively, for all chromatin states of the 15-state model of *Arabidopsis thaliana*. (H) Neighbourhood Enrichment analysis of chromatin states of *Arabidopsis* 15-state model representing the fold enrichment for each state at fixed positions relative to the anchor position (TSS and TTS). (I) Flow diagram showing the overlap (in bp) between the chromatin states of extensive 26-state model and the 15-state model *Arabidopsis thaliana*. The colour code for the flow diagram represents the colours for the states of the 15-state model.

(TIF)

S4 Fig. Chromatin state dynamics across gene bodies in *Arabidopsis* and *Marchantia*. (A, B) Bar plot of the proportion of chromatin states for 20 bins across coding regions of 19102 protein-coding genes with experimentally derived TSS and TTS from Ivanov et al., 2021 ordered by transcripts per million (TPM) (A) and coefficient of variation (CV) (B) in *Arabidopsis* leaves. (C, D) Bar plot of the proportion of chromatin states for 20 bins across coding regions of 19102 protein-coding genes with experimentally derived TSS and TTS from Ivanov et al., 2021 and the 8104 protein-coding genes with TAIR10 derived TSS and TTS ordered by transcripts per million (TPM) (C) and coefficient of variation (CV) (D) in *Arabidopsis* leaves. (E, F) Bar plot of the proportion of chromatin states for 20 bins across coding regions of all 27206 protein-coding genes with TAIR10 derived TSS and TTS ordered by transcripts per million (TPM) (E) and coefficient of variation (CV) (F) in *Arabidopsis* leaves. (G, H) Bar plot of the proportion of chromatin states for 20 bins across coding regions of genes ordered by transcripts per million (TPM) (G) and coefficient of variation (CV) (H) *Arabidopsis* leaves. (I, J) Bar plot of the proportion of chromatin states for 20 bins across coding regions of genes ordered by transcripts per million (TPM) (I) and coefficient of variation (CV) (J) *Marchantia* thallus.

(TIF)

S5 Fig. TF binding distribution at TSS and developmental expression patterns in *Arabidopsis*. (A) Histogram of TF binding distribution relative to the experimentally derived transcription starting site (TSS) of 19102 protein-coding genes (Ivanov et al., 2021) of *Arabidopsis thaliana* TFs from different clusters defined in Fig 4F. (B) Histogram of TF binding distribution relative to the experimentally derived transcription starting site (TSS) of 19102 protein-coding genes (Ivanov et al., 2021) and TAIR10 derived TSS for the remaining 8104 protein-coding genes of *Arabidopsis thaliana* TFs from different clusters defined in Fig 4F. (C) Histogram of TF binding distribution relative to the TAIR10 derived transcription starting site (TSS) of all 27206 protein-coding genes of *Arabidopsis thaliana* TFs from different clusters defined in Fig 4F. (D) Line plots showing the median expression levels of the TFs belonging to the 5 clusters shown in Fig 4F across the broad developmental stages of *Arabidopsis thaliana* development.

(TIF)

S6 Fig. Model of transcription factor interplay with facultative and open chromatin to control transcription rates.

The histone icon was adapted from Bioicons: https://bioicons.com/icons/cc-by-4.0/Nucleic_acids/DBCLS/histone.svg. Acknowledgement: histone icon by DBCLS <https://togotv.dbcls.jp/en/pics.html> is licensed under CC-BY 4.0 Unported <https://creativecommons.org/licenses/by/4.0/>.

(TIF)

S7 Fig. Impact of TSS/TTS annotation sources on chromatin state enrichment patterns in *Arabidopsis*. (A) Strand-aware density of TSS and TES shifts (bp) between the Ivanov et al. (2021) calls and TAIR10 annotations. Kernel density

estimates (y-axis: probability density, 1/bp) show the distribution of per-gene shifts (in bp) between Ivanov et al. (2021) calls and TAIR10 protein-coding gene annotations for *Arabidopsis thaliana*. (B) Neighbourhood enrichment analysis of the chromatin states on 19102 protein-coding genes with experimentally derived TSS and TTS from Ivanov et al., 2021. (C) Neighbourhood enrichment analysis of the chromatin states on 19102 genes with experimentally derived TSS and TTS from Ivanov et al., 2021 and the 8104 protein-coding genes with TAIR10 derived TSS and TTS. (D) Neighbourhood enrichment analysis of the chromatin states on all 27206 protein-coding genes with TAIR10 derived TSS and TTS. (E) Jaccard similarity index estimating the similarity in TSS and TTS positions between the 19102 updated coordinates from Ivanov et al., 2021 and 19102 + (8104 coordinates from TAIR10) compared with TAIR10 TSS/TTS coordinates. The higher the jaccard index, the higher similarity between the two coordinates.

(TIF)

S1 Table. Datasets used to define *Arabidopsis* chromatin states model.

(XLSX)

S2 Table. Datasets used to define *Marchantia* chromatin states model.

(XLSX)

S3 Table. List of *Arabidopsis* Transcription Factors enriched in each of the 5 clusters and related metadata.

(XLSX)

S4 Table. List of Transcription Factors studied in *Marchantia* and related metadata.

(XLSX)

S5 Table. (A–E) Gene Ontology lists for cluster 1–5 of *Arabidopsis*.

(XLSX)

S1 Data. BED files containing the chromatin states segments data files of *Arabidopsis* 26 and 15–state models and *Marchantia* 15–state model.

(BED)

S2 Data. MP 5 segments.

(BED)

S3 Data. AT 15 segments.

(BED)

Acknowledgments

We thank the entire Berger group, for their insightful, considerate, and helpful discussions, as well as Matt Watson for editorial assistance during the preparation of the manuscript. We thank the Molecular Biology Service and Media Kitchen for a constant supply of plates, basic reagents, and cloning and sequencing services, as well as the Peptide Synthesis Service. Additionally, we thank the Vienna BioCenter Core Facilities, in particular the Next Generation Sequencing facility for their advice and swift handling of all our requests and helpful discussions. For open access purposes, the authors have applied a CC BY public copyright license to any author accepted manuscript version arising from this submission.

Author contributions

Conceptualization: Frederic Berger, Facundo Romani.

Data curation: Vikas Shukla, Elin Axelsson, Jim Haseloff, Facundo Romani.

Formal analysis: Vikas Shukla, Elin Axelsson, Facundo Romani.

Funding acquisition: Tetsuya Hisanaga, Frederic Berger, Facundo Romani.

Investigation: Vikas Shukla, Elin Axelsson, Tetsuya Hisanaga, Facundo Romani.

Methodology: Vikas Shukla, Elin Axelsson, Tetsuya Hisanaga, Facundo Romani.

Project administration: Frederic Berger, Facundo Romani.

Resources: Vikas Shukla, Elin Axelsson, Tetsuya Hisanaga, Facundo Romani.

Software: Vikas Shukla, Elin Axelsson, Facundo Romani.

Supervision: Frederic Berger.

Validation: Vikas Shukla, Elin Axelsson, Facundo Romani.

Visualization: Vikas Shukla, Facundo Romani.

Writing – original draft: Vikas Shukla, Frederic Berger, Facundo Romani.

Writing – review & editing: Vikas Shukla, Elin Axelsson, Frederic Berger, Facundo Romani.

References

1. Filion GJ, van Bommel JG, Braunschweig U, Talhout W, Kind J, Ward LD, et al. Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell*. 2010;143(2):212–24. <https://doi.org/10.1016/j.cell.2010.09.009> PMID: 20888037
2. Ernst J, Kellis M. Chromatin-state discovery and genome annotation with ChromHMM. *Nat Protoc*. 2017;12(12):2478–92. <https://doi.org/10.1038/nprot.2017.124> PMID: 29120462
3. Kharchenko PV, Alekseyenko AA, Schwartz YB, Minoda A, Riddle NC, Ernst J, et al. Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature*. 2011;471(7339):480–5. <https://doi.org/10.1038/nature09725> PMID: 21179089
4. Slattery M, Zhou T, Yang L, Dantas Machado AC, Gordân R, Rohs R. Absence of a simple code: how transcription factors read the genome. *Trends Biochem Sci*. 2014;39(9):381–99. <https://doi.org/10.1016/j.tibs.2014.07.002> PMID: 25129887
5. Ernst J, Kheradpour P, Mikkelsen TS, Shores N, Ward LD, Epstein CB, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*. 2011;473(7345):43–9. <https://doi.org/10.1038/nature09906> PMID: 21441907
6. Ernst J, Kellis M. Interplay between chromatin state, regulator binding, and regulatory motifs in six human cell types. *Genome Res*. 2013;23(7):1142–54. <https://doi.org/10.1101/gr.144840.112> PMID: 23595227
7. Magnani L, Ballantyne EB, Zhang X, Lupien M. PBX1 genomic pioneer function drives ER α signaling underlying progression in breast cancer. *PLOS Genet*. 2011;7(11):e1002368. <https://doi.org/10.1371/journal.pgen.1002368>
8. Zaret KS, Mango SE. Pioneer transcription factors, chromatin dynamics, and cell fate control. *Curr Opin Genet Dev*. 2016;37:76–81. <https://doi.org/10.1016/j.gde.2015.12.003> PMID: 26826681
9. Iwafuchi-Doi M, Zaret KS. Cell fate control by pioneer transcription factors. *Development*. 2016;143(11):1833–7. <https://doi.org/10.1242/dev.133900> PMID: 27246709
10. Wang J, Wang GG. No easy way out for EZH2: its pleiotropic, noncanonical effects on gene regulation and cellular function. *Int J Mol Sci*. 2020;21(24):9501. <https://doi.org/10.3390/ijms21249501>
11. Bieluszewski T, Xiao J, Yang Y, Wagner D. PRC2 activity, recruitment, and silencing: a comparative perspective. *Trends Plant Sci*. 2021;26(11):1186–98.
12. Pelayo MA, Yamaguchi N, Ito T. One factor, many systems: the floral homeotic protein AGAMOUS and its epigenetic regulatory mechanisms. *Curr Opin Plant Biol*. 2021;61:102009. <https://doi.org/10.1016/j.pbi.2021.102009> PMID: 33640614
13. Dent SYR. KAT tales: Functions of Gcn5 and PCAF lysine acetyltransferases in SAGA and ATAC. *J Biol Chem*. 2024;300(10):107744. <https://doi.org/10.1016/j.jbc.2024.107744> PMID: 39222683
14. Soffers JHM, Workman JL. The SAGA chromatin-modifying complex: the sum of its parts is greater than the whole. *Genes Dev*. 2020;34(19–20):1287–303.
15. Jamge B, Lorković ZJ, Axelsson E, Osakabe A, Shukla V, Yelagandula R, et al. Histone variants shape chromatin states in Arabidopsis. *eLife*. 2023;12:RP87714.
16. Sequeira-Mendes J, Aragüez I, Peiró R, Mendez-Giraldez R, Zhang X, Jacobsen SE, et al. The functional topography of the Arabidopsis genome is organized in a reduced number of linear motifs of chromatin states. *Plant Cell*. 2014;26(6):2351–66. <https://doi.org/10.1105/tpc.114.124578> PMID: 24934173
17. Sun Y, Dong L, Zhang Y, Lin D, Xu W, Ke C, et al. 3D genome architecture coordinates trans and cis regulation of differentially expressed ear and tassel genes in maize. *Genome Biol*. 2020;21(1):143. <https://doi.org/10.1186/s13059-020-02063-7> PMID: 32546248

18. Baker K, Dhillon T, Colas I, Cook N, Milne I, Milne L, et al. Chromatin state analysis of the barley epigenome reveals a higher-order structure defined by H3K27me1 and H3K27me3 abundance. *Plant J*. 2015;84(1):111–24. <https://doi.org/10.1111/tpj.12963> PMID: [26255869](#)
19. Heyndrickx KS, Van de Velde J, Wang C, Weigel D, Vandepoele K. A functional and evolutionary perspective on transcription factor binding in *Arabidopsis thaliana*. *Plant Cell*. 2014;26(10):3894–910. <https://doi.org/10.1105/tpc.114.130591> PMID: [25361952](#)
20. Lai X, Blanc-Mathieu R, Grand Vuillemin L, Huang Y, Stigliani A, Lucas J, et al. The LEAFY floral regulator displays pioneer transcription factor properties. *Molecular Plant*. 2021;14(5):829–37. <https://doi.org/10.1016/j.molp.2021.01.001>
21. Pajoro A, Madrigal P, Muiño JM, Matus JT, Jin J, Mecchia MA, et al. Dynamics of chromatin accessibility and gene regulation by MADS-domain transcription factors in flower development. *Genome Biol*. 2014;15(3):R41. <https://doi.org/10.1186/gb-2014-15-3-r41> PMID: [24581456](#)
22. Wu MF, Sang Y, Bezhani S, Yamaguchi N, Han SK, Li Z, et al. SWI2/SNF2 chromatin remodeling ATPases overcome polycomb repression and control floral organ identity with the LEAFY and SEPALLATA3 transcription factors. *Proc Natl Acad Sci U S A*. 2012;109(9):3576–81.
23. Swift J, Coruzzi GM. A matter of time - How transient transcription factor interactions create dynamic gene regulatory networks. *Biochim Biophys Acta Gene Regul Mech*. 2017;1860(1):75–83. <https://doi.org/10.1016/j.bbagr.2016.08.007> PMID: [27546191](#)
24. Alvarez JM, Brooks MD, Swift J, Coruzzi GM. Time-based systems biology approaches to capture and model dynamic gene regulatory networks. *Annu Rev Plant Biol*. 2021;72:105–31. <https://doi.org/10.1146/annurev-arplant-081320-090914> PMID: [33667112](#)
25. Kudron MM, Victorsen A, Gevirtzman L, Hillier LW, Fisher WW, Vafeados D, et al. The ModERN resource: genome-wide binding profiles for hundreds of *Drosophila* and *Caenorhabditis elegans* transcription factors. *Genetics*. 2018;208(3):937–49.
26. de Sousa F, Foster PG, Donoghue PCJ, Schneider H, Cox CJ. Nuclear protein phylogenies support the monophyly of the three bryophyte groups (Bryophyta Schimp.). *New Phytol*. 2019;222(1):565–75. <https://doi.org/10.1111/nph.15587> PMID: [30411803](#)
27. Montgomery SA, Tanizawa Y, Galik B, Wang N, Ito T, Mochizuki T, et al. Chromatin organization in early land plants reveals an ancestral association between H3K27me3, transposons, and constitutive heterochromatin. *Curr Biol*. 2020;30(4):573–88. <https://doi.org/10.1016/j.cub.2019.12.034>
28. Hisanaga T, Wu S, Schafran P, Axelsson E, Akimcheva S, Dolan L, et al. The ancestral chromatin landscape of land plants. *New Phytol*. 2023;240(5):2085–101. <https://doi.org/10.1111/nph.19311> PMID: [37823324](#)
29. Yelagandula R, Osakabe A, Axelsson E, Berger F, Kawashima T. Genome-wide profiling of histone modifications and histone variants in *Arabidopsis thaliana* and *Marchantia polymorpha*. *Methods Mol Biol*. 2017;1610:93–106. https://doi.org/10.1007/978-1-4939-7003-2_7 PMID: [28439859](#)
30. Tanizawa Y, Mochizuki T, Yagura M, Sakamoto M, Fujisawa T, Kawamura S, et al. MarpolBase: Genome database for *Marchantia polymorpha* featuring high quality reference genome sequences [Internet]. bioRxiv. 2025. [cited 2025 Jul 4]. Available from: <https://www.biorxiv.org/content/10.1101/2025.03.30.646155v1>
31. Yin X, Romero-Campero FJ, de Los Reyes P, Yan P, Yang J, Tian G, et al. H2AK121ub in *Arabidopsis* associates with a less accessible chromatin state at transcriptional regulation hotspots. *Nat Commun*. 2021;12(1):315. <https://doi.org/10.1038/s41467-020-20614-1> PMID: [33436613](#)
32. Takahashi H, Kato S, Murata M, Carninci P. CAGE (cap analysis of gene expression): a protocol for the detection of promoter and transcriptional networks. *Methods Mol Biol*. 2012;786:181–200. https://doi.org/10.1007/978-1-61779-292-2_11 PMID: [21938627](#)
33. Shu J, Chen C, Thapa RK, Bian S, Nguyen V, Yu K, et al. Genome-wide occupancy of histone H3K27 methyltransferases CURLY LEAF and SWINGER in *Arabidopsis* seedlings. *Plant Direct*. 2019;3(1):e00100. <https://doi.org/10.1002/pld3.100> PMID: [31245749](#)
34. Zhou Y, Romero-Campero FJ, Gómez-Zambrano Á, Turck F, Calonje M. H2A monoubiquitination in *Arabidopsis thaliana* is generally independent of LHP1 and PRC2 activity. *Genome Biol*. 2017;18(1):69. <https://doi.org/10.1186/s13059-017-1197-z> PMID: [28403905](#)
35. Schubert D, Primavesi L, Bishopp A, Roberts G, Doonan J, Jenuwein T, et al. Silencing by plant Polycomb-group genes requires dispersed trimethylation of histone H3 at lysine 27. *EMBO J*. 2006;25(19):4638–49.
36. Chanvivattana Y, Bishopp A, Schubert D, Stock C, Moon Y-H, Sung ZR, et al. Interaction of Polycomb-group proteins controlling flowering in *Arabidopsis*. *Development*. 2004;131(21):5263–76. <https://doi.org/10.1242/dev.01400> PMID: [15456723](#)
37. O'Malley RC, Huang S-SC, Song L, Lewsey MG, Bartlett A, Nery JR, et al. Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell*. 2016;165(5):1280–92. <https://doi.org/10.1016/j.cell.2016.04.038> PMID: [27203113](#)
38. Brooks MD, Juang C-L, Katari MS, Alvarez JM, Pasquino A, Shih H-J, et al. ConnectTF: a platform to integrate transcription factor-gene interactions and validate regulatory networks. *Plant Physiol*. 2021;185(1):49–66. <https://doi.org/10.1093/plphys/kiaa012> PMID: [33631799](#)
39. Romani F, Moreno JE. Molecular mechanisms involved in functional macroevolution of plant transcription factors. *New Phytol*. 2021;230(4):1345–53. <https://doi.org/10.1111/nph.17161> PMID: [33368298](#)
40. Franco-Zorrilla JM, López-Vidriero I, Carrasco JL, Godoy M, Vera P, Solano R. DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proc Natl Acad Sci U S A*. 2014;111(6):2367–72. <https://doi.org/10.1073/pnas.1316278111> PMID: [24477691](#)
41. Wong F, Gunawardena J. Gene Regulation in and out of Equilibrium. *Annu Rev Biophys*. 2020;49:199–226.
42. Guilfoyle TJ, Hagen G. Auxin response factors. *Curr Opin Plant Biol*. 2007;10(5):453–60. <https://doi.org/10.1016/j.pbi.2007.08.014> PMID: [17900969](#)
43. Voicheck Y, Hristova G, Mollá-Morales A, Weigel D, Nordborg M. Widespread position-dependent transcriptional regulatory sequences in plants. *Nat Genet*. 2024;56(10):2238–46. <https://doi.org/10.1038/s41588-024-01907-3> PMID: [39266765](#)
44. Dietz K-J, Vogel MO, Viehhauser A. AP2/EREBP transcription factors are part of gene regulatory networks and integrate metabolic, hormonal and environmental signals in stress acclimation and retrograde signalling. *Protoplasma*. 2010;245(1–4):3–14. <https://doi.org/10.1007/s00709-010-0142-8> PMID: [20411284](#)

45. Kizis D, Lumberras V, Pagès M. Role of AP2/EREBP transcription factors in gene regulation during abiotic stress. *FEBS Lett.* 2001;498(2–3):187–9. [https://doi.org/10.1016/s0014-5793\(01\)02460-7](https://doi.org/10.1016/s0014-5793(01)02460-7) PMID: [11412854](#)
46. Sridhar VV, Surendrarao A, Liu Z. APETALA1 and SEPALLATA3 interact with SEUSS to mediate transcription repression during flower development. *Development.* 2006;133(16):3159–66. <https://doi.org/10.1242/dev.02498> PMID: [16854969](#)
47. Hu Y, Lai Y, Chen X, Zhou D-X, Zhao Y. Distribution pattern of histone marks potentially determines their roles in transcription and RNA processing in rice. *J Plant Physiol.* 2020;249:153167. <https://doi.org/10.1016/j.jplph.2020.153167> PMID: [32353606](#)
48. Leng X, Thomas Q, Rasmussen SH, Marquardt S. A g(enomic)p(ositioning)s(ystem) for plant RNAPII transcription. *Trends Plant Sci.* 2020;25(8):744–64.
49. Hisanaga T, Romani F, Wu S, Kowar T, Wu Y, Lintermann R, et al. The Polycomb repressive complex 2 deposits H3K27me3 and represses transposable elements in a broad range of eukaryotes. *Curr Biol.* 2023;33(20):4367–80. <https://doi.org/10.1016/j.cub.2023.08.020>
50. Navarrete C, Montgomery SA, Mendieta J, Lara-Astiaso D, Sebé-Pedrós A. Diversity and evolution of chromatin regulatory states across eukaryotes [Internet]. *bioRxiv.* 2025 [cited 2025 Jun 26]. Available from: <https://www.biorxiv.org/content/10.1101/2025.03.17.643675v1>
51. Tamburri S, Conway E, Pasini D. Polycomb-dependent histone H2A ubiquitination links developmental disorders with cancer. *Trends Genet.* 2022;38(4):333–52. <https://doi.org/10.1016/j.tig.2021.07.011> PMID: [34426021](#)
52. Loubiere V, Martinez A-M, Cavalli G. Cell fate and developmental regulation dynamics by polycomb proteins and 3D genome architecture. *Bioessays.* 2019;41(3):e1800222. <https://doi.org/10.1002/bies.201800222> PMID: [30793782](#)
53. de Potter B, Raas MWD, Seidl MF, Verrijzer CP, Snel B. Uncoupled evolution of the polycomb system and deep origin of non-canonical PRC1. *Commun Biol.* 2023;6(1):1144. <https://doi.org/10.1038/s42003-023-05501-x> PMID: [37949928](#)
54. Wang Q, Shen WH. Chromatin modulation and gene regulation in plants: insight about PRC1 function. *Biochem Soc Trans.* 2018;46(4):957–66.
55. Yip KY, Cheng C, Bhardwaj N, Brown JB, Leng J, Kundaje A, et al. Classification of human genomic regions based on experimentally determined binding sites of more than 100 transcription-related factors. *Genome Biol.* 2012;13(9):R48. <https://doi.org/10.1186/gb-2012-13-9-r48> PMID: [22950945](#)
56. Luzete-Monteiro E, Zaret KS. Structures and consequences of pioneer factor binding to nucleosomes. *Curr Opin Struct Biol.* 2022;75:102425. <https://doi.org/10.1016/j.sbi.2022.102425> PMID: [35863165](#)
57. Carminati M, Vecchia L, Stoos L, Thomä NH. Pioneer factors: emerging rules of engagement for transcription factors on chromatinized DNA. *Curr Opin Struct Biol.* 2024;88:102875. <https://doi.org/10.1016/j.sbi.2024.102875> PMID: [38991237](#)
58. Jin R, Klasfeld S, Zhu Y, Fernandez Garcia M, Xiao J, Han S-K, et al. LEAFY is a pioneer transcription factor and licenses cell reprogramming to floral fate. *Nat Commun.* 2021;12(1):626. <https://doi.org/10.1038/s41467-020-20883-w> PMID: [33504790](#)
59. Strader L, Weijers D, Wagner D. Plant transcription factors - being in the right place with the right company. *Curr Opin Plant Biol.* 2022;65:102136. <https://doi.org/10.1016/j.pbi.2021.102136> PMID: [34856504](#)
60. Insights into Land Plant Evolution Garnered from the *Marchantia polymorpha* Genome - University of California Berkeley [Internet]. [cited 2025 Nov 2]. Available from: https://search.library.berkeley.edu/discovery/fulldisplay/cdi_swepub_primary_oai_DiVA_org_uu_333225/01UCS_BER:UCB
61. Quinlan AR. BEDTools: the swiss-army tool for genome feature analysis. *Curr Protoc Bioinformatics.* 2014;47:11.12.1-34. <https://doi.org/10.1002/0471250953.bi1112s47> PMID: [25199790](#)
62. Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data. 2010. Available from: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
63. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9(4):357–9. <https://doi.org/10.1038/nmeth.1923> PMID: [22388286](#)
64. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics.* 2009;25(16):2078–9. <https://doi.org/10.1093/bioinformatics/btp352> PMID: [19505943](#)
65. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research.* 2016;44(W1):W160–5.
66. Riemondy KA, Sheridan RM, Gillen A, Yu Y, Bennett CG, Hesselberth JR. valr: Reproducible genome interval analysis in R [Internet]. *F1000Research.* 2017 [cited 2025 Feb 12]. Available from: <https://f1000research.com/articles/6-1025>
67. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the Tidyverse. *J Open Source Softw.* 2019;4(43):1686. <https://doi.org/10.21105/joss.01686>
68. Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize Implements and enhances circular visualization in R. *Bioinformatics.* 2014;30(19):2811–2. <https://doi.org/10.1093/bioinformatics/btu393> PMID: [24930139](#)
69. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics.* 2016;32(18):2847–9.
70. Cheng C-Y, Krishnakumar V, Chan AP, Thibaud-Nissen F, Schobel S, Town CD. Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant J.* 2017;89(4):789–804. <https://doi.org/10.1111/tpj.13415> PMID: [27862469](#)

71. Ivanov M, Sandelin A, Marquardt S. TranscriptomeReconstructoR: data-driven annotation of complex transcriptomes. *BMC Bioinformatics*. 2021;22(1):290. <https://doi.org/10.1186/s12859-021-04208-2> PMID: [34058980](https://pubmed.ncbi.nlm.nih.gov/34058980/)
72. Bourguet P, Picard CL, Yelagandula R, Pélissier T, Lorković ZJ, Feng S, et al. The histone variant H2A.W and linker histone H1 co-regulate heterochromatin accessibility and DNA methylation. *Nat Commun*. 2021;12(1):2683. <https://doi.org/10.1038/s41467-021-22993-5> PMID: [33976212](https://pubmed.ncbi.nlm.nih.gov/33976212/)
73. Bourguet P, Yelagandula R, To TK, Osakabe A, Alishe A, Lu RJH, et al. The histone variant H2A.W cooperates with chromatin modifications and linker histone H1 to maintain transcriptional silencing of transposons in Arabidopsis [Internet]. *bioRxiv*. 2022 [cited 2025 Feb 13]. Available from: <https://www.biorxiv.org/content/10.1101/2022.05.31.493688v1>
74. Kawamura S, Romani F, Yagura M, Mochizuki T, Sakamoto M, Yamaoka S, et al. MarpolBase expression: a web-based, comprehensive platform for visualization and analysis of transcriptomes in the liverwort *Marchantia polymorpha*. *Plant Cell Physiol*. 2022;63(11):1745–55.
75. Ruprecht C, Proost S, Hernandez-Coronado M, Ortiz-Ramirez C, Lang D, Rensing SA, et al. Phylogenomic analysis of gene co-expression networks reveals the evolution of functional modules. *Plant J*. 2017;90(3):447–65. <https://doi.org/10.1111/tpj.13502> PMID: [28161902](https://pubmed.ncbi.nlm.nih.gov/28161902/)
76. Song L, Huang S-SC, Wise A, Castanon R, Nery JR, Chen H, et al. A transcription factor hierarchy defines an environmental stress response network. *Science*. 2016;354(6312):aag1550. <https://doi.org/10.1126/science.aag1550> PMID: [27811239](https://pubmed.ncbi.nlm.nih.gov/27811239/)
77. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. *Bioinformatics*. 2011;27(7):1017–8.
78. Yu G, Wang L-G, He Q-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*. 2015;31(14):2382–3. <https://doi.org/10.1093/bioinformatics/btv145> PMID: [25765347](https://pubmed.ncbi.nlm.nih.gov/25765347/)
79. Obayashi T, Aoki Y, Tadaka S, Kagaya Y, Kinoshita K. ATTED-II in 2018: a plant coexpression database based on investigation of the statistical property of the mutual rank index. *Plant Cell Physiol*. 2018;59(1):e3. <https://doi.org/10.1093/pcp/pcx191> PMID: [29216398](https://pubmed.ncbi.nlm.nih.gov/29216398/)
80. Jin J, Tian F, Yang D-C, Meng Y-Q, Kong L, Luo J, et al. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res*. 2017;45(D1):D1040–5. <https://doi.org/10.1093/nar/gkw982> PMID: [27924042](https://pubmed.ncbi.nlm.nih.gov/27924042/)
81. Thomas PD, Ebert D, Muruganujan A, Mushayahama T, Albou L-P, Mi H. PANTHER: making genome-scale phylogenetics accessible to all. *Protein Sci*. 2022;31(1):8–22. <https://doi.org/10.1002/pro.4218> PMID: [34717010](https://pubmed.ncbi.nlm.nih.gov/34717010/)