DISSERTATION | DOCTORAL THESIS

Titel | Title

The Interplay Between Chromatin States and Transcription Factors Fine-Tune Gene Expression in Land Plants

verfasst von | submitted by

Vikas Shukla

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD)

Wien | Vienna, 2025

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 794 620 490
Dissertationsgebiet lt. Studienblatt | Field of study as it appears on the student record sheet:	Molekulare Biologie
Betreut von | Supervisor:	Dr. Frederic Berger PhD

तत्वमसि

You-are-that

Acknowledgements

Dear Friends, Colleagues, and Family,

This thesis represents not just my individual effort, but the collective support, wisdom, and encouragement of many remarkable people who have shaped this journey in countless ways. As I reflect on the path that led to this work, I am filled with gratitude for the mentors, colleagues, friends, and family who championed me, challenged me, and stood by me through every triumph and setback.

I would like to begin by expressing my deepest gratitude to my supervisor, **Frédéric Berger**, whose unwavering support has been the cornerstone of this entire journey. Fred's belief in my potential gave me the confidence to tackle ambitious questions and push boundaries I thought impossible. His door was always open for discussions, whether I needed help interpreting puzzling results or simply reassurance during the inevitable moments of doubt that every PhD student faces. His ability to see the bigger picture while helping me navigate the intricate details of chromatin biology has shaped both this thesis and my growth as a scientist.

Throughout this research journey, Fred provided the perfect balance of intellectual freedom and supportive guidance. His enthusiasm for science is infectious, and his commitment to fostering a collaborative lab environment made even the most challenging days manageable. I am profoundly grateful for his mentorship, patience, and genuine care for my development both as a researcher and as a person.

I owe immense gratitude to **Elin**, my desk neighbor and dearest colleague. Sitting right next to me, Elin became not just a colleague but a constant source of support, laughter, and expertise. Her incredible sense of humor brightened even the most frustrating debugging sessions, and her willingness to help with every single challenge I encountered—no matter how small or complex—has been truly extraordinary. I lost count of how many times I turned to her with a panicked 'Elin, can I show you something?' only to have her calmly walk me through the solution. Whether I was wrestling with a stubborn script, puzzling over unexpected results, or simply needed someone to bounce ideas off, Elin was always there with her sharp insights and infectious laughter. Her combination of technical brilliance and genuine kindness made our shared office space a place I looked forward to being in every day.

I am grateful to **Pierre**, who became my go-to person for absolutely everything. Countless times, I would spin 180 degrees in my chair to ask Pierre yet another question. His extraordinary knowledge and exemplary attention to detail saved me from countless mistakes and dead ends. What made these interactions even more special was how they often evolved with Elin inevitably joining our discussions, and suddenly our corner of the office would transform into an impromptu problem-solving session. These three-way conversations were magical; no matter how stuck I felt when I first turned around in my chair, I would always walk away with a clear, well-formulated plan of action.

I would also like to thank **Zdravko**, whose knowledge, directness, and genuine kindness have been invaluable throughout my PhD. Zdravko has this wonderful ability to cut straight to the heart of any problem while remaining incredibly easy to talk to. What touched me most was his habit of checking on me randomly—he would simply appear at my desk to make sure I wasn't stuck somewhere, offering help before I even realized I needed it. His straightforward approach and willingness to share his expertise, combined with his genuine care for my progress, provided exactly the kind of support that kept me moving forward when research felt overwhelming. Zdravko's blend of scientific rigor and human warmth made him someone I could always count on for both technical guidance and reassurance.

I am deeply grateful to **Chung-Hyun**, whose caring and supportive nature brought such warmth to my PhD experience. He regularly checked on me and was always ready to help when I needed it most. Some of my fondest memories are sharing tea during cold winter days with him and the dinners he kindly took me to many times. Chung-Hyun's thoughtfulness and generosity created a sense of community that made even the most challenging periods feel not so bad.

I'm grateful to **Jian Yi**, my fellow PhD companion, who will be defending just one day after me. Sharing this final stretch of the PhD journey with someone who truly understands the experience has been invaluable. I also want to thank **Coleen**, **Barbara**, and **Akhilesh** whose fun-loving spirits and camaraderie made the lab feel like a true community. It’s you people and **Giovanni**, our official Jägermeister supplier who made the last GMI retreat party epic.

Thanks to **Zach**, whose hot takes kept our discussions lively and thought-provoking, while his knowledge and helpfulness made him someone I could always turn to for solid

advice. I'm grateful to **Keiichi**, **Kanno**, **Hiromu**, and **Felix** for their friendship, support, and the countless ways they contributed to making our lab environment both productive and enjoyable. I want to thank **Sveta**, an extremely kind person who was always happy to help and brought such warmth to our lab community.

Among past lab members, I want to thank **Tetsuya**, who was incredibly supportive during the early days of my PhD, and **Arie**, with whom I had countless engaging discussions that shaped my thinking. I'm grateful to **Facundo Romani**, my collaborator and co-author on my second publication, for his partnership and contributions to our shared research.

I want to thank the **VBC PhD programme**, including **Chris**, **Eva**, and **Chiara**, for providing not only amazing facilities and resources but also fostering a vibrant campus environment filled with inspiring people and collaborative opportunities. I'm grateful to my PhD TAC members, **Liam Dolan** and **Christa Bücker**, for their valuable guidance, insightful feedback, and support throughout my doctoral journey.

I want to thank my amazing girlfriend **Yukti**, for her unwavering love, patience, and support throughout this journey. Her incredible sense of humor kept me laughing through everything. Thank you for going to the same breakfast places over and over again just because I die in trying new places, for planning amazing vacations that were always perfectly enjoyable, and for making the entire PhD journey not just bearable but extremely fun. Here is a message for you: जय कढ़ी चावल! I also want to thank **Anastasia**, Yukti's friend, with whom together we shared many fun weekends in Graz. Thank you both for organizing Diwali party this year! I want to also thank my flat mates **Karim** and **Chako** for making home a place of comfort and friendship, and for all the shared moments that made life in Vienna so much more enjoyable.

Finally, I want to express my deepest gratitude to my mother, whose unconditional love and unwavering belief in me have been my greatest source of strength throughout this journey.

With deepest appreciation,

Vikas

November 25, 2025, Vienna

Abstract

Chromatin organization shapes gene expression through histone variants and modifications to create fine-tuned regulatory environments readable by transcriptional effectors like transcription factors (TFs). These regulatory environments are organized into chromatin states: the most likely combinations of histone variants and modifications that define genomic regions. Complex interactions between TFs and chromatin states determine gene expression patterns. While well-characterized in animals, chromatin states and their interplay with TFs in land plants remained poorly understood. This thesis presents a comprehensive analysis of chromatin organization and its interactions with TFs in land plants.

Using ChromHMM with 27 histone modifications and variants, we identified 26 distinct chromatin states in *Arabidopsis thaliana*. Through systematic analyses, we demonstrated that H2A variants are major determinants of chromatin state identity: H2A.W marks constitutive heterochromatin, H2A.Z defines facultative heterochromatin, and H2A/H2A.X enrich euchromatin, challenging the view that histone modifications primarily drive chromatin organization.

Comparative analysis between *Arabidopsis* and *Marchantia* (450+ million years diverged) showed remarkable chromatin state conservation, including H2A variant-domain relationships, indicating early establishment of fundamental regulatory mechanisms in land plants.

We developed a TF activity scoring system integrating binding data with gene co-expression to distinguish functional TF binding occurrences. This revealed three TF groups based on chromatin preferences and identified potential candidate pioneer transcription factors binding repressive chromatin states.

This work establishes chromatin states as a conserved plant regulatory framework. Histone variants serve as molecular landmarks organizing modifications into functional domains, while TFs navigate through specialized binding preferences. These findings provide comprehensive understanding of plant gene regulation, revealing evolutionary conservation of chromatin-mediated transcriptional control in land plants.

Zusammenfassung

Die Chromatinorganisation prägt die Genexpression durch Histonvarianten und -modifikationen, um fein abgestimmte regulatorische Umgebungen zu schaffen, die von transkriptionellen Effektoren wie Transkriptionsfaktoren (TFs) gelesen werden können. Diese regulatorischen Umgebungen sind in Chromatinzustände organisiert: die wahrscheinlichsten Kombinationen von Histonvarianten und -modifikationen, die genomische Regionen definieren. Komplexe Wechselwirkungen zwischen TFs und Chromatinzuständen bestimmen Genexpressionsmuster. Während diese bei Tieren gut charakterisiert sind, blieben Chromatinzustände und ihr Zusammenspiel mit TFs bei Landpflanzen schlecht verstanden. Diese Dissertation präsentiert eine umfassende Analyse der Chromatinorganisation und ihrer Wechselwirkungen mit TFs bei Landpflanzen.

Unter Verwendung von ChromHMM mit 27 Histonmodifikationen und -varianten identifizierten wir 26 verschiedene Chromatinzustände in Arabidopsis thaliana. Durch systematische Analysen zeigten wir, dass H2A-Varianten wichtige Determinanten der Chromatinzustandsidentität sind: H2A.W markiert konstitutives Heterochromatin, H2A.Z definiert fakultatives Heterochromatin, und H2A/H2A.X reichern Euchromatin an, was die Ansicht in Frage stellt, dass Histonmodifikationen primär die Chromatinorganisation antreiben.

Eine vergleichende Analyse zwischen Arabidopsis und Marchantia (über 450 Millionen Jahre divergiert) zeigte eine bemerkenswerte Konservierung der Chromatinzustände, einschließlich der H2A-Varianten-Domänen-Beziehungen, was auf eine frühe Etablierung fundamentaler regulatorischer Mechanismen bei Landpflanzen hinweist.

Wir entwickelten ein TF-Aktivitätsbewertungssystem, das Bindungsdaten mit Genkoexpression integriert, um funktionelle TF-Bindungsereignisse zu unterscheiden. Dies enthüllte drei TF-Gruppen basierend auf Chromatinpräferenzen und identifizierte potenzielle Kandidaten für Pionier-Transkriptionsfaktoren, die repressive Chromatinzustände binden.

Diese Arbeit etabliert Chromatinzustände als ein konserviertes pflanzliches regulatorisches Framework. Histonvarianten dienen als molekulare Orientierungspunkte, die Modifikationen in funktionelle Domänen organisieren, während TFs durch spezialisierte Bindungspräferenzen navigieren. Diese Erkenntnisse bieten ein

umfassendes Verständnis der pflanzlichen Genregulation und enthüllen die evolutionäre Konservierung chromatin-vermittelter transkriptioneller Kontrolle bei Landpflanzen.

Table of Contents

ACKNOWLEDGEMENTS II
ABSTRACT V
ZUSAMMENFASSUNG VI

1. CHROMATIN ORGANIZATION AND FUNCTION 1
2. CHROMATIN STATES 25
3. TRANSCRIPTION FACTOR–CHROMATIN INTERACTION 34
4. LIST OF PUBLICATIONS 46
5. PUBLICATION I 47
6. PUBLICATION II 87
7. DISCUSSION 143

EPILOGUE 156

Chapter 1

1. Chromatin Organization and Function

Contents

1.1 INTRODUCTION 1

1.2 CHROMATIN: FROM BARRIER TO REGULATORY PLATFORM 2

1.3 TYPES OF CHROMATIN: EUCHROMATIN AND HETEROCHROMATIN 2

1.4 MOLECULAR DETERMINANTS OF CHROMATIN FUNCTION: HISTONE PTMS 3

1.4.1 HISTONE PTMS ASSOCIATED WITH EUCHROMATIN 4

H3K36 methylation *4*

H3K4 methylation *5*

Histone Acetylations *5*

1.4.2 HISTONE PTMS ASSOCIATED WITH FACULTATIVE HETEROCHROMATIN 6

H3K27me3 and Polycomb Repressive Complex 2 (PRC2) *6*

H2AKub and Polycomb Repressive Complex 1 (PRC1) *6*

1.4.3 HISTONE PTMS ASSOCIATED WITH CONSTITUTIVE HETEROCHROMATIN 7

1.5 MOLECULAR DETERMINANTS OF CHROMATIN FUNCTION: HISTONE VARIANTS 8

1.5.1 VARIANTS OF HISTONE H3 9

H3.3 and H3.1 ..... *9*

CenH3 or CENP-A *10*

1.5.2 VARIANTS OF HISTONE H2A 11

H2A.Z *11*

H2A.X *12*

H2A.W and macroH2A *12*

1.5.3 VARIANTS OF HISTONE H2B AND H4 13

1.6 CONCLUSION 13

BIBLIOGRAPHY 14

1.1 Introduction

Genomes carry the instructions to build an organism, encoding not only the blueprint for cellular components but also the regulatory logic that governs gene expression across development and in response to environmental cues. The evolution of the eukaryotic nucleus, which compartmentalizes these genomes within a membrane-bound organelle, represents one of the most transformative events in the history of life, fundamentally reshaping how genetic information is organized, accessed, and regulated. The compartmentalization of genomes created a fundamental biophysical challenge of packaging meters of DNA into a space measuring only micrometers in diameter. Central to this solution are histones, a family of small, basic proteins that are among the most highly conserved in eukaryotes, characterized by their positive charge, which enables tight association with the negatively charged phosphate backbone of DNA. Eukaryotic chromatin contains four core histone proteins: H2A, H2B, H3, and H4. Two copies each of these core histones organize into an octamer (Kornberg, 1974) that wraps

around ~147 base pairs (bp) of DNA to form a nucleosome (Luger et al., 1997). Adjacent nucleosomes are connected by linker DNA, whose length varies between 20 and 80 bp depending on genomic context and can be associated with the linker histone H1. This arrangement of nucleosomes connected by linker DNA forms a 10 nm fiber, often described as a 'beads-on-a-string' structure, which represents the first level of chromatin compaction (Kornberg & Klug, 1981).

1.2 Chromatin: From Barrier to Regulatory Platform

While the nucleosome represents a robust solution to the packaging problem, early *in vitro* studies revealed a fundamental challenge: nucleosomes pose a significant barrier to transcription machinery. Reconstituted nucleosomes assembled on DNA templates strongly inhibit the progression of RNA polymerase II (Pol II), with the histone octamer creating a physical obstacle that must be overcome for transcription to proceed (Lorch et al., 1987; Wasylyk et al., 1979). This observation raised a critical question: how do eukaryotic cells reconcile the need for genome compaction with the requirement for dynamic access to genetic information?

The answer lies in chromatin—a complex of proteins, RNA, and DNA that constitutes the physiological state of the genome, integrating multiple layers of information including histone variants and post-translational modifications (PTMs), DNA methylation, nucleosome positioning and remodeling complexes, and proteins and RNAs involved in higher-order chromatin interactions. Collectively, these features establish chromatin as a sophisticated regulatory platform that determines when, where, and to what extent genetic information is accessed.

1.3 Types of Chromatin: Euchromatin and Heterochromatin

The term "chromatin" itself has historical roots in microscopy, coined by Walther Flemming in the 1880s to describe the readily stainable material observed within the nucleus (Flemming, 1882). This staining property reflected fundamental differences in chromatin organization that would later prove central to understanding gene regulation. In the 1920s, Emil Heitz built upon these observations, distinguishing between two distinct types of chromatin based on their staining intensity and behavior through the cell cycle (Heitz, 1928). He termed the lightly stained, less condensed regions "euchromatin" and

the densely stained, highly condensed regions "heterochromatin."

These early morphological distinctions have since been refined through molecular and genomic approaches, revealing that the structural differences observed by Heitz correspond to functionally distinct chromatin types. Euchromatin represents open, accessible chromatin that is generally associated with active transcription and permissive to the binding of Transcription Factors (TFs) and regulatory machinery. In contrast, heterochromatin can be further subdivided into two functionally distinct categories. Facultative heterochromatin encompasses genomic regions that are currently transcriptionally repressed but retain the potential to become active in response to developmental or environmental signals. Constitutive heterochromatin, on the other hand, maintains stable repression of genomic elements that must remain silenced, including Transposable Elements (TEs) and repetitive sequences.

1.4 Molecular Determinants of Chromatin Function: Histone PTMs

The crystal structure of the nucleosome core particle, solved by Luger and colleagues in 1997, provided crucial insights into how histones interact with DNA and revealed key features relevant to chromatin regulation (Luger et al., 1997). While the globular domains of the core histones form a structured octamer around which DNA wraps, the N-terminal and C-terminal tails of the histones protrude outward from the nucleosome core. These histone tails are unstructured, highly accessible, and enriched in basic amino acids particularly Lysines (K), Arginines (A), and Serines (S), that serve as substrates for a diverse array of chemical modifications.

A critical breakthrough in understanding chromatin regulation came in 1964 when Vincent Allfrey and colleagues discovered that histone acetylation correlates with transcriptional activity (Allfrey et al., 1964). Using isolated calf thymus nuclei, they demonstrated that actively transcribing chromatin was enriched in acetylated histones, while transcriptionally silent regions contained predominantly unacetylated histones. This seminal observation established that chemical modifications of histones could modulate chromatin function and gene expression, fundamentally transforming our understanding of transcriptional regulation. The discovery that histone acetylation is a reversible process, catalyzed by histone acetyltransferases (HATs) and histone deacetylases (HDACs) (Turner, 2000),

further reinforced the concept that chromatin structure is dynamically regulated in response to cellular signals.

The discovery of histone methylation followed a similar trajectory. In the late 1960s, Murray and colleagues identified methylated lysine residues in histones, though the functional significance of this modification remained unclear for decades (Murray, 1964; Paik & Kim, 1967). Unlike acetylation, which uniformly correlates with transcriptional activation, histone methylation proved to be more complex, with different methylation sites associated with either gene activation or repression depending on their location and degree of methylation. A major conceptual advance came with the identification of the first histone methyltransferase, SUV39H1, which specifically methylates histone H3 at lysine 9 (H3K9), a modification associated with heterochromatin formation and gene silencing (Rea et al., 2000). Subsequent research identified additional histone Post-translational Modifications (PTMs), including phosphorylation, ubiquitination, SUMOylation, and others, each with distinct regulatory roles.

These discoveries established that histone PTMs are catalyzed by specific enzymatic machinery and can exert distinct functional outcomes on chromatin structure and gene expression. This enzymatic machinery is commonly classified into three functional categories: "writers" that deposit modifications (such as HATs and histone methyltransferases), "erasers" that remove them (such as HDACs and histone demethylases), and "readers" that recognize and bind specific modifications to recruit downstream effector complexes. In the following sections, prominent histone PTMs associated with distinct chromatin types are discussed in detail.

1.4.1 Histone PTMs Associated with Euchromatin

H3K36 methylation

Histone H3 lysine 36 methylation (H3K36me) was first identified in the early 2000s as a modification enriched within the coding regions of actively transcribed genes (Bannister et al., 2005; Krogan et al., 2003). The discovery that Set2, a histone methyltransferase in yeast, physically interacts with the elongating form of RNA polymerase II (Pol II) provided crucial mechanistic insight into how this modification is deposited (Krogan et al., 2003). Set2 specifically recognizes the phosphorylated C-terminal domain (CTD) of RNA polymerase II (Pol II), ensuring that H3K36 methylation is deposited co-transcriptionally on gene bodies. This coupling between

transcription and H3K36 methylation establishes the modification as a hallmark of active transcription, with H3K36me3 (trimethylation) showing particularly strong enrichment in the 3' regions of actively transcribed genes.

H3K4 methylation

Histone H3 lysine 4 methylation (H3K4me) emerged as one of the earliest characterized histone methylation marks associated with transcriptional activation. The modification has deep evolutionary roots, with the Trithorax group (TrxG) proteins in *Drosophila* being among the first identified H3K4 methyltransferases, functioning as antagonists to Polycomb-mediated repression (See section 1.4.2) and maintaining active gene expression during development (Schuettengruber et al., 2007). In *Saccharomyces cerevisiae*, the discovery of Set1, the catalytic subunit of the COMPASS (Complex of Proteins Associated with Set1) complex, established the enzymatic machinery responsible for H3K4 methylation (Miller et al., 2001; Roguev et al., 2001). Unlike H3K36me3, which marks gene bodies, H3K4me3 shows a distinct distribution pattern, with prominent enrichment at transcription start sites (TSSs) and promoters of actively transcribed genes (Santos-Rosa et al., 2002; Schneider et al., 2004). Mechanistically, H3K4me3 can recruit the TFIID complex through its TAF3 subunit in some species, facilitating transcription initiation, though this interaction is not universally conserved across all eukaryotes.

In addition to H3K4me3, the monomethylated form (H3K4me1) exhibits distinct genomic distributions depending on the organism. In mammals, H3K4me1 together with H3K27ac serves as a hallmark of active enhancers (Oya et al., 2022). However, in *Arabidopsis*, which lacks the extensive enhancer landscapes characteristic of mammalian genomes, H3K4me1 is predominantly found within gene bodies, suggesting organism-specific functional roles for this modification.

Histone Acetylations

As discussed earlier, the discovery by Allfrey and colleagues in 1964 established the foundational link between histone acetylation and transcriptional activity (Allfrey et al., 1964). Acetylation involves the addition of an acetyl group to lysine residues, neutralizing their positive charge and thereby weakening histone-DNA interactions. This charge neutralization is thought to promote a more open chromatin structure that facilitates access by transcriptional machinery. Among the various acetylation marks, H3K9ac and H3K14ac are particularly

enriched at active promoters and enhancers, serving as robust indicators of transcriptionally active chromatin (Karmodiya et al., 2012). The dynamic nature of histone acetylation is maintained by the opposing activities of histone acetyltransferases (HATs), which deposit acetyl groups, and histone deacetylases (HDACs), which remove them. This reversible modification system enables rapid modulation of chromatin accessibility in response to developmental and environmental cues.

1.4.2 Histone PTMs Associated with Facultative Heterochromatin

H3K27me3 and Polycomb Repressive Complex 2 (PRC2)

Histone H3 lysine 27 trimethylation (H3K27me3) was identified as a repressive chromatin mark deposited by Polycomb Repressive Complex 2 (PRC2), with the catalytic activity residing in the Enhancer of Zeste homolog (E(z) in *Drosophila*, EZH2 in mammals) (Cao et al., 2002; Czermin et al., 2002; Müller et al., 2002). The discovery of H3K27me3 provided a molecular explanation for Polycomb-mediated gene silencing, a phenomenon first observed genetically in *Drosophila* decades earlier. In mammals and flowering plants, H3K27me3 plays a critical role in development, mediating the stable but reversible repression of lineage-inappropriate genes and developmental regulators (X. Zhang et al., 2007). Unlike the permanent silencing associated with constitutive heterochromatin, H3K27me3-marked regions in these organisms retain the capacity to be reactivated in response to developmental signals, establishing this modification as the defining hallmark of facultative heterochromatin.

However, recent studies in early-diverging land plants have revealed predominant association of H3K27me3 with transposable elements (TEs) and marking constitutive heterochromatin, suggesting an evolutionary specialization of this mark as a facultative heterochromatin regulator in higher plants and animals (Hisanaga et al., 2023).

H2AKub and Polycomb Repressive Complex 1 (PRC1)

Histone H2A lysine 119 monoubiquitination (H2AK119ub, also known as H2AK121ub in *Arabidopsis*) was identified as the modification catalyzed by Polycomb Repressive Complex 1 (PRC1), with the RING1A/B subunits providing the E3 ubiquitin ligase activity (de Napoles et al., 2004; Wang et al., 2004). Initially, PRC1 was thought to function downstream of PRC2,

with H3K27me3 serving as a recruitment signal for PRC1 through the chromodomain of the Polycomb protein. However, subsequent studies revealed a more complex relationship between these two complexes. While canonical PRC1 recruitment does depend on PRC2-mediated H3K27me3, variant PRC1 complexes can be recruited independently of PRC2 and can even facilitate PRC2 recruitment in some contexts (X. Zhang et al., 2007). Additionally, H2AK121ub is not exclusively associated with repressed chromatin; it has also been detected on the promoters of some actively transcribing genes, suggesting both PRC2-dependent and PRC2-independent roles for this modification in chromatin regulation. The intricate interplay between PRC1 and PRC2, and the context-dependent functions of H2AK121ub, are discussed in greater detail later in this thesis.

1.4.3 Histone PTMs Associated with Constitutive Heterochromatin

Histone H3 lysine 9 methylation (H3K9me) emerged as the canonical mark of constitutive heterochromatin following the identification of SUV39H1 as the first histone methyltransferase capable of catalyzing H3K9 methylation (Rea et al., 2000). In mammals, H3K9me3 functions as a binding platform for Heterochromatin Protein 1 (HP1), which recognizes this modification through its chromodomain. The multivalent interactions between HP1 proteins and H3K9me3-marked nucleosomes drive liquid-liquid phase separation, resulting in the formation of condensed heterochromatin domains that exclude transcriptional machinery and maintain stable gene silencing (Larson et al., 2017; Strom et al., 2017).

In *Arabidopsis*, however, H3K9me2 follows a distinct mechanistic pathway. Rather than recruiting HP1 homologs, H3K9me2 is recognized by the plant-specific DNA methyltransferases CHROMOMETHYLASE 2 (CMT2) and CHROMOMETHYLASE 3 (CMT3), which deposit DNA methylation in the CHG context (where H represents A, T, or C) (Du et al., 2012; Stroud et al., 2014). This DNA methylation, in turn, recruits the H3K9 methyltransferase SUVH family proteins (such as SUVH4/KYP) through their SRA domains, which recognize methylated DNA and deposit H3K9me2, thereby establishing a self-reinforcing feedback loop between DNA methylation and H3K9 methylation (Johnson et al., 2002).

Beyond H3K9 methylation, other histone modifications contribute to constitutive heterochromatin identity, including H3K27me1 and H4K20me1. In flowering plants, H3K27me1 is deposited by the *Arabidopsis* Trithorax-Related proteins

ATXR5 and ATXR6, which specifically target heterochromatic regions and play crucial roles in maintaining genome stability (Jacob et al., 2009; Raynaud et al., 2006). In contrast, the enzyme responsible for depositing H4K20me1 in plants remains unknown.

Collectively, these histone PTMs establish distinct chromatin environments that define euchromatin, facultative heterochromatin, and constitutive heterochromatin, each characterized by specific combinations of modifications that dictate accessibility, transcriptional activity, and epigenetic stability across the genome.

1.5 Molecular Determinants of Chromatin Function: Histone Variants

The diversity in chromatin organization is further expanded through the incorporation of histone variants—non-allelic isoforms of the canonical core histones that differ in their amino acid sequences. While the canonical histones (H2A, H2B, H3, and H4) are expressed primarily during S phase and deposited during DNA replication, histone variants are typically expressed throughout the cell cycle and can be incorporated into chromatin in a replication-independent manner. Some core histone variants have also been shown to be deposited in a cell cycle dependent manner including the centromeric histone H3.

The existence of histone variants was first recognized in the 1970s and 1980s through biochemical fractionation studies that identified histones with distinct electrophoretic mobilities and amino acid compositions (Franklin & Zweidler, 1977; West & Bonner, 1980). However, the functional significance of these variants remained largely enigmatic until molecular and genomic approaches revealed their specialized roles in chromatin organization and gene regulation.

The functional importance of histone variants became apparent through two key observations. First, structural and biochemical analyses demonstrated that the amino acid differences between variants and their canonical counterparts confer unique properties to nucleosomes, including altered stability, modified DNA-histone interactions, and differential recruitment of chromatin-associated factors. Second, genome-wide mapping studies revealed that histone variants are not randomly distributed but instead show remarkable specificity for specific chromatin contexts, with different variants localizing to distinct genomic regions such as actively transcribed genes, regulatory elements, or specialized

chromosomal structures (Loppin & Berger, 2020).

The discovery of these specific localization patterns prompted investigations into the mechanisms underlying variant deposition, leading to the identification of dedicated histone chaperone complexes that recognize and deposit specific variants at their target genomic locations. Additionally, ATP-dependent chromatin remodeling complexes have been shown to facilitate the exchange or removal of specific histone variants, further contributing to the dynamic regulation of variant-containing nucleosomes (Clapier et al., 2017). These distinct biochemical properties, localization patterns, and dedicated deposition machinery enable histone variants to establish functionally specialized chromatin domains essential for chromatin regulation. In the following sections, the major histone variants and their specific roles in chromatin organization are discussed in detail.

1.5.1 Variants of Histone H3

Histone H3 exists in multiple variant forms that differ in their expression patterns, genomic localization, and functional roles. The major H3 variants include the replication-coupled H3.1, the replication-independent H3.3, and the highly specialized centromeric variant CENP-A/CenH3.

H3.3 and H3.1

The first histone H3 variants to be characterized were H3.1 and H3.3, initially distinguished in the early 1980s based on their differential expression patterns and amino acid sequences (Wells et al., 1987; Wells & Kedes, 1985). H3.1 represents the replication-coupled canonical histone, expressed during S phase and deposited behind the replication fork to assemble nucleosome on newly synthesized DNA. H3.3 differs from H3.1 by only a few amino acids (four in mammals and *Arabidopsis*, five in Drosophila) but exhibits replication-independent expression and deposition, allowing its incorporation throughout the cell cycle.

In mammals, genome-wide mapping studies revealed distinct localization patterns for these variants. H3.3 is enriched at actively transcribed gene bodies, promoters, enhancers, and regulatory elements, correlating with regions of high transcriptional activity and nucleosome turnover (Ahmad & Henikoff, 2002; Mikkelsen et al., 2007). By contrast, H3.1 is more uniformly distributed across the genome, reflecting its role in bulk chromatin assembly during DNA replication.

In *Arabidopsis*, the localization patterns of H3 variants show both similarities and notable differences compared to mammals. *Arabidopsis* encodes three

H3.3-like variants (HTR4, HTR5, and HTR8), multiple H3.1-like variants (HTR1, HTR2, HTR3, HTR9, HTR13) and additional variant H3.6, H3.10 H3.11, H3.14 and H3.15 expressed either in a cell specific manner or in response to a stimulus (Ingouff et al., 2010a; Nunez-Vazquez et al., 2025). Like mammals, *Arabidopsis* H3.3 variants are enriched at actively transcribed genes towards the 3' end of gene body that also associate with gene body CpG methylation and show replication-independent deposition (Stroud et al., 2014; Wollmann et al., 2012). However, studies have also revealed H3.1 enrichment at heterochromatic regions and transposable elements in *Arabidopsis*, suggesting additional specialized roles in transcriptional silencing (Ingouff et al., 2010b; Stroud et al., 2012). This observation is partially explained by the fact that H3.1 serves as the preferred substrate for H3K27me1 deposition by ATXR5 and ATXR6, linking variant identity to heterochromatin maintenance (Jacob et al., 2009; Stroud et al., 2012).

The distinct localization patterns of H3.1 and H3.3 are achieved through variant-specific deposition machinery. H3.1 is deposited by the Chromatin Assembly Factor 1 (CAF-1) complex, which associates with the replication machinery through interactions with PCNA (Proliferating Cell Nuclear Antigen), ensuring replication-coupled nucleosome assembly (Smith & Stillman, 1989; Verreault et al., 1996). H3.3, on the other hand, is deposited by dedicated chaperone complexes: the HIRA complex mediates H3.3 deposition at genic regions and regulatory elements in a transcription-coupled manner, while the DAXX-ATRX complex deposits H3.3 at telomeres and pericentric heterochromatin (Goldberg et al., 2010; Tagami et al., 2004). In *Arabidopsis*, similar mechanisms operate, with the HIRA complex facilitating H3.3 deposition at active genes (Duc et al., 2017; Nie et al., 2014; Stroud et al., 2012).

CenH3 or CENP-A

The centromeric histone H3 variant, known as CENP-A in mammals and CenH3 in plants, represents one of the most divergent histone variants. CENP-A was first identified in the late 1980s as a centromere-specific protein (Earnshaw & Rothfield, 1985; Palmer et al., 1987). Unlike H3.1 and H3.3, CENP-A shows substantial sequence divergence from canonical H3 (Dawe & Henikoff, 2006) and localizes exclusively to centromeric chromatin, where it is essential for kinetochore assembly and chromosome segregation (Black & Cleveland, 2011; Talbert et al., 2002).

Recent studies in *Arabidopsis* have revealed unexpected connections between H3 variant deposition and other chromatin components. Loss of CAF-1 results not only in depletion of H3.1 but also in loss of the plant-specific H2A variant H2A.W and reduction of H3K9me2 (Benoit et al., 2019), suggesting a functional link between H3 and H2A variant deposition mechanisms and their roles in heterochromatin maintenance.

1.5.2 Variants of Histone H2A

Histone H2A is the most diverse of the core histones, with numerous variants that have evolved specialized functions in chromatin regulation. Unlike H3, which has relatively few variants, H2A variants show greater sequence divergence and functional specialization across eukaryotes. The major H2A variants include H2A.Z, which regulates transcription and nucleosome stability; H2A.X, which plays a critical role in DNA damage response; macroH2A, involved in transcriptional repression and X-chromosome inactivation in mammals; and the plant-specific H2A.W, associated with heterochromatin and transposable element silencing. Each of these variants confers distinct biochemical properties to nucleosomes and localizes to specific genomic regions, enabling precise control over chromatin function.

H2A.Z

H2A.Z was among the first histone variants to be identified, initially discovered in the 1980s as a highly conserved H2A variant present across eukaryotes (West & Bonner, 1980). Structurally, H2A.Z is distinguished by an extended acidic patch on the nucleosome surface, which alters nucleosome-nucleosome interactions and affects chromatin compaction (Ausió & Abbott, 2002; Marzluff et al., 2002; Osakabe et al., 2018). The functional importance of H2A.Z is underscored by its essentiality in many mammals and the severe growth defects observed upon its loss in yeast and *Arabidopsis*, indicating critical roles in development and cellular function (Carr et al., 1994; Faast et al., 2001; March-Díaz et al., 2008).

Genome-wide mapping studies revealed that H2A.Z shows preferential localization at the +1 nucleosome immediately downstream of transcription start sites, where it influences transcriptional regulation and nucleosome positioning (Guillemette et al., 2005; H. Zhang et al., 2005).H2A.Z deposition is mediated by the SWR1 complex (SWR1-C in yeast, SRCAP or p400 in mammals), an ATP-dependent chromatin remodeling complex that catalyzes the exchange of canonical H2A-H2B dimers with H2A.Z-H2B

dimers (Kobor et al., 2004; Mizuguchi et al., 2004).

In *Arabidopsis*, H2A.Z similarly localizes to the +1 nucleosomes of actively expressed genes. However, *Arabidopsis* H2A.Z also exhibits a distinctive localization pattern at stress-responsive genes, where it co-occupies gene bodies together with the repressive H3K27me3 mark, suggesting a role in maintaining these genes in facultative heterochromatin state that can be rapidly activated upon environmental stimuli (Coleman-Derr & Zilberman, 2012; Zilberman et al., 2008). Additionally, H2A.Z localization shows an anticorrelation with CpG DNA methylation, indicating that these two chromatin features are largely mutually exclusive in the *Arabidopsis* genome (Zilberman et al., 2008).

H2A.X

H2A.X was identified in the 1990s as a variant of H2A distinguished by a unique C-terminal extension containing an SQ motif (Mannironi et al., 1989; Redon et al., 2002). This seemingly minor sequence difference confers a specialized and critical function in the DNA damage response. Upon the formation of DNA double-strand breaks (DSBs), the serine residue within the SQ motif is rapidly phosphorylated by the PI3K-related kinases ATM, ATR, and DNA-PKcs, generating the phosphorylated form known as γH2A.X (Burma et al., 2001; Rogakou et al., 1998). This phosphorylation event serves as an early signal for DNA damage, recruiting DNA repair machinery and checkpoint proteins to the sites of DSBs. The spreading of γH2A.X over megabase-sized chromatin domains flanking the break site amplifies the damage signal and facilitates the assembly of repair complexes, making H2A.X essential for maintaining genome stability (Rogakou et al., 1999).

H2A.W and macroH2A

Specialized H2A variants have evolved in plants and animals that associate with transcriptional repression and heterochromatin. MacroH2A was first identified in mammals in the 1990s as an unusually large histone variant, approximately three times the size of canonical H2A (Pehrson & Fried, 1992). MacroH2A possesses a large C-terminal macro domain that extends outward from the nucleosome core and is thought to mediate interactions with chromatin-modifying complexes (Karras et al., 2005; Pehrson & Fuji, 1998). MacroH2A localizes to the inactive X chromosome in female mammals and other transcriptionally repressed regions, where it contributes to stable heterochromatin maintenance (Changolkar & Pehrson, 2006; Costanzi & Pehrson, 1998).

Similarly in plants, H2A.W was identified as a specialized variant that localizes to constitutive heterochromatin and transposable elements (Yelagandula et al., 2014). H2A.W contains a characteristic KSPKK motif in its C-terminal tail, which is essential for its heterochromatic localization and function (Yelagandula et al., 2014). In *Arabidopsis*, H2A.W shows strong enrichment at transposable elements and constitutive heterochromatin, co-localizing with repressive marks such as H3K9me2 and DNA methylation (Bourguet et al., 2021; Yelagandula et al., 2014).

While histone chaperones such as FACT and NAP1 are known to deposit H2A-H2B dimers during replication in a variant-independent manner, the mechanisms governing replication-independent deposition of macroH2A and H2A.W remain poorly understood. Despite their independent evolutionary origins, both macroH2A and H2A.W play critical roles in heterochromatin regulation and genome stability (Jenness et al., 2018; Osakabe et al., 2021).

Together with other histone variants, macroH2A and H2A.W exemplify how lineage-specific chromatin components have evolved to address common regulatory challenges in genome organization.

1.5.3 Variants of Histone H2B and H4

H2B variants were first identified through biochemical studies and have been shown to exhibit highly tissue-specific expression and functions in mammals. In *Arabidopsis*, tissue-specific expression patterns have also been observed, with ten H2B variants encoded in the genome (Jiang et al., 2020). Among these, H2B.3 is notably depleted from heterochromatic regions, while H2B.2 shows enrichment at heterochromatin, suggesting distinct roles in chromatin organization (Buttress et al., 2022; Jiang et al., 2020). However, the precise mechanisms by which these variants regulate transcription and chromatin function remain largely unknown.

Histone H4 is the most conserved of all core histones, and accordingly, H4 variants are exceedingly rare. To date, very few functionally distinct H4 variants have been characterized including the H4.G in humans and others in some parasitic species, reflecting the stringent structural constraints on H4 function within the nucleosome core.

1.6 Conclusion

The organization of eukaryotic genomes into chromatin represents a sophisticated regulatory system that extends far beyond simple DNA

packaging. Through the combinatorial action of histone post-translational modifications, histone variants, and their associated deposition and remodeling machinery, chromatin establishes functionally distinct domains that participate in the regulation of gene expression, maintain genome stability, and enable cellular responses to developmental and environmental signals. The interplay between these chromatin features—from the activating marks of euchromatin to the repressive landscapes of facultative and constitutive heterochromatin—creates a dynamic regulatory framework that is both stable enough to maintain cellular identity and flexible enough to permit transcriptional reprogramming. However, chromatin does not function in isolation. The accessibility and regulatory potential of chromatin states are ultimately interpreted through their interactions with transcription factors, which must navigate and respond to the chromatin landscape to execute gene regulatory programs. In land plants, which have evolved unique chromatin features and regulatory strategies, understanding how chromatin states interface with transcription factor binding and activity represents a critical frontier in deciphering the logic of gene regulation. The following chapter explores these interactions, examining how chromatin organization and transcription factor function are integrated to control gene expression in plant genomes.

Bibliography

Ahmad, K., & Henikoff, S. (2002). The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. *Molecular Cell*, *9*(6), 1191–1200. https://doi.org/10.1016/s1097-2765(02)00542-7

Allfrey, V. G., Faulkner, R., & Mirsky, A. E. (1964). Acetylation and methylation of histones and their possible role in the regulation of rna synthesis*. *Proceedings of the National Academy of Sciences*, *51*(5), 786–794. https://doi.org/10.1073/pnas.51.5.786

Ausió, J., & Abbott, D. W. (2002). The many tales of a tail: Carboxyl-terminal tail heterogeneity specializes histone H2A variants for defined chromatin function. *Biochemistry*, *41*(19), 5945–5949. https://doi.org/10.1021/bi020059d

Bannister, A. J., Schneider, R., Myers, F. A., Thorne, A. W., Crane-Robinson, C., & Kouzarides, T. (2005). Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. *The Journal of Biological Chemistry*, *280*(18), 17732–17736. https://doi.org/10.1074/jbc.M500796200

Benoit, M., Simon, L., Desset, S., Duc, C., Cotterell, S., Poulet, A., Le Goff, S., Tatout, C., & Probst, A. V. (2019). Replication-coupled histone H3.1 deposition determines nucleosome composition and heterochromatin dynamics during Arabidopsis seedling development. *The New Phytologist*, *221*(1), 385–398. https://doi.org/10.1111/nph.15248

Black, B. E., & Cleveland, D. W. (2011). Epigenetic centromere propagation and the nature of CENP-a nucleosomes. *Cell*, *144*(4), 471–479. https://doi.org/10.1016/j.cell.2011.02.002

Bourguet, P., Picard, C. L., Yelagandula, R., Pélissier, T., Lorković, Z. J., Feng, S., Pouch-Pélissier, M.-N., Schmücker, A., Jacobsen, S. E., Berger, F., & Mathieu, O. (2021). The histone variant H2A.W and linker histone H1 co-regulate heterochromatin accessibility and DNA methylation. *Nature Communications*, *12*(1), 2683. https://doi.org/10.1038/s41467-021-22993-5

Burma, S., Chen, B. P., Murphy, M., Kurimasa, A., & Chen, D. J. (2001). ATM phosphorylates histone H2AX in response to DNA double-strand breaks. *The Journal of Biological Chemistry*, *276*(45), 42462–42467. https://doi.org/10.1074/jbc.C100466200

Buttress, T., He, S., Wang, L., Zhou, S., Saalbach, G., Vickers, M., Li, G., Li, P., & Feng, X. (2022). Histone H2B.8 compacts flowering plant sperm through chromatin phase separation. *Nature*, *611*(7936), 614–622. https://doi.org/10.1038/s41586-022-05386-6

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R. S., & Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science (New York, N.Y.)*, *298*(5595), 1039–1043. https://doi.org/10.1126/science.1076997

Carr, A. M., Dorrington, S. M., Hindley, J., Phear, G. A., Aves, S. J., & Nurse, P. (1994). Analysis of a histone H2A variant from fission yeast: Evidence for a role in chromosome stability. *Molecular and General Genetics MGG*, *245*(5), 628–635. https://doi.org/10.1007/BF00282226

Changolkar, L. N., & Pehrson, J. R. (2006). macroH2A1 histone variants are depleted on active genes but concentrated on the inactive X chromosome. *Molecular and Cellular Biology*, *26*(12), 4410–4420. https://doi.org/10.1128/MCB.02258-05

Clapier, C. R., Iwasa, J., Cairns, B. R., & Peterson, C. L. (2017). Mechanisms of

action and regulation of ATP-dependent chromatin-remodelling complexes. *Nature Reviews Molecular Cell Biology*, *18*(7), 407–422. https://doi.org/10.1038/nrm.2017.26

Coleman-Derr, D., & Zilberman, D. (2012). Deposition of histone variant H2A.Z within gene bodies regulates responsive genes. *PLoS Genetics*, *8*(10), e1002988. https://doi.org/10.1371/journal.pgen.1002988

Costanzi, C., & Pehrson, J. R. (1998). Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. *Nature*, *393*(6685), 599–601. https://doi.org/10.1038/31275

Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A., & Pirrotta, V. (2002). Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell*, *111*(2), 185–196. https://doi.org/10.1016/s0092-8674(02)00975-3

Dawe, R. K., & Henikoff, S. (2006). Centromeres put epigenetics in the driver's seat. *Trends in Biochemical Sciences*, *31*(12), 662–669. https://doi.org/10.1016/j.tibs.2006.10.004

de Napoles, M., Mermoud, J. E., Wakao, R., Tang, Y. A., Endoh, M., Appanah, R., Nesterova, T. B., Silva, J., Otte, A. P., Vidal, M., Koseki, H., & Brockdorff, N. (2004). Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. *Developmental Cell*, *7*(5), 663–676. https://doi.org/10.1016/j.devcel.2004.10.005

Du, J., Zhong, X., Bernatavichute, Y. V., Stroud, H., Feng, S., Caro, E., Vashisht, A. A., Terragni, J., Chin, H. G., Tu, A., Hetzel, J., Wohlschlegel, J. A., Pradhan, S., Patel, D. J., & Jacobsen, S. E. (2012). Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. *Cell*, *151*(1), 167–180. https://doi.org/10.1016/j.cell.2012.07.034

Duc, C., Benoit, M., Détourné, G., Simon, L., Poulet, A., Jung, M., Veluchamy, A., Latrasse, D., Le Goff, S., Cotterell, S., Tatout, C., Benhamed, M., & Probst, A. V. (2017). Arabidopsis ATRX Modulates H3.3 Occupancy and Fine-Tunes Gene Expression. *The Plant Cell*, *29*(7), 1773–1793. https://doi.org/10.1105/tpc.16.00877

Earnshaw, W. C., & Rothfield, N. (1985). Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. *Chromosoma*, *91*(3–4), 313–321. https://doi.org/10.1007/BF00328227

Faast, R., Thonglairoam, V., Schulz, T. C., Beall, J., Wells, J. R., Taylor, H., Matthaei, K., Rathjen, P. D., Tremethick, D. J., & Lyons, I. (2001). Histone variant H2A.Z is required for early mammalian development. *Current Biology: CB*, *11*(15), 1183–1187. https://doi.org/10.1016/s0960-9822(01)00329-3

Franklin, S. G., & Zweidler, A. (1977). Non-allelic variants of histones 2a, 2b and 3 in mammals. *Nature*, *266*(5599), 273–275. https://doi.org/10.1038/266273a0

Goldberg, A. D., Banaszynski, L. A., Noh, K.-M., Lewis, P. W., Elsaesser, S. J., Stadler, S., Dewell, S., Law, M., Guo, X., Li, X., Wen, D., Chapgier, A., DeKelver, R. C., Miller, J. C., Lee, Y.-L., Boydston, E. A., Holmes, M. C., Gregory, P. D., Greally, J. M., … Allis, C. D. (2010). Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell*, *140*(5), 678–691. https://doi.org/10.1016/j.cell.2010.01.003

Guillemette, B., Bataille, A. R., Gévry, N., Adam, M., Blanchette, M., Robert, F., & Gaudreau, L. (2005). Variant Histone H2A.Z Is Globally Localized to the Promoters of Inactive Yeast Genes and Regulates Nucleosome Positioning. *PLOS Biology*, *3*(12), e384. https://doi.org/10.1371/journal.pbio.0030384

Heitz: "Das" Heterochromatin der Moose—Google Scholar. (n.d.). Retrieved November 10, 2025, from https://scholar.google.com/scholar_lookup?&title=&journal=Jahrb.%20Wiss.%20Bot.&volume=69&pages=762-818&publication_year=1928&author=Heitz%2CE&author=Heterochromatin%20der%20Moose%2CI

Hisanaga, T., Wu, S., Schafran, P., Axelsson, E., Akimcheva, S., Dolan, L., Li, F.-W., & Berger, F. (2023). The ancestral chromatin landscape of land plants. *New Phytologist*, *240*(5), 2085–2101. https://doi.org/10.1111/nph.19311

Ingouff, M., Rademacher, S., Holec, S., Soljić, L., Xin, N., Readshaw, A., Foo, S. H., Lahouze, B., Sprunck, S., & Berger, F. (2010a). Zygotic resetting of the HISTONE 3 variant repertoire participates in epigenetic reprogramming in Arabidopsis. *Current Biology: CB*, *20*(23), 2137–2143. https://doi.org/10.1016/j.cub.2010.11.012

Ingouff, M., Rademacher, S., Holec, S., Soljić, L., Xin, N., Readshaw, A., Foo, S. H., Lahouze, B., Sprunck, S., & Berger, F. (2010b). Zygotic resetting of the HISTONE 3 variant repertoire participates in epigenetic reprogramming in Arabidopsis. *Current Biology: CB*, *20*(23), 2137–2143. https://doi.org/10.1016/j.cub.2010.11.012

Jacob, Y., Feng, S., LeBlanc, C. A., Bernatavichute, Y. V., Stroud, H., Cokus, S., Johnson, L. M., Pellegrini, M., Jacobsen, S. E., & Michaels, S. D. (2009). ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing. *Nature Structural & Molecular Biology*, *16*(7), 763–768. https://doi.org/10.1038/nsmb.1611

Jenness, C., Giunta, S., Müller, M. M., Kimura, H., Muir, T. W., & Funabiki, H. (2018). HELLS and CDCA7 comprise a bipartite nucleosome remodeling complex defective in ICF syndrome. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(5), E876–E885. https://doi.org/10.1073/pnas.1717509115

Jiang, D., Borg, M., Lorković, Z. J., Montgomery, S. A., Osakabe, A., Yelagandula, R., Axelsson, E., & Berger, F. (2020). The evolution and functional divergence of the histone H2B family in plants. *PLoS Genetics*, *16*(7), e1008964. https://doi.org/10.1371/journal.pgen.1008964

Johnson, L. M., Cao, X., & Jacobsen, S. E. (2002). Interplay between Two Epigenetic Marks: DNA Methylation and Histone H3 Lysine 9 Methylation. *Current Biology*, *12*(16), 1360–1367. https://doi.org/10.1016/S0960-9822(02)00976-4

Karmodiya, K., Krebs, A. R., Oulad-Abdelghani, M., Kimura, H., & Tora, L. (2012). H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. *BMC Genomics*, *13*, 424. https://doi.org/10.1186/1471-2164-13-424

Karras, G. I., Kustatscher, G., Buhecha, H. R., Allen, M. D., Pugieux, C., Sait, F., Bycroft, M., & Ladurner, A. G. (2005). The macro domain is an ADP-ribose binding module. *The EMBO Journal*, *24*(11), 1911–1920. https://doi.org/10.1038/sj.emboj.7600664

Kobor, M. S., Venkatasubrahmanyam, S., Meneghini, M. D., Gin, J. W., Jennings, J. L., Link, A. J., Madhani, H. D., & Rine, J. (2004). A Protein Complex Containing the Conserved Swi2/Snf2-Related ATPase Swr1p Deposits Histone Variant H2A.Z into Euchromatin. *PLOS Biology*, *2*(5), e131. https://doi.org/10.1371/journal.pbio.0020131

Kornberg, R. D. (1974). Chromatin Structure: A Repeating Unit of Histones and DNA. *Science*, *184*(4139), 868–871. https://doi.org/10.1126/science.184.4139.868

Kornberg, R. D., & Klug, A. (1981). The nucleosome. *Scientific American*,

244(2), 52–64. https://doi.org/10.1038/scientificamerican0281-52

Krogan, N. J., Kim, M., Tong, A., Golshani, A., Cagney, G., Canadien, V., Richards, D. P., Beattie, B. K., Emili, A., Boone, C., Shilatifard, A., Buratowski, S., & Greenblatt, J. (2003). Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II. *Molecular and Cellular Biology*, *23*(12), 4207–4218. https://doi.org/10.1128/MCB.23.12.4207-4218.2003

Larson, A. G., Elnatan, D., Keenen, M. M., Trnka, M. J., Johnston, J. B., Burlingame, A. L., Agard, D. A., Redding, S., & Narlikar, G. J. (2017). Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature*, *547*(7662), 236–240. https://doi.org/10.1038/nature22822

Loppin, B., & Berger, F. (2020). Histone Variants: The Nexus of Developmental Decisions and Epigenetic Memory. *Annual Review of Genetics*, *54*, 121–149. https://doi.org/10.1146/annurev-genet-022620-100039

Lorch, Y., LaPointe, J. W., & Kornberg, R. D. (1987). Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. *Cell*, *49*(2), 203–210. https://doi.org/10.1016/0092-8674(87)90561-7

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, *389*(6648), 251–260. https://doi.org/10.1038/38444

Mannironi, C., Bonner, W. M., & Hatch, C. L. (1989). H2A.X. a histone isoprotein with a conserved C-terminal sequence, is encoded by a novel mRNA with both DNA replication type and polyA 3' processing signals. *Nucleic Acids Research*, *17*(22), 9113–9126. https://doi.org/10.1093/nar/17.22.9113

March-Díaz, R., García-Domínguez, M., Lozano-Juste, J., León, J., Florencio, F. J., & Reyes, J. C. (2008). Histone H2A.Z and homologues of components of the SWR1 complex are required to control immunity in Arabidopsis. *The Plant Journal*, *53*(3), 475–487. https://doi.org/10.1111/j.1365-313X.2007.03361.x

Marzluff, W. F., Gongidi, P., Woods, K. R., Jin, J., & Maltais, L. J. (2002). The human and mouse replication-dependent histone genes. *Genomics*, *80*(5), 487–498.

Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-

K., Koche, R. P., Lee, W., Mendenhall, E., O'Donovan, A., Presser, A., Russ, C., Xie, X., Meissner, A., Wernig, M., Jaenisch, R., … Bernstein, B. E. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, *448*(7153), 553–560. https://doi.org/10.1038/nature06008

Miller, T., Krogan, N. J., Dover, J., Erdjument-Bromage, H., Tempst, P., Johnston, M., Greenblatt, J. F., & Shilatifard, A. (2001). COMPASS: A complex of proteins associated with a trithorax-related SET domain protein. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(23), 12902–12907. https://doi.org/10.1073/pnas.231473398

Mizuguchi, G., Shen, X., Landry, J., Wu, W.-H., Sen, S., & Wu, C. (2004). ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science (New York, N.Y.)*, *303*(5656), 343–348. https://doi.org/10.1126/science.1090701

Müller, J., Hart, C. M., Francis, N. J., Vargas, M. L., Sengupta, A., Wild, B., Miller, E. L., O'Connor, M. B., Kingston, R. E., & Simon, J. A. (2002). Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. *Cell*, *111*(2), 197–208. https://doi.org/10.1016/s0092-8674(02)00976-5

Murray, K. (1964). The Occurrence of iε-N-Methyl Lysine in Histones. *Biochemistry*, *3*(1), 10–15. https://doi.org/10.1021/bi00889a003

Nie, X., Wang, H., Li, J., Holec, S., & Berger, F. (2014). The HIRA complex that deposits the histone H3.3 is conserved in Arabidopsis and facilitates transcriptional dynamics. *Biology Open*, *3*(9), 794–802. https://doi.org/10.1242/bio.20148680

Nunez-Vazquez, R., Madeira, S., Rodríguez-Casillas, L., Gomez-Martinez, D., Desvoyes, B., & Gutierrez, C. (2025). The histone variant H3.14 is an early player in the abiotic stress response of Arabidopsis. *Developmental Cell*, *60*(21), 2931-2945.e7. https://doi.org/10.1016/j.devcel.2025.06.019

Osakabe, A., Jamge, B., Axelsson, E., Montgomery, S. A., Akimcheva, S., Kuehn, A. L., Pisupati, R., Lorković, Z. J., Yelagandula, R., Kakutani, T., & Berger, F. (2021). The chromatin remodeler DDM1 prevents transposon mobility through deposition of histone variant H2A.W. *Nature Cell Biology*, *23*(4), 391–400. https://doi.org/10.1038/s41556-021-00658-1

Osakabe, A., Lorkovic, Z. J., Kobayashi, W., Tachiwana, H., Yelagandula, R., Kurumizaka, H., & Berger, F. (2018). Histone H2A variants confer specific

properties to nucleosomes and impact on chromatin accessibility. *Nucleic Acids Research*, *46*(15), 7675–7685. https://doi.org/10.1093/nar/gky540

Oya, S., Takahashi, M., Takashima, K., Kakutani, T., & Inagaki, S. (2022). Transcription-coupled and epigenome-encoded mechanisms direct H3K4 methylation. *Nature Communications*, *13*(1), 4521. https://doi.org/10.1038/s41467-022-32165-8

Paik, W. K., & Kim, S. (1967). Enzymatic methylation of protein fractions from calf thymus nuclei. *Biochemical and Biophysical Research Communications*, *29*(1), 14–20. https://doi.org/10.1016/0006-291X(67)90533-5

Palmer, D. K., O'Day, K., Wener, M. H., Andrews, B. S., & Margolis, R. L. (1987). A 17-kD centromere protein (CENP-A) copurifies with nucleosome core particles and with histones. *The Journal of Cell Biology*, *104*(4), 805–815. https://doi.org/10.1083/jcb.104.4.805

Pehrson, J. R., & Fried, V. A. (1992). MacroH2A, a core histone containing a large nonhistone region. *Science (New York, N.Y.)*, *257*(5075), 1398–1400. https://doi.org/10.1126/science.1529340

Pehrson, J. R., & Fuji, R. N. (1998). Evolutionary conservation of histone macroH2A subtypes and domains. *Nucleic Acids Research*, *26*(12), 2837–2842. https://doi.org/10.1093/nar/26.12.2837

Raynaud, C., Sozzani, R., Glab, N., Domenichini, S., Perennes, C., Cella, R., Kondorosi, E., & Bergounioux, C. (2006). Two cell-cycle regulated SET-domain proteins interact with proliferating cell nuclear antigen (PCNA) in Arabidopsis. *The Plant Journal: For Cell and Molecular Biology*, *47*(3), 395–407. https://doi.org/10.1111/j.1365-313X.2006.02799.x

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z. W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C. P., Allis, C. D., & Jenuwein, T. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature*, *406*(6796), 593–599. https://doi.org/10.1038/35020506

Redon, C., Pilch, D., Rogakou, E., Sedelnikova, O., Newrock, K., & Bonner, W. (2002). Histone H2A variants H2AX and H2AZ. *Current Opinion in Genetics & Development*, *12*(2), 162–169. https://doi.org/10.1016/s0959-437x(02)00282-4

Rogakou, E. P., Boon, C., Redon, C., & Bonner, W. M. (1999). Megabase chromatin domains involved in DNA double-strand breaks in vivo. *The*

Journal of Cell Biology, *146*(5), 905–916. https://doi.org/10.1083/jcb.146.5.905

Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova, V. S., & Bonner, W. M. (1998). DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *The Journal of Biological Chemistry*, *273*(10), 5858–5868. https://doi.org/10.1074/jbc.273.10.5858

Roguev, A., Schaft, D., Shevchenko, A., Pijnappel, W. W. M. P., Wilm, M., Aasland, R., & Stewart, A. F. (2001). The Saccharomyces cerevisiae Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. *The EMBO Journal*, *20*(24), 7137–7148. https://doi.org/10.1093/emboj/20.24.7137

Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C. T., Schreiber, S. L., Mellor, J., & Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. *Nature*, *419*(6905), 407–411. https://doi.org/10.1038/nature01080

Schneider, R., Bannister, A. J., Weise, C., & Kouzarides, T. (2004). Direct binding of INHAT to H3 tails disrupted by modifications. *The Journal of Biological Chemistry*, *279*(23), 23859–23862. https://doi.org/10.1074/jbc.C400151200

Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., & Cavalli, G. (2007). Genome regulation by polycomb and trithorax proteins. *Cell*, *128*(4), 735–745. https://doi.org/10.1016/j.cell.2007.02.009

Smith, S., & Stillman, B. (1989). Purification and characterization of CAF-I, a human cell factor required for chromatin assembly during DNA replication in vitro. *Cell*, *58*(1), 15–25. https://doi.org/10.1016/0092-8674(89)90398-x

Strom, A. R., Emelyanov, A. V., Mir, M., Fyodorov, D. V., Darzacq, X., & Karpen, G. H. (2017). Phase separation drives heterochromatin domain formation. *Nature*, *547*(7662), 241–245. https://doi.org/10.1038/nature22989

Stroud, H., Do, T., Du, J., Zhong, X., Feng, S., Johnson, L., Patel, D. J., & Jacobsen, S. E. (2014). Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. *Nature Structural & Molecular Biology*, *21*(1), 64–72. https://doi.org/10.1038/nsmb.2735

Stroud, H., Otero, S., Desvoyes, B., Ramírez-Parra, E., Jacobsen, S. E., & Gutierrez, C. (2012). Genome-wide analysis of histone H3.1 and H3.3 variants in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(14), 5370–5375.

https://doi.org/10.1073/pnas.1203145109

Tagami, H., Ray-Gallet, D., Almouzni, G., & Nakatani, Y. (2004). Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. *Cell*, *116*(1), 51–61. https://doi.org/10.1016/s0092-8674(03)01064-x

Talbert, P. B., Masuelli, R., Tyagi, A. P., Comai, L., & Henikoff, S. (2002). Centromeric localization and adaptive evolution of an Arabidopsis histone H3 variant. *The Plant Cell*, *14*(5), 1053–1066. https://doi.org/10.1105/tpc.010425

Turner, B. M. (2000). Histone acetylation and an epigenetic code. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, *22*(9), 836–845. https://doi.org/10.1002/1521-1878(200009)22:9%253C836::AID-BIES9%253E3.0.CO;2-X

Verreault, A., Kaufman, P. D., Kobayashi, R., & Stillman, B. (1996). Nucleosome assembly by a complex of CAF-1 and acetylated histones H3/H4. *Cell*, *87*(1), 95–104. https://doi.org/10.1016/s0092-8674(00)81326-4

Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R. S., & Zhang, Y. (2004). Role of histone H2A ubiquitination in Polycomb silencing. *Nature*, *431*(7010), 873–878. https://doi.org/10.1038/nature02985

Wasylyk, B., Thevenin, G., Oudet, P., & Chambon, P. (1979). Transcription of in vitro assembled chromatin by Escherichia coli RNA polymerase. *Journal of Molecular Biology*, *128*(3), 411–440. https://doi.org/10.1016/0022-2836(79)90095-0

Wells, D., Hoffman, D., & Kedes, L. (1987). Unusual structure, evolutionary conservation of non-coding sequences and numerous pseudogenes characterize the human H3.3 histone multigene family. *Nucleic Acids Research*, *15*(7), 2871–2889. https://doi.org/10.1093/nar/15.7.2871

Wells, D., & Kedes, L. (1985). Structure of a human histone cDNA: Evidence that basally expressed histone genes have intervening sequences and encode polyadenylylated mRNAs. *Proceedings of the National Academy of Sciences of the United States of America*, *82*(9), 2834–2838. https://doi.org/10.1073/pnas.82.9.2834

West, M. H., & Bonner, W. M. (1980). Histone 2A, a heteromorphous family of eight protein species. *Biochemistry*, *19*(14), 3238–3245. https://doi.org/10.1021/bi00555a022

Wollmann, H., Holec, S., Alden, K., Clarke, N. D., Jacques, P.-É., & Berger, F. (2012). Dynamic Deposition of Histone

Variant H3.3 Accompanies Developmental Remodeling of the Arabidopsis Transcriptome. *PLOS Genetics*, *8*(5), e1002658. https://doi.org/10.1371/journal.pgen.1002658

Yelagandula, R., Stroud, H., Holec, S., Zhou, K., Feng, S., Zhong, X., Muthurajan, U. M., Nie, X., Kawashima, T., Groth, M., Luger, K., Jacobsen, S. E., & Berger, F. (2014). The histone variant H2A.W defines heterochromatin and promotes chromatin condensation in Arabidopsis. *Cell*, *158*(1), 98–109. https://doi.org/10.1016/j.cell.2014.06.006

Zhang, H., Roberts, D. N., & Cairns, B. R. (2005). Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss. *Cell*, *123*(2), 219–231. https://doi.org/10.1016/j.cell.2005.08.036

Zhang, X., Clarenz, O., Cokus, S., Bernatavichute, Y. V., Pellegrini, M., Goodrich, J., & Jacobsen, S. E. (2007). Whole-Genome Analysis of Histone H3 Lysine 27 Trimethylation in Arabidopsis. *PLoS Biology*, *5*(5), e129. https://doi.org/10.1371/journal.pbio.0050129

Zilberman, D., Coleman-Derr, D., Ballinger, T., & Henikoff, S. (2008). Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. *Nature*, *456*(7218), 125–129. https://doi.org/10.1038/nature07324

Chapter 2

2. Chromatin States

Contents

2.1 INTRODUCTION 25

2.2 CHROMATIN STATES: DECODING COMBINATORIAL PATTERNS 26

2.2.1 MATHEMATICAL FOUNDATION OF CHROMATIN STATE DISCOVERY 26
2.2.2 FUNCTIONAL CHARACTERIZATION OF CHROMATIN STATES 27
2.2.3 INTERPRETATION OF CHROMATIN STATES 29

2.3 CHROMATIN STATES IN PLANTS 29

2.3.1 CHROMATIN STATES OF *ARABIDOPSIS* 29
2.3.2 CHROMATIN STATES OF OTHER LAND PLANTS 30

2.4 CHROMATIN STATES AND TRANSCRIPTIONAL CONTROL 31

BIBLIOGRAPHY 32

2.1 Introduction

Chromatin organization of eukaryotic genomes resembles a complex molecular code, where the post-translational modifications (PTMs) and variants of histone create distinct signatures across different genomic regions. While Chapter 1 explored how individual chromatin components: from specific histone marks to chromatin remodeling complexes, contribute to transcriptional regulation, a more intriguing picture emerges when we consider their collective behavior. Each histone PTM and variant has been shown to associate with specific genomic features: H3K4me3 marks active promoters, H3K27me3 silences developmental genes, and H2A.Z destabilizes nucleosomes at regulatory elements. Yet, these individual associations represent only fragments of a larger regulatory puzzle.

The mechanistic basis underlying the regulatory effects of histone PTMs and variants lies in their ability to alter nucleosome properties and modulate DNA accessibility. For instance, acetylation of histone tails generally reduces histone-DNA interactions and promotes chromatin decompaction,

while methylation marks indirectly either activate or repress transcription depending on their specific residue and context. Similarly, histone variants such as H3.3 and H2A.Z confer unique biophysical properties to nucleosomes, influencing their stability and interactions with regulatory proteins.

This mechanistic understanding leads to a crucial realization: transcriptional regulation is not governed by individual PTMs or variants acting in isolation, but rather by their combinatorial presence and interactions within chromatin domains. The regulatory output of a genomic locus depends on the specific combination of modifications present, their relative abundance, and their spatial organization along the chromatin fiber. This combinatorial logic suggests that to truly understand chromatin-mediated gene regulation, we require methodological approaches that can capture and analyze the preferential combinatorial behavior of histone PTMs and variants across the genome. How can we systematically identify and characterize these recurring combinations of chromatin features, and what functional significance does these combinatorial patterns hold for transcriptional control?

2.2 Chromatin States: Decoding Combinatorial Patterns

The quest to understand combinatorial patterns in chromatin modifications began with early observations that certain histone marks frequently co-occurred at specific genomic loci. Early attempts to systematically analyze these combinatorial patterns relied on pairwise correlation analyses and manual clustering approaches. The co-occurrence of different histone marks across the genome was used to identify regions with similar modification profiles. However, these approaches were limited by their inability to handle the complexity of multiple overlapping modifications and their subjective nature in defining boundaries between different chromatin domains. The breakthrough came with the development of unsupervised machine learning approaches, culminating in the ChromHMM algorithm developed by Ernst and Kellis in 2012 (Ernst & Kellis, 2012).

2.2.1 Mathematical Foundation of Chromatin State Discovery

ChromHMM employs a Hidden Markov Model (HMM) framework to systematically identify recurring combinatorial patterns of chromatin marks across the genome. The fundamental assumption underlying this

approach is that the genome can be segmented into discrete states, each characterized by a specific combination of chromatin features. In the HMM framework, these chromatin states represent the "hidden" states that cannot be directly observed, while the presence or absence of histone modifications at genomic bins represents the "observable" emissions.

The mathematical foundation of ChromHMM rests on two key probability distributions: the emission probabilities and the transition probabilities. Emission probabilities define the likelihood of observing a particular combination of histone marks given a specific chromatin state. For each state k and modification m, the emission probability ($e_{k,m}$) represents the probability that modification m is present in state k. Transition probabilities ($t_{i,j}$) defines the likelihood of transitioning from chromatin state i to state j in adjacent genomic bins, capturing the spatial organization and stability of chromatin domains.

The ChromHMM algorithm uses the Baum-Welch expectation-maximization algorithm to iteratively optimize these parameters and identify the most likely sequence of chromatin states across the genome. The process begins with random initialization of emission and transition probabilities, followed by iterative refinement until convergence. The optimal number of states is typically determined through model selection criteria that balance model complexity with explanatory power.

Through this computational framework, ChromHMM defines chromatin states as genomically recurring, combinatorial patterns of histone modifications that represent functionally distinct regulatory elements. Each chromatin state is characterized by its unique emission signature: the specific combination and relative enrichment of histone marks, and its genomic distribution and functional associations. Importantly, chromatin states are not simply clusters of similar modifications but rather represent discrete regulatory units with distinct biological functions, from active promoters and enhancers to various forms of heterochromatin.

2.2.2 Functional Characterization of Chromatin States

A critical aspect of chromatin state analysis is that the ChromHMM algorithm does not inherently determine the optimal number of states, and this decision rests with the user. The algorithm can theoretically segment the genome into any number of states, from a few broad categories to hundreds of highly specific patterns. However, the challenge lies in finding the appropriate balance between model complexity and

functional relevance. Too few states may oversimplify the chromatin landscape and miss important regulatory distinctions, while too many states can create overly granular divisions that lack clear biological meaning or statistical power.

This balance necessitates systematic functional characterization of the identified chromatin states to validate their biological significance and guide model selection. Several complementary strategies have emerged as standard approaches for functionally characterizing chromatin states and determining their regulatory relevance.

Overlap with known regulatory elements provides the most direct validation of chromatin state functionality. States are assessed for their enrichment at well-characterized genomic features such as transcription start sites (TSS), gene bodies, enhancers, silencers, and repetitive elements. States that show strong enrichment at promoters, for instance, can be confidently annotated as promoter-associated states, while those enriched at intergenic regions with enhancer-like histone signatures likely represent enhancer states.

Correlations with gene expression offer insights into the transcriptional activity associated with different chromatin states. States found at promoters and gene bodies can be ranked by their association with high, moderate, or low gene expression levels, revealing the spectrum from active to repressed chromatin environments. This analysis helps distinguish between different levels of transcriptional activity and identifies states associated with transcriptional silencing.

DNA accessibility patterns provide additional functional context, as measured by techniques such as ATAC-seq or DNase-seq. Active regulatory states typically show high accessibility, while repressed states exhibit low accessibility. This correlation helps validate the functional predictions made based on histone modification patterns alone.

DNA methylation levels offer another layer of functional characterization, particularly for identifying heterochromatic and silenced states. States associated with high DNA methylation often correspond to constitutive heterochromatin or silenced repetitive elements, while unmethylated states may represent active or poised regulatory regions.

Genomic distribution patterns reveal the spatial organization and prevalence of different chromatin states across chromosomes and genomic features. Some states may be predominantly

found in gene-rich euchromatic regions, while others cluster in pericentromeric heterochromatin or telomeric regions, providing insights into their functional roles in genome organization.

Through the integration of these functional characterization approaches, one can confidently annotate chromatin states with biologically meaningful labels and select models that capture functionally relevant chromatin organization patterns.

2.2.3 Interpretation of Chromatin States

With the standard bin size of 200 bp used in chromatin state analyses, each genomic bin approximates the DNA wrapped around a single nucleosome. In this context, a chromatin state represents the probability of specific histone PTMs and variants being present on a nucleosome, providing insights into the types of nucleosomes and their abundance and localization patterns across the genome.

For the purposes of this thesis, a **chromatin state** is defined as:

"The probabilistic combination of histone post-translational modifications and variants present on a 200 bp genomic bin, corresponding roughly to a single nucleosome."

This definition emphasizes that chromatin states represent statistical descriptions of nucleosome modification patterns and provides a framework for understanding the combinatorial logic of chromatin organization at nucleosome resolution.

2.3 Chromatin States in Plants

The application of chromatin state analysis to plant genomes began with pioneering work in *Arabidopsis thaliana*, building upon successful implementations in animal systems (Roudier et al., 2011).

2.3.1 Chromatin states of *Arabidopsis*

The systematic analysis of chromatin states in plants began with the pioneering work of (Sequeira-Mendes et al., 2014), who applied ChromHMM to ChIP-seq data from *Arabidopsis thaliana* seedlings and identified nine distinct chromatin states based on eight histone modifications (H3K4me1, H3K4me2, H3K4me3, H3K9me2, H3K27me1, H3K27me3, H3K36me3, and H4K20me1). This foundational study revealed that plant genomes, like their animal counterparts, exhibit discrete combinatorial patterns of histone modifications that correlate with specific genomic features and transcriptional activities. The authors demonstrated that active genes were associated with H3K4me3 and H3K36me3 marks, while repressed regions showed enrichment

for H3K27me3, establishing the first comprehensive chromatin state map for a plant species.

However, this initial framework had a notable limitation: the model did not incorporate all histone variants, which were later found to contribute significantly to chromatin state definition (Jamge et al., 2023). Additionally, the analysis was restricted to only eight histone marks. Recognizing these gaps, we addressed these limitations through a comprehensive analysis that greatly expanded the scope of *Arabidopsis* chromatin state analysis (Jamge et al., 2023; see manuscript 1). Our study integrated a total of 27 histone post-translational modifications and variants, including crucial histone variants that had been overlooked in previous analyses. This expanded approach revealed previously uncharacterized chromatin states and demonstrated the importance of histone variants in defining distinct chromatin environments. The detailed methodology and findings of this work are presented as a complete manuscript later in this thesis, illustrating how the inclusion of histone variants can refine our understanding of chromatin state biology in plants.

2.3.2 Chromatin states of other land plants

Following the establishment of chromatin state analysis in *Arabidopsis*, efforts to characterize chromatin states in other plant species have been limited but revealing. Studies in rice (*Oryza sativa*) have demonstrated that the distribution patterns of histone marks potentially determine their roles in transcription and RNA processing, suggesting conservation of chromatin organization principles across flowering plants (Hu et al., 2020). Similarly, analyses in other angiosperm species have indicated that chromatin states described in *Arabidopsis* are broadly representative of flowering plants (Roudier et al., 2011; Wrightsman et al., 2022).

However, to truly understand the evolutionary conservation of chromatin states across land plants, it was essential to examine species that diverged much earlier in plant evolution. This led us to investigate chromatin states in *Marchantia polymorpha*, a model liverwort that represents the bryophyte lineage. The evolutionary distance between bryophytes and vascular plants makes *Marchantia* an ideal model for exploring the conservation of chromatin states and highlighting the evolution of fundamental chromatin regulatory mechanisms.

Marchantia shares similar genome organization with other bryophyte species (hornworts and mosses), which together form the monophyletic group of bryophytes that diverged from vascular plants approximately 450 million years ago (de Sousa et al., 2019). This substantial evolutionary distance, combined with distinct genome organization between bryophytes and flowering plants, provides a powerful comparative framework for understanding which aspects of chromatin organization represent ancient, conserved features versus more recent evolutionary innovations.

The choice of *Marchantia* was further motivated by several practical considerations. First, comprehensive histone modification profiles were already available for this species (Montgomery et al., 2020), providing the necessary foundation for chromatin state analysis. Second, the comparison between haploid gametophytic tissues of bryophytes and diploid sporophytic tissues of angiosperms offers unique insights into whether chromatin organization reflects ancestral patterns that predate the evolution of complex life cycles in land plants.

Our analysis not only revealed striking conservation of chromatin state composition and their association with transcription between *Marchantia* and *Arabidopsis*, despite their evolutionary distance and distinct genome organizations but also highlighted key differences. This conservation suggests that the fundamental principles of chromatin-mediated gene regulation were established early in land plant evolution and have been maintained across major evolutionary transitions (see manuscript 2).

2.4 Chromatin States and Transcriptional Control

The comprehensive characterization of chromatin states across diverse plant species reveals a fundamental layer of genome organization that provides the regulatory foundation for transcriptional control. However, chromatin states alone do not determine gene expression outcomes, but they establish the regulatory potential and accessibility landscape upon which transcription factors operate. The distinct histone modification signatures that define each chromatin state create specific molecular environments that can either facilitate or restrict transcription factor binding and activity. Active chromatin states, characterized by marks such as H3K4me3 and accessible chromatin, provide permissive environments for transcription factor recruitment, while repressive states marked by H3K27me3 or heterochromatic signatures present

barriers that must be overcome for transcriptional activation. This raises critical questions about the mechanistic interplay between chromatin states and transcription factors: How do transcription factors navigate and interpret the chromatin state landscape? Can transcription factors actively remodel chromatin states, or are they primarily constrained by pre-existing chromatin environments?

Understanding these dynamic interactions between chromatin states and transcription factors is essential for deciphering the complete regulatory logic that governs plant gene expression and development. In the following chapter (and in manuscript 2), we explore the molecular mechanisms by which transcription factors recognize, bind to, and potentially remodel chromatin states, examining how these regulatory proteins navigate the complex chromatin landscape to orchestrate precise spatiotemporal gene expression patterns in plants.

Bibliography

de Sousa, F., Foster, P. G., Donoghue, P. C. J., Schneider, H., & Cox, C. J. (2019). Nuclear protein phylogenies support the monophyly of the three bryophyte groups (Bryophyta Schimp.). *New Phytologist*, *222*(1), 565–575. https://doi.org/10.1111/nph.15587

Ernst, J., & Kellis, M. (2012). ChromHMM: Automating chromatin-state discovery and characterization. *Nature Methods*, *9*(3), 215–216. https://doi.org/10.1038/nmeth.1906

Hu, Y., Lai, Y., Chen, X., Zhou, D.-X., & Zhao, Y. (2020). Distribution pattern of histone marks potentially determines their roles in transcription and RNA processing in rice. *Journal of Plant Physiology*, *249*, 153167. https://doi.org/10.1016/j.jplph.2020.153167

Jamge, B., Lorković, Z. J., Axelsson, E., Osakabe, A., Shukla, V., Yelagandula, R., Akimcheva, S., Kuehn, A. L., & Berger, F. (2023). Histone variants shape chromatin states in Arabidopsis. *eLife*, *12*, RP87714. https://doi.org/10.7554/eLife.87714

Montgomery, S. A., Tanizawa, Y., Galik, B., Wang, N., Ito, T., Mochizuki, T., Akimcheva, S., Bowman, J. L., Cognat, V., Maréchal-Drouard, L., Ekker, H., Hong, S.-F., Kohchi, T., Lin, S.-S., Daisy Liu, L.-Y., Nakamura, Y., Valeeva, L. R., Shakirov, E. V., Shippen, D. E., … Berger, F. (2020). Chromatin Organization in Early Land Plants Reveals an Ancestral Association between H3K27me3, Transposons, and Constitutive Heterochromatin. *Current Biology : CB*, *30*(4), 573-588.e7.

https://doi.org/10.1016/j.cub.2019.12.015

Roudier, F., Ahmed, I., Bérard, C., Sarazin, A., Mary-Huard, T., Cortijo, S., Bouyer, D., Caillieux, E., Duvernois-Berthet, E., Al-Shikhley, L., Giraut, L., Després, B., Drevensek, S., Barneche, F., Dèrozier, S., Brunaud, V., Aubourg, S., Schnittger, A., Bowler, C., … Colot, V. (2011). Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. *The EMBO Journal*, *30*(10), 1928–1938. https://doi.org/10.1038/emboj.2011.103

Sequeira-Mendes, J., Aragüez, I., Peiró, R., Mendez-Giraldez, R., Zhang, X., Jacobsen, S. E., Bastolla, U., & Gutierrez, C. (2014). The Functional Topography of the Arabidopsis Genome Is Organized in a Reduced Number of Linear Motifs of Chromatin States. *The Plant Cell*, *26*(6), 2351–2366. https://doi.org/10.1105/tpc.114.124578

Wrightsman, T., Marand, A. P., Crisp, P. A., Springer, N. M., & Buckler, E. S. (2022). Modeling chromatin state from sequence across angiosperms using recurrent convolutional neural networks. *The Plant Genome*, *15*(3), e20249. https://doi.org/10.1002/tpg2.20249

Chapter 3

3. Transcription Factor–Chromatin Interaction

Contents

3.1 INTRODUCTION 34

3.2 TRANSCRIPTION FACTOR FUNCTION AND DNA RECOGNITION 34

3.3 CHROMATIN AS A BARRIER AND REGULATORY PLATFORM 35

3.4 PIONEER TFS: OVERCOMING THE CHROMATIN BARRIERS 36

3.5 FUNCTIONAL AND NON-FUNCTIONAL TF BINDINGS 37

3.6 METHODS TO STUDY TF–CHROMATIN STATE INTERACTIONS 38

- 3.6.1 Genome-wide TF Binding Approaches 38
- 3.6.2 Integration with Chromatin State Models 39
- 3.6.3 TF Activity Scores distinguish Functional TF Binding 40

3.7 CONCLUSION 41

BIBLIOGRAPHY 42

3.1 Introduction

Transcription factors (TFs) are sequence-specific DNA-binding proteins that regulate gene expression by recognizing and binding to cis-regulatory elements in the genome. Upon binding, TFs recruit or stabilize the transcriptional machinery, including RNA Polymerase II and associated co-factors, to initiate or modulate transcription (Wong & Gunawardena, 2020). In eukaryotes, however, DNA is not freely accessible but is packaged into chromatin, creating a fundamental challenge: how do TFs access their target sites within this complex nucleoprotein structure?

3.2 Transcription Factor Function and DNA Recognition

TFs typically contain modular domains that enable their regulatory functions. The DNA-binding domain (DBD) recognizes specific DNA sequences, often 6-12 base pairs in length, through direct readout of the DNA major and minor grooves. Different TF families possess distinct DBD architectures such as basic helix-loop-helix (bHLH), MYB, AP2/EREBP, WRKY, and MADS-box

domains, each with characteristic DNA-binding specificities (Franco-Zorrilla et al., 2014a). Beyond the DBD, TFs contain regulatory domains that mediate transcriptional activation or repression, often through recruitment of chromatin-modifying complexes or direct interaction with the transcriptional machinery. Many plant TFs also harbor intrinsically disordered regions (IDRs) whose functions are increasingly recognized in transcriptional regulation and protein-protein interactions.

According to the classical model of TF function, TFs bind to accessible DNA sequences in promoters and enhancers, where they facilitate the assembly of the Pre-Initiation Complex (PIC) and stabilize RNA Polymerase II engagement (Wong & Gunawardena, 2020). This model, however, was largely developed from studies using naked DNA templates and does not fully account for the chromatin context in which TFs must operate *in vivo*.

3.3 Chromatin as a Barrier and Regulatory Platform

As discussed in Chapters 1 and 2, chromatin is organized into distinct states characterized by specific combinations of histone variants and post-translational modifications. These chromatin states create varying degrees of accessibility to the underlying DNA. Nucleosomes, the fundamental repeating units of chromatin, present a significant physical barrier to TF binding. *In vitro* studies have demonstrated that nucleosome assembly on DNA templates strongly inhibits TF binding and transcription factor-mediated activation (Lorch et al., 1987; Wasylyk et al., 1979). The wrapping of ~147 base pairs of DNA around the histone octamer occludes many potential TF binding sites and reduces the accessibility of DNA to regulatory proteins.

However, chromatin is not merely a static barrier but functions as a dynamic regulatory platform. Chromatin accessibility varies across the genome, with nucleosome-free regions (NFRs) at promoters and enhancers providing accessible DNA for TF binding (Heyndrickx et al., 2014). These NFRs are often maintained by the combined action of chromatin remodelers, histone modifying enzymes, and the binding of TFs themselves, creating a dynamic equilibrium between nucleosome occupancy and accessibility.

Histone modifications play crucial roles in modulating chromatin accessibility and TF binding. Acetylation of histone tails, particularly H3K9ac and H3K14ac, neutralizes positive charges and weakens histone-DNA interactions, generally promoting a more open

chromatin structure permissive to TF binding. Conversely, methylation marks can have diverse associations with TF binding or chromatin accessibility: H3K4me3 at promoters is associated with active transcription and TF binding, while H3K27me3 deposited by Polycomb Repressive Complex 2 (PRC2) marks facultative heterochromatin and is generally associated with transcriptional repression and reduced TF accessibility (Chanvivattana et al., 2004). Constitutive heterochromatin, marked by H3K9me2 (and H3K27me1 in land plants), represents the most repressive chromatin environment, largely excluding TF binding.

Histone variants add an additional layer of complexity to TF-chromatin interactions. As demonstrated in Chapter 2, H2A variants are major determinants of chromatin state identity. H2A.Z is enriched at promoters and in facultative heterochromatin, where it may influence nucleosome stability and TF access. The plant-specific variant H2A.W, enriched in constitutive heterochromatin, creates a highly repressive environment through its C-terminal KSPK motif that strengthens DNA-histone interactions (Bourguet et al., 2021; Osakabe et al., 2021). In contrast, the canonical H2A and H2A.X variants are associated with actively transcribed euchromatin and appear more permissive to TF binding.

3.4 Pioneer TFs: Overcoming the Chromatin Barriers

While most TFs preferentially bind to accessible chromatin, a specialized class of TFs called pioneer factors can bind to nucleosomal DNA and initiate chromatin remodeling (Iwafuchi-Doi & Zaret, 2016; Zaret & Mango, 2016). Pioneer factors possess unique structural features that enable them to recognize their target sequences even when wrapped around nucleosomes. Upon binding, pioneer factors recruit chromatin remodeling complexes, such as SWI/SNF family ATPases, which can slide, evict, or restructure nucleosomes to increase DNA accessibility (Luzete-Monteiro & Zaret, 2022).

In plants, the floral regulator LEAFY (LFY) is the best-characterized pioneer factor. LFY can bind to nucleosomal DNA at its target genes APETALA1 and AGAMOUS, and recruits SWI/SNF chromatin remodelers to increase chromatin accessibility and enable binding of secondary TFs (Jin et al., 2017; Lai et al., 2021). LFY can bind its target DNA sequences even when they are wrapped within nucleosomes, a property that allows it to initiate chromatin opening in otherwise

inaccessible regions. Mechanistically, LFY displaces linker histone H1 and recruits SWI/SNF chromatin remodelers, thereby facilitating the binding of additional transcription factors and activating floral meristem identity genes. Several MADS-box TFs have also been proposed as candidate pioneer factors based on their ability to bind chromatin in repressive contexts and their roles in major developmental transitions (Pajoro et al., 2014). However, compared to the extensive characterization of pioneer factors in animal systems particularly in stem cell biology and early development, the identification and mechanistic understanding of plant pioneer factors remain limited to the TF LEAFY (Yamaguchi, 2021).

3.5 Functional and Non-Functional TF Bindings

A major challenge in understanding TF function is that genome-wide TF binding assays, such as ChIP-seq and DAP-seq, often reveal thousands of binding events, many of which do not result in measurable changes in gene expression (Alvarez et al., 2021; Swift & Coruzzi, 2017). This raises fundamental questions: What determines whether a TF binding event is functional? How does chromatin context influence TF activity versus passive binding?

Several factors may contribute to non-functional binding. First, TFs may bind to accessible DNA sequences that match their recognition motif but lack the appropriate chromatin context or co-factors necessary for transcriptional regulation. Second, highly accessible regions, particularly at active promoters, may accumulate multiple TF binding events in what are termed "highly occupied target" (HOT) regions, where binding may be redundant or non-functional (Yip et al., 2012). Third, the chromatin state itself may determine whether TF binding is productive: binding to open euchromatin may lead to transcriptional activation, while binding to facultative heterochromatin may represent poised or non-productive interactions.

In animals, chromatin state models have proven valuable for distinguishing functional modes of TF activity and understanding how chromatin context influences TF function (Ernst et al., 2011; Ernst & Kellis, 2017; Slattery et al., 2014). These studies have shown that different TF families preferentially associate with specific chromatin states, and that chromatin state can predict whether TF binding will be activating, repressive, or non-functional. However, until recently, plants have lacked a comparable framework for understanding TF-chromatin state relationships.

3.6 Methods to Study TF–Chromatin State Interactions

3.6.1 Genome-wide TF Binding Approaches

Understanding how transcription factors interact with chromatin states requires comprehensive maps of TF binding across the genome. Several complementary approaches have been developed to identify TF binding sites, each with distinct advantages and limitations.

Chromatin Immunoprecipitation sequencing (ChIP-seq) is the gold standard for mapping TF binding *in vivo* (Johnson et al., 2007). This technique uses formaldehyde to crosslink proteins to DNA in living cells, followed by chromatin fragmentation and immunoprecipitation with TF-specific antibodies. The co-precipitated DNA fragments are then sequenced to identify genomic regions bound by the TF of interest. ChIP-seq captures TF binding in its native chromatin context, reflecting the influence of nucleosome positioning, histone modifications, and chromatin accessibility on TF occupancy. However, ChIP-seq requires high-quality, specific antibodies for each TF, which limits its scalability and has resulted in binding data for only a subset of plant TFs.

In recent years, additional methods such as **Cleavage Under Targets and Tagmentation (CUT&Tag)** and **Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq)** have become widely used to map TF binding and chromatin accessibility at high resolution. CUT&Tag enables efficient profiling of TF binding with low input material and high signal-to-noise by tethering a transposase to antibody-bound TFs (Kaya-Okur et al., 2019), while ATAC-seq identifies accessible chromatin regions that are potential TF binding sites (Lu et al., 2019), providing a complementary view of the regulatory landscape.

DNA Affinity Purification sequencing (DAP-seq) offers a high-throughput alternative that bypasses the need for antibodies (O'Malley et al., 2016). In this approach, TFs are expressed in vitro and incubated with fragmented genomic DNA, allowing the TF to bind to its recognition sequences. Bound DNA is then purified and sequenced. DAP-seq can rapidly generate binding profiles for hundreds of TFs and is not constrained by chromatin accessibility, potentially revealing binding sites that may be occluded by nucleosomes *in vivo*. However, because DAP-seq is performed on naked DNA, it does not capture the influence of chromatin

structure on TF binding and may identify binding sites that are inaccessible or non-functional *in vivo*.

Position Weight Matrix (PWM) predictions provide a computational approach to identify potential TF binding sites based on DNA sequence alone. PWMs are derived from known TF binding sites and represent the probability of each nucleotide occurring at each position within the binding motif. These matrices can be scanned across genomic sequences to predict TF binding sites using tools such as FIMO (Find Individual Motif Occurrences) from the MEME Suite (Grant et al., 2011). PWM predictions are particularly valuable for species where experimental TF binding data are limited, such as *Marchantia polymorpha*. However, PWM-based predictions do not account for chromatin context, co-factor requirements, or TF expression, resulting in many false-positive predictions.

For our study (see publication II), we compiled a comprehensive dataset of TF binding profiles in *Arabidopsis thaliana* from publicly available ChIP-seq and DAP-seq experiments (Heyndrickx et al., 2014; O'Malley et al., 2016; Song et al., 2016). This dataset encompasses binding profiles for approximately 300 TFs, representing diverse TF families including WRKY, bHLH, bZIP, MYB, AP2/EREBP, MADS-box, and others. For *Marchantia polymorpha*, where experimental TF binding data are scarce, we combined 24 available DAP-seq experiments with PWM predictions. PWMs were derived from *Arabidopsis* protein-binding microarray data (Franco-Zorrilla et al., 2014b) and mapped to orthologous *Marchantia* TFs based on the conservation of DNA-binding specificity across land plants (Romani & Moreno, 2021).

Each approach provides complementary information: ChIP-seq reveals *in vivo* binding constrained by chromatin, DAP-seq identifies potential binding sites without chromatin constraints, and PWM predictions enable comparative analysis across species. By integrating these datasets with the chromatin state models described in Chapter 2, we can systematically investigate how chromatin context influences TF binding and activity.

3.6.2 Integration with Chromatin State Models

To understand how chromatin states influence TF binding, we integrated genome-wide TF binding data with the chromatin state annotations established in Chapter 2. For each TF, binding peaks identified by ChIP-seq, DAP-seq, or PWM predictions were assigned to

chromatin states based on their genomic coordinates. When a TF binding peak overlapped multiple chromatin states, the peak was assigned to the state occupying its center position, ensuring unambiguous state assignment.

We calculated TF occupancy scores to quantify the preference of each TF for each chromatin state, following the approach developed for human TF studies (Ernst & Kellis, 2013). The occupancy score represents the enrichment or depletion of TF binding in each chromatin state relative to what would be expected by chance. Specifically, for each TF and chromatin state (s), the occupancy score was calculated as:

Occupancy score = (a/b) / (c/d)

where:

a = total number of base pairs within TF peaks that overlap state (s)

b = total number of base pairs within all TF peaks genome-wide

c = total number of base pairs occupied by state (s) in the genome

d = total genome size

An occupancy score greater than 1 indicates enrichment of TF binding in that chromatin state, while a score less than 1 indicates depletion. This normalization accounts for differences in the genomic coverage of different chromatin states and the total number of binding sites for each TF, enabling fair comparisons across TFs and states.

3.6.3 TF Activity Scores distinguish Functional TF Binding

While TF occupancy scores reveal where TFs bind across chromatin states, they do not distinguish functional binding events that lead to transcriptional regulation from passive or non-productive binding. To address this limitation, we developed a TF activity score that integrates TF binding data with gene co-expression information.

For each TF, we first identified its putative target genes based on the proximity of TF binding peaks to transcription start sites. We then extracted lists of genes positively and negatively co-expressed with each TF from the ATTED-II database (Obayashi et al., 2018), which provides co-expression relationships derived from thousands of RNA-seq experiments spanning diverse developmental stages and environmental conditions. Using Fisher's exact test, we calculated the enrichment of co-expressed genes among the TF's putative targets within each chromatin state, comparing this to a random set of genes of equal size. The TF activity score was computed as:

$$\text{TF activity score} = -\log_{10}(\text{p-value}) + \log_2(\text{odds ratio} + 1)$$

This score was then normalized to the maximum score for each TF. High activity scores indicate chromatin states where TF binding is associated with coordinated gene expression changes, suggesting functional regulatory interactions. This approach reduces technical biases, particularly from *in vitro* binding methods, and highlights TF-chromatin interactions that are functionally relevant for gene regulation.

3.7 Conclusion

The integration of genome-wide TF binding data with chromatin state models provides a powerful framework for understanding gene regulation in plants. By combining experimental approaches that capture TF binding in different contexts: ChIP-seq for *in vivo* binding, DAP-seq for potential binding sites, and computational predictions for comparative analysis, with quantitative measures of chromatin state preference and functional activity, we can move beyond simple catalogs of TF binding sites to a mechanistic understanding of how chromatin context shapes TF function. The TF activity score, which filters functional from non-functional binding by incorporating gene co-expression data, represents a critical advance in distinguishing productive regulatory interactions from passive binding events. This analytical framework reveals that plant TFs are not uniform in their chromatin requirements: some TFs are restricted to highly accessible chromatin, while others can engage nucleosomal DNA in repressive chromatin states, suggesting the existence of pioneer-like factors that drive chromatin remodeling and developmental transitions. By applying these approaches across evolutionarily distant species such as *Arabidopsis* and *Marchantia*, we can identify conserved principles of TF-chromatin interactions that are fundamental to land plant biology. The following publications present the empirical foundation for these concepts: the first demonstrates how histone variants, particularly H2A variants, shape chromatin states and their dynamic regulation by chromatin remodelers such as DDM1; the second applies these chromatin state models to systematically characterize TF-chromatin state relationships in *Arabidopsis* and *Marchantia*, revealing distinct classes of TFs with specific chromatin preferences and identifying candidate pioneer factors that may orchestrate major developmental transitions in plants.

Bibliography

Alvarez, J. M., Brooks, M. D., Swift, J., & Coruzzi, G. M. (2021). Time-Based Systems Biology Approaches to Capture and Model Dynamic Gene Regulatory Networks. *Annual Review of Plant Biology*, *72*, 105–131. https://doi.org/10.1146/annurev-arplant-081320-090914

Bourguet, P., Picard, C. L., Yelagandula, R., Pélissier, T., Lorković, Z. J., Feng, S., Pouch-Pélissier, M.-N., Schmücker, A., Jacobsen, S. E., Berger, F., & Mathieu, O. (2021). The histone variant H2A.W and linker histone H1 co-regulate heterochromatin accessibility and DNA methylation. *Nature Communications*, *12*(1), 2683. https://doi.org/10.1038/s41467-021-22993-5

Chanvivattana, Y., Bishopp, A., Schubert, D., Stock, C., Moon, Y.-H., Sung, Z. R., & Goodrich, J. (2004). Interaction of Polycomb-group proteins controlling flowering in Arabidopsis. *Development (Cambridge, England)*, *131*(21), 5263–5276. https://doi.org/10.1242/dev.01400

Ernst, J., & Kellis, M. (2013). Interplay between chromatin state, regulator binding, and regulatory motifs in six human cell types. *Genome Research*, *23*(7), 1142–1154. https://doi.org/10.1101/gr.144840.112

Ernst, J., & Kellis, M. (2017). Chromatin-state discovery and genome annotation with ChromHMM. *Nature Protocols*, *12*(12), 2478–2492. https://doi.org/10.1038/nprot.2017.124

Ernst, J., Kheradpour, P., Mikkelsen, T. S., Shoresh, N., Ward, L. D., Epstein, C. B., Zhang, X., Wang, L., Issner, R., Coyne, M., Ku, M., Durham, T., Kellis, M., & Bernstein, B. E. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, *473*(7345), 43–49. https://doi.org/10.1038/nature09906

Franco-Zorrilla, J. M., López-Vidriero, I., Carrasco, J. L., Godoy, M., Vera, P., & Solano, R. (2014a). DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(6), 2367–2372. https://doi.org/10.1073/pnas.1316278111

Franco-Zorrilla, J. M., López-Vidriero, I., Carrasco, J. L., Godoy, M., Vera, P., & Solano, R. (2014b). DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(6), 2367–2372. https://doi.org/10.1073/pnas.1316278111

Grant, C. E., Bailey, T. L., & Noble, W. S. (2011). FIMO: Scanning for occurrences of a given motif. *Bioinformatics (Oxford, England)*, *27*(7), 1017–1018. https://doi.org/10.1093/bioinformatics/btr064

Heyndrickx, K. S., Van de Velde, J., Wang, C., Weigel, D., & Vandepoele, K. (2014). A functional and evolutionary perspective on transcription factor binding in Arabidopsis thaliana. *The Plant Cell*, *26*(10), 3894–3910. https://doi.org/10.1105/tpc.114.130591

Iwafuchi-Doi, M., & Zaret, K. S. (2016). Cell fate control by pioneer transcription factors. *Development (Cambridge, England)*, *143*(11), 1833–1837. https://doi.org/10.1242/dev.133900

Jin, J., Tian, F., Yang, D.-C., Meng, Y.-Q., Kong, L., Luo, J., & Gao, G. (2017). PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Research*, *45*(D1), D1040–D1045. https://doi.org/10.1093/nar/gkw982

Johnson, D. S., Mortazavi, A., Myers, R. M., & Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. *Science (New York, N.Y.)*, *316*(5830), 1497–1502. https://doi.org/10.1126/science.1141319

Kaya-Okur, H. S., Wu, S. J., Codomo, C. A., Pledger, E. S., Bryson, T. D., Henikoff, J. G., Ahmad, K., & Henikoff, S. (2019). CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nature Communications*, *10*(1), 1930. https://doi.org/10.1038/s41467-019-09982-5

Lai, X., Blanc-Mathieu, R., GrandVuillemin, L., Huang, Y., Stigliani, A., Lucas, J., Thévenon, E., Loue-Manifel, J., Turchi, L., Daher, H., Brun-Hernandez, E., Vachon, G., Latrasse, D., Benhamed, M., Dumas, R., Zubieta, C., & Parcy, F. (2021). The LEAFY floral regulator displays pioneer transcription factor properties. *Molecular Plant*, *14*(5), 829–837. https://doi.org/10.1016/j.molp.2021.03.004

Lorch, Y., LaPointe, J. W., & Kornberg, R. D. (1987). Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. *Cell*, *49*(2), 203–210. https://doi.org/10.1016/0092-8674(87)90561-7

Lu, Z., Marand, A. P., Ricci, W. A., Ethridge, C. L., Zhang, X., & Schmitz, R. J. (2019). The prevalence, evolution and chromatin signatures of plant regulatory elements. *Nature Plants*, *5*(12), 1250–1259. https://doi.org/10.1038/s41477-019-0548-z

Luzete-Monteiro, E., & Zaret, K. S. (2022). Structures and consequences of

pioneer factor binding to nucleosomes. *Current Opinion in Structural Biology*, *75*, 102425. https://doi.org/10.1016/j.sbi.2022.102425

Obayashi, T., Aoki, Y., Tadaka, S., Kagaya, Y., & Kinoshita, K. (2018). ATTED-II in 2018: A Plant Coexpression Database Based on Investigation of the Statistical Property of the Mutual Rank Index. *Plant and Cell Physiology*, *59*(1), e3. https://doi.org/10.1093/pcp/pcx191

O'Malley, R. C., Huang, S.-S. C., Song, L., Lewsey, M. G., Bartlett, A., Nery, J. R., Galli, M., Gallavotti, A., & Ecker, J. R. (2016). Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. *Cell*, *165*(5), 1280–1292. https://doi.org/10.1016/j.cell.2016.04.038

Osakabe, A., Jamge, B., Axelsson, E., Montgomery, S. A., Akimcheva, S., Kuehn, A. L., Pisupati, R., Lorković, Z. J., Yelagandula, R., Kakutani, T., & Berger, F. (2021). The chromatin remodeler DDM1 prevents transposon mobility through deposition of histone variant H2A.W. *Nature Cell Biology*, *23*(4), 391–400. https://doi.org/10.1038/s41556-021-00658-1

Pajoro, A., Madrigal, P., Muiño, J. M., Matus, J. T., Jin, J., Mecchia, M. A., Debernardi, J. M., Palatnik, J. F., Balazadeh, S., Arif, M., Ó'Maoiléidigh, D. S., Wellmer, F., Krajewski, P., Riechmann, J.-L., Angenent, G. C., & Kaufmann, K. (2014). Dynamics of chromatin accessibility and gene regulation by MADS-domain transcription factors in flower development. *Genome Biology*, *15*(3), R41. https://doi.org/10.1186/gb-2014-15-3-r41

Romani, F., & Moreno, J. E. (2021). Molecular mechanisms involved in functional macroevolution of plant transcription factors. *The New Phytologist*, *230*(4), 1345–1353. https://doi.org/10.1111/nph.17161

Slattery, M., Zhou, T., Yang, L., Dantas Machado, A. C., Gordân, R., & Rohs, R. (2014). Absence of a simple code: How transcription factors read the genome. *Trends in Biochemical Sciences*, *39*(9), 381–399. https://doi.org/10.1016/j.tibs.2014.07.002

Song, L., Huang, S. C., Wise, A., Castanon, R., Nery, J. R., Chen, H., Watanabe, M., Thomas, J., Bar-Joseph, Z., & Ecker, J. R. (2016). A transcription factor hierarchy defines an environmental stress response network. *Science*, *354*(6312), aag1550. https://doi.org/10.1126/science.aag1550

Swift, J., & Coruzzi, G. M. (2017). A matter of time—How transient

transcription factor interactions create dynamic gene regulatory networks. *Biochimica Et Biophysica Acta. Gene Regulatory Mechanisms*, *1860*(1), 75–83. https://doi.org/10.1016/j.bbagrm.2016.08.007

Wasylyk, B., Thevenin, G., Oudet, P., & Chambon, P. (1979). Transcription of in vitro assembled chromatin by Escherichia coli RNA polymerase. *Journal of Molecular Biology*, *128*(3), 411–440. https://doi.org/10.1016/0022-2836(79)90095-0

Wong, F., & Gunawardena, J. (2020). Gene Regulation in and out of Equilibrium. *Annual Review of Biophysics*, *49*, 199–226. https://doi.org/10.1146/annurev-biophys-121219-081542

Yamaguchi, N. (2021). LEAFY, a Pioneer Transcription Factor in Plants: A Mini-Review. *Frontiers in Plant Science*, *12*, 701406. https://doi.org/10.3389/fpls.2021.701406

Yip, K. Y., Cheng, C., Bhardwaj, N., Brown, J. B., Leng, J., Kundaje, A., Rozowsky, J., Birney, E., Bickel, P., Snyder, M., & Gerstein, M. (2012). Classification of human genomic regions based on experimentally determined binding sites of more than 100 transcription-related factors. *Genome Biology*, *13*(9), R48. https://doi.org/10.1186/gb-2012-13-9-r48

Zaret, K. S., & Mango, S. E. (2016). Pioneer transcription factors, chromatin dynamics, and cell fate control. *Current Opinion in Genetics & Development*, *37*, 76–81. https://doi.org/10.1016/j.gde.2015.12.003

Chapter 4

4. List of Publications

Publication I

Jamge, B., Lorković, Z. J., Axelsson, E., Osakabe, A., **Shukla, V.**, Yelagandula, R., Akimcheva, S., Kuehn, A. L., & Berger, F. (2023). Histone variants shape chromatin states in Arabidopsis. eLife, 12, RP87714.

https://doi.org/10.7554/eLife.87714.3

Contribution: Resources, Data curation, Software, Formal analysis.

The manuscript was published in eLife on July 19th, 2023.

Publication II

Shukla, V., Axelsson, E., Hisanaga, T., Haseloff, J., Berger, F., & Romani, F. (2025). Conservation of chromatin states and their association with transcription factors in land plants (p. 2025.07.12.664529). bioRxiv.

https://doi.org/10.1101/2025.07.12.664529

Contribution: Methodology, Validation, Formal Analysis, Investigation, Data Curation, Writing – original draft, and Visualization.

The manuscript was posted on bioRxiv on July 17th, 2025. The revised manuscript is currently under consideration for publishing with the journal as of November 24, 2025.

Histone variants shape chromatin states in Arabidopsis

Jamge, B., Lorković, Z. J., Axelsson, E., Osakabe, A., **Shukla, V.**, Yelagandula, R., Akimcheva, S., Kuehn, A. L., & Berger, F. (2023). Histone variants shape chromatin states in Arabidopsis. eLife, 12, RP87714.

https://doi.org/10.7554/eLife.87714.3

Contribution: Resources, Data curation, Software, Formal analysis.

The manuscript was published in eLife on July 19th, 2023.

RESEARCH ARTICLE

Histone variants shape chromatin states in Arabidopsis

Bhagyshree Jamge[1,2†], Zdravko J Lorković[1†], Elin Axelsson[1†], Akihisa Osakabe[1,3,4], Vikas Shukla[1,2], Ramesh Yelagandula[1,5‡], Svetlana Akimcheva[1], Annika Luisa Kuehn[1], Frédéric Berger[1]*

[1]Gregor Mendel Institute, Austrian Academy of Sciences, Vienna BioCenter, Vienna, Austria; [2]Vienna BioCenter, Vienna, Austria; [3]Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Hongo, Bunkyo-ku, Tokyo, Japan; [4]PRESTO, Japan Science and Technology Agency, Honcho, Kawaguchi, Japan; [5]Institute of Molecular Biotechnology, IMBA, Dr. Bohr-Gasse 3, Vienna, Austria

***For correspondence:** Frederic.berger@gmi.oeaw.ac.at

[†]These authors contributed equally to this work

Present address: [‡]Laboratory of Epigenetics, Cell fate & Disease, Centre for DNA Fingerprinting and Diagnostics (CDFD), Hyderabad, India

Competing interest: The authors declare that no competing interests exist.

Funding: See page 21

Preprint posted 10 March 2023
Sent for Review 22 March 2023
Reviewed preprint posted 18 May 2023
Reviewed preprint revised 04 July 2023
Version of Record published 19 July 2023

Reviewing Editor: Paula Casati, Center of Photosynthetic And Biochemical Studies (CEFOBI), Argentina

Summary How different intrinsic sequence variations and regulatory modifications of histones combine in nucleosomes remain unclear. To test the importance of histone variants in the organization of chromatin we investigated how histone variants and histone modifications assemble in the *Arabidopsis thaliana* genome. We showed that a limited number of chromatin states divide euchromatin and heterochromatin into several subdomains. We found that histone variants are as significant as histone modifications in determining the composition of chromatin states. Particularly strong associations were observed between H2A variants and specific combinations of histone modifications. To study the role of H2A variants in organizing chromatin states we determined the role of the chromatin remodeler DECREASED IN DNA METHYLATION (DDM1) in the organization of chromatin states. We showed that the loss of DDM1 prevented the exchange of the histone variant H2A.Z to H2A.W in constitutive heterochromatin, resulting in significant effects on the definition and distribution of chromatin states in and outside of constitutive heterochromatin. We thus propose that dynamic exchanges of histone variants control the organization of histone modifications into chromatin states, acting as molecular landmarks.

eLife assessment

This study presents an **important** description on the dynamics of histone variant exchange controlling the organization of the chromatin state of the Arabidopsis genome, combining the analysis of histone variants, histone modification, and chromatin states. The evidence supporting the claims of the authors is **compelling**. This work will be of great interest to those in the field of epigenetics and chromatin biology.

Introduction

Eukaryotic genomes are packaged into chromatin, a structure defined by repeating units of ~147 bp of DNA wrapped around a protein complex known as the nucleosome (***Luger et al., 1997***). Nucleosomes contain two copies of each of the core histones H2A, H2B, H3, and H4. Histone variants have arisen through the divergence of their intrinsically disordered loop and tail regions and dozens of histone variants in the H2A, H2B, and H3 families have been identified in eukaryotes (***Loppin and Berger, 2020***; ***Talbert and Henikoff, 2021***). These variants regulate the properties of nucleosomes (***Chakravarthy et al., 2004***; ***Horikoshi et al., 2013***; ***Koyama and Kurumizaka, 2018***; ***Tachiwana et al., 2012***) and affect transcription (***Rudnizky et al., 2016***; ***Subramanian et al., 2015***; ***Weber and***

Henikoff, 2014; *Wollmann et al., 2012*; *Wollmann et al., 2017*). Animals and plants have evolved unique H2A variants associated with transcriptional repression (*Loppin and Berger, 2020*; *Talbert and Henikoff, 2021*) and with gametes (*Osakabe and Molaro, 2023*). In vascular plants, the H2A.W variants are distinguished by a C-terminal KSPK motif (*Lei et al., 2021*; *Bourguet et al., 2021*; *Yelagandula et al., 2014*). In synergy with H3K9 methylation, H2A.W favors heterochromatic silencing by directly altering the interaction between H2A.W and DNA (*Bourguet et al., 2021*; *Osakabe et al., 2021*; *Bourguet et al., 2021*; *Yelagandula et al., 2014*). The roles of the recently recognized H2B variants (*Jiang et al., 2020*; *Raman et al., 2022*) in transcription remains unknown. However, H2B variants might differ in their capacity to be ubiquitinated, an important modification for RNA polymerase II elongation (*Feng and Shen, 2014*; *Kim et al., 2009*; *Minsky et al., 2008*). Histone variants thus represent complex, diverse nucleosome components that could regulate and organize the functional activities of chromatin.

Histones are subject to a wide range of post-translational modifications (*Talbert and Henikoff, 2021*; *Tessarz and Kouzarides, 2014*). Applying a Hidden Markov Model (ChromHMM) (*Ernst and Kellis, 2012*; *Ernst and Kellis, 2017*) to genomic profiles of histone modifications revealed chromatin states that reflect the combination of post-transcriptional histone modifications present in mammals, *Drosophila*, *Arabidopsis*, and rice (*Ernst and Kellis, 2012*; *Ernst and Kellis, 2017*; *Kharchenko et al., 2011*; *Liu et al., 2018*; *Roudier et al., 2011*; *Sequeira-Mendes et al., 2014*). Chromatin states distinguish the major domains of chromatin within the genome, comprising euchromatin (with active genes), facultative heterochromatin (with repressed genes), and constitutive heterochromatin (with transposons and repeats) in animals and plants. They also highlight specific features such as promoters and enhancers.

Compared with the wealth of data on the relationship between histone modifications and transcription, the role of histone variants in this process remains to be explored. In *Arabidopsis*, there is a clear correlation between transcript levels and enrichment of the variant H3.3 (*Stroud et al., 2012*; *Wollmann et al., 2012*; *Wollmann et al., 2017*) and H2A and H2A.X (*Yelagandula et al., 2014*) on gene bodies marked by H3K36me2 (*Leng et al., 2020*). In contrast, H2A.Z is enriched on repressed genes marked by H3K27me3 (*Carter et al., 2018*), and the histone variant H2A.W is present on transposons and repeats marked by H3K9me2 (*Osakabe et al., 2021*; *Yelagandula et al., 2014*). Although it was suggested that histone variants might have a global role in defining chromatin states (*Hake and Allis, 2006*; *Henikoff et al., 2004*; *Weber and Henikoff, 2014*) the complex, intimate associations between histone variants and modifications have not been fully characterized in plants or mammals.

Here, we analyzed how all thirteen histone variants expressed in vegetative tissues associate with twelve prominent histone modifications to form chromatin states in the model flowering plant *Arabidopsis thaliana*. Our findings indicate that H2A variants are major factors that differentiate euchromatin, facultative heterochromatin, and constitutive heterochromatin. This hypothesis is supported by in silico analyses and the mis-assembly of chromatin states caused by the deregulation of the exchange of H2A variants.

Results

Co-occurrence of H3 modifications and histone variants in nucleosomes

In *Arabidopsis*, homotypic nucleosomes containing a single type of H2A variant are prevalent (*Osakabe et al., 2018*). However, it was not determined whether H3 variants also assemble in a homotypic manner and if H2A variants preferably assemble with a specific H3 variant or histone modification. To answer these questions, we immunoprecipitated mononucleosomes (*Figure 1—figure supplement 1A-C*) from transgenic *Arabidopsis* lines expressing HA-tagged *HTR13* (H3.1) and *HTR5* (H3.3) under the control of their native promoters in the respective mutant backgrounds (*Jiang and Berger, 2017*). Because tagged H3 ran slower than endogenous H3 on SDS-PAGE (*Figure 1—figure supplement 1B*), this approach allowed us to analyze the composition of H3.1 and H3.3 nucleosomes by mass spectrometry (MS) and immunoblotting. Spectral counts of peptides covering regions around lysine 27 (K27), where H3.1 and H3.3 can be distinguished by MS analysis, revealed that ~60% of H3.1 and ~42% of H3.3 nucleosomes contained both H3 variants, (*Figure 1A*). We analyzed H3.1 and H3.3 nucleosomes for the presence of H2A variants by western blotting and found that neither H3.1 nor H3.3 was preferentially associated with a specific H2A variant (*Figure 1B*). This was further confirmed

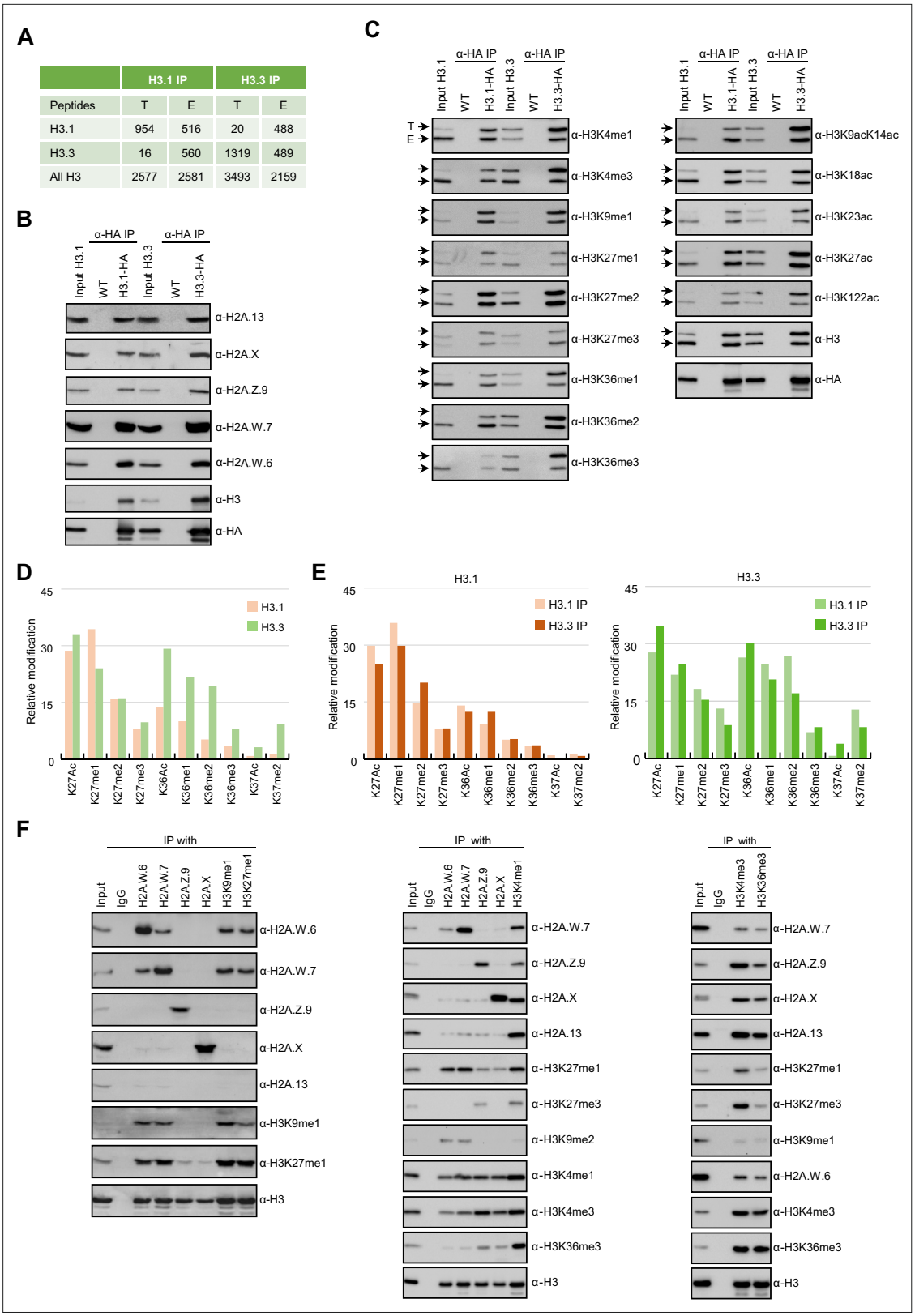


Figure 1. Biochemical analysis of the association between histone variants and histone marks. (**A**) Histones H3.1 and H3.3 form homotypic and heterotypic nucleosomes. Spectral counts of H3.1- and H3.3-specific peptides in the respective immunoprecipitations (T – transgenic, E – endogenous H3.1 and H3.3). (**B**) H2A variants do not preferentially associate with H3.1- or H3.3-containing nucleosomes. HA-tagged H3.1 and H3.3 mononucleosomes were immunoprecipitated with HA agarose and analyzed for the presence of H2A variants by immunoblotting. (**C**) Histone H3 marks

Figure 1 continued on next page

Figure 1 continued

are present on both H3.1 and H3.3. HA-tagged H3.1 and H3.3 mononucleosomes were immunoprecipitated with HA agarose and analyzed for the presence of H3 marks by immunoblotting. Arrows indicate transgenic (**T**) and endogenous (**E**) H3. (**D**) Mass spectrometry (MS) analysis of cumulative H3K27, H3K36, and H3K37 modifications on H3.1 and H3.3. All measured spectra corresponding to H3.1 and H3.3 peptides from both IPs were used for analysis. (**E**) Relative abundance of H3K27, H3K36, and H3K37 modifications on H3.1 variant analyzed separately from MS data of H3.1 and H3.3 purified nucleosomes (left panel). Relative abundance of H3K27, H3K36, and H3K37 modifications on H3.3 variant analyzed separately from MS data of H3.1 and H3.3 purified nucleosomes (right panel). (**F**) Co-occurrence of H2A variants and H3 marks. Mononucleosomes were immunoprecipitated with the indicated antibodies and analyzed for the presence of H2A variants and H3 marks by western blotting. Original pictures of the gels are provided in ***Figure 1—source data 1***, ***Figure 1—source data 2*** and ***Figure 1—source data 3***.

The online version of this article includes the following source data and figure supplement(s) for figure 1:

Source data 1.

Source data 2. The data contains the orginal images of the gels.

Source data 3. The data contains the orginal images of the gels.

Figure supplement 1. Biochemical analysis of the association between histone variants and histone marks.

Figure supplement 1—source data 1. The data contains the orginal images of the gels.

by MS analysis of nucleosomes pulled down with H2A variant-specific antibodies (***Figure 1—figure supplement 1D***). Thus, by contrast with H2A variants, H3 variants do not necessarily form homotypic nucleosomes, despite the presence of H3 variant-specific deposition mechanisms (***Loppin and Berger, 2020***; ***Nie et al., 2014***; ***Probst, 2022***), and they do not associate with a specific H2A variant.

Next, we explored the potential link between H3 modifications and H3 variants. We found that all the methylation and acetylation marks surveyed were present on transgenic (T) and endogenous (E) H3 in both H3.1 and H3.3 immunoprecipitations (***Figure 1C***), although there was some degree of preferential association of H3K4me, H3K36me, and H3 acetylation with H3.3 and of H3K27me1 with H3.1 (***Figure 1C***). H3.1 and H3.3 modifications were also analyzed by MS by using information from both transgenic and endogenous H3 in respective immunoprecipitations. This approach unambiguously distinguishes the modifications of peptides covering lysines 27 (H3K27) and 36 (H3K36) originating from the endogenous or the transgenic copies of H3.1 and H3.3 (***Figure 1—figure supplement 1E***). We put our focus on modifications at these two positions, because they are diagnostic of constitutive heterochromatin (H3K27me1), facultative heterochromatin (H3K27me3), and actively transcribed euchromatin (H3K36me). In contrast to previous reports (***Johnson et al., 2004***; ***Zhang et al., 2007***) we identified similar levels of H3K27me2 and H3K27me3 marks on both H3.1 and H3.3 whereas the constitutive heterochromatin mark H3K27me1 was higher on H3.1 (***Figure 1D and E***; ***Figure 1—figure supplement 1E-G***). High levels of H3K27me1 on H3.3 were unexpected considering the specificity of H3K27 methyltransferases ATXR5/6 for H3.1 (***Jacob et al., 2014***) and are the likely product of the demethylation of H3K27me3 by the JUMONJI H3K27 demethylases (***Gan et al., 2015***; ***Liu et al., 2010***) or intermediates of methylation by the Polycomb Repressive Complex 2. Acetylation and all three methylation states at H3K36 that are associated with active chromatin were likewise found on both H3.1 and H3.3 but with higher enrichment on H3.3 (***Figure 1D and E***; ***Figure 1—figure supplement 1E-G***). This is consistent with western blot analyses (***Figure 1C***) and with the established role of H3.3 as a replacement variant during transcription (***Delaney and Almouzni, 2023***). The levels of H3K27 and H3K36 modifications on H3.1 and H3.3 displayed the same trends irrespective of whether they were precipitated with either H3.1 or H3.3 nucleosomes, contrasting with the predominant enrichment of H3K37 modifications on H3.3 (***Figure 1E***). The H3 region covering K4, K9, K14, K18, and K23 residues cannot be distinguished between H3.1 and H3.3 by either bottom-up MS or western blotting. Therefore, we analyzed H3K9me1 and H3K9me2, which mark constitutive heterochromatin in plants, only on transgenic H3 and found that both marks were similarly distributed among H3.1 and H3.3 (***Figure 1—figure supplement 1H***) as reported previously (***Johnson et al., 2004***). Also, acetylation of H3K9, H3K14, H3K18, H3K23 and combinations thereof were very similar in H3.1 and H3.3 immunoprecipitations (***Figure 1—figure supplement 1I-K***). Modifications on H3K4 were not analyzed by MS because our experimental setup produced too short peptides to be captured by nanoLC. As the total numbers of measured spectra covering H3 modifications were nearly identical for H3.1 and H3.3 (***Figure 1—figure supplement 1F***) our MS together with western blot data suggest that the vast majority of H3 modifications are not H3 variant specific to the notable exception of H3K36 and H3K37

modifications that are associated preferentially but not uniquely with H3.3. This preference for H3.3 might be related to the presence of substitutions at position 41 specific that distinguishes H3.1 from H3.3 in plants (***Borg et al., 2021***; ***Probst, 2022***). These findings suggested that in the heterotypic nucleosomes the same modifications are found on either H3 variants present irrespective of their identity.

Unlike H3 variants, homotypic nucleosomes containing either H2A.W, H2A.Z, H2A, or H2A.X showed marked enrichment of specific histone modifications (***Figure 1F***). Histone modifications in constitutive heterochromatin (H3K9me1, H3K9me2, and H3K27me1) were primarily associated with H2A.W consistent with their synergistic impact on silencing transposons (***Bourguet et al., 2021***). The modification H3K27me3 (facultative heterochromatin) was detected only in H2A.Z nucleosomes, and H3K36me3 (marking euchromatin) was primarily detected in H2A, H2A.X, and H2A.Z nucleosomes. Other modifications, such as H3K4me1 and H3K4me3, displayed complex patterns of association with H2A variants, among themselves, and with other H3 modifications (***Figure 1F***). Surprisingly, H3K4me1 and H3K4me3 co-precipitated high levels of H3K27me1, H3K27me3, and H3K36me3 and low levels of H3K9me1 and H3K9me2 (***Figure 1F***) showing that repressive and active H3 marks co-exist on the same nucleosomes or potentially on the same H3 tail. Similar complex interplay between H3 marks has also been observed in mouse ES cells (***Schwämmle et al., 2016***). Therefore, H2A but not H3 variants form homotypic nucleosomes that preferentially carry specific histone modifications associated with either the transcriptional status of protein-coding genes or transposons.

Associations of histone variants with the chromatin states

We performed ChIP-seq using *Arabidopsis* seedlings to identify the combinations of all histone variants present in somatic cells and their associations with twelve prominent histone modifications. We included the most abundant isoforms of H2A.W (H2A.W.6 and H2A.W.7), H2A.Z (H2A.Z.9 and H2A.Z.11), and H2A (H2A.2 and H2A.13) in our analysis. We used specific polyclonal antibodies to detect each H2A variant (see ***Figure 2—figure supplement 1***) for the demonstration of antibodies against H2A.Z.11 and H2A.2 and (***Osakabe et al., 2021***; ***Osakabe et al., 2018***). The algorithm ChromHMM (***Ernst and Kellis, 2017***) was used to define chromatin states in *Arabidopsis* (see Methods for details). We chose to analyze the 26 chromatin states model, which proved to be most parsimonious (***Figure 2—figure supplement 2A***; see Methods). Chromatin states were clustered based on the emission probability for each modification and histone variant (***Figure 2A***) and were distinguished based on distinct combinations of enrichment amongst the 27 chromatin components analyzed (***Figure 2A***). Chromatin states occupied different proportions of the genome (***Figure 2—figure supplement 2B***) and showed distinct relative abundance of transposons, repeats, and elements of protein-coding genes (***Figure 2B***). For each chromatin state, we measured the average level of transcriptional activity (***Figure 2C***), the degree of enrichment in CG methylation (***Figure 2D***)**,** the degree of enrichment in CHG and CHH methylation (***Figure 2—figure supplement 2C, D***), the level of accessibility by DNase I-seq (***Figure 2E***), the nucleosome density determined by MNase-seq (***Figure 2—figure supplement 2E***), and the length in base pairs (***Figure 2—figure supplement 2F***). Chromatin states H1–H6 showed low accessibility, the highest levels of DNA methylation, and primarily occupied transcriptionally inactive transposons and repeats, thus representing constitutive heterochromatin. The states F1–F6 were associated with low transcriptional activity and were present over genes and pseudogenes, indicating that they represented facultative heterochromatin. As defined by state occupancy, constitutive and facultative heterochromatin composed 17% and 20% of the genome, respectively (***Figure 2—figure supplement 2B***), corresponding to previous estimates (***Roudier et al., 2011***). The states E1–E11 occupied expressed genes and thus comprised euchromatin. Three states (I1–I3) showed the lowest nucleosome density and highest accessibility and occupied a large fraction of non-coding RNAs (***Figure 2—figure supplement 2G***) and a quarter of untranslated regions of genes (3′UTR and 5′UTR) (***Figure 2B***). Hence, the states I1–I3 were classified as intergenic. We conclude that specific groups of chromatin states define constitutive heterochromatin (H1-6), facultative heterochromatin (F1-6), euchromatin (E1-11), and intergenic regions (I1-3).

Overall, the chromatin states recapitulate the preferential associations between H2A variants and histone modifications observed in mononucleosomes (***Figure 1***; ***Figure 1—figure supplement 1***). H2A.W and H3K9me1 are the determining marks of all six heterochromatic states. H2A.Z, with the polycomb histone modifications H2AK121ub and H3K27me3, are the hallmark of facultative

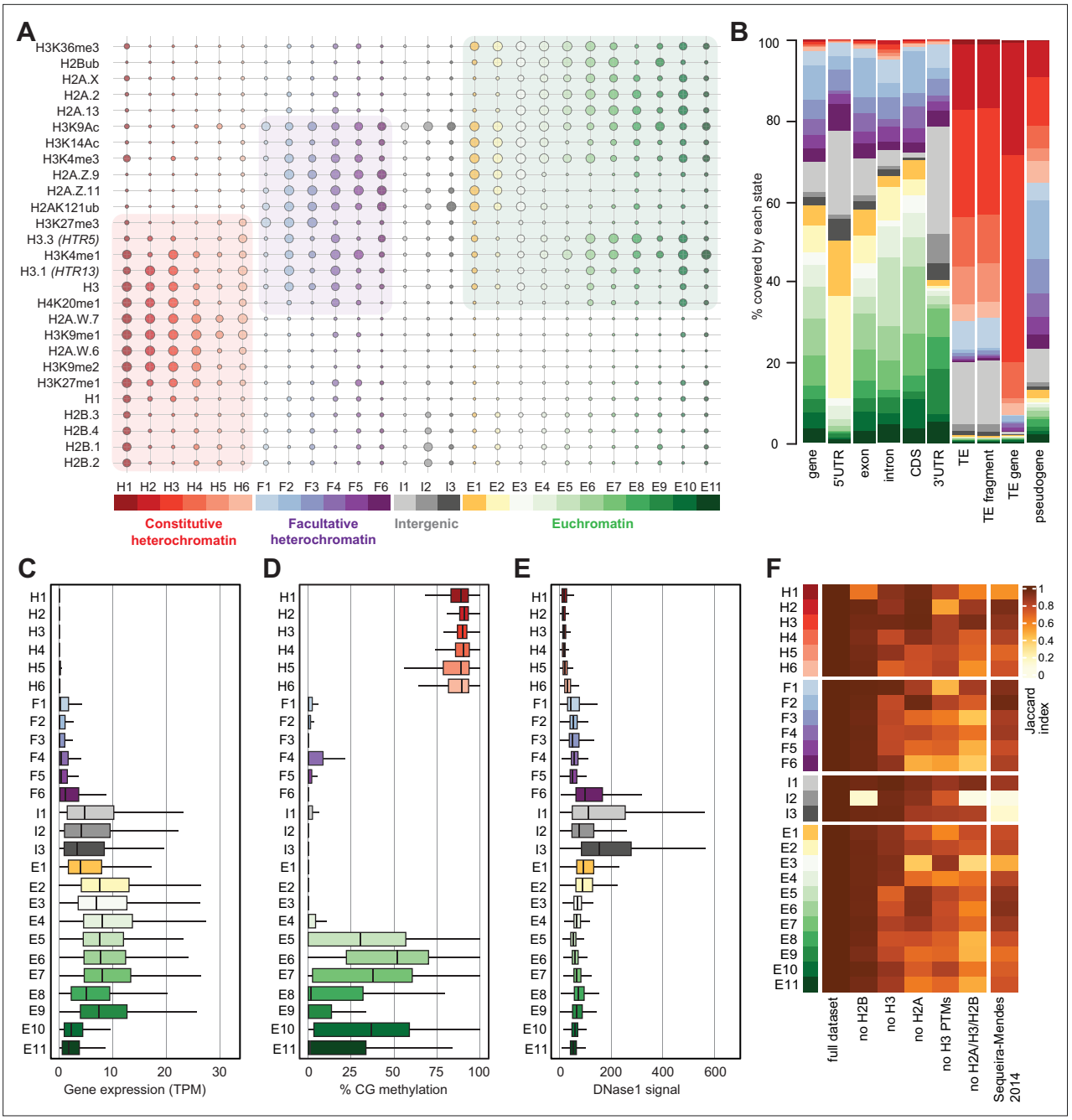


Figure 2. Histone variants define chromatin states in *Arabidopsis thaliana*. (**A**) Bubble plot showing the emission probabilities for histone modifications/variants across the 26 chromatin states. (The size of the bubble represents the emission probability ranging from 0 to 1). The colors are ascribed for each type of chromatin. (**B**) Stacked bar plot showing the overlap between annotated genomic features and chromatin states. (**C**) Box plot showing the expression of protein-coding genes overlapping with each chromatin state in Transcripts per Million (TPM). (**D**) Box plot showing levels of CG methylation across chromatin states. (**E**) Box plot comparing DNase I-seq read coverage across chromatin states representing chromatin accessibility. (**F**) 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. The comparison with a 9-state model (***Sequeira-Mendes et al., 2014***) did not include CG content, DNA methylation, H4K5ac, and H3K4me2 which were not used in the 26-state model. H2B did not seem to make a significant contribution. Excluding H3 did reduce effectively the Jaccard index for both the 26 and the 9 chromatin states model, which both included H3.1 and H3.3. The major differences between the 26 and the 9 chromatin states model are in the intergenic states. Overall H2A variants affected most strongly the Jaccard index.

The online version of this article includes the following source data and figure supplement(s) for figure 2:

Source data 1. This source data contain the original pictures of the gels.

Figure supplement 1. Validation of H2A.2 and H2A.

Figure 2 continued on next page

Figure 2 continued

Figure supplement 1—source data 1. This source data contain the original pictures of the gels.

Figure supplement 2. Properties of chromatin states in *Arabidopsis thaliana*.

heterochromatin states. H2A and H2A.X associated with H2Bub and H3K36me3 mark euchromatic states. Contrasting with the strong association between H2A variants and the three major domains of chromatin, much less prominent associations are observed between histone modifications and either H3 or H2B variants.

To examine the importance of H2A variants in the definition of chromatin states, we compared the 26 chromatin states defined in our study with the 9 states defined in a previous study that did not include the comprehensive set of histone variants present in *Arabidopsis* chromatin (***Sequeira-Mendes et al., 2014***; ***Figure 2—figure supplement 2H***). The blocks of heterochromatic states H1–H6 corresponded to the previously identified states 8 and 9, which were also defined as heterochromatin in a previous study (***Sequeira-Mendes et al., 2014***). States F1–F6 tended to map to the previously defined facultative heterochromatin states 5 and 6, although state F6 was split among the three states 2, 4, and 6. Similarly, although there was a broad correspondence between euchromatin states E1– E11 with states 1, 3, and 7, there were noticeable differences. Overall, several newly defined states were not associated with the corresponding types of chromatin domains described in previous studies. Previously defined states 2, 4, and 6 contained elements belonging to multiple newly defined chromatin states. To test the dissimilarity between the 9-state and the 26-state models we calculated a 26-state model based on the chromatin marks used to define the 9-state model. The difference between the resulting matrices were measured by the Jaccard index (***Figure 2F***). This showed an overall coherence between the two models, with only mis-assignment of chromatin identity by the 9-state model of chromatin in intergenic regions, H1, H5, H6, and a lower diversity of euchromatin states. Altogether, the addition of the comprehensive set of histone variants in the 26-state model provides a more refined and coherent classification of elements of chromatin than the 9-state model defined primarily based on chromatin modifications and the variants H3.1 and H3.3, suggesting the importance of histone variants in the definition of chromatin states.

To evaluate the contribution of individual components in defining the chromatin states, we recalculated chromatin states after excluding either H2B variants, H3 variants, H2A variants, H3 modifications (no H3 PTMs) or all histone variants (no H2A/H3/H2B) from the dataset used to learn the 26-state model. The difference between the resulting matrices of chromatin states was measured by the Jaccard index (***Figure 2F***). Removing H2B variants did not affect most chromatin states, except for H1 and I2, which showed high emissions probabilities for H2B variants (***Figure 2A***). In contrast, H3 variants showed larger contributions in defining specific states, and removing all H2A variants caused the loss of several states marked by a Jaccard index <0.7 (***Figure 2F***). Eventually, removing all histone variants had a stronger impact than removing all H3 modifications from the model. In summary, our computational and biochemical analyses led to conclude that, among histone variants, H2A variants have the strongest effect on defining the chromatin states by their association with histone modifications.

Perturbation of H2A variant dynamics affects the genomic localization of chromatin states

We considered the idea that the dynamics of the exchange of H2A variants would affect the chromatin states. To test this idea, we studied the impact of the loss of the chromatin remodeler DDM1, which binds and controls the deposition of H2A.W (***Osakabe et al., 2021***), on the definition of chromatin states. In the *ddm1* mutant, TEs show a loss of DNA methylation, H3K9 methylation, and linker histone H1 as well as an increase in H3K27me3 (***Jeddeloh et al., 1998***; ***Kakutani et al., 1999***; ***Osakabe et al., 2021***; ***Rougée et al., 2021***). This suggested that the loss of H2A.W deposition had a profound impact on chromatin states.

In our previous study, we used the leaves of five-week-old plants to show the impact of *ddm1* on the profiles of H2A.W.6, H2A.X, H1, H3K9me2, H3K36me3 and H3K27me3 (***Osakabe et al., 2021***). This study showed that DDM1 causes the deposition of H2A.W.6 to heterochromatin and we thus extended this investigation to the two other marks of heterochromatin H3K9me1, H2A.W.7 and

H2A.Z.9 and H2A.Z.11 in leaves (11 profiles in total) of *ddm1* and wild-type. Because these profiles were acquired on leaves while seedlings were used to define the 26-state model in the wild-type, we first considered whether the development stage could affect the definition of chromatin states in the wild-type. To test this idea, we built a concatenated chromHMM model based on the profiles of the same set of histone variants and modifications from 10-day-old seedlings and leaves of five-week-old plants (***Figure 3—figure supplement 1A***). Although this model had only 15 chromatin states we observed an association between H2A variants and histone modifications comparable to the model describing the 26-state model (***Figure 2A***). The absence of most typical euchromatic histone modifications in the model likely explained the strong reduction of the number of states of euchromatin compared with the 26-state model (***Figure 3—figure supplement 1A***). The complexity of the different heterochromatin states were preserved in the concatenated model, supporting that the set of H2A variants and histone modifications used to study *ddm1* are sufficient to represent the complexity of the heterochromatin affected directly by *ddm1*. The composition of the chromatin states did not vary significantly between seedlings and leaves and each state occupied a similar proportion of the genome in seedlings or leaves to the exception of state 5 present primarily in leaves and state 13 only present in seedlings (***Figure 3—figure supplement 1A*** right column with green bars). However, as expected by the dissimilar transcriptomes of these two developmental stages euchromatic states occupied distinct genes in seedlings and leaves (***Figure 3—figure supplement 1B***). This indicated that neither tissue type nor developmental stage has a dramatic effect on the definition of chromatin states in the wild-type.

To directly compare chromatin states between *ddm1* and wild-type, we built a concatenated chromHMM model (***Ernst and Kellis, 2017***). The concatenated model created a common set of 16 chromatin states that are shared between wild-type and *ddm1* (***Figure 3A***) and identified the percentage of the genome occupied by each chromatin states for each genotype individually (see Methods) (***Figure 3A*,** and bar plot on the left). The states were classified as constitutive heterochromatin (hI-hIII), facultative heterochromatin (fI-fVI), euchromatin (eI-eIII), and intergenic (iI) (***Figure 3A***), based on their emission probability and genomic overlap with states from those groups in the 26-state model (see Methods). The presence of chromatin profiles from *ddm1* caused three additional mixed states in the 16-state model, which represents regions covered by facultative and constitutive heterochromatin (fhI) or by a combination of intergenic and constitutive heterochromatin (ihI,ihII) in the 26-state model. Because several euchromatic histone PTMs were not included in the new model, only three states represented euchromatin (eI-eIII) compared with the 11 euchromatic states in the 26-state model. In the 16-state concatenated model with *ddm1* the complexity of facultative (fI-fVI) was preserved but the complexity of heterochromatin states was reduced to three (hI-hIII) compared with the six heterochromatin states in the wild-type 26-state model. Overall, *ddm1* caused a perturbation of the chromatin states associated with constitutive heterochromatin and did not significantly affect the group similarity for facultative heterochromatin and euchromatin states between the 16- and 26- state models, with genomic overlaps between 68 and 100% for the non-mixed states (***Figure 3—figure supplement 1C***).

To compare the chromatin state assignments between *ddm1* and wild-type genomes, we measured the Jaccard index (***Figure 3B***; ***Figure 3—figure supplement 1D***) and overlap (***Figure 3C***) of each state between the wild-type and *ddm1*. The strongest differences, in terms of genomic coverage, between *ddm1* and wild-type were observed in constitutive and facultative heterochromatin (***Figure 3D***). Accordingly, the regions marked by heterochromatin (red states h) became covered by states typical of facultative heterochromatin (blue states f) or intergenic states (gray states iI, ihII) (***Figure 3—figure supplement 1C and E***). To a lesser extent, regions covered by facultative state fVI and euchromatic state eIII were converted to intergenic state iI. Hence, the loss of DDM1 broadly affected the association of chromatin states with regions covered with constitutive heterochromatin in the wild-type. Accordingly, chromatin states were changed over TEs, including TE fragments and TE genes (assigned to TE families in TAIR10 annotation) but did not change over protein-coding genes (***Figure 3—figure supplement 1F***). Overall, the constitutive heterochromatin present over TEs in wild-type was replaced in *ddm1* by chromatin states found in intergenic, facultative heterochromatin, and euchromatin in wild-type (***Figure 3E and F*; *Figure 3—figure supplement 1E, F***). In addition, we observed that *ddm1* affected not only constitutive heterochromatin but also regions of euchromatin and facultative heterochromatin that adopted chromatin states distinct from these found in the wild-type. suggesting

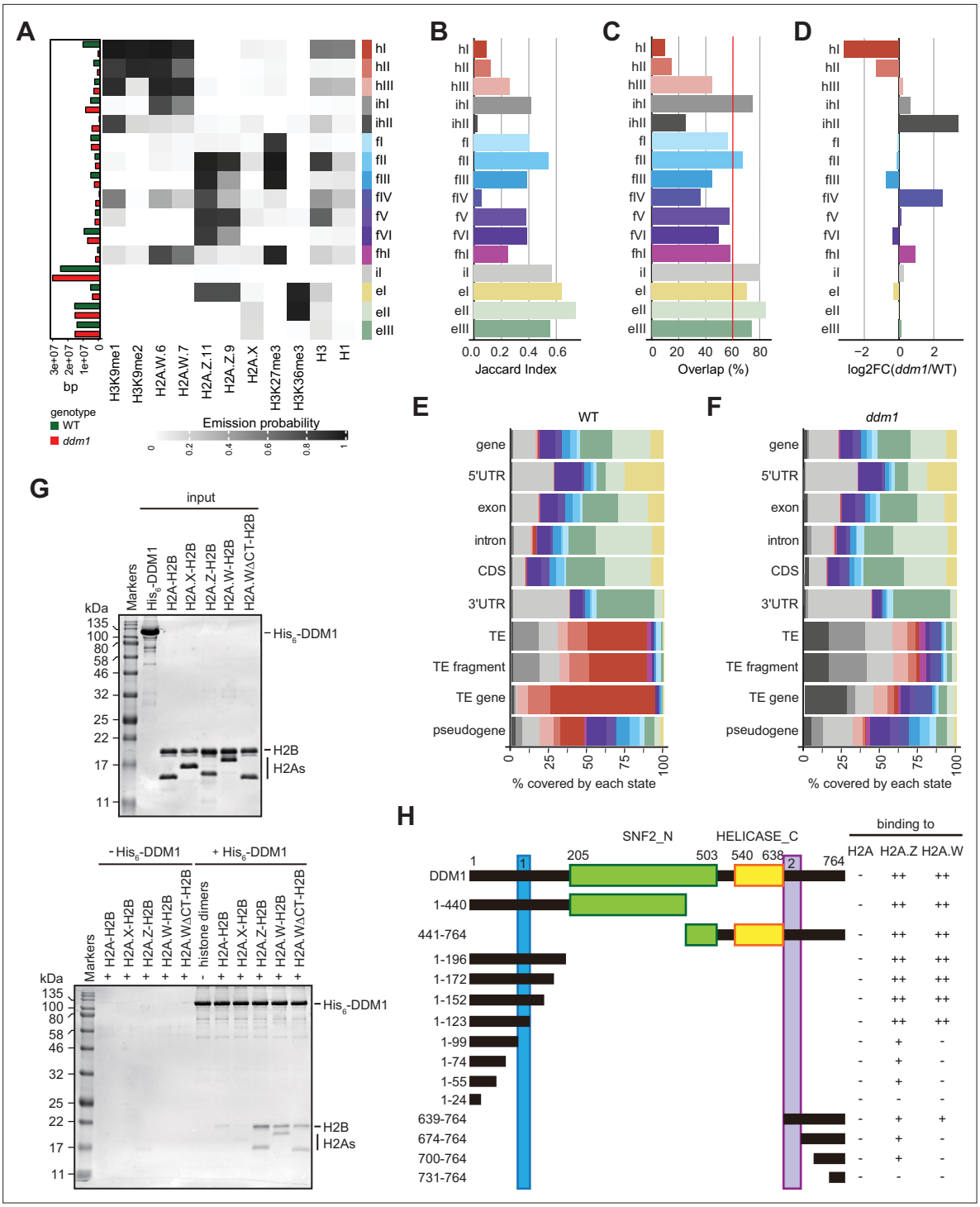


Figure 3. Decreased in DNA Methylation (DDM1) loss of function disrupts chromatin states in *Arabidopsis thaliana*. (**A**) Heatmap showing the emission probability for each mark/variant across the 16 chromatin states of the concatenated wild-type and *ddm1* mutant model. The bar plot on the left represents the proportion of the genome covered by each state in wild-type (green) and in *ddm1* (red). (**B**) Bar plot showing the Jaccard indices between the state assignments in wild-type and *ddm1* mutant. (**C**) Bar plot showing the state assignment overlap between the wild-type and *ddm1* for each chromatin state. The red vertical line represents the genome-wide overlap (62.2%). (**D**) Bar plot showing the log2 fold changes of the proportion of genome covered by each state across the *ddm1* genome compared to the wild-type. (**E**) Stacked bar plot showing the overlap between annotated genomic features and chromatin states from the concatenated model in wild-type. (**F**) Stacked bar plot showing the overlap between annotated genomic features and chromatin states from the concatenated model in *ddm1* mutant. (**G**) DDM1 interaction with H2A.W and H2A.Z. Coomassie-stained 15% SDS-PAGE gel showing input protein samples (top panel) used for in vitro pull-down (bottom panel) with His_6-tagged DDM1 and histone

Figure 3 continued on next page

Figure 3 continued

dimers. The lane ΔCTH2A.W shows that the deletion of the C-terminal tail of H2A.W does not influence binding to DDM1. (**H**) Summary of the pull-down assays to identify regions in DDM1 binding to H2A.W and H2A.Z. Blue and purple boxes indicate the H2A.W binding regions in DDM1 identified by previous work (***Osakabe et al., 2021***). Original pictures of the gels are provided in ***Figure 3—source data 1***.

The online version of this article includes the following source data and figure supplement(s) for figure 3:

Source data 1. The data contains the orginal images of the gels.

Figure supplement 1. Decreased in DNA Methylation (DDM1) loss of function disrupts chromatin states in *Arabidopsis thaliana*.

Figure supplement 2. Interaction of Decreased in DNA Methylation (DDM1) and DDM1 deletion mutants with histone variants.

Figure supplement 2—source data 1. The souce data file contain the original pcitures of the gels.

indirect effects of the disruption of constitutive heterochromatin caused directly by the loss of DDM1 (***Figure 3—figure supplement 1F***). We thus concluded that the loss of DDM1 causes a profound perturbation of both the definition of chromatin states and their distribution.

The H2A.W-binding domains of DDM1 bind H2A.Z

The conversion of chromatin states occupied by H2A.W in wild-type to chromatin states occupied by other H2A variants in *ddm1* suggested that DDM1 could also control the dynamics of other H2A variants. We have previously shown that DDM1 specifically binds H2A.W but neither H2A.X nor H1 linker histones (***Osakabe et al., 2021***). To determine if DDM1 binds to H2A or H2A.Z in addition to H2A.W, we performed a Ni-NTA agarose pulldown assay with recombinant histone heterodimers composed of H2B and one of the four H2A variants. We found that His_6-tagged DDM1 co-precipitated heterodimers containing H2A.W and H2A.Z but not H2A or H2A.X (***Figure 3G***). This demonstrated that DDM1 bound H2A.W and H2A.Z but not H2A and H2A.X. Notably, the C-terminal tail of H2A.W was dispensable for DDM1 binding (***Figure 3G***). To test whether the conserved motifs involved in H2A.W binding also bind H2A.Z, we prepared a series of DDM1 fragments fused with either His_6- or GST-tag. Our pulldown assays showed that fragments containing residues 100–123 or 639–673 were able to bind H2A.W and H2A.Z, whereas some additional regions of DDM1 showed weak binding only to H2A.Z (***Figure 3H***; ***Figure 3—figure supplement 2A-C***). Thus, DDM1 uses the same conserved sites to bind H2A.Z and H2A.W.

Exchange of chromatin states defined by H2A.W and H2A.Z is not directly associated with transcription

H2A.W evolved in land plants independently from macroH2A in animals (***Osakabe and Molaro, 2023***). macroH2A is also associated with heterochromatin although not as strictly as H2A.W (***Sun and Bernstein, 2019***). The deposition of macroH2A depends at least in part on the DDM1 murine ortholog LSH (***Ni et al., 2020***), and it was proposed that LSH binds and exchanges H2A to macroH2A (***Ni and Muegge, 2021***). Finding that DDM1 uses the same sites to bind both H2A.Z and H2A.W suggested that DDM1 controls the deposition of H2A.W and H2A.Z. To test this hypothesis, we surveyed the enrichment of H2A variants and histone PTMs over TEs in *ddm1* and compared this with the wild-type. To prevent carry-over mutations from generations of self-fertilized *ddm1* homozygous mutants we used leaves from *ddm1* homozygous mutant plants segregated from heterozygous *ddm1* mutants. We observed an overall enrichment of H2A.Z over TE genes in *ddm1*, supporting the idea that DDM1 promotes the removal of H2A.Z at TE genes where DDM1 deposits H2A.W in the wild-type (***Figure 4A***). Remarkably, non-transcribed TE genes also accumulated H2A.Z (***Figure 4A***) supporting the conclusion that the replacement of H2A.W by H2A.Z observed in *ddm1* mutants is not governed by transcription. Transcribed TE genes showed enrichment in H2A.Z at their TSS (***Figure 4A***) but also H2A.X on the gene body (***Figure 4—figure supplement 1A***). Transcription of TEs in *ddm1* did not affect the degree of enrichment for H3K27me3 of facultative heterochromatin but with increasing levels of transcription, there was increased enrichment of H3K36me3, resulting in a chromatin profile typical of expressed protein-coding genes (***Figure 4—figure supplement 1B***). Accordingly, strongly expressed TE genes showed an accumulation of euchromatic states, in contrast with non-expressed TE genes that became covered with states typical of facultative heterochromatin, irrespective of the type of constitutive heterochromatin observed on these TE genes in wild-type (***Figure 4B and C***; ***Figure 4—figure supplement 2A***). The change of chromatin states did not correlate with the length

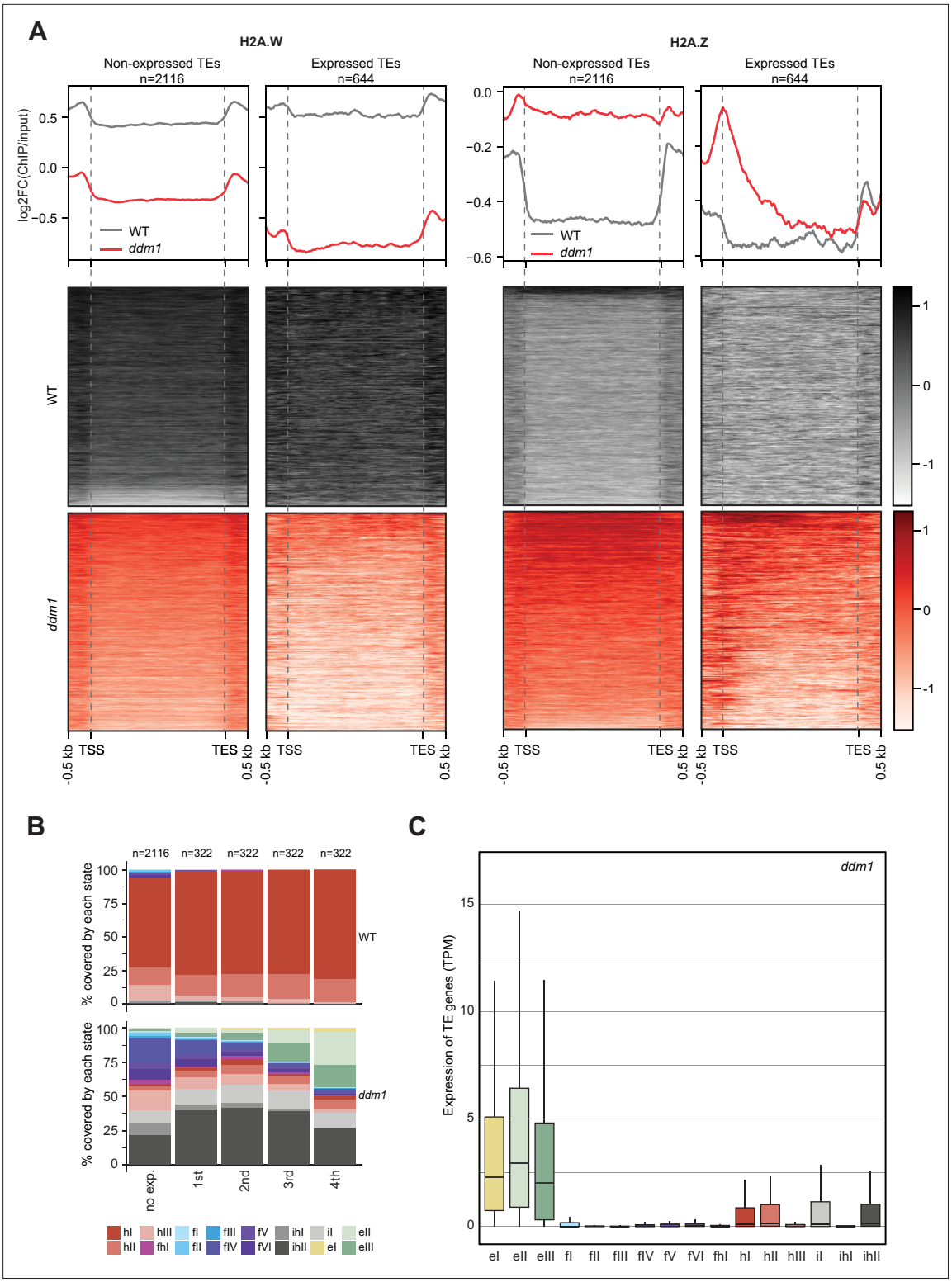


Figure 4. Impact of expression on chromatin states over TE genes in *ddm1*. (**A**) Enrichment profiles of H2A.W.6 and H2A.Z.9 over TE genes in *ddm1*. TE genes were grouped by expression in *ddm1* mutant. Out of the 3901 TE genes in the *Arabidopsis* genome annotation, 497 were excluded because they showed expression in wild-type, 2116 TE genes showed no expression in *ddm1* (non-expressed TE genes) while 1288 TE genes were expressed. Because many of these TE genes showed very low expression levels, we divided the expressed TEs into 4 quartiles (322 TE genes each) based on their expression values where the 1st quartile contains TE genes with the lowest expression and the 4th quartile contains TE genes with the highest expression. Given that the TE genes in 1st and 2nd quartile showed nominal expression values, we placed only TE genes in 3rd and 4th quartile (644 TE

Figure 4 continued on next page

Figure 4 continued

genes) in the category of expressed TEs. n represents the number of TE genes in each group. (**B**) Stacked bar plots of the proportion of states in wild-type (top panel) and in *ddm1* (bottom panel) overlapping TE genes grouped by expression in *ddm1*. (**C**) Box plot showing the expression of TE genes overlapping the 16 concatenated model states in *ddm1*.

The online version of this article includes the following figure supplement(s) for figure 4:

Figure supplement 1. Analyses of the parameters that could correlate with the chromatin states with TE expression in *ddm1*.

Figure supplement 2. Analyses of the parameters that could correlate with the chromatin states with TE expression in decreased in DNA Methylation (*ddm1*).

of the TE gene (***Figure 4—figure supplement 2B***) or the homogeneity of the chromatin landscape (***Figure 4—figure supplement 2C***). We conclude that at TE genes, DDM1 prevents the replacement of H2A.W by H2A.Z in a transcription-independent manner. Whether the replacement of H2A.W by H2A.Z in *ddm1* is caused by a direct exchange comparable to the exchange of H2A by H2A.Z catalyzed by SWR1 (***Ranjan et al., 2020***) or is the result of a cascade of activities of chaperones and remodelers is unclear. Yet, the alteration in the dynamics of H2A.W and H2A.Z in *ddm1* perturbs the allocation of chromatin states at constitutive heterochromatin and in other regions of chromatin. In some TEs, the new chromatin state acquired in *ddm1* permits transcription, which is then accompanied by a further exchange of H2A.Z to H2A and H2A.X and further loss of marks of constitutive heterochromatin and H1 and gain of marks of euchromatin (***Figure 4—figure supplement 1***).

Discussion

We report that histone modifications and histone variants combine into chromatin states that subdivide the intergenic space, the non-protein-coding space, and the protein-coding genic space of the genome into biologically significant subunits. These chromatin states summarize the prevalent organization of chromatin across cell types and stages of vegetative development from seedlings to mature leaves. Our data suggest that minor variations in chromatin states occur between vegetative cell types. We predict that highly specialized cell types such as gametes (***Borg et al., 2020***; ***Borg et al., 2021***) with specific histone variants and unusually high or low levels of chromatin regulators harbor specific chromatin states that are not reflected in our study. Calculation of chromatin states and biochemical analyses demonstrate the strongest associations between nucleosomes homotypic for H2A variants and histone modifications while nucleosomes are primarily heterotypic for H3 variants with a weaker association with histone modifications.

The prominent role of H2A variants in the organization of chromatin states is supported by systematic evaluation of the individual contribution of histone variants, H3 marks, and a combination thereof (***Figure 2F***) and the impact of the loss of the remodeler DDM1, which deposits H2A.W to maintain TE silencing (***Osakabe et al., 2021***). We further show that the DDM1 binding sites for H2A.W also bind H2A.Z, but not H2A or H2A.X. In *ddm1*, we observe a net replacement of H2A.W by H2A.Z. The effect can be interpreted in several manners. DDM1 could exchange directly H2A.Z to H2A.W. This model would support the idea of the convergent evolution of DDM1 and its orthologs LSH and HELLS in mammals with LSH catalyzing the exchange between replication dependent H2A and macroH2A (***Berger et al., 2023***; ***Ni and Muegge, 2021***). An alternative model is that the replacement of H2A.W by H2A.Z is indirect in *ddm1*. For example, H2A.Z might be deposited because of the loss of DNA methylation in *ddm1*, which is consistent with the well-established mutually exclusive patterns of H2A.Z and DNA methylation (***Zilberman et al., 2008***). It is also possible that the binding of DDM1 to H2A.Z is relevant in a different context that is not illustrated by our study. Whichever mechanism is involved it is unlikely that on its own, the loss of H2A.W in *ddm1* causes the deposition of H2A.Z on constitutive heterochromatin because this is not observed in mutants lacking H2A.W (***Bourguet et al., 2021***). The resulting enrichment of H2A.Z in *ddm1* is accompanied by a corresponding exchange of constitutive to facultative chromatin states over TEs. Notably, this does not require transcription, as TEs that remain silent in *ddm1* are likewise occupied by a facultative heterochromatin marked by H2A.Z and H3K27me3. When TEs are strongly expressed in *ddm1*, they also acquire euchromatin states marked by H2A and H2A.X over their gene body, while H2A.Z remains confined to the transcription start site. Hence, H2A variants appear to be important in shaping the deposition of histone modifications and defining chromatin states.

Overall, our observations suggest that chromatin writers and erasers operate with different affinities and efficacies based on the H2A variants in the nucleosome, an idea that is supported by previous reports. The inhibition of the H2B ubiquitin ligase by the H2A.Z tail (***Surface et al., 2016***) is sufficient to explain the anticorrelation between H2Bub and H2A.Z over gene bodies. H2A.Z is the major substrate for H2AK121Ub deposited by PRC1, as indicated by their co-occurrence within chromatin states as well as previous biochemical analyses (***Gómez-Zambrano et al., 2019***). PRC2 strongly prefers arrays of H2A.Z as a substrate (***Wang et al., 2018***), thus supporting the link between H2A.Z and H3K27me3 (***Carter et al., 2018***). We propose to expand these examples to the more general idea that the nature of the H2A variant controls the activity of the machinery that either deposits or erases the histone modifications. Such a model would explain the importance of histone variants in the establishment of chromatin states.

How are chromatin states maintained? In mammalian cells, there is a broad maintenance of the patterns of H3 modifications by recycling H3 and H4 at the replication fork (***Reverón-Gómez et al., 2018***). As plant cells divide, H3.1 sustains the deposition of H3K27me3 (***Jiang and Berger, 2017***) and H3K27me1 (***Jacob et al., 2014***) thus providing a positive feedback loop to sustain the polycomb repressive state and constitutive heterochromatin in dividing cells, respectively. Recent data in mammalian cells show that the recycling of H2A-H2B after replication provides another positive feedback loop to maintain the patterns of H2A variants (***Flury et al., 2023***). Such replication-dependent maintenance mechanisms no longer operate in non-dividing cells, providing opportunities for changing the allocation of chromatin states. This is illustrated by several studies in *Arabidopsis*. The positive feedback loops that maintain histone modifications become inactive and reversion to active chromatin states becomes possible by the deposition of H3.3 and eviction of H2A.Z by INO80 in response to temperature (***Zhao et al., 2023***) or to light (***Willige et al., 2023***). Mechanisms coupling H3.3 and H2A.Z dynamics are likely general for the regulation of responsive genes that are covered with facultative chromatin states (***Long et al., 2023***). Conversely, on expressed genes, acetylated histones assist in the docking of the chromatin remodeler SWI2/SNF2-related 1 (SWR1), which deposits H2A.Z at the transcription start sites (TSS) of expressed genes (***Aslam et al., 2019***; ***Bieluszewski et al., 2022***).

Although H3.1 and H3.3 do not strongly differentiate the three main classes of chromatin, they likely refine their definition. A relative enrichment of H3.1 with constitutive heterochromatin modifications (***Johnson et al., 2004***) were previously noted and confirmed in our study. Accordingly, the preferred substrate for the deposition of the heterochromatic mark H3K27me1 is H3.1 (***Jacob et al., 2014***). We also observed a relationship between H3.3 and H3K36me1/2, pointing to a possible preference of H3.3 for the deposition of H3K36me or for the demethylation of H3K36me3. In addition, it is becoming apparent that the dynamics of H3.3 and H2A.Z are coordinated in *Arabidopsis* (***Zhao et al., 2023***) and mammals (***Yang et al., 2022***). We thus propose that as yet unknown mechanisms link the deposition of H2A and H3 variants, resulting in specific types of nucleosomes that are sufficient to orchestrate the deposition of PTMs in distinct chromatin states. These will be read, interpreted, and further modified by silencing mechanisms in heterochromatin and the transcriptional machinery in euchromatin, leading to a changing chromatin landscape responding to the activity of the cell and its responses to the environment.

Methods

Key resources table

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Antibody	anti-H2A.X.3/5 (Rabbit polyclonal)	***Yelagandula et al., 2014***		WB: 1 µg/ml 5 µg per ChIP 10 µg per IP
Antibody	anti-H2A.13 (Rabbit polyclonal)	***Yelagandula et al., 2014***		WB: 1 µg/ml 5 µg per ChIP
Antibody	anti-H2A.2 (Rabbit polyclonal)	This work (See Materials and methods)		WB: 1 µg/ml 5 µg per ChIP

Continued on next page

Continued

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Antibody	anti-H2A.Z.9 (Rabbit polyclonal)	***Yelagandula et al., 2014***		WB: 1 µg/ml 5 µg per ChIP 10 µg per IP
Antibody	anti-H2A.Z.11 (Rabbit polyclonal)	This work (See Materials and methods)		WB: 1 µg/ml 5 µg per ChIP
Antibody	anti-H2A.W.6 (Rabbit polyclonal)	***Yelagandula et al., 2014***		WB: 1 µg/ml 5 µg per ChIP 10 µg per IP
Antibody	anti-H2A.W.7 (Rabbit polyclonal)	***Lorković et al., 2017***		WB: 1 µg/ml 5 µg per ChIP 10 µg per IP
Antibody	anti-H3 (Rabbit polyclonal)	Abcam	Cat# ab1791, RRID:AB_302613	WB: 0,5 µg/ml 5 µg per ChIP
Antibody	anti- H3K36me3 (Rabbit polyclonal)	Abcam	Cat# ab9050, RRID:AB_306966	WB: 1 µg/ml 5 µg per ChIP 10 µg per IP
Antibody	anti-H3K27me3 (Rabbit polyclonal)	Millipore	Cat# 07–449, RRID:AB_310624	WB: 1 µg/ml 5 µg per ChIP
Antibody	anti-H3K4me3 (Rabbit polyclonal)	Abcam	Cat# ab8580, RRID:AB_306649	WB: 1 µg/ml 5 µg per ChIP 10 µg per IP
Antibody	anti-H3K4me1 (Rabbit polyclonal)	Abcam	Cat# ab8895, RRID:AB_306847	WB: 1 µg/ml 5 µg per ChIP 10 µg per IP
Antibody	anti-H3K27me1 (Rabbit polyclonal)	Millipore	Cat# 17–643, RRID:AB_1587128	WB: 1 µg/ml 5 µl per ChIP 20 µl per IP
Antibody	anti-H1 (Rabbit polyclonal)	Agrisera	AS11 1801	5 µg per ChIP
Antibody	anti-H4K20me1 (Rabbit polyclonal)	Abcam	Cat# ab9051, RRID:AB_306967	5 µg per ChIP
Antibody	anti-H3K9me1 (Rabbit polyclonal)	Abcam	Cat# ab8896, RRID:AB_732929	WB: 1 µg/ml 5 µg per ChIP 10 µg per IP
Antibody	anti-H3K9me2 (Mouse monoclonal)	Abcam	Cat# ab1220, RRID:AB_449854	WB: 1 µg/ml 5 µg per ChIP
Antibody	anti-H3K9ac (rabbit polyclonal)	Millipore	Cat# 17–615, RRID:AB_1163437	WB: 1 µg/ml
Antibody	anti-H3K18ac (rabbit polyclonal)	Abcam	Cat# ab1191, RRID:AB_298692	WB: 1 µg/ml
Antibody	anti-H3K23ac (rabbit polyclonal)	Abcam	Cat# ab47813, RRID:AB_880444	WB: 1 µg/ml
Antibody	anti-H3K27ac (rabbit polyclonal)	Abcam	Cat# ab4729, RRID:AB_2118291	WB: 1 µg/ml
Antibody	anti-H3K122ac (rabbit polyclonal)	Abcam	Cat# ab33309, RRID:AB_942262	WB: 1 µg/ml
Antibody	anti-H3K27me2 (rabbit polyclonal)	Abcam	Cat# ab24684, RRID:AB_448222	WB: 1 µg/ml
Antibody	anti-H3K36me2 (rabbit polyclonal)	Abcam	Cat# ab9049, RRID:AB_1280939	WB: 1 µg/ml

Continued on next page

Continued

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Antibody	anti-HA (Rat monoclonal)	Roche	Cat# 11867423001, RRID:AB_390918	WB: 0,5 µg/ml
Antibody	Rabbit IgG (monoclonal)	Abcam	Cat# ab171870, RRID:AB_2687657	10 µg per IP
Commercial assay or kit	Nugen Ovation Ultralow V2 DNA-Seq library prep kit	NuGen	Cat# 0344	
Commercial assay or kit	NEBNext Ultra II DNA library prep kit for Illumina	New England Biolabs	Cat# E7645L	
Commercial assay or kit	Dynabeads Protein A	Invitrogen	Cat# 10746713	Used for ChIP
Commercial assay or kit	Protein A Mag Separose	GE Healthcare (Cytiva)	Cat# 28951378	Used for IP
Commercial assay or kit	Anti-HA Affinity matrix	Roche	Cat# 11815016001	Used for IP
Software, algorithm	ChromHMM	***Ernst and Kellis, 2012***; ***Ernst and Kellis, 2017***	RRID:SCR_018141	
Software, algorithm	Trim Galore	DOI:10.5281/zenodo.5127898.	RRID:SCR_011847	
Software, algorithm	Bowtie 2	***Langmead and Salzberg, 2012***	RRID:SCR_016368	
Software, algorithm	Picard	broadinstitute.github.io/picard/	RRID:SCR_006525	
Software, algorithm	Deeptools	***Ramírez et al., 2016***	RRID:SCR_016366	

Generation of antibodies, isolation of nuclei, MNase digestion, immunoprecipitation, SDS-PAGE, and western blotting

Antibodies against H2A.Z.11 (KGLVAAKTMAANKDKC) and H2A.2 (CPKKAGSSKPTEED) peptides were raised in rabbits (Eurogentec) and purified by a peptide affinity column. Purified IgG fractions were tested for specificity on nuclear extracts from WT and knock-out lines or with overexpressed histone variants (***Figure 2—figure supplement 1***). Specificities of custom-made polyclonal antibodies against *Arabidopsis* H2A.Z.9, H2A.X, H2A.W.6, H2A.13, H2A.W.7, H2Bs, and linker histone H1 was validated in previous publications (***Jiang et al., 2020***; ***Lorković et al., 2017***; ***Osakabe et al., 2018***; ***Yelagandula et al., 2014***).

For MNase digestion followed by immunoprecipitation, nuclei were isolated from 4 g of 2–3 weeks old leaves and the procedure as described in ***Lorković et al., 2017*** was followed. Isolated nuclei were washed once in 1 ml of N buffer (15 mM Tris-HCl pH 7.5, 60 mM KCl, 15 mM NaCl, 5 mM $MgCl_2$, 1 mM $CaCl_2$, 250 mM sucrose, 1 mM DTT, 10 mM ß-glycerophosphate) supplemented with protease inhibitors (Roche). After spinning for 5 min at 1800 × *g* at 4 °C nuclei were re-suspended in N buffer to a volume of 1 ml. Twenty microliters of MNase (0.1 u/µl) (Sigma-Aldrich) were added to each tube and incubated for 15 min at 37 °C and during the incubation nuclei were mixed four times by inverting the tubes. MNase digestion was stopped on ice by addition of 110 µl of MNase stop solution (100 mM EDTA, 100 mM EGTA). Nuclei were lysed by the addition of 110 µl of 5 M NaCl (final concentration of 500 mM NaCl). The suspension was mixed by inverting the tubes and they were then kept on ice for 15 min. Extracts were cleared by centrifugation for 10 min at 20,000 × *g* at 4 °C. Supernatants were collected and centrifuged again as above. For each immunoprecipitation extract, an equivalent of 4 g of leaf material was used, usually in a volume of 1 ml. To control MNase digestion efficiency, 100 µl of the extract was kept for DNA extraction. Antibodies, including non-specific IgG from rabbits, were bound to protein A magnetic beads (GE Healthcare) and then incubated with MNase extracts overnight at 4 °C. Beads were washed two times with N buffer without sucrose, containing 300 mM NaCl, followed by three washes with N buffer containing 500 mM NaCl. Beads were re-suspended in 150 µl of 1 × Laemmli loading buffer in 0.2 × PBS.

Proteins were resolved on 15% SDS-PAGE, transferred to a Protran nitrocellulose membrane (GE Healthcare) and analyzed by western blotting using standard procedures. The blots were developed with an enhanced chemiluminescence kit (Thermo Fisher Scientific) and signals acquired by a

ChemiDoc instrument (BioRad). All primary histone variant-specific and H3 marks-specific antibodies were used at 1 µg/ml dilution. H3-specific antibody was used at 1:5000 dilution. Rat anti-HA antibody (Roche 3F10) was used at 1:2000 dilution. Secondary antibodies were goat anti-rabbit IgG (BioRad) and goat-anti rat IgG (Sigma-Aldrich), both at a 1:10,000 dilution.

Mass spectrometry

For mass spectrometry, immunoprecipitated nucleosomes were resolved on a 4–20% gradient gels (Serva) and silver-stained. Histone H3 bands were excised, reduced, alkylated, in-gel trypsin or LysC digested, and processed for MS. The nano HPLC system used was a Dionex UltiMate 3000 HPLC RSLC nano system (Thermo Fisher Scientific, Amsterdam, Netherlands) coupled to a Q Exactive mass spectrometer (Thermo Fisher Scientific, Bremen, Germany), equipped with a Proxeon nanospray source (Thermo Fisher Scientific, Odense, Denmark). Peptides were loaded onto a trap column (Thermo Fisher Scientific, Amsterdam, Netherlands, PepMap C18, 5 mm 3300 mm ID, 5 mm particles, 100 Å pore size) at a flow rate of 25 ml/min using 0.1% TFA as the mobile phase. After 10 min, the trap column was switched in line with the analytical column (Thermo Fisher Scientific, Amsterdam, Netherlands, PepMap C18, 500 mm 375 mm ID, 2 mm, 100 Å). Peptides were eluted using a flow rate of 230 nl/min and a binary 2 hr gradient. The gradient starts with the mobile phases: 98% solution A (water/formic acid, 99.9/0.1, v/v) and 2% solution B (water/acetonitrile/formic acid, 19.92/80/0.08, v/v/v), increases to 35% of solution B over the next 120 min, followed by a gradient in 5 min to 90% of solution B, stays there for 5 min and decreases in 5 min back to the gradient 98% of solution A and 2% of solution B for equilibration at 30 °C. The Q Exactive HF mass spectrometer was operated in data-dependent mode, using a full scan (m/z range 380–1500, nominal resolution of 60,000, target value 1E6) followed by MS/MS scans of the 10 most abundant ions. MS/MS spectra were acquired using a normalized collision energy of 27%, an isolation width of 1.4 m/z, and a resolution of 30.000, and the target value was set to 1E5. Precursor ions selected for fragmentation (exclude charge states 1, 7, 8, >8) were put on a dynamic exclusion list for 40 s. Additionally, the minimum AGC target was set to 5E3, and the intensity threshold was calculated to be 4.8E4. The peptide match feature was set to preferred, and the exclude isotopes feature was enabled. For peptide identification, the RAW files were loaded into Proteome Discoverer (version 2.1.0.81, Thermo Scientific). The resulting MS/MS spectra were searched against *Arabidopsis thaliana* histone H3 sequences (seven sequences; 951 residues) using MS Amanda v2.1.5.8733 (***Dorfer et al., 2014***). The following search parameters were used: Beta-methylthiolation on cysteine was set as a fixed modification, oxidation on methionine, deamidation of asparagine and glutamine, acetylation on lysine, phosphorylation on serine, threonine, and tyrosine, methylation and di-methylation on lysine and threonine, tri-methylation on lysine, and ubiquitinylation on lysine were set as variable modifications. The peptide mass tolerance was set to ±5 ppm, and the fragment mass tolerance was set to ±15 ppm. The maximal number of missed cleavages was set to 2. The result was filtered to 1% FDR at the peptide level using the Percolator algorithm integrated with Thermo Proteome Discoverer. The localization of the phosphorylation sites within the peptides was performed with the tool ptmRS, which is based on the tool phosphoRS (***Taus et al., 2011***).

Peptides diagnostic for H3.1 and H3.3 covering positions K27 and K36 (see ***Figure 1—figure supplement 1E***) was used for the analysis of modifications. All peptides covering these two positions were selected in H3.1 and H3.3 immunoprecipitation samples and analyzed for the presence of modifications with the threshold for modification probability set to 95% or higher. Relative modification levels were expressed as the number of modified peptides (***Figure 1—figure supplement 1G***) divided by the total number of peptides (***Figure 1—figure supplement 1F***) that were measured for each lysine position resulting in total modification levels for H3.1 and H3.3 (see ***Figure 1D***). We also analyzed the same data by splitting H3.1 and H3.3 specific peptides in each immunoprecipitation and obtained highly similar trends for H3.1 and H3.3 irrespective of whether they were precipitated with H3.1 or H3.3 (***Figure 1E***). Histone acetylation was analyzed by selecting all peptides covering indicated positions and expressed as relative acetylation levels in each immunoprecipitation without differentiating H3.1 and H3.3 variants (***Figure 1—figure supplement 1I and J***). We also analyzed H3K9, H3K14, and H3K18 acetylation from peptides derived from transgenic copy alone because these data reflect modification levels of each H3 variant and obtained highly similar levels (***Figure 1—figure supplement 1K***) as without differentiation between H3.1 and H3.3.

Purification of recombinant *Arabidopsis* DDM1 and its truncations

His_6-tagged DDM1 and its truncations (DDM1(1–440), DDM1(1–196), DDM1(441–764)), GST-tagged DDM1 truncations, DDM1(1–24), DDM1(1–55), DDM1(1–74), DDM1(1–99), DDM1(1–123), DDM1(1–152), DDM1(1–172), DDM1(1–196), were expressed and purified as described previously (***Osakabe et al., 2021***).

Purification and reconstitution of recombinant *Arabidopsis* histone heterodimers

To purify and reconstitute *Arabidopsis* histone dimers, recombinant *Arabidopsis* histones H2A.13, H2A.X.3, H2A.Z.9, H2A.W.6, H2A.W.6 ΔCT (aa 1–128), and H2B.9 were expressed and purified as previously described (***Osakabe et al., 2013***; ***Tachiwana et al., 2010***; ***Tanaka et al., 2004***). The H2A-H2B heterodimers were reconstituted and purified as previously described (***Osakabe et al., 2013***).

Pull-down assay

The pull-down assay using Ni-NTA and GS4B beads for His_6-tagged DDM1 or truncations and GST-tagged DDM1 truncations were performed as described previously (***Osakabe et al., 2021***). Proteins precipitated with beads were analyzed by 15% SDS-PAGE stained with Coomassie Brilliant Blue.

Plant material for ChIP-seq

Col-0 wild-type (WT) *Arabidopsis thaliana* seeds were stratified at 4 °C in the dark for three days. Seeds were germinated on ½ MS sterilized plates in the growth chamber under long day (LD) conditions (21 °C; 16 hr light/8 hr dark). After 10 days seedling tissue was freshly harvested. For ChIP-seq from leaf tissue, Col-0 wild-type (WT), and seeds from *ddm1*/+ plants were stratified at 4 °C in the dark for three days. To prevent carry-over mutations from generations of *ddm1* homozygous mutants we used leaves from *ddm1* homozygous mutant plants segregated from heterozygous *ddm1*/+ mutants. Plants were grown on soil in the growth chamber under a LD conditions (21 °C; 16 hr light/8 hr dark) and leaf tissue was harvested 5 weeks after germination.

Chromatin immunoprecipitation (ChIP)

For ChIP nuclei isolation was performed using 10-day-old seedlings (WT Col-0) or leaves (WT Col-0, *ddm1*). The procedure for nuclei isolation and chromatin immunoprecipitation was performed as described in ***Osakabe et al., 2021***. Freshly harvested tissues (0.3 g of tissue was used for each immunoprecipitation) were fixed with 1% formaldehyde for 15 min and the cross-linking reaction was stopped by the addition of 125 mM glycine. Crosslinked tissues were frozen and ground with a mortar and pestle in liquid nitrogen to obtain a fine powder. Ground tissues were resuspended in M1 buffer (10 mM sodium phosphate pH 7.0, 100 mM NaCl, 1 M hexylene glycol, 10 mM 2-mercaptoethanol, and protease inhibitor (Roche)), and the suspension was filtered through miracloth. Nuclei were precipitated by centrifugation and washed six times with M2 buffer (10 mM sodium phosphate pH 7.0, 100 mM NaCl, 1 M hexylene glycol, 10 mM $MgCl_2$, 0.5% Triton X-100, 10 mM 2-mercaptoethanol, and protease inhibitor), and then further washed once with M3 buffer (10 mM sodium phosphate pH 7.0, 100 mM NaCl, 10 mM 2-mercaptoethanol, and protease inhibitor). Nuclei pellets were rinsed and resuspended in a sonication buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.1% SDS, and protease inhibitor) and sonicated with a Covaris E220 High Performance Focused Ultrasonicator for 15 min at 4 °C (Duty factor 5.0; PIP 140.0; Cycles per Burst 200) in 1 ml Covaris milliTUBE. After chromatin shearing, the debris was removed by centrifugation, and the solutions containing chromatin fragments were diluted with three times the volume of ChIP dilution buffer (16.7 mM Tris-HCl pH 8.0, 167 mM NaCl, 1.2 mM EDTA, 1.1% Triton X-100, 0.01% SDS, and protease inhibitor). After dilution, protein A/G magnetic beads (50 µl for one gram of tissue; Thermo Fisher Scientific) were added to sheared chromatin and incubated at 4 °C for 1 hr with rotation. Pre-cleared samples were collected and incubated with 5 µg of in-house prepared anti-H2A.X.3/5, anti-H2A.13, anti-H2A.2, anti-H2A.Z.9, anti-H2A.Z.11, anti-H2A.W.6, anti-H2A.W.7, anti-H1, and anti-H3 (Abcam, ab1791), anti-H3K36me3 (Abcam, ab9050), anti-H3K27me3 (Millipore, 07–449), anti-H3K4me3 (Abcam, ab8580), anti-H3K4me1 (Abcam, ab8895), anti-H3K27me1 (Millipore 17–643), anti-H4K20me1 (Abcam, ab9051), anti-H3K9me1 (Abcam, ab8896), or anti-H3K9me2 (Abcam, ab1220) antibodies at 4°C overnight with rotation. After incubation, samples were mixed with 30 µl of protein A/G magnetic beads, incubated at 4°C for 3 hr

with rotation, washed twice with low salt buffer (20 mM Tris-HCl pH 8.0, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, and 0.1% SDS), once with high salt buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 2 mM EDTA, 1% Triton X-100, and 0.1% SDS), once with LiCl buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.25 M LiCl, 1% IGEPAL CA-630, and 0.1% deoxycholic acid), and twice with TE buffer (10 mM Tris-HCl pH 8.0 and 1 mM EDTA). Immunoprecipitated DNA was eluted by adding 500 µl of elution buffer (1% SDS and 0.1 M $NaHCO_3$), incubated at 65°C for 15 min, and mixed with 51 µl of reverse crosslink buffer (40 mM Tris-HCl pH 8.0, 0.2 M NaCl, 10 mM EDTA, 0.04 mg/ml proteinase K (Thermo Fisher Scientific)). The reaction mixture was then incubated at 45°C for 3 hr and subsequently at 65°C for 16 hr. After crosslink reversal, DNA was treated with 10 µg of RNaseA (Thermo Fisher Scientific), incubated at room temperature for 30 min, and purified with a MinElute PCR purification kit (Qiagen).

ChIP-seq library prep and data processing

For ChIP-seq, libraries were prepared either with Nugen Ovation Ultralow V2 DNA-Seq library prep kit (NuGen) or NEBNext Ultra II DNA library prep kit for Illumina (New England Biolabs) following the manufacturer's instructions. The libraries were sequenced with an Illumina Hiseq 2000 to generate single-end 50 bp reads. For alignment and quality check of sequenced samples, bedtools v2.27.1 (***Quinlan, 2014***) was used to convert the raw BAM files to fastq. FastQC v0.11.8 (https://qubeshub.org/resources/fastqc) was used to generate quality reports for all sequencing data. Reads were trimmed using trim_galore v0.6.5 (DOI:10.5281/zenodo.5127898) (*trim_galore --dont_gzip --stringency 1 --fastqc --length 5 $*) and then aligned to the TAIR10 *Arabidopsis* genome using Bowtie2 v2.4.1 (***Langmead and Salzberg, 2012***) with default settings, Picard v2.22.8 (https://broadinstitute.github.io/picard/) was used to remove duplicated reads. Deeptools v3.1.2 (***Ramírez et al., 2016***) was used to examine correlations between the ChIP samples. The bamCompare function from Deeptools was used to normalize ChIP samples to Input or H3 and to generate log2 ratio (ChIP/(Input or H3)) bigwig files.

ChIP-seq data processing for published data

Publicly available ChIP-seq data for H2B variants were downloaded from GEO GSE151166 (***Jiang et al., 2020***). ChIP-seq data for H3 variants were downloaded from GEO GSE34840 (***Stroud et al., 2012***). ChIP-seq data for H3K14Ac, H3K9Ac was downloaded from GEO GSE89768 (***Kim et al., 2016***). ChIP-seq data for H2AK121Ub and H3K27me3 was downloaded from GEO GSE89357 (***Zhou et al., 2017***). Raw data was downloaded and processed as described above. Quality control metrics for all samples are included in source file 1.

Defining chromatin states

Aligned ChIP-seq BAM files for chromatin features were generated as described above. Aligned BAM files were then converted to BED format using bedtools bamtobed v2.27.1 (***Quinlan, 2014***). These genome-wide BED files were then used for learning chromatin states by the multivariate Hidden Markov Models. For the extensive wild-type model, we used the BinarizeBed and LearnModel programs from ChromHMM (***Ernst and Kellis, 2017***) with default settings and window size of 200 bp. We generated models from 2 to 50 chromatin states. To compare emission parameters of models with different numbers of states, the command CompareModel from ChromHMM was used. Thus, we selected a reference model with the largest number of states (50) and compared it with other models with lower number of states (from 49 to 2). This resulted in a correlation matrix between each state of the reference model and any state of the other models being compared (***Figure 2—figure supplement 2***). This comparison revealed that the correlation dropped for models including less than 13 states, suggesting that some of the biologically relevant states were not resolved. We thus concluded that at least 13 states should be used. However, ChromHMM (***Ernst and Kellis, 2017***) does not establish the optimal state number as it does not consider genomic features which are associated with specific biological function. Therefore, we analyzed the association between chromatin states with genomic features in all models ranging from 13 to 30 states. As a result, for example, we observed that models with more than 13 states could resolve biologically meaningful heterochromatic states (H2 to H6) (H2 to H4 associated with pericentromeric repeats/CMT2 associated repeats; H5 and H6 associated with chromosome arm repeats/RdDM repeats) (as shown in ***Jamge et al., 2022***, Figure 3). Furthermore, only with a 26-state model, we observed a clearly defined novel state H1 with H2B enrichment, largely

deprived of TE genes and located in the closest proximity to centromeres. Increasing the number of states to 27 and above gave rise to an additional chromatin state that could no longer be associated with any distinct genomic feature. Hence, we decided that, with the set of chromatin components included, a 26-state model is optimal for our analysis.

For the mixed seedling and leaf data, and for the mixed wild-type and *ddm1* mutant data we used concatenated ChromHMM models (***Ernst and Kellis, 2017***). Those models use the data from both tissues/genotypes to build a common model. To this end, we used the BinarizeBed and LearnModel with the default settings and window size of 200 bp but with the tissue or genotype as additional information. The number of states in the final model was decided on as described for the extensive model.

Analysis of sub-epigenome models

The ChromHMM command EvalSubset was used to compare models where histone variants or modifications were excluded from the model. Five models were generated and evaluated against the full model: no H2B variants, no H3 variants, no H2A variants, no histone modifications, and no variants (H2A/H3/H2B). The diagonals of the resulting confusion matrices (representing the Jaccard indices) were extracted and visualized using the R package ComplexHeatmap v2.10.0 (***Gu et al., 2016***).

Analysis of chromatin states

The emission matrices from the ChromHMM models were read into R 4.1.2 (https://www.R-project.org/.) and plotted using the R packages ggplot2 from tidyverse (doi:10.21105/joss.01686) and package ComplexHeatmap v2.10.0 (***Gu et al., 2016***). The state assignment files from ChromHMM for the different models and tissues/genotypes were read in and analyzed using packages from tidyverse and valr v0.6.6 (***Riemondy et al., 2017***). To compare the wild-type and *ddm1* assignments from the concatenated model, separate files were joined. To this end, for each (200 bp) bin of the genome, the information about which state had been assigned in wild-type and in *ddm1*, respectively was combined. The state assignment datasets were then overlapped, using the function bed_intersect from valr, with regions for genomic features defined in the file `Araport11_GFF3_genes_transposons.Jun2016.gff` downloaded from (https://www.arabidopsis.org/download_files/Genes/Araport11_genome_release/archived/Araport11_GFF3_genes_transposons.Jun2016.gff.gz). The transcriptome, methylome, and DNase I datasets were integrated to generate plots using R and the ggplot2 package from tidyverse. All box plots show the data excluding outliers.

Comparison of states from extensive and concatenated models

To compare the states from the two tissue models with the chromatin types defined in the extensive model, the states assigned to seedlings by the two tissue model were overlapped with the chromatin types from the extensive model. For the comparison of the wild-type and *ddm1* mutant model states with the chromatin types from the extensive model, the states assigned to the wild-type was used. The genomic overlaps were calculated using bed_intersect from valr v0.6.6 (***Riemondy et al., 2017***). States overlapping one chromatin type with >66% and all others with <20% were assigned the largest overlapping chromatin type. In other cases, the state was classified as mixed (***Figure 3—figure supplement 1B***; ***Figure 3—figure supplement 1C***).

Analysis of chromatin state changes in *ddm1* mutant

The Jaccard index for each state was calculated as the total length of regions assigned to that state in both wild-type and mutant divided by the combined length of all regions belonging to that state in at least one of the two genotypes. The overlap was calculated as the total length of regions assigned to that state in both wild-type and mutant divided by the total length of regions assigned to that state in wild-type. The size fold change was calculated as the total length of regions assigned to that state in *ddm1* mutant divided by the total length of all regions assigned to the state in the wild-type.

Or, formally, if $\boldsymbol{B_{w,m}}$ is the total number of bins that are assigned to state *w* in the wild-type and state *m* in *ddm1*, then the (JI) Jaccard index, (O) overlap, and (FC) size change for states can be calculated as:

$$JI_s = \frac{B_{s,s}}{B_{s,.} + B_{.,s} - B_{s,s}},\ O_s = \frac{B_{s,s}}{B_{s,.}},\ FC_s = log_2\left(\frac{B_{.,s}}{B_{s,.}}\right)$$

The plots were generated using R and the ggplot2 package from tidyverse. The metaplots were generated using Deeptools v3.1.2 (***Ramírez et al., 2016***). Using the bamCompare function of Deeptools v3.1.2 bigWig files were generated by normalizing ChIP samples with input/H3. These bigWig files were used for plotting the metaplot heatmap (***Figure 4A***) and profile plots (***Figure 4—figure supplement 1A and B***) using deeptools v3.1.2 plotHeatmap and plotProfile functions, respectively.

RNA-seq of WT seedlings

Total RNA was extracted with Spectrum plant total RNA kit (Sigma Aldrich) from 10 day seedlings of WT Col-0. A DNA-free DNA removal kit was used to remove contaminated DNA (Thermo Fisher Scientific). RNA-seq poly-A libraries were generated with NEBNext UltraII directional RNA library prep kit for Illumina (New England Biolabs) following the manufacturer's instructions. The libraries were sequenced with Illumina Hiseq 2500 to generate single-end 50 bp reads. Samples were prepared from three independent biological replicates. The RNA-seq data was processed as described in ***Osakabe et al., 2021***.

Methylation data

BS-seq data for WT Col-0 was downloaded from GSE39901 (***Stroud et al., 2013***). BS-seq reads were processed using the nf-core pipeline (***Ewels et al., 2023***) as described in ***Pisupati et al., 2023***. Cutadapt v2.10 (***Martin, 2011***) was used to trim the adaptors (default parameters). Trimmed reads were then mapped to TAIR10 (Col-0) assembly using bismark v0.2.1 (***Krueger and Andrews, 2011***) allowing mismatches to 0.5. Methylation calls were performed using methylpy v1.3.7 (***Schultz et al., 2016***) on the aligned bam files. Methylation call bed files were used to calculate average methylation over chromatin state bed regions.

DNase I-seq

We downloaded processed DNaseI - seq bigwig files data from GEO series GSE53322 for WT Col-0 (GSM1289358) (***Sullivan et al., 2014***). Bedtools map v2.3 (***Quinlan, 2014***) was used to calculate the averaged signal over bed regions in each chromatin state. Box plots were generated in R v3.5 using ggplot2 to compare the accessibility across chromatin states.

RNA-seq data analysis of wild-type and *ddm1* mutant leaves

We used the wild-type and *ddm1* mutant RNA-seq data from GSE150435. The data was processed as described in ***Osakabe et al., 2021***. Before grouping the TE genes into five groups we first excluded all TE genes showing any expression in wild-type (tpm >0). TE genes without expression (tpm <0.1) in *ddm1* mutants were put in the 'no expression' group. The remaining TE genes were divided into four quartiles based on the tpm values in the *ddm1* mutant.

Acknowledgements

FB acknowledges support from the next-generation sequencing and PlantS facilities at the Vienna BioCenter Core Facilities (VBCF), the BioOptics facility, the Proteomics facility, and Molecular Biology Services from the Institute for Molecular Pathology (IMP), the Molecular Biology Services at the GMI. We thank all members of the Berger Laboratory for their technical help. We thank J Matthew Watson for insightful discussions and critical reading of the manuscript. This work was supported by the Japan Society for the Promotion of Science (JSPS) Overseas Research Fellowships [to AO], the Austrian Science Fund (FWF): M2539-B21 [to AO], P32054, P30802, and P33380 [to FB], and DK1238 [to BJ], the Austrian Academy of Sciences [to FB, ZJL, EA., SA, RP and RY], JSPS Grant-in-Aid for Research Activity Start-up (JP21K20628), JSPS Grant-in-Aid for Transformative Research Areas (JP22H05172 and JP22H05178), and the Japan Science and Technology Agency (JST) PRESTO (JPMJPR20K3) [to AO].

Additional information

Funding

Funder	Grant reference number	Author
Austrian Science Fund	M2539-B21	Akihisa Osakabe
Austrian Science Fund	P32054	Frédéric Berger
Austrian Science Fund	P30802	Frédéric Berger
Austrian Science Fund	P33380	Frédéric Berger
Austrian Science Fund	DK1238	Bhagyshree Jamge
Japan Society for the Promotion of Science	JP21K20628	Akihisa Osakabe
Japan Society for the Promotion of Science	JP22H05172	Akihisa Osakabe
Japan Society for the Promotion of Science	JP22H05178	Akihisa Osakabe
Japan Science and Technology Agency	JPMJPR20K3	Akihisa Osakabe
Austrian Academy of Sciences		Frédéric Berger Zdravko J Lorković Elin Axelsson Svetlana Akimcheva Ramesh Yelagandula
Japan Society for the Promotion of Science	Overseas Research Fellowships	Akihisa Osakabe

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Bhagyshree Jamge, Formal analysis, Validation, Investigation, Visualization, Methodology, Writing - review and editing; Zdravko J Lorković, Data curation, Formal analysis, Validation, Investigation, Visualization, Methodology, Writing - original draft, Writing - review and editing; Elin Axelsson, Resources, Data curation, Software, Formal analysis, Supervision, Validation, Investigation, Visualization, Methodology, Project administration, Writing - review and editing; Akihisa Osakabe, Data curation, Formal analysis, Validation, Investigation, Visualization, Methodology, Writing - review and editing; Vikas Shukla, Resources, Data curation, Software, Formal analysis; Ramesh Yelagandula, Annika Luisa Kuehn, Investigation; Svetlana Akimcheva, Resources, Methodology, Project administration; Frédéric Berger, Conceptualization, Supervision, Funding acquisition, Writing - original draft, Project administration, Writing - review and editing

Author ORCIDs

Elin Axelsson http://orcid.org/0000-0003-4382-1880
Vikas Shukla http://orcid.org/0000-0001-8739-8610
Ramesh Yelagandula http://orcid.org/0000-0001-7418-4322
Frédéric Berger http://orcid.org/0000-0002-3609-8260

Peer review material

Reviewer #1 (Public Review): https://doi.org/10.7554/eLife.87714.3.sa1
Reviewer #2 (Public Review): https://doi.org/10.7554/eLife.87714.3.sa2
Reviewer #3 (Public Review): https://doi.org/10.7554/eLife.87714.3.sa3
Reviewer #4 (Public Review): https://doi.org/10.7554/eLife.87714.3.sa4
Author Response https://doi.org/10.7554/eLife.87714.3.sa5

Additional files

Supplementary files

- Supplementary file 1. Mapping statistics. The file includes output statistics for mapping and trimming of all data included in the manuscript.
- MDAR checklist

Data availability

The genome-wide sequencing (ChIP-seq) generated for this study as well as published datasets (ChIP-seq, RNA-seq) utilized to support the findings in this study have been deposited on NCBI's Gene Expression Omnibus (GEO) with the accession number of subseries GSE226469, GSE231398, GSE150434 and GSE150433. Super series associated with this study and published data are GSE231408 and GSE150436. Source data for all the main figures as well as supplementary figures have been provided with this manuscript. All the code used to analyze the genome-wide sequencing data presented in this study, as described in the Methods are available at https://github.com/Gregor-Mendel-Institute/jamge_states_2023 (copy archived at swh:1:rev:5254163528f971a97c4c514daca96f01b11e545a).

The following datasets were generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Jamge B	2023	Chromatin immunoprecipitation DNA-sequencing (ChIP-seq) for histone H2A variants and histone modifications in WT Col-0 seedlings	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE226469	NCBI Gene Expression Ominbus, GSE226469
Jamge B	2023	ChIP-seq for Histone variants and Histone modifictaions in WT and ddm1(-/-) in 1st generation Leaves	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE231398	NCBI Gene Expression Omnibus, GSE231398
Jamge B	2023	ChIP-seq for Histone variants and Histone modifications	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE231408	NCBI Gene Expression Omnibus, GSE231408

The following previously published datasets were used:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Jamge B	2021	The chromatin remodeler DDM1 silences transposons through deposition of the histone variant H2A.W	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE150436	NCBI Gene Expression Omnibus, GSE150436
Jamge B, Osakabe A	2021	RNA-seq for WT and ddm1(-/-) to evaluate transposon expression in ddm1 1st generation	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE150433	NCBI Gene Expression Omnibus, GSE150433
Jamge B, Osakabe A	2021	ChIP-seq for Histone Variant H2A.W, H2A.X, H1 and H3K9me2 in WT and ddm1(-/-) to quantify loss of H2A.W in ddm1 1st generation	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE150434	NCBI Gene Expression Omnibus, GSE150434

References

Aslam M, Fakher B, Jakada BH, Cao S, Qin Y. 2019. Swr1 Chromatin remodeling complex: A key transcriptional regulator in plants. *Cells* **8**:1621. DOI: https://doi.org/10.3390/cells8121621, PMID: 31842357

Berger F, Muegge K, Richards EJ. 2023. Seminars in cell and development biology on Histone variants Remodelers of H2A variants associated with Heterochromatin. *Seminars in Cell & Developmental Biology* **135**:93–101. DOI: https://doi.org/10.1016/j.semcdb.2022.02.026, PMID: 35249811

Bieluszewski T, Sura W, Dziegielewski W, Bieluszewska A, Lachance C, Kabza M, Szymanska-Lejman M, Abram M, Wlodzimierz P, De Winne N, De Jaeger G, Sadowski J, Côté J, Ziolkowski PA. 2022. Nua4 and H2A.Z control environmental responses and Autotrophic growth in Arabidopsis. *Nature Communications* **13**:277. DOI: https://doi.org/10.1038/s41467-021-27882-5, PMID: 35022409

Borg M, Jacob Y, Susaki D, LeBlanc C, Buendía D, Axelsson E, Kawashima T, Voigt P, Boavida L, Becker J, Higashiyama T, Martienssen R, Berger F. 2020. Targeted Reprogramming of H3K27Me3 Resets epigenetic memory in plant paternal Chromatin. *Nature Cell Biology* **22**:621–629. DOI: https://doi.org/10.1038/s41556-020-0515-y, PMID: 32393884

Borg M, Papareddy RK, Dombey R, Axelsson E, Nodine MD, Twell D, Berger F. 2021. Epigenetic Reprogramming Rewires transcription during the alternation of generations in Arabidopsis. *eLife* **10**:e61894. DOI: https://doi.org/10.7554/eLife.61894, PMID: 33491647

Bourguet P, Picard CL, Yelagandula R, Pélissier T, Lorković ZJ, Feng S, Pouch-Pélissier M-N, Schmücker A, Jacobsen SE, Berger F, Mathieu O. 2021. The Histone variant H2A.W and Linker Histone H1 Co-regulate Heterochromatin accessibility and DNA methylation. *Nature Communications* **12**:2683. DOI: https://doi.org/10.1038/s41467-021-22993-5, PMID: 33976212

Brunson JC, Read QD. 2020. ggalluvial: Alluvial Plots in ggplot2. R Package Version 0.12.4. http://corybrunson.github.io/ggalluvial/

Carter B, Bishop B, Ho KK, Huang R, Jia W, Zhang H, Pascuzzi PE, Deal RB, Ogas J. 2018. The Chromatin Remodelers PKL and Pie1 act in an epigenetic pathway that determines H3K27Me3 homeostasis in Arabidopsis. *The Plant Cell* **30**:1337–1352. DOI: https://doi.org/10.1105/tpc.17.00867, PMID: 29802212

Chakravarthy S, Bao Y, Roberts VA, Tremethick D, Luger K. 2004. Structural characterization of Histone H2A variants. *Cold Spring Harbor Symposia on Quantitative Biology* **69**:227–234. DOI: https://doi.org/10.1101/sqb.2004.69.227, PMID: 16117653

Delaney K, Almouzni G. 2023. Transcription-coupled H3.3 recycling: A link with Chromatin States. *Seminars in Cell & Developmental Biology* **135**:13–23. DOI: https://doi.org/10.1016/j.semcdb.2022.05.003, PMID: 35595602

Dorfer V, Pichler P, Stranzl T, Stadlmann J, Taus T, Winkler S, Mechtler K. 2014. MS Amanda, a universal identification algorithm optimized for high accuracy tandem mass spectra. *Journal of Proteome Research* **13**:3679–3684. DOI: https://doi.org/10.1021/pr500202e, PMID: 24909410

Ernst J, Kellis M. 2012. Chromhmm: automating Chromatin-state discovery and characterization. *Nature Methods* **9**:215–216. DOI: https://doi.org/10.1038/nmeth.1906, PMID: 22373907

Ernst J, Kellis M. 2017. Chromatin-state discovery and genome annotation with Chromhmm. *Nature Protocols* **12**:2478–2492. DOI: https://doi.org/10.1038/nprot.2017.124, PMID: 29120462

Ewels P, Hüther P, Sateesh_Peri NS, Miller E, nf-core bot, Hörtenhuber M, Peltzer A, Sven F, Alneberg J, Di Tommaso P, Garcia MU, Davenport C, Ajith V, Krueger F, Patel H, Alessia GD, Syme R, Céline N. 2023. Nf-core/Methylseq. [2.4.0]. Zenodo. https://doi.org/10.5281/zenodo.8029942

Feng J, Shen WH. 2014. Dynamic regulation and function of Histone Monoubiquitination in plants. *Frontiers in Plant Science* **5**:83. DOI: https://doi.org/10.3389/fpls.2014.00083, PMID: 24659991

Flury V, Reverón-Gómez N, Alcaraz N, Stewart-Morgan KR, Wenger A, Klose RJ, Groth A. 2023. Recycling of modified H2A-H2B provides short-term memory of Chromatin States. *Cell* **186**:1050–1065. DOI: https://doi.org/10.1016/j.cell.2023.01.007, PMID: 36750094

Gan ES, Xu Y, Ito T. 2015. Dynamics of H3K27Me3 methylation and Demethylation in plant development. *Plant Signaling & Behavior* **10**:e1027851. DOI: https://doi.org/10.1080/15592324.2015.1027851, PMID: 26313233

Gómez-Zambrano Á, Merini W, Calonje M. 2019. The repressive role of Arabidopsis H2A.Z in transcriptional regulation depends on Atbmi1 activity. *Nature Communications* **10**:2828. DOI: https://doi.org/10.1038/s41467-019-10773-1, PMID: 31249301

Gu Z, Eils R, Schlesner M. 2016. Complex Heatmaps reveal patterns and correlations in multidimensional Genomic data. *Bioinformatics* **32**:2847–2849. DOI: https://doi.org/10.1093/bioinformatics/btw313, PMID: 27207943

Hake SB, Allis CD. 2006. Histone H3 variants and their potential role in indexing mammalian Genomes: the "H3 Barcode hypothesis. *PNAS* **103**:6428–6435. DOI: https://doi.org/10.1073/pnas.0600803103, PMID: 16571659

Henikoff S, McKittrick E, Ahmad K. 2004. Epigenetics, Histone H3 variants, and the inheritance of Chromatin States. *Cold Spring Harbor Symposia on Quantitative Biology* **69**:235–243. DOI: https://doi.org/10.1101/sqb.2004.69.235, PMID: 16117654

Horikoshi N, Sato K, Shimada K, Arimura Y, Osakabe A, Tachiwana H, Hayashi-Takanaka Y, Iwasaki W, Kagawa W, Harata M, Kimura H, Kurumizaka H. 2013. Structural polymorphism in the L1 loop regions of human H2A.Z.1 and H2A.Z.2. *Acta Crystallographica. Section D, Biological Crystallography* **69**:2431–2439. DOI: https://doi.org/10.1107/S090744491302252X, PMID: 24311584

Jacob Y, Bergamin E, Donoghue MTA, Mongeon V, LeBlanc C, Voigt P, Underwood CJ, Brunzelle JS, Michaels SD, Reinberg D, Couture JF, Martienssen RA. 2014. Selective methylation of Histone H3 variant H3.1 regulates Heterochromatin replication. *Science* **343**:1249–1253. DOI: https://doi.org/10.1126/science.1248357, PMID: 24626927

Jamge B, Lorkovic ZJ, Axelsson E, Yelagandula R, Akimcheva S, Berger F. 2022. Transcriptional Activity Is Shaped by Chromatin Landscapes in Arabidopsis. *bioRvix*. DOI: https://doi.org/10.1101/2022.06.02.494419

Jeddeloh JA, Bender J, Richards EJ. 1998. The DNA methylation locus Ddm1 is required for maintenance of gene silencing in Arabidopsis. *Genes & Development* **12**:1714–1725. DOI: https://doi.org/10.1101/gad.12.11.1714, PMID: 9620857

Jiang D, Berger F. 2017. DNA replication-coupled Histone modification maintains Polycomb gene silencing in plants. *Science* **357**:1146–1149. DOI: https://doi.org/10.1126/science.aan4965, PMID: 28818970

Jiang D, Borg M, Lorković ZJ, Montgomery SA, Osakabe A, Yelagandula R, Axelsson E, Berger F. 2020. The evolution and functional divergence of the Histone H2B family in plants. *PLOS Genetics* **16**:e1008964. DOI: https://doi.org/10.1371/journal.pgen.1008964, PMID: 32716939

Johnson L, Mollah S, Garcia BA, Muratore TL, Shabanowitz J, Hunt DF, Jacobsen SE. 2004. Mass Spectrometry analysis of Arabidopsis Histone H3 reveals distinct combinations of post-Translational modifications. *Nucleic Acids Research* **32**:6511–6518. DOI: https://doi.org/10.1093/nar/gkh992, PMID: 15598823

Kakutani T, Munakata K, Richards EJ, Hirochika H. 1999. Meiotically and Mitotically stable inheritance of DNA Hypomethylation induced by Ddm1 Mutation of *Arabidopsis thaliana*. *Genetics* **151**:831–838. DOI: https://doi.org/10.1093/genetics/151.2.831, PMID: 9927473

Kharchenko PV, Alekseyenko AA, Schwartz YB, Minoda A, Riddle NC, Ernst J, Sabo PJ, Larschan E, Gorchakov AA, Gu T, Linder-Basso D, Plachetka A, Shanower G, Tolstorukov MY, Luquette LJ, Xi R, Jung YL, Park RW, Bishop EP, Canfield TK, et al. 2011. Comprehensive analysis of the Chromatin landscape in *Drosophila melanogaster*. *Nature* **471**:480–485. DOI: https://doi.org/10.1038/nature09725, PMID: 21179089

Kim J, Guermah M, McGinty RK, Lee JS, Tang Z, Milne TA, Shilatifard A, Muir TW, Roeder RG. 2009. Rad6-mediated transcription-coupled H2B Ubiquitylation directly stimulates H3K4 methylation in human cells. *Cell* **137**:459–471. DOI: https://doi.org/10.1016/j.cell.2009.02.027, PMID: 19410543

Kim YJ, Wang R, Gao L, Li D, Xu C, Mang H, Jeon J, Chen X, Zhong X, Kwak JM, Mo B, Xiao L, Chen X. 2016. POWERDRESS and Hda9 interact and promote Histone H3 Deacetylation at specific Genomic sites in Arabidopsis. *PNAS* **113**:14858–14863. DOI: https://doi.org/10.1073/pnas.1618618114, PMID: 27930340

Koyama M, Kurumizaka H. 2018. Structural diversity of the Nucleosome. *Journal of Biochemistry* **163**:85–95. DOI: https://doi.org/10.1093/jb/mvx081, PMID: 29161414

Krueger F, Andrews SR. 2011. Bismark: a flexible Aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**:1571–1572. DOI: https://doi.org/10.1093/bioinformatics/btr167, PMID: 21493656

Langmead B, Salzberg SL. 2012. Fast Gapped-read alignment with Bowtie 2. *Nature Methods* **9**:357–359. DOI: https://doi.org/10.1038/nmeth.1923, PMID: 22388286

Lei B, Capella M, Montgomery SA, Borg M, Osakabe A, Goiser M, Muhammad A, Braun S, Berger F. 2021. A synthetic approach to reconstruct the evolutionary and functional innovations of the plant Histone variant H2A.W. *Current Biology* **31**:182–191. DOI: https://doi.org/10.1016/j.cub.2020.09.080, PMID: 33096036

Leng X, Thomas Q, Rasmussen SH, Marquardt S. 2020. A G(Enomic)P(Ositioning)S(Ystem) for plant RNAPII transcription. *Trends in Plant Science* **25**:744–764. DOI: https://doi.org/10.1016/j.tplants.2020.03.005, PMID: 32673579

Liu C, Lu F, Cui X, Cao X. 2010. Histone methylation in higher plants. *Annual Review of Plant Biology* **61**:395–420. DOI: https://doi.org/10.1146/annurev.arplant.043008.091939, PMID: 20192747

Liu Y, Tian T, Zhang K, You Q, Yan H, Zhao N, Yi X, Xu W, Su Z. 2018. PCSD: a plant Chromatin state database. *Nucleic Acids Research* **46**:D1157–D1167. DOI: https://doi.org/10.1093/nar/gkx919, PMID: 29040761

Long J, Carter B, Johnson ET, Ogas J. 2023. Contribution of the Histone variant H2A.Z to expression of responsive genes in plants. *Seminars in Cell & Developmental Biology* **135**:85–92. DOI: https://doi.org/10.1016/j.semcdb.2022.04.006, PMID: 35474148

Loppin B, Berger F. 2020. Histone variants: the nexus of developmental decisions and epigenetic memory. *Annual Review of Genetics* **54**:121–149. DOI: https://doi.org/10.1146/annurev-genet-022620-100039, PMID: 32857637

Lorković ZJ, Park C, Goiser M, Jiang D, Kurzbauer MT, Schlögelhofer P, Berger F. 2017. Compartmentalization of DNA damage response between Heterochromatin and Euchromatin is mediated by distinct H2A Histone variants. *Current Biology* **27**:1192–1199. DOI: https://doi.org/10.1016/j.cub.2017.03.002, PMID: 28392109

Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. 1997. Crystal structure of the Nucleosome core particle at 2.8 A resolution. *Nature* **389**:251–260. DOI: https://doi.org/10.1038/38444, PMID: 9305837

Martin M. 2011 Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal* **17**:10–12. DOI: https://doi.org/10.14806/ej.17.1.200

Minsky N, Shema E, Field Y, Schuster M, Segal E, Oren M. 2008. Monoubiquitinated H2B is associated with the transcribed region of highly expressed genes in human cells. *Nature Cell Biology* **10**:483–488. DOI: https://doi.org/10.1038/ncb1712, PMID: 18344985

Ni K, Ren J, Xu X, He Y, Finney R, Braun SMG, Hathaway NA, Crabtree GR, Muegge K. 2020. LSH mediates gene repression through Macroh2A deposition. *Nature Communications* **11**:5647. DOI: https://doi.org/10.1038/s41467-020-19159-0, PMID: 33159050

Ni K, Muegge K. 2021. LSH Catalyzes ATP-driven exchange of Histone variants Macroh2A1 and Macroh2A2. *Nucleic Acids Research* **49**:8024–8036. DOI: https://doi.org/10.1093/nar/gkab588, PMID: 34223906

Nie X, Wang H, Li J, Holec S, Berger F. 2014. The HIRA complex that deposits the Histone H3.3 is conserved in Arabidopsis and facilitates transcriptional Dynamics. *Biology Open* **3**:794–802. DOI: https://doi.org/10.1242/bio.20148680, PMID: 25086063

Osakabe A, Tachiwana H, Takaku M, Hori T, Obuse C, Kimura H, Fukagawa T, Kurumizaka H. 2013. Vertebrate Spt2 is a novel Nucleolar Histone chaperone that assists in Ribosomal DNA transcription. *Journal of Cell Science* **126**:1323–1332. DOI: https://doi.org/10.1242/jcs.112623, PMID: 23378026

Osakabe A, Lorkovic ZJ, Kobayashi W, Tachiwana H, Yelagandula R, Kurumizaka H, Berger F. 2018. Histone H2A variants confer specific properties to Nucleosomes and impact on Chromatin accessibility. *Nucleic Acids Research* **46**:7675–7685. DOI: https://doi.org/10.1093/nar/gky540, PMID: 29945241

Osakabe A, Jamge B, Axelsson E, Montgomery SA, Akimcheva S, Kuehn AL, Pisupati R, Lorković ZJ, Yelagandula R, Kakutani T, Berger F. 2021. The Chromatin Remodeler Ddm1 prevents Transposon mobility through deposition of Histone variant H2A.W. *Nature Cell Biology* **23**:391–400. DOI: https://doi.org/10.1038/s41556-021-00658-1, PMID: 33833428

Osakabe A, Molaro A. 2023. Histone renegades: unusual H2A Histone variants in plants and animals. *Seminars in Cell & Developmental Biology* **135**:35–42. DOI: https://doi.org/10.1016/j.semcdb.2022.05.001, PMID: 35570098

Pisupati R, Nizhynska V, Morales AM, Nordborg M. 2023 On the Causes of Gene-Body Methylation Variation in *Arabidopsis Thaliana*. *bioRxiv*: . DOI: https://doi.org/10.1101/2022.12.04.519028

Probst AV. 2022. Deposition and Eviction of Histone variants define functional Chromatin States in plants. *Current Opinion in Plant Biology* **69**:102266. DOI: https://doi.org/10.1016/j.pbi.2022.102266, PMID: 35981458

Quinlan AR. 2014. Bedtools: the Swiss-army tool for genome feature analysis. *Current Protocols in Bioinformatics* **47**:11. DOI: https://doi.org/10.1002/0471250953.bi1112s47, PMID: 25199790

Raman P, Rominger MC, Young JM, Molaro A, Tsukiyama T, Malik HS. 2022. Novel classes and evolutionary turnover of Histone H2B variants in the mammalian Germline. *Molecular Biology and Evolution* **39**:msac019. DOI: https://doi.org/10.1093/molbev/msac019, PMID: 35099534

Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. 2016. Deeptools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research* **44**:W160–W165. DOI: https://doi.org/10.1093/nar/gkw257, PMID: 27079975

Ranjan A, Nguyen VQ, Liu S, Wisniewski J, Kim JM, Tang X, Mizuguchi G, Elalaoui E, Nickels TJ, Jou V, English BP, Zheng Q, Luk E, Lavis LD, Lionnet T, Wu C. 2020. Live-cell single particle imaging reveals the role of RNA polymerase II in Histone H2A.Z Eviction. *eLife* **9**:e55667. DOI: https://doi.org/10.7554/eLife.55667, PMID: 32338606

Reverón-Gómez N, González-Aguilera C, Stewart-Morgan KR, Petryk N, Flury V, Graziano S, Johansen JV, Jakobsen JS, Alabert C, Groth A. 2018. Accurate recycling of parental Histones reproduces the Histone modification landscape during DNA replication. *Molecular Cell* **72**:239–249. DOI: https://doi.org/10.1016/j.molcel.2018.08.010, PMID: 30146316

Riemondy KA, Sheridan RM, Gillen A, Yu Y, Bennett CG, Hesselberth JR. 2017. Valr: reproducible genome interval analysis in R. *F1000Research* **6**:1025. DOI: https://doi.org/10.12688/f1000research.11997.1, PMID: 28751969

Roudier F, Ahmed I, Bérard C, Sarazin A, Mary-Huard T, Cortijo S, Bouyer D, Caillieux E, Duvernois-Berthet E, Al-Shikhley L, Giraut L, Després B, Drevensek S, Barneche F, Dèrozier S, Brunaud V, Aubourg S, Schnittger A, Bowler C, Martin-Magniette ML, et al. 2011. Integrative Epigenomic mapping defines four main Chromatin States in Arabidopsis. *The EMBO Journal* **30**:1928–1938. DOI: https://doi.org/10.1038/emboj.2011.103, PMID: 21487388

Rougée M, Quadrana L, Zervudacki J, Hure V, Colot V, Navarro L, Deleris A. 2021. Polycomb mutant partially suppresses DNA Hypomethylation-associated phenotypes in Arabidopsis. *Life Science Alliance* **4**:e202000848. DOI: https://doi.org/10.26508/lsa.202000848, PMID: 33443101

Rudnizky S, Bavly A, Malik O, Pnueli L, Melamed P, Kaplan A. 2016. H2A.Z controls the stability and mobility of Nucleosomes to regulate expression of the LH genes. *Nature Communications* **7**:12958. DOI: https://doi.org/10.1038/ncomms12958, PMID: 27653784

Schultz MD, He Y, Whitaker JW, Hariharan M, Mukamel EA, Leung D, Rajagopal N, Nery JR, Urich MA, Chen H, Lin S, Lin Y, Jung I, Schmitt AD, Selvaraj S, Ren B, Sejnowski TJ, Wang W, Ecker JR. 2016. Corrigendum: human body Epigenome maps reveal Noncanonical DNA methylation variation. *Nature* **530**:242. DOI: https://doi.org/10.1038/nature16179, PMID: 26605523

Schwämmle V, Sidoli S, Ruminowicz C, Wu X, Lee CF, Helin K, Jensen ON. 2016. Systems level analysis of Histone H3 post-Translational modifications (Ptms) reveals features of PTM Crosstalk in Chromatin regulation. *Molecular & Cellular Proteomics* **15**:2715–2729. DOI: https://doi.org/10.1074/mcp.M115.054460, PMID: 27302890

Sequeira-Mendes J, Aragüez I, Peiró R, Mendez-Giraldez R, Zhang X, Jacobsen SE, Bastolla U, Gutierrez C. 2014. The functional topography of the Arabidopsis genome is organized in a reduced number of linear motifs of Chromatin States. *The Plant Cell* **26**:2351–2366. DOI: https://doi.org/10.1105/tpc.114.124578, PMID: 24934173

Stroud H, Otero S, Desvoyes B, Ramírez-Parra E, Jacobsen SE, Gutierrez C. 2012. Genome-wide analysis of Histone H3.1 and H3.3 variants in *Arabidopsis thaliana*. *PNAS* **109**:5370–5375. DOI: https://doi.org/10.1073/pnas.1203145109, PMID: 22431625

Stroud H, Greenberg MVC, Feng S, Bernatavichute YV, Jacobsen SE. 2013. Comprehensive analysis of silencing Mutants reveals complex regulation of the Arabidopsis Methylome. *Cell* **152**:352–364. DOI: https://doi.org/10.1016/j.cell.2012.10.054, PMID: 23313553

Subramanian V, Fields PA, Boyer LA. 2015. H2A.Z: a molecular Rheostat for transcriptional control. *F1000prime Reports* **7**:01. DOI: https://doi.org/10.12703/P7-01, PMID: 25705384

Sullivan AM, Arsovski AA, Lempe J, Bubb KL, Weirauch MT, Sabo PJ, Sandstrom R, Thurman RE, Neph S, Reynolds AP, Stergachis AB, Vernot B, Johnson AK, Haugen E, Sullivan ST, Thompson A, Neri FV, Weaver M, Diegel M, Mnaimneh S, et al. 2014. Mapping and Dynamics of regulatory DNA and transcription factor

networks in *A. thaliana*. *Cell Reports* **8**:2015–2030. DOI: https://doi.org/10.1016/j.celrep.2014.08.019, PMID: 25220462

Sun Z, Bernstein E. 2019. Histone variant Macroh2A: from Chromatin deposition to molecular function. *Essays in Biochemistry* **63**:59–74. DOI: https://doi.org/10.1042/EBC20180062, PMID: 31015383

Surface LE, Fields PA, Subramanian V, Behmer R, Udeshi N, Peach SE, Carr SA, Jaffe JD, Boyer LA. 2016. H2A.Z.1 Monoubiquitylation Antagonizes Brd2 to maintain poised Chromatin in Escs. *Cell Reports* **14**:1142–1155. DOI: https://doi.org/10.1016/j.celrep.2015.12.100, PMID: 26804911

Tachiwana H, Kagawa W, Osakabe A, Kawaguchi K, Shiga T, Hayashi-Takanaka Y, Kimura H, Kurumizaka H. 2010. Structural basis of instability of the Nucleosome containing a Testis-specific Histone variant, human H3t. *PNAS* **107**:10454–10459. DOI: https://doi.org/10.1073/pnas.1003064107, PMID: 20498094

Tachiwana H, Kagawa W, Kurumizaka H. 2012. Comparison between the CENP-A and Histone H3 structures in Nucleosomes. *Nucleus* **3**:6–11. DOI: https://doi.org/10.4161/nucl.18372, PMID: 22127263

Talbert PB, Henikoff S. 2021. Histone variants at a glance. *Journal of Cell Science* **134**:jcs244749. DOI: https://doi.org/10.1242/jcs.244749, PMID: 33771851

Tanaka Y, Tawaramoto-Sasanuma M, Kawaguchi S, Ohta T, Yoda K, Kurumizaka H, Yokoyama S. 2004. Expression and purification of recombinant human Histones. *Methods* **33**:3–11. DOI: https://doi.org/10.1016/j.ymeth.2003.10.024, PMID: 15039081

Taus T, Köcher T, Pichler P, Paschke C, Schmidt A, Henrich C, Mechtler K. 2011. Universal and confident Phosphorylation site localization using phosphoRS. *Journal of Proteome Research* **10**:5354–5362. DOI: https://doi.org/10.1021/pr200611n, PMID: 22073976

Tessarz P, Kouzarides T. 2014. Histone core modifications regulating Nucleosome structure and Dynamics. *Nature Reviews. Molecular Cell Biology* **15**:703–708. DOI: https://doi.org/10.1038/nrm3890, PMID: 25315270

Wang Y, Long H, Yu J, Dong L, Wassef M, Zhuo B, Li X, Zhao J, Wang M, Liu C, Wen Z, Chang L, Chen P, Wang QF, Xu X, Margueron R, Li G. 2018. Histone variants H2A.Z and H3.3 Coordinately regulate Prc2-dependent H3K27Me3 deposition and gene expression regulation in mES cells. *BMC Biology* **16**:107. DOI: https://doi.org/10.1186/s12915-018-0568-6, PMID: 30249243

Weber CM, Henikoff S. 2014. Histone variants: dynamic punctuation in transcription. *Genes & Development* **28**:672–682. DOI: https://doi.org/10.1101/gad.238873.114, PMID: 24696452

Willige BC, Zander M, Yoo CY, Phan A, Garza RM, Wanamaker SA, He Y, Nery JR, Chen H, Chen M, Ecker JR, Chory J. 2023. Author correction: PHYTOCHROME-INTERACTING factors trigger environmentally responsive Chromatin Dynamics in plants. *Nature Genetics* **55**:355. DOI: https://doi.org/10.1038/s41588-023-01299-w, PMID: 36690790

Wollmann H, Holec S, Alden K, Clarke ND, Jacques PÉ, Berger F. 2012. Dynamic deposition of Histone variant H3.3 accompanies developmental remodeling of the Arabidopsis Transcriptome. *PLOS Genetics* **8**:e1002658. DOI: https://doi.org/10.1371/journal.pgen.1002658, PMID: 22570629

Wollmann H, Stroud H, Yelagandula R, Tarutani Y, Jiang D, Jing L, Jamge B, Takeuchi H, Holec S, Nie X, Kakutani T, Jacobsen SE, Berger F. 2017. The Histone H3 variant H3.3 regulates gene body DNA methylation in *Arabidopsis thaliana*. *Genome Biology* **18**:94. DOI: https://doi.org/10.1186/s13059-017-1221-3, PMID: 28521766

Yang Y, Zhang L, Xiong C, Chen J, Wang L, Wen Z, Yu J, Chen P, Xu Y, Jin J, Cai Y, Li G. 2022. HIRA complex Presets transcriptional potential through coordinating depositions of the Histone variants H3.3 and H2A.Z on the poised genes in mESCs. *Nucleic Acids Research* **50**:191–206. DOI: https://doi.org/10.1093/nar/gkab1221, PMID: 34893908

Yelagandula R, Stroud H, Holec S, Zhou K, Feng S, Zhong X, Muthurajan UM, Nie X, Kawashima T, Groth M, Luger K, Jacobsen SE, Berger F. 2014. The Histone variant H2A.W defines Heterochromatin and promotes Chromatin condensation in Arabidopsis. *Cell* **158**:98–109. DOI: https://doi.org/10.1016/j.cell.2014.06.006, PMID: 24995981

Zhang K, Sridhar VV, Zhu J, Kapoor A, Zhu JK. 2007. Distinctive core Histone post-Translational modification patterns in *Arabidopsis thaliana*. *PLOS ONE* **2**:e1210. DOI: https://doi.org/10.1371/journal.pone.0001210, PMID: 18030344

Zhao F, Xue M, Zhang H, Li H, Zhao T, Jiang D. 2023. Coordinated Histone variant H2A.Z Eviction and H3.3 deposition control plant Thermomorphogenesis. *The New Phytologist* **238**:750–764. DOI: https://doi.org/10.1111/nph.18738, PMID: 36647799

Zhou Y, Romero-Campero FJ, Gómez-Zambrano Á, Turck F, Calonje M. 2017. H2A Monoubiquitination in *Arabidopsis thaliana* is generally independent of Lhp1 and Prc2 activity. *Genome Biology* **18**:69. DOI: https://doi.org/10.1186/s13059-017-1197-z, PMID: 28403905

Zilberman D, Coleman-Derr D, Ballinger T, Henikoff S. 2008. Histone H2A.Z and DNA methylation are mutually antagonistic Chromatin marks. *Nature* **456**:125–129. DOI: https://doi.org/10.1038/nature07324, PMID: 18815594

Histone variants shape chromatin states in Arabidopsis

Bhagyshree Jamge[1,2,*], Zdravko J. Lorković[1,*], Elin Axelsson[1,*], Akihisa Osakabe[1,3,4], Vikas Shukla[1,2], Ramesh Yelagandula[1,5,6], Svetlana Akimcheva[1], Annika Luisa Kuehn[1], and Frédéric Berger[1**]

Supplemental figures and legends

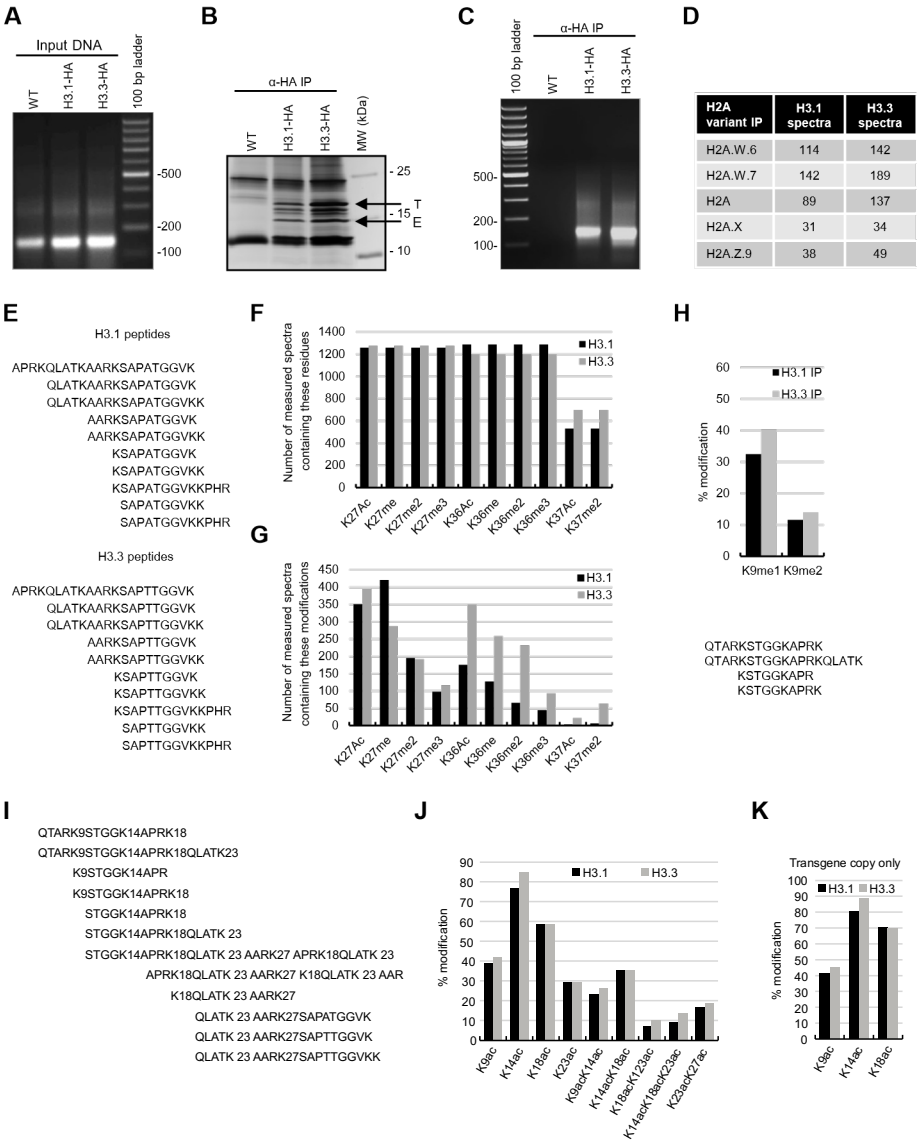


Figure 1 figure supplement 1

Figure 1 figure supplement 1. Biochemical analysis of the association between histone variants and histone marks. (**A**) Analysis of DNA in input samples used for H3.1 and H3.3 mononucleosome immunoprecipitation after MNase digestion. (**B**) Silver-stained 8-20% gradient gel of immunoprecipitated H3.1 and H3.3 mononucleosomes. Arrows indicate transgenic (T) and

endogenous (E) H3 proteins. Molecular weight markers are indicated on the right. (**C**) Analysis of DNA extracted from immunoprecipitated H3.1 and H3.3 mononucleosomes. (**D**) Spectral counts of H3.1- and H3.3-specific peptides (as indicated in panel **E**) in mononucleosomes immunoprecipitated with H2A variant-specific antibodies. (**E**) H3.1- and H3.3-specific peptides used for the analysis of H3K27, H3K36, and H3K37 modifications. (**F**) Number of MS measured spectra that covered the indicated residues in H3.1 and H3.3 mononucleosomes. (**G**) Number of measured MS spectra containing the indicated modifications in H3.1 and H3.3 mononucleosomes. (**H**) (top) Relative methylation levels of H3K9me1 and H3K9me2 on H3.1 and H3.3 revealed from transgenic copies of the respective H3 variants. (bottom) Sequences of peptides used to evaluate H3K9me1 and H3K9me2 by MS. (**I**) Sequences of peptides used to evaluate H3 acetylation by MS. (**J**) Relative acetylation levels of the indicated H3 residues in immunoprecipitated H3.1 and H3.3 mononucleosomes without differentiating H3.1 and H3.3 variants. (**K**) Relative acetylation levels of the indicated H3 residues on H3.1 and H3.3 revealed from transgenic copies of the respective H3 variants.

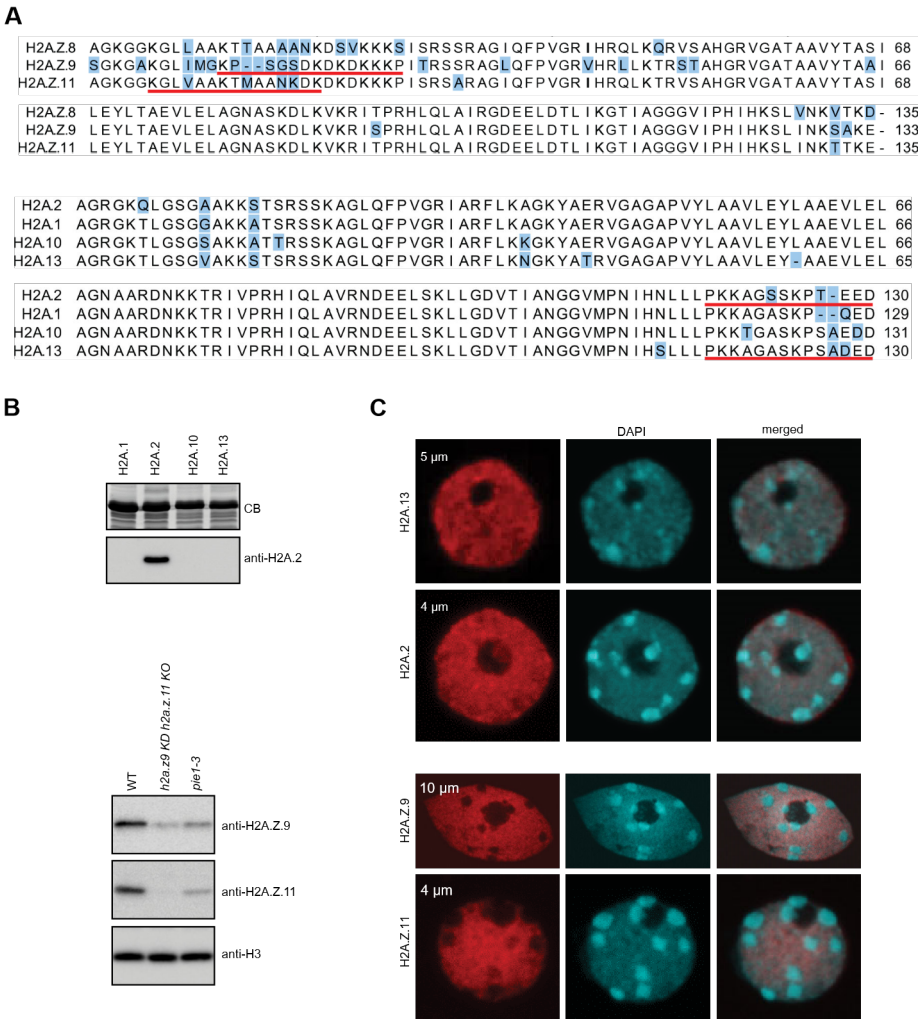


Figure 2 figure supplement 1

Figure 2 figure supplement 1. Validation of H2A.2 and H2A.Z.11 polyclonal antibodies. (**A**) Sequence alignments of three H2A.Z and four H2A histone variants. Positions that differ between variants are highlighted in blue. Underlined are sequences of peptides used for immunization. (**B**) Antibodies against H2A.2 were tested with four H2A variants expressed in bacteria. Relevant part of Coomassie stained gel with overexpressed H2As is shown on the top. As shown on the bottom panel antibody raised against H2A.2 peptide recognized only H2A.2. Antibodies against H2A.Z.11

were tested on nuclear extracts from WT, *h2a.z.9* knock-down/*h2a.z.11* knock out (*h2a.z.9 KD h2a.z.11 KO*; this line still expresses H2A.Z.8), and *pie1-3* (mutant in large subunit of the SWR1 complex responsible for deposition of H2A.Z) lines. The H2A.Z.11 antibodies did not recognize any protein in the *h2a.z.9 KD h2a.z.11* KO line indicating its specificity. In *pie1-3* line, both H2A.Z.9 and H2A.Z.11 were detected but at reduced levels as compared to WT which is in line with the inefficient deposition of H2A.Z in the absence of SWR1. The H3 antibody served as the loading control. (**C**) Immunostaining of WT nuclei with indicated antibodies.

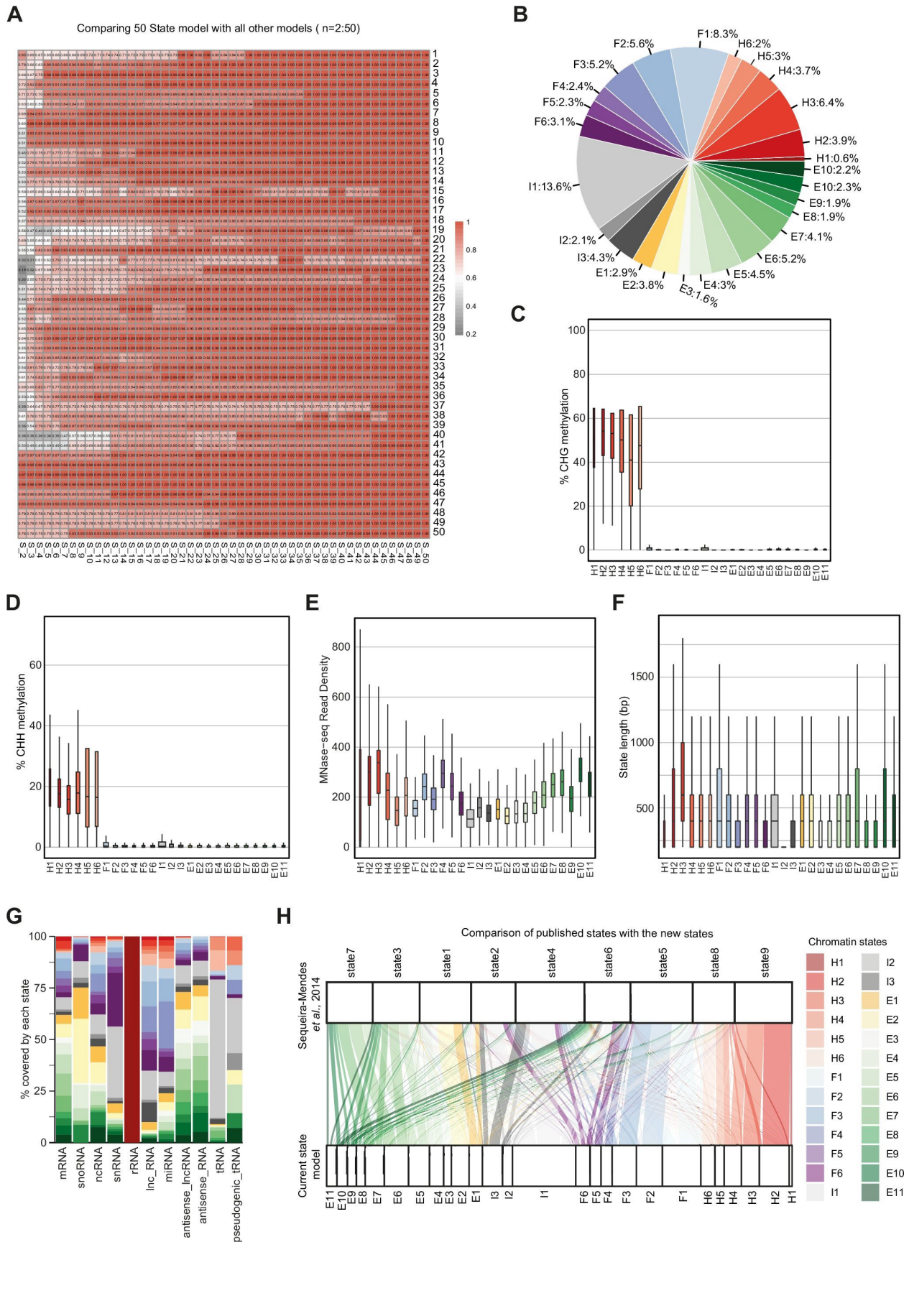

A
Comparing 50 State model with all other models (n=2:50)
B
F1:8.3%
H6:2%
H5:3%
H4:3.7%
H3:6.4%
H2:3.9%
H1:0.6%
E10:2.2%
E10:2.3%
E9:1.9%
E8:1.9%
E7:4.1%
E6:5.2%
E5:4.5%
E4:3%
E3:1.6%
E2:3.8%
E1:2.9%
I3:4.3%
I2:2.1%
I1:13.6%
F6:3.1%
F5:2.3%
F4:2.4%
F3:5.2%
F2:5.6%
C
% CHG methylation
D
% CHH methylation
E
MNase−seq Read Density
F
State length (bp)
G
% covered by each state
mRNA
snoRNA
ncRNA
snRNA
rRNA
lnc_RNA
miRNA
antisense_lncRNA
antisense_RNA
tRNA
pseudogenic_tRNA
H
Comparison of published states with the new states
Sequeira-Mendes et al., 2014
Current state model
state7
state3
state1
state2
state4
state6
state5
state8
state9
Chromatin states
H1
H2
H3
H4
H5
H6
F1
F2
F3
F4
F5
F6
I1
I2
I3
E1
E2
E3
E4
E5
E6
E7
E8
E9
E10
E11

Figure 2 figure supplement 2. Properties of chromatin states in *Arabidopsis thaliana.* (**A**) Heatmap showing the correlation (0 to 1) between emission parameters of each model. The Y axis represents the distinct states in the reference state model (n=50). The X-axis represents models including 2 to 49 states compared to the 50-state model. The correlation plotted here is the maximum correlation between each state of the reference model and any state of the other models being compared. (**B**) Pie chart showing the proportion of the genome covered by each state. (**C, D**) Box plots showing levels of CHG (**C**) and CHH (**D**) methylation across states. (**E**) Box plot showing nucleosome occupancy (MNase-seq read density) across states. (**F**) Box plots showing the length (bp) of chromatin states. (**G**) Stacked bar plot showing the overlap between non-protein-coding gene features and chromatin states. (**H**) Flow diagram showing the overlap (in bp) between published states (Sequeira-Mendes et al., 2014) and the current states. The color code for the flow diagram represents the chromatin states in the current model.

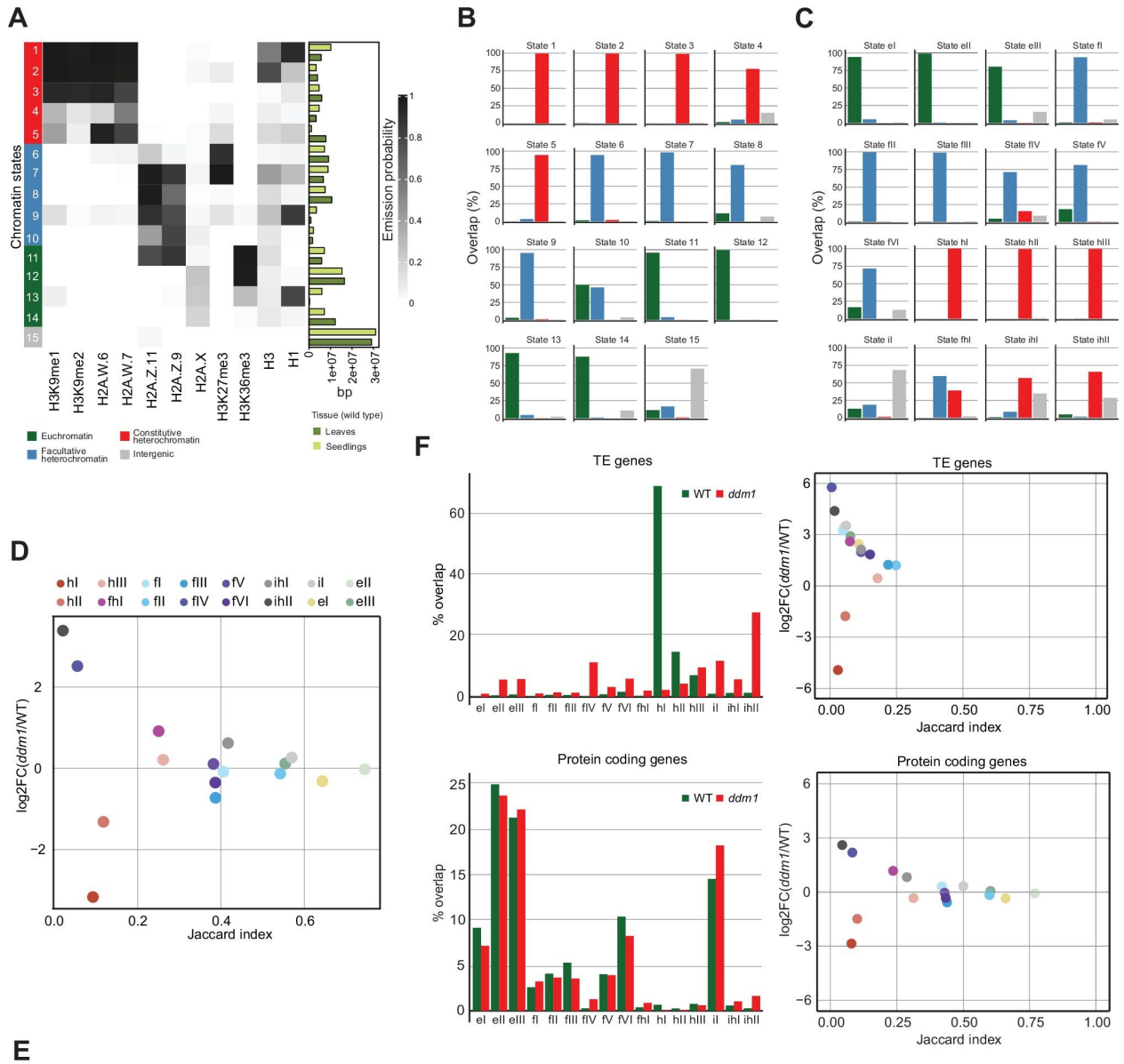

A
Chromatin states
H3K9me1
H3K9me2
H2A.W.6
H2A.W.7
H2A.Z.11
H2A.Z.9
H2A.X
H3K27me3
H3K36me3
H3
H1
Emission probability
bp
Tissue (wild type)
Leaves
Seedlings
Euchromatin
Facultative heterochromatin
Constitutive heterochromatin
Intergenic
B
Overlap (%)
C
D
log2FC(ddm1/WT)
Jaccard index
F
TE genes
Protein coding genes
% overlap
WT
ddm1
E
WT
ddm1

Figure 3 figure supplement 1. DDM1 loss of function disrupts chromatin states in *Arabidopsis thaliana.* (**A**) Heatmap showing the emission probability of histone mark/variant across the 15 chromatin states of the concatenated model computed with chromatin profiles from seedlings and leaves. The bar plot on the right represents the proportion of the genome covered by each state in seedlings (light green) and in leaves (dark green). (**B**) Bar plots of the genomic overlap between the seedling and leaf concatenated model states and states from the four chromatin types in the 26-state model. 14 states show overlaps with predominantly one chromatin type, whereas only state 10 overlaps with two types in similar proportions. (**C**) Bar plots of the genomic overlap between the wild type and *ddm1* concatenated model states and states from the four chromatin types in the 26-state model. 13 states overlap with predominantly one chromatin type, whereas 3 states overlap with two types. In **B** and **C**, color codes for four major groups of chromatin states are as in panel **A**. (**D**) Scatter plot of the Jaccard index *vs*. the log2 fold change in proportion of genome for the 16 states from the wild type and *ddm1* concatenated model. For each state the Jaccard index was calculated for the genomic regions covered in wild type and in *ddm1*. The log2 fold change was calculated using the total number of base pairs covered by each state in *ddm1* mutant compared to in the wild type (see also methods). (**E**) Alluvial plot showing the chromatin state changes between the wild type (top row) and *ddm1* (bottom row). The plot was generated using the R package ggalluviel v0.12.5 (http://corybrunson.github.io/ggalluvial/). (**F**) Scatter plots of the Jaccard index *vs.* the log2 fold change in proportion of genome for the 16 states from the wild type and *ddm1* concatenated model (see **Figure 3 figure supplement 1D** and Methods) for TE genes and protein coding genes separately (right panels). In addition, the percentages of the combined length of the TE/protein coding genes covered by each state are shown as bar plots on the left. Color codes for chromatin states are as in panel **D**.

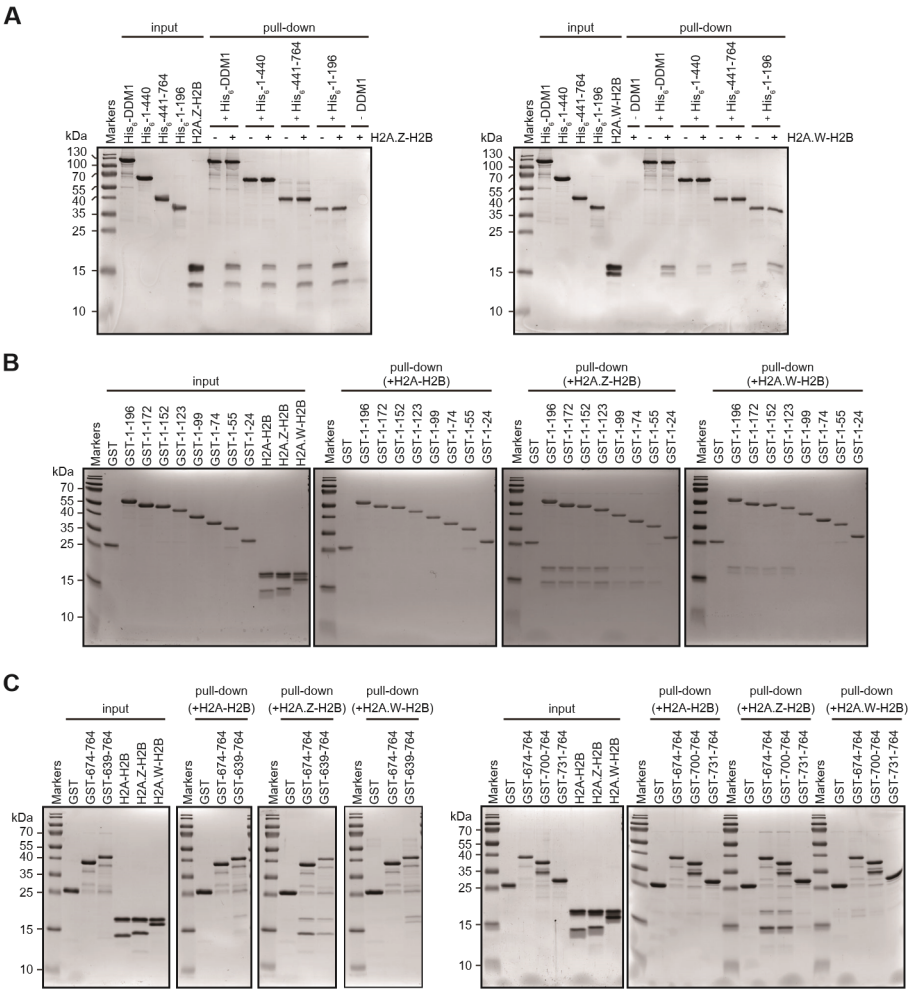


Figure 3 figure supplement 2

Figure 3 figure supplement 2. Interaction of DDM1 and DDM1 deletion mutants with histone variants. (**A-C**) Identification of H2A.W and H2A.Z binding sites in DDM1. His_6- (**A**) or GST- (**B** and **C**) tagged DDM1 fragments, as indicated with amino acids numbers, were incubated with indicated H2A-H2B dimers and after washing samples were analyzed on 15% SDS-PAGE and stained with Coomassie.

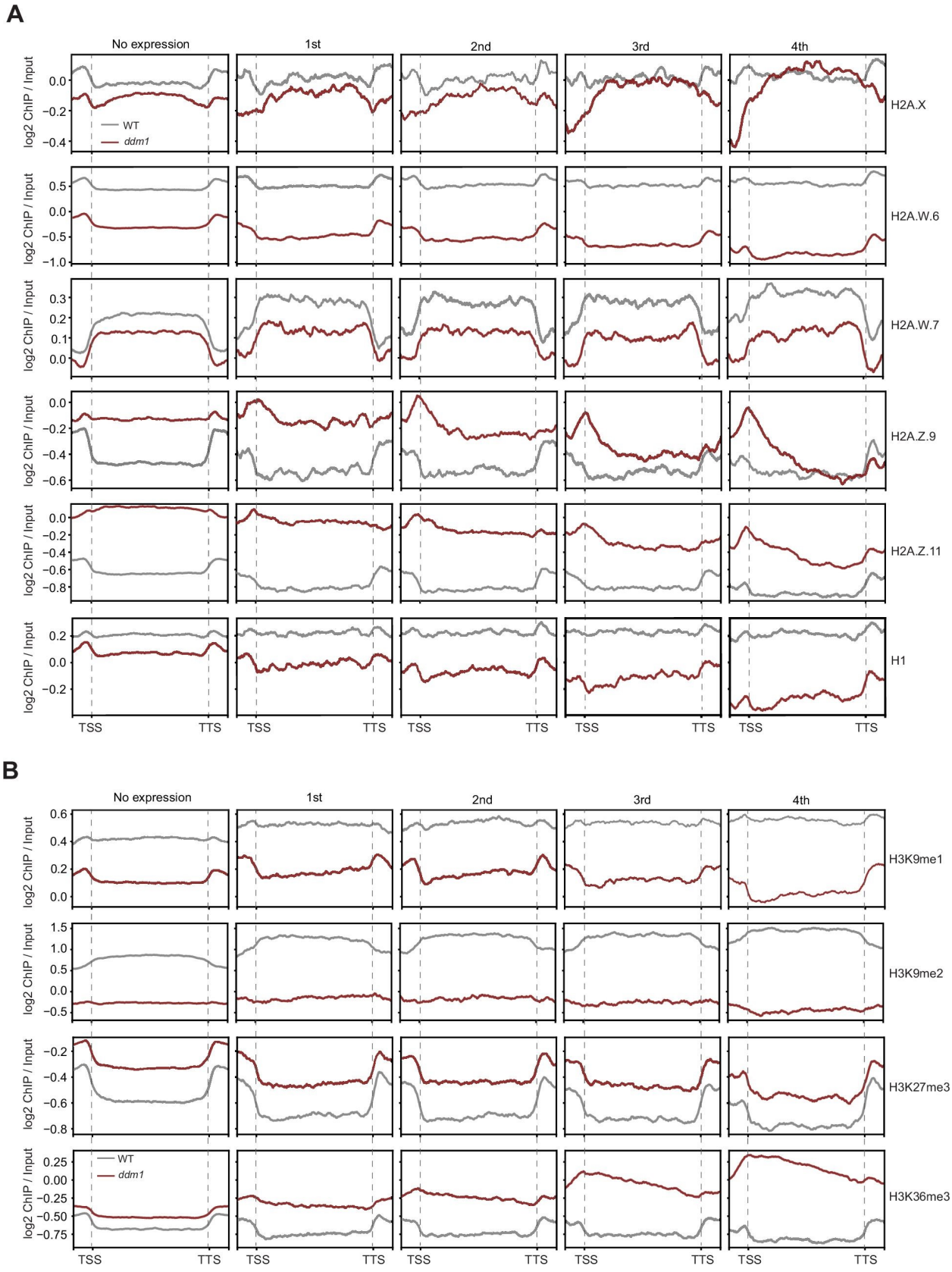

A
No expression
1st
2nd
3rd
4th
log2 ChIP / Input
WT
ddm1
H2A.X
H2A.W.6
H2A.W.7
H2A.Z.9
H2A.Z.11
H1
TSS
TTS
B
H3K9me1
H3K9me2
H3K27me3
H3K36me3

Figure 4 figure supplement 1. Analyses of the parameters that could correlate the chromatin states with TE expression in *ddm1.* Enrichment profiles of histone variants (**A**) and H3 modifications (**B**) plotted over groups of TE genes based on their expression in *ddm1* as defined in the legend for **Figure 4A** and Methods.

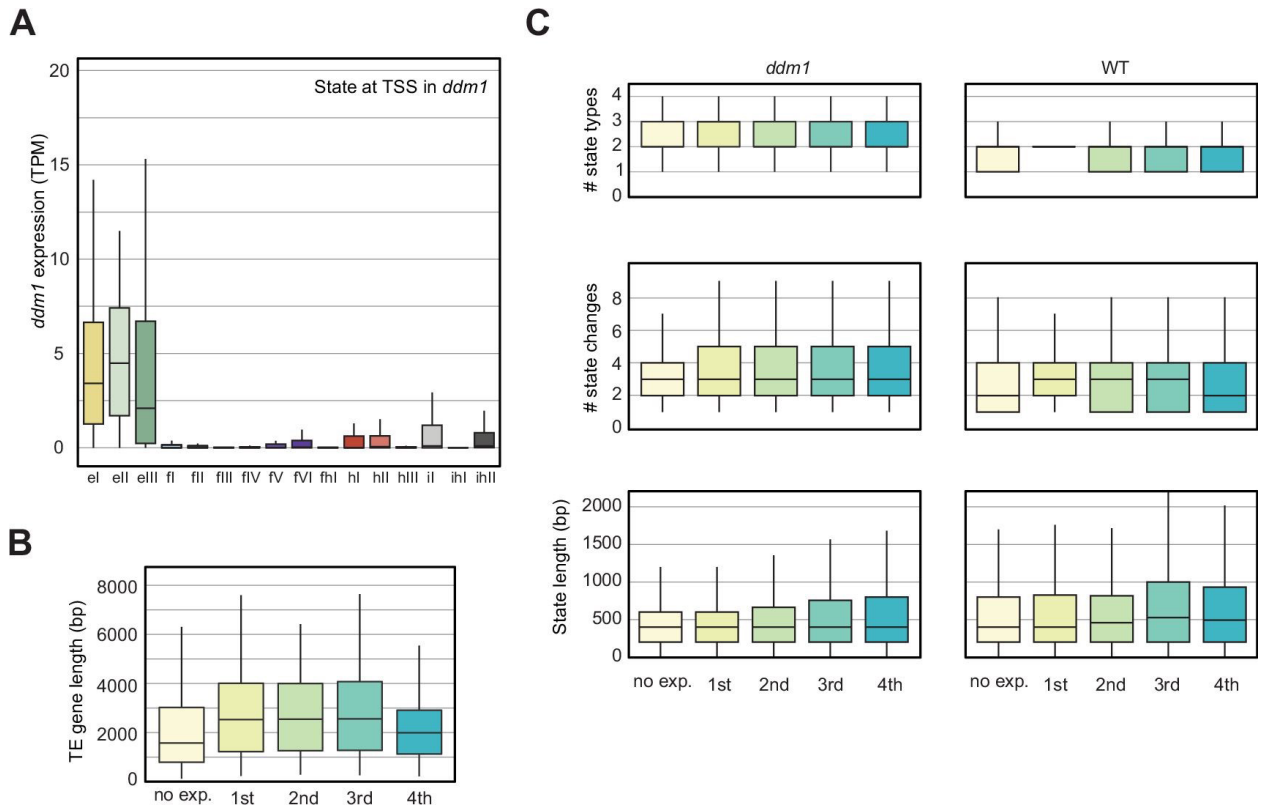


Figure 4 figure supplement 2. Analyses of the parameters that could correlate the chromatin states with TE expression in *ddm1.* (**A**) Box plot showing the expression of the TE genes grouped by the state overlapping the TSS in *ddm1*. See also **Figure 4C**. (**B**) Box plot showing the TE gene length distributions across the 5 different TE gene expression groups. (**C**) Box plots showing state statistics for TE genes per expression group in *ddm1* and wild type (WT). "# state types" (top) is the number of different kinds of states present over the TE genes. "# state changes" (middle) is the number of times there is a transition from one state to another and the bottom shows the length of the states overlapping the TE genes. No trend can be seen across expression groups, whereas all TE genes tend to have a larger diversity of states, more frequent transitions between states and hence shorter stretches of each state.

Conservation of chromatin states and their association with transcription factors in land plants

Shukla, V., Axelsson, E., Hisanaga, T., Haseloff, J., Berger, F., & Romani, F. (2025). Conservation of chromatin states and their association with transcription factors in land plants (p. 2025.07.12.664529). bioRxiv.

https://doi.org/10.1101/2025.07.12.664529

Contribution: Methodology, Validation, Formal Analysis, Investigation, Data Curation, Writing – original draft, and Visualization.

The manuscript was posted on bioRxiv on July 17th, 2025. The revised manuscript is currently under consideration for publishing with the journal as of November 24, 2025.

Conservation of chromatin states and their association with transcription factors in land plants

Vikas Shukla[1,2], Elin Axelsson[1], Tetsuya Hisanaga[1,3], Jim Haseloff[4], Frédéric Berger[1], and Facundo Romani[4].

[1]Gregor Mendel Institute, Austrian Academy of Sciences, Vienna, 1030, 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, 8916-5 Takayama, Ikoma, Nara 630-0192, Japan

[4]Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, United Kingdom.

*Corresponding authors. Emails: frederic.berger@gmi.oeaw.ac.at and fr391@cam.ac.uk

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. The association with the +1 nucleosome defines a list of candidate pioneer factors we know little about in plants, compared to their important roles in animal stem cells and early development.

KEYWORDS

Chromatin, evolution, transcription factors, land plants, *Arabidopsis*, *Marchantia*

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). 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 applications outside *Arabidopsis thaliana* 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, only a handful of TFs have been proposed to act as pioneer factors in plants. The best characterized 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 candidate pioneer factors. We propose that TF–chromatin relationships follow similar principles in both species, supporting the conservation of regulatory strategies across land plants.

METHODS

Profiling of *M. polymorpha* H2A variants using ChIP-seq

ChIP experiments were performed using a previously described protocol with some modifications (26). Two-week-old thalli started from gemmae of *M. 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 $MgCl_2$, 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 $MgCl_2$, 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 $MgCl_2$, 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, 16.7mM Tris-HCl pH

8.0, 167 mM NaCl). Diluted chromatin was cleared by incubating with proteinA/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 $NaHCO_3$) 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 [source table 1]. 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 [source table 1].

Raw BAM files were converted to FASTQ files using the “bamtofastq” sub-command of the “BEDTools suite” v2.27.1 (27). Sequencing quality of the raw files was evaluated using quality reports generated by FastQC v0.11.5 (28). 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 (29) and further processed using SAMtools v1.9 (30) and BEDTools v2.27.1 (27). 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 (31). 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 200 bp 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 to 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) [Supplementary Figure S1E]. A follow up on all models between 23 to 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.

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* 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*.

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 (32). The results were plotted as stacked bar plots using the Tidyverse (33) package of R. [Figures 1B, 2B, and 3B]

The NeighbourhoodEnrichment command of ChromHMM v1.23 (2) was used to calculate the enrichment of each state relative to the Transcription Start Site (TSS) 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 (33) package of R. [Supplementary Figures S1A, S2A, and S3A]

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 (33) package of R showing the preferential localization of heterochromatic states near the centromere region. [Supplementary Figures S1F and S2F]

We identified the states of constitutive heterochromatin by their strong enrichment in the regions containing Transposable Elements (TE), TE genes and TE fragments [Figures 1B, 2B, and 3B] along with high emission probabilities of know heterochromatic marks including H2A.W, H3K9me1/2 among others. [Figures 1A, 2A, and 3A] 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*. [Supplementary Figures S1B, S2C, and S3B]. 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. [Supplementary Figures S1A, S2A, and S3A]

(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 peri-centromere in case of *Arabidopsis* states. [Supplementary Figures S1F and S2F]

(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 "MpI" 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 (33) package of R. [Figures 1A, 2A, and Supplementary Figure 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 (33) package of R. [Supplementary Figures S1G, S2A, and S3C]

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" (34) package of R. Similar analysis also led to the comparison of the *Arabidopsis* 26 and 15-state models. [Supplementary Figures S1E and S3I]

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 (35) package of R. [Supplementary Figure S2G]

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 (36) while the RNA-seq data was

downloaded from (37). All data was aligned to TAIR10 using bowtie2 v2.4.1 (29) and further processed using SAMtools v1.9 (30) and BEDTools v2.27.1(27). 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 (32). The results were plotted as box plots using the Tidyverse (33) package of R [Figure 1A, and Supplementary Figures S1A-D, S3D-G].

The Bisulfite-seq, ATAC-seq, and RNA-seq data for *Marchantia* was downloaded from (38) 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 [Figure 2A, and Supplementary Figures S2B-E].

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 (39) for *Marchantia* and EvoRepro for *Arabidopsis* (40). 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 (38) with data from (38) and (41).

Transcription Factor Occupancy and Chromatin State Association

To analyse TF occupancy, we compiled publicly available *Arabidopsis* thaliana ChIP-seq and DAP-seq BED files from multiple studies (19,42,43). For *Marchantia* polymorpha, we inferred TF binding sites by mapping position weight matrices (PWMs) from (42) to the *Marchantia* v7.1 genome using FIMO (MEME Suite) (44). 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 (45) with default parameters. Then, co-expression data from ATTEDII (46) (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 this contingency table:

Count (genes with peaks + co expressed)	count (genes with peaks)

Count (co expressed)	27205 for *Arabidopsis or* 18328 for *Marchantia*

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

For each method we generated an $N_{TF} \times N_{state}$ 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 (42) and PlantTFDB (47). GO term enrichment analysis of biological processes in *Arabidopsis* was performed using PANTHER (48) with default parameters.

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 [Figure 1A] and were associated with diverse levels of transcripts [Supplementary Figure S1A], DNA methylation [Supplementary Figures S1B – D], and were differentially associated with transposable elements, genes or intergenic regions [Figure 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) [Figure 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 [Supplementary Figure S1E].

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 [Supplementary Figure S1F]. 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) [Supplementary Figure S1F]. This constitutive heterochromatin represents ca. 20% of the genome [Supplementary Figure S1G].

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. [Figure 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. Each nucleosome occupies ca. 200 bp and the states E1, E2, and E3 showed enrichment in positions corresponding 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 [Figures 1A, 1C and Supplementary Figure S1B], states E6, E7, and E8 occupied the end of the gene body [Figure 1C]. Ploting 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 [Figure 1D]. A succession of facultative heterochromatin states F3 – F5 occupied gene bodies [Figure 1C] although their organization was less complex compared with euchromatin states. The facultative

heterochromatin states were primarily associated with the intergenic states I1 to I3, based on their enrichment in H3K27me3 and H2AK121ub, low accessibility, and low gene expression [Figure 1A and Supplementary Figure S1A] and altogether occupied circa 30% of the genome [Supplementary Figure S1G]. Contrasting with these graded patterns of 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 TES (Fig 1C), but shared enrichment in linker histone H1 and were associated with genes expressed at very low levels [Figures 1A, 1C and Supplementary Figure 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 (49). To the set of publicly available histone-modification enrichment profiles (38,50), 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* (26). Enrichment of chromatin features was mapped onto the newest telomere-to-telomere chromosome-level genome assembly v7.1 (51). Using ChromHMM we described fifteen chromatin states [Figure 2A; see Methods] which were annotated based on their differential enrichment in different genomic features

[Figure 2B]. Intergenic states covered approximately a quarter of the genome [Supplementary Figure 2A]. Chromatin states associated with specific levels of gene expression [Supplementary Figure S2B], chromatin accessibility [Figure 2A], and DNA methylation [Supplementary Figure S2C - E]. Based on the enrichment of known constitutive heterochromatin marks, including H3K9me1/2 and H3K27me1, we ascribed six distinct states to constitutive heterochromatin (MpH1 - MpH6) [Figure 2A] that covered 39% of the genome [Supplementary Figure S2A]. These states were also distinguished by the lowest accessibility, enrichment in the H2A variant H2A.M.2 [Figure 2A], the highest degree of DNA methylation [Figures S2C - E] and were primarily present on transposable elements, with a minor representation on protein coding genes [Figure 2B]. The hallmark of heterochromatin states is their enrichment in H2A.M.2 and H3K9me2, but they are divided into three groups [Figure 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 covered 16 % of the genome [Figures 2A, 2B and Supplementary Figures S2A, and S2B] and the state MpF1 also enriched in H3K27me3. As in *Arabidopsis*, facultative heterochromatin states were also enriched in H3K4me1 and H3K4me3. The neighbourhood analyses centred on TSS and TTS showed that the association of

intergenic states MpI1 upstream of MpF1, both enriched in H3K27me3, representing one third of facultative heterochromatin with low accessibility occupying primarily repressed genes [Figures 2A-C and Supplementary Figures S2A, S2B]. The remaining two thirds of facultative heterochromatin of *Marchantia* was occupied by the states MpI2 and MpF2, marked by an enrichment of H2A.Z and H2AUb but not H3K27me3 and occupying more accessible chromatin and genes less repressed than genes covered by F1 [Figures 2A, 2C, 2D and Supplementary Figure S2B]. MpI2 occupies 19% of the genome and altogether facultative heterochromatin occupies 37% of the genome which is much higher than in *Arabidopsis*.

Regions with the highest gene expression represented euchromatin [Supplementary Figure S2A] and comprised the intergenic state MpI3 and four genic euchromatin states E1 – E4 covering altogether with 24 % of *Marchantia* genome [Figures 2A, 2B and Supplementary Figure S2A]. The neighbourhood analyses centred on TSS and TTS showed that the intergenic state MpI3 corresponds to a region located in intergenic regions upstream of the TSS [Figures 2B, 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. The most accessible, non-intergenic state is MpE1, which marked the +1 nucleosome of most protein-coding genes of *Marchantia* [Figures 2A, 2B]. This state, with strong enrichment of H2A.Z and H3K36me3, is present in the first few nucleosomes of active genes while the states MpE2 to MpE4 successively occupy the gene body and are enriched in H2A.X [Figures 2C, 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 [Supplementary Figure S2F]. 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.

To directly compare the chromatin states of *Arabidopsis* and *Marchantia*, we first calculated a 15-states model in *Arabidopsis* based on the same 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 [Figure 1A, Supplementary Figure S3A and S3I]. To compare the 15 chromatin states of *Arabidopsis* and *Marchantia*, we performed hierarchical clustering of emission probabilities of the two models [Supplementary Figure S2G]. 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 MpI2 and MpF2 were more similar to accessibility of euchromatin states than in *Arabidopsis* [Figure 2A, Supplementary Figure S3A] suggesting an ambivalent role for this subtype of facultative heterochromatin in *Marchantia*. To further compare the regulation by PRC2 in

both species, we analysed chromatin landscapes in the *Marchantia* Mp*knox2* Mpe*(z)1 double* mutant, which exhibited reduced PRC2 activity (41,50). Peaks enriched in H3K27me3 in the wild type were depleted in the mutant deprived of PRC2, but the loss of PRC2 activity in *Marchantia* did not affect the deposition of H2AUb [Figure 2E] presumably by PRC1 (52). These findings suggest a strong 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. The bulk of the modification H3K27me3 was associated with transposable elements and only to a minor fraction of facultative heterochromatin. 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 (50) compared with *Arabidopsis* (53,54). 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. This is in contrast with *Arabidopsis*, where the intergenic state I4 symmetrically occupies these two regions [Figure 2C]. This further suggests that in *Marchantia*, the orientation of genes defines distinct chromatin environment 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

[Figure 3A]. In the case of *Arabidopsis*, I3 was also depleted in highly expressed genes and enriched in genes with low expression [Figures 3A and 3D]. The E1 state, 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 [Figures 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 [Figures 3B and 3E]. In contrast, euchromatin states showed the opposite trend. A similar pattern was observed when analyzing chromatin states within coding regions, with a much more prominent participation of euchromatin states [Supplementary Figure S4].

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 [Figures 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 [Figures 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. 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 [Figure 4A], to a large extent consistent with the accessibility constraints [Figure 1A]. Prediction of TF binding sites using DAP-seq, which does not depend on chromatin accessibility (42), exhibited a broader distribution across chromatin states with some preference for intergenic states [Figure 4B]. In contrast,

PWM-based TF binding predictions, which only rely on DNA binding motifs, displayed a nearly uniform distribution [Figure 4C].

The observed differences between experimental approaches highlight the need for integrative analyses to fully resolve TF-chromatin state relationships in plants. TF occupancy combining data from ChIP-seq and DAP-seq was evaluated using an approach similar to that applied in human TF studies (5). To extend our analysis, we assessed TF activity, incorporating co-expression data as a proxy for functional interactions (55). 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 [Figures 4D and 4E], which is the main factor that differentiated ChiP-seq from DAP-seq and PWM [Figures 4A - 4E]. With this method to estimate TF activity, the scores of TF occupancy and activity converged. To look at different patterns of chromatin preferences among TFs, we kept ChIP-seq and DAP-seq data for ~300 TFs in *Arabidopsis* (after filtering out TFs with low scores of occupancy and activity). Using hierarchical clustering, we identified distinct associations between TFs and chromatin states [Figure 4F] with distinct binding preferences relative to the TSS [Figure 4G].

Unlike *Arabidopsis*, *Marchantia* lacks a comprehensive dataset of ChIP-seq or DAP-seq experiments. However, previous studies have shown that TF DNA binding preferences are highly conserved across land plants (56), 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 (57) to serve as an alternative approach to infer TF binding. As with *Arabidopsis*, PWM-derived occupancy was evenly distributed across chromatin states [Supplementary Figure S2H]. However, when TF activity was estimated using co-expression, a striking shift emerged [Supplementary Figure S2I]. The most accessible chromatin states mI2 and mF2 were the most enriched, followed by mE2, mirroring the patterns observed in *Arabidopsis*. A clustering analysis revealed further preferential association of clusters of TFs with states enriched either before the TSS (I1 and I2) or after the TSS (F2 and E2) [Supplementary Figure S2J]. 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 clusters [Figure 4H]. The largest group (Cluster 3, 32% of all TFs) was associated primarily with I4 and bound upstream of the TSS [Figure 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 [Figure 4H and Supplementary Table 3], likely following the classic mode of RNA Polymerase II recruitment and stabilization of the pre-initiation complex (58). A GO term analysis indicated that TFs of cluster 3 primarily participated in biosynthetic processes, photosynthesis and circadian rhythm, and other housekeeping functions [Supplementary

Table 7]. 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 [Supplementary Table 9], including ARF (59). Most TFs from cluster 4 clearly bound after the TSS and several of these were associated with the control of the cell cycle [Supplementary Table 8], including E2FA, which is a typical marker of cell division. Clusters 4 and 5 also included the recently described GATC binding TFs (60), a large fraction of the AP2/EREBP family involved in stress response (61,62) 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 [Figures 1A and 4F]. Cluster 2 contained TFs primarily controlling development, and more particularly root and vascular development [Supplementary Table 6]. 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 [Figure 4F]. Many of these TFs binding sites were located upstream of the TSS [Figure 4G] and are thus expected to bind to typical promoters. In contrast, cluster 1 TFs primarily associated with the first nucleosome downstream of the TSS [Figure 4G]. 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 (22,63) [Figure 4F and Supplementary Table 5], 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 and 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 (50). Similarly, chromatin states described in *Arabidopsis* are representative of flowering plants (64,65). 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 (66). 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. Overall, together with the similarities in the organization of histone modifications among bryophytes (50) and flowering plants (15,65) 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 (67). Specifically, we show a broad 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 (68,69). 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 (70). Our results and the differences in the phenotypes of the mutants in PRC1 and PRC2 in *Arabidopsis* (71) 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 [Supplemental Figure S5]. 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, generating highly occupied target regions (HOT) 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 (72,73). 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 (20,74). MADS-domain TFs, bZIP TFs, LEAFY COTYLEDON1 (LEC1), TERMINAL FLOWER 1, FLOWERING LOCUS T, and MONOPTEROUS, have been proposed as candidate pioneer factors (75) and most belong to cluster 1 or 2, suggesting that these clusters define a new list of candidate 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 candidate pioneer TFs in plants.

RESOURCE AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Frédéric Berger (frederic.berger@gmi.oeaw.ac.at).

DATA AND CODE AVAILABILITY

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.

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: FB, and FR; Methodology: VS, FR, EA, and TH; Validation: VS, FR, and EA; Formal Analysis: VS, FR, and EA; Investigation: VS, FR, EA, and TH; Data Curation: VS, FR, EA, and TH; Writing – original draft: FB, VS, and FR; Visualization: ZHH, KS, and TW; Supervision: FB; Project Administration: FB, and FR; Funding Acquisition: FB, TH and FR.

FUNDING

This project was funded by the Austrian academy of sciences, and the following grants from the Austrian Research Fund FWF P32054, P36231, PAT1104523 and PAT6138924 to FB, funding from the European Union's Framework Programme for Research and

Innovation Horizon 2020 (2014–2020) under the Marie Curie Skłodowska Grant Agreement no. 847548 (VIP2) to TH. F.R. is a Leverhulme Early Career Fellow (ECF-2023-534) funded by the Leverhulme Trust and the Isaac Newton Trust (23.08(f)).

DECLARATION OF INTERESTS

The authors declare no conflict of interest regarding this manuscript.

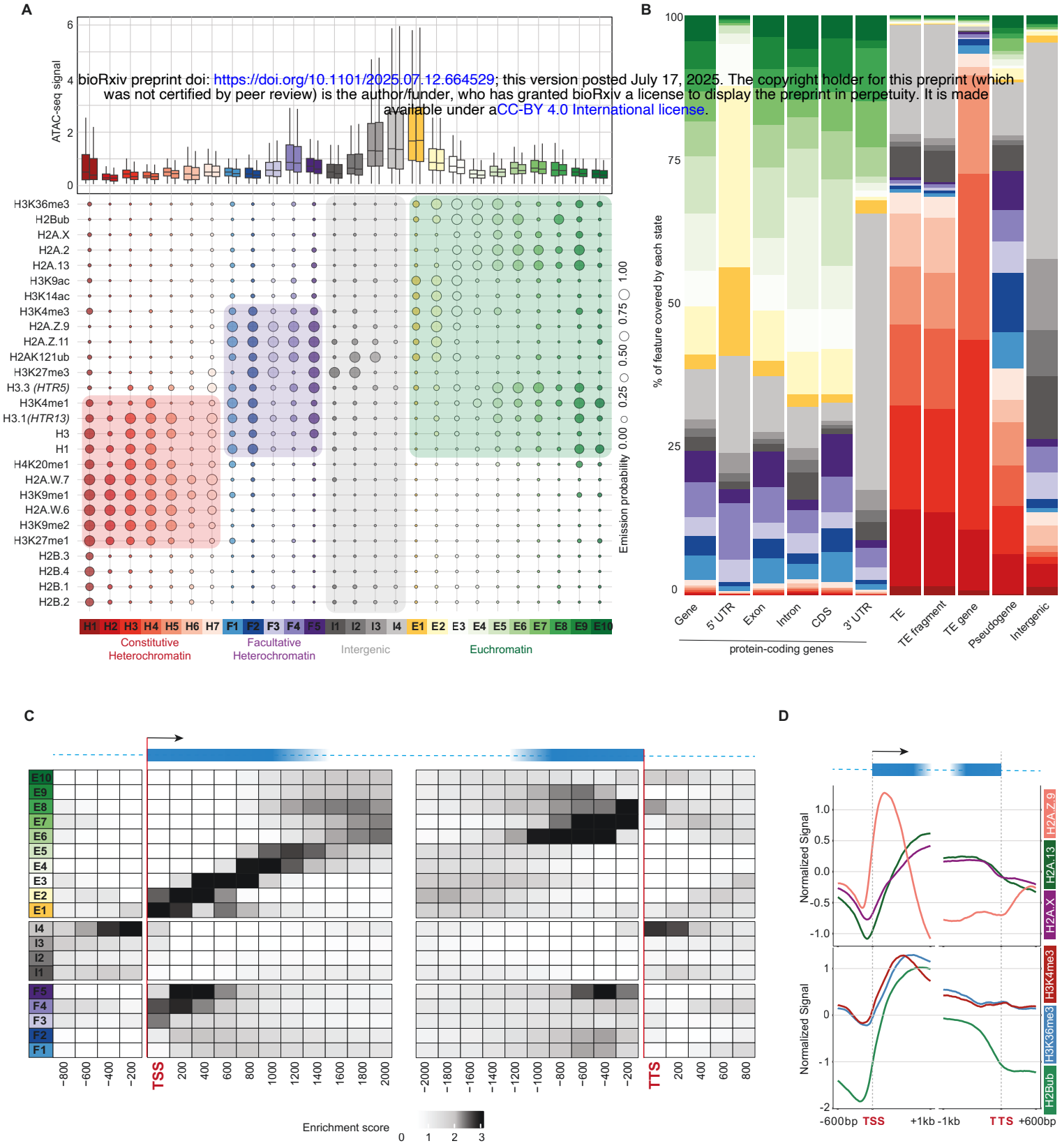


Figure-1

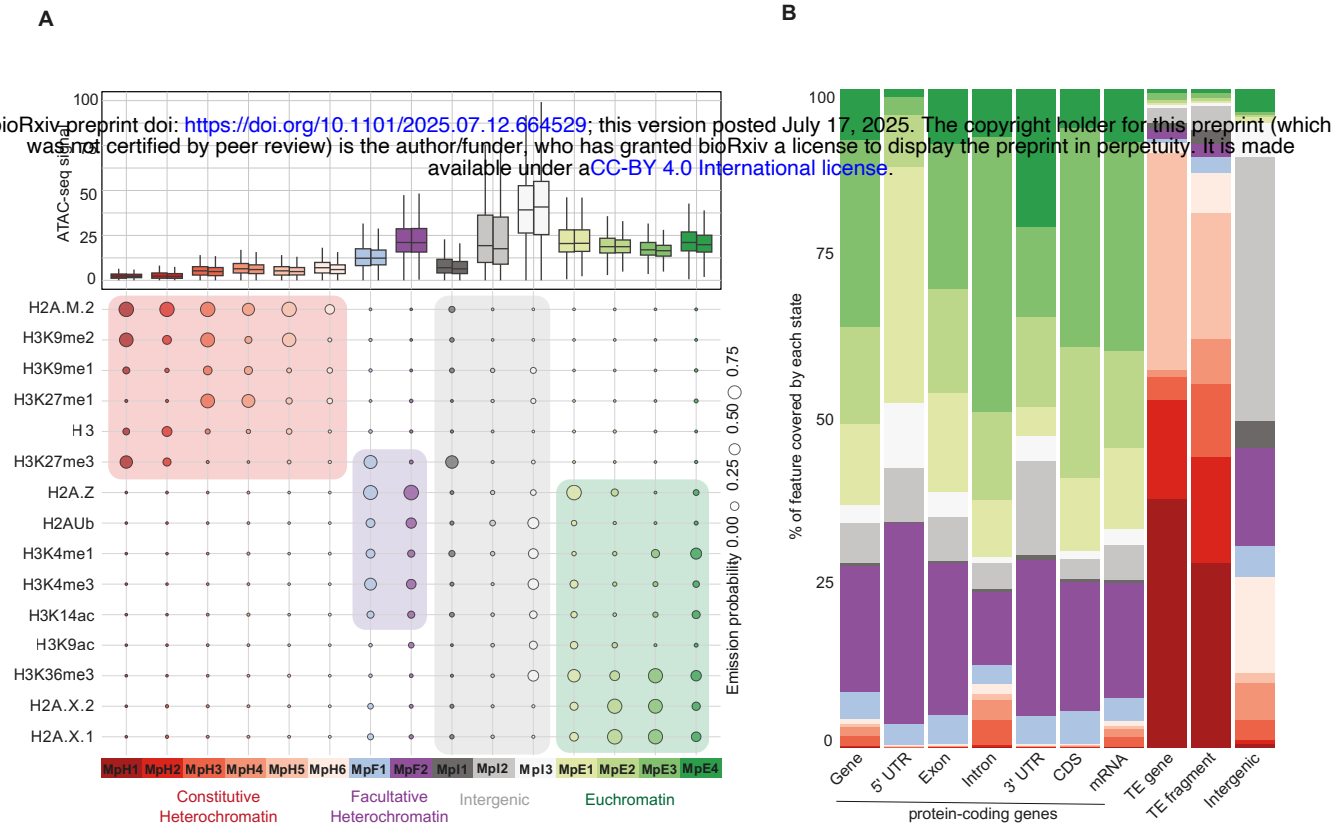
Figure 1. Chromatin states of *Arabidopsis thaliana*

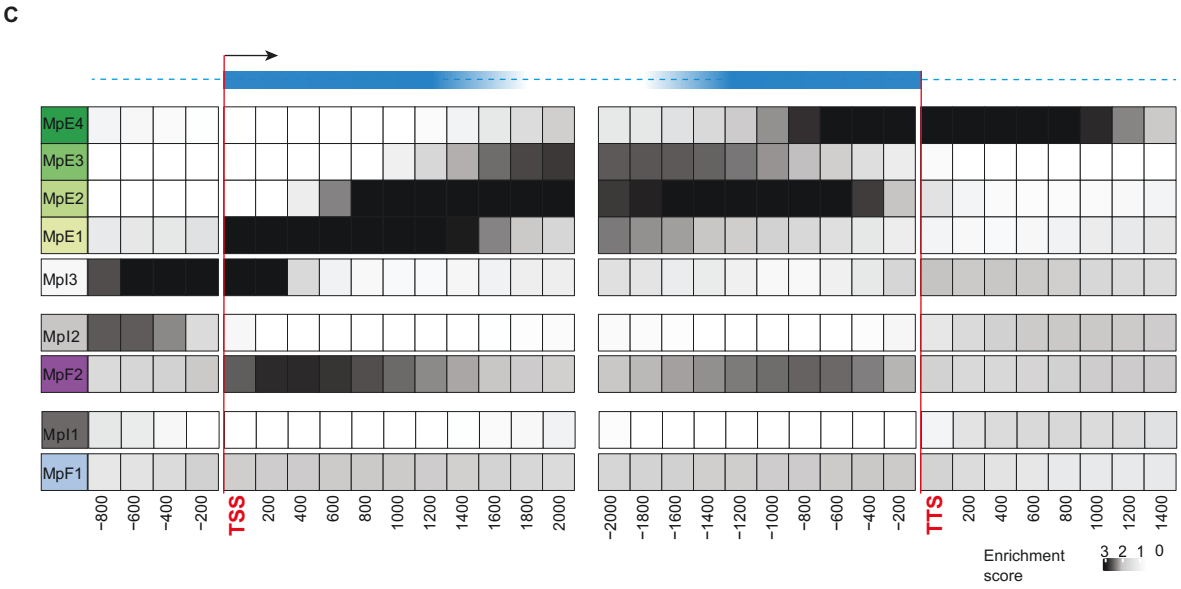
(A) Bubble plot showing the emission probabilities for histone modifications/variants across the 26 chromatin states of *Arabidopsis thaliana*. The size of the bubble represents the emission probability ranging from 0 to 1. Coloured rectangles demarcate the classification of states into major domains 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.

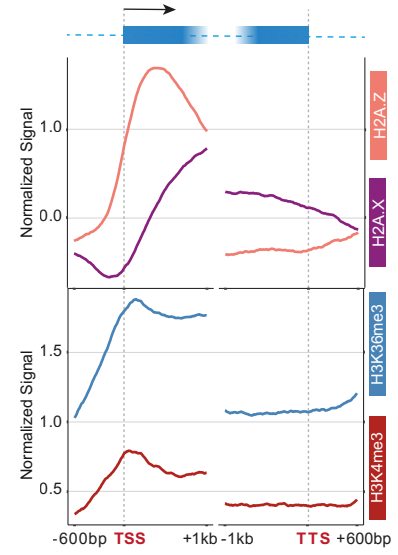
(B) Stacked bar plot showing the overlap between annotated genomic features and chromatin states.

(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)







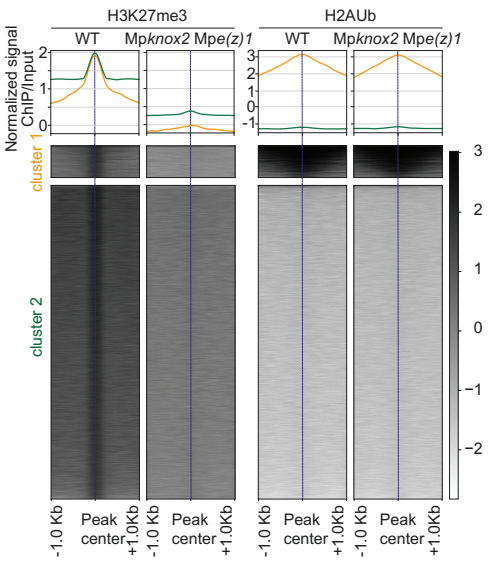


Figure-2

Figure 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 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 MpE1-MpE4. (n = 10715 / 17944 protein coding genes)

(E) Enrichment heatmap and metaplots showing the apparent disassociation between H3K27me3 and H2AKUb marks in wild type and Mp*knox2* Mp*e(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 cluster 1 represents peaks of H3K27me3 shared with H2AKUb while cluster 2 represents majority of H3K27me3 peaks that do not overlap with H2AKUb.

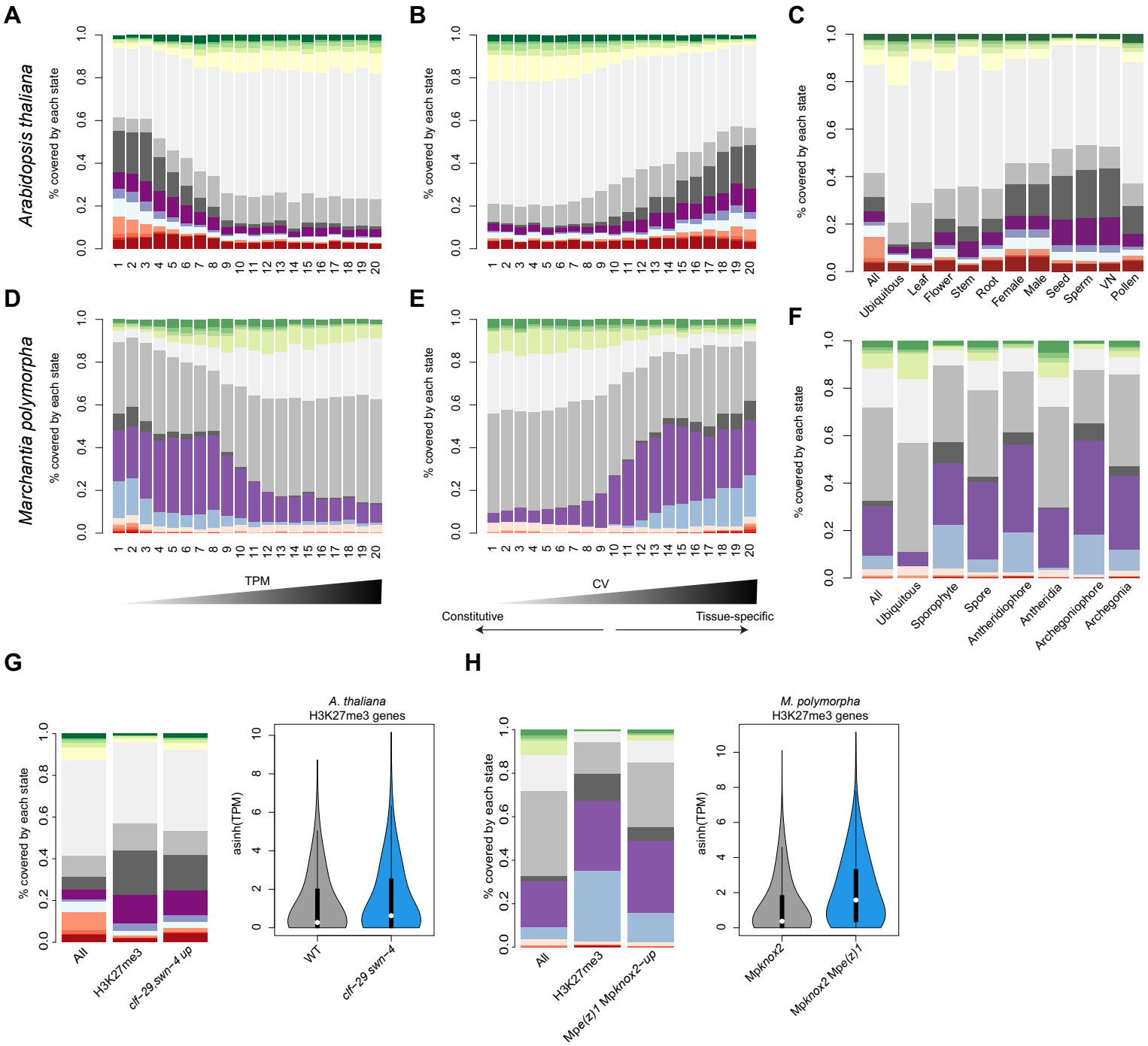


Figure-3

Figure 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* **(D)** 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/swn* 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, asinh(TPM) of expression values for the same group of genes are shown for both mutant and the WT. The colour code for the diagram represents the chromatin states in the 15 states models.

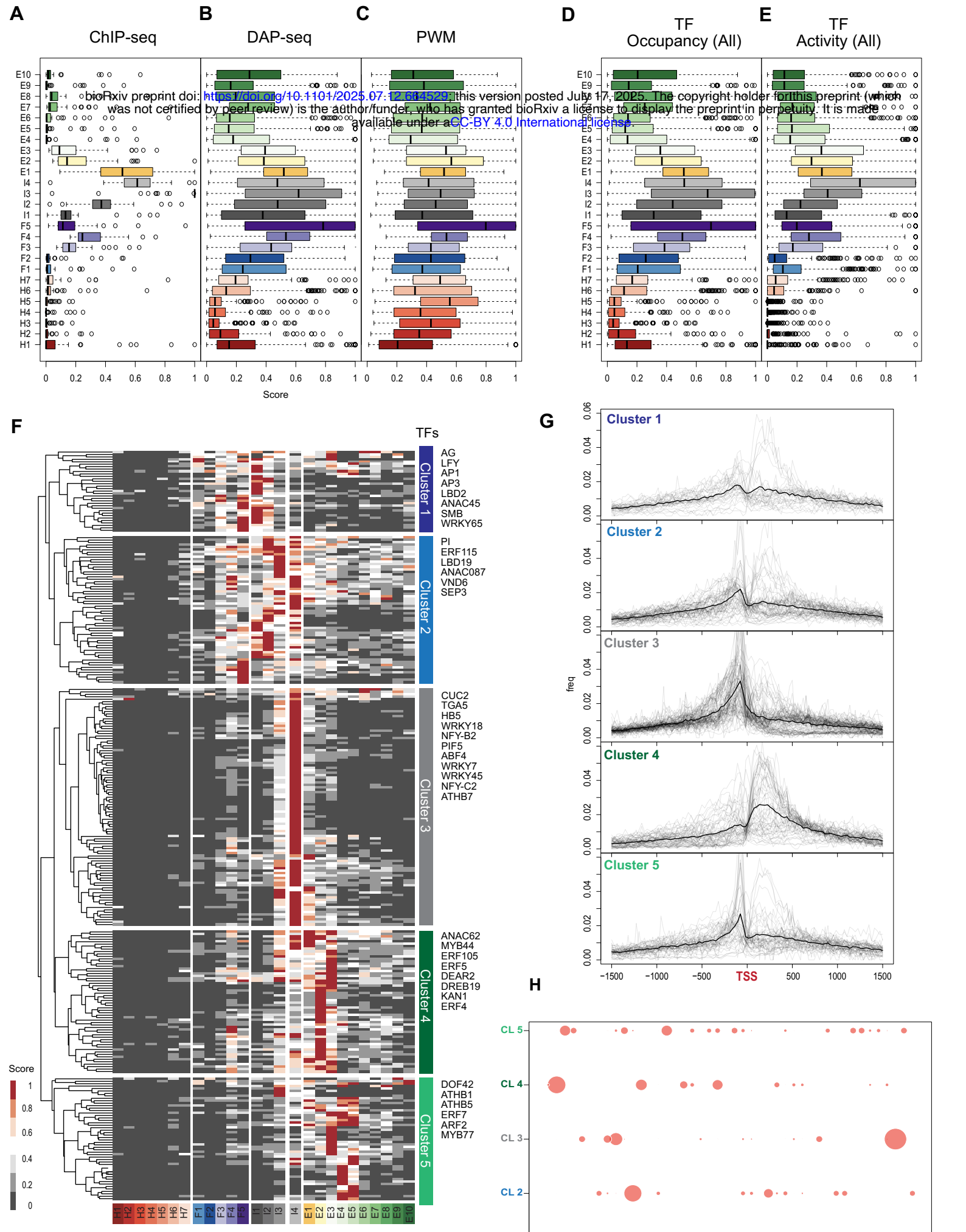


Figure-4

Figure 4. Transcription factors activity shows heterogenous chromatin state preferences in *Arabidopsis*

(A-C) Boxplot of TF occupancy scores based on ChiP-seq **(A)**, DAP-seq **(B)**, or position weight matrixes (PWM) motifs **(C)** in *Arabidopsis*.

(D, E) Comparison of TF occupancy **(D)** and TF activity **(E)** scores across different chromatin states. The later scores are calculated as the statistical enrichment of TF binding over putative targets (co-expressed genes).

(F) Heatmap and hierarchical clustering of TF activity scores for ChIP-seq and DAP-seq experiments 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.

(G) Histogram of TF binding distribution relative to the transcription starting site (TSS) for *Arabidopsis thaliana* TFs from different clusters defined in figure-H.

(H) Bubble plot of Fisher's Test of TF families across clusters defined in figure-H.

REFERENCES

1. Filion GJ, van Bemmel 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 Oct 15;143(2):212–24.

2. Ernst J, Kellis M. Chromatin-state discovery and genome annotation with ChromHMM. Nat Protoc. 2017 Dec;12(12):2478–92.

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 Mar 24;471(7339):480–5.

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 Sep;39(9):381–99.

5. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature. 2011 May;473(7345):43–9.

6. Ernst J, Kellis M. Interplay between chromatin state, regulator binding, and regulatory motifs in six human cell types. Genome Res. 2013 Jul;23(7):1142–54.

7. Magnani L, Ballantyne EB, Zhang X, Lupien M. PBX1 Genomic Pioneer Function Drives ERα Signaling Underlying Progression in Breast Cancer. PLOS Genet. 2011 Nov 17;7(11):e1002368.

8. Zaret KS, Mango SE. Pioneer transcription factors, chromatin dynamics, and cell fate control. Curr Opin Genet Dev. 2016 Apr;37:76–81.

9. Iwafuchi-Doi M, Zaret KS. Cell fate control by pioneer transcription factors. Dev Camb Engl. 2016 Jun 1;143(11):1833–7.

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 Dec 14;21(24):9501.

11. Bieluszewski T, Xiao J, Yang Y, Wagner D. PRC2 activity, recruitment, and silencing: a comparative perspective. Trends Plant Sci. 2021 Nov 1;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 Jun;61:102009.

13. Dent SYR. KAT tales: Functions of Gcn5 and PCAF lysine acetyltransferases in SAGA and ATAC. J Biol Chem. 2024 Oct;300(10):107744.

14. Soffers JHM, Workman JL. The SAGA chromatin-modifying complex: the sum of its parts is greater than the whole. Genes Dev. 2020 Oct 1;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. Casati P, Kleine-Vehn J, editors. eLife. 2023 Jul 19;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 Jun;26(6):2351–66.

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 Jun 16;21(1):143.

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 Cell Mol Biol. 2015 Oct;84(1):111–24.

19. Heyndrickx KS, de Velde JV, Wang C, Weigel D, Vandepoele K. A Functional and Evolutionary Perspective on Transcription Factor Binding in Arabidopsis thaliana. Plant Cell. 2014 Oct 1;26(10):3894–910.

20. Lai X, Blanc-Mathieu R, GrandVuillemin L, Huang Y, Stigliani A, Lucas J, et al. The LEAFY floral regulator displays pioneer transcription factor properties. Mol Plant. 2021 May 3;14(5):829–37.

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 Mar 3;15(3):R41.

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 Feb 28;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 Jan;1860(1):75–83.

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 Jun 17;72:105–31.

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 Mar 1;208(3):937–49.

26. 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 Clifton NJ. 2017;1610:93–106.

27. Quinlan AR. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Curr Protoc Bioinforma. 2014;47(1):11.12.1-11.12.34.

28. Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data. 2010; Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

29. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012 Apr;9(4):357–9.

30. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009 Aug 15;25(16):2078–9.

31. 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 Res. 2016 Jul 8;44(W1):W160–5.

32. 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

33. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. Welcome to the Tidyverse. J Open Source Softw. 2019 Nov 21;4(43):1686.

34. Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize Implements and enhances circular visualization in R. Bioinforma Oxf Engl. 2014 Oct;30(19):2811–2.

35. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics. 2016 Sep 15;32(18):2847–9.

36. 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 May 11;12(1):2683.

37. 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]. p. 2022.05.31.493688. Available from: https://www.biorxiv.org/content/10.1101/2022.05.31.493688v1

38. 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 Feb 24;30(4):573-588.e7.

39. 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 Nov 1;63(11):1745–55.

40. 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.

41. 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 Jan;3(1):e00100.

42. O'Malley RC, Huang S shan C, Song L, Lewsey MG, Bartlett A, Nery JR, et al. Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. Cell. 2016 May 19;165(5):1280–92.

43. Song L, Huang S shan C, Wise A, Castanon R, Nery JR, Chen H, et al. A transcription factor hierarchy defines an environmental stress response network. Science. 2016 Nov 4;354(6312):aag1550.

44. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. Bioinformatics. 2011 Apr 1;27(7):1017–8.

45. Yu G, Wang LG, He QY. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics. 2015 Jul 1;31(14):2382–3.

46. 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 Jan 1;59(1):e3.

47. Jin J, Tian F, Yang DC, Meng YQ, 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 Jan 4;45(D1):D1040–5.

48. Thomas PD, Ebert D, Muruganujan A, Mushayahama T, Albou LP, Mi H. PANTHER: Making genome-scale phylogenetics accessible to all. Protein Sci. 2022;31(1):8–22.

49. 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.

50. Hisanaga T, Wu S, Schafran P, Axelsson E, Akimcheva S, Dolan L, et al. The ancestral chromatin landscape of land plants. New Phytol. 2023 Dec;240(5):2085–101.

51. 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]. p. 2025.03.30.646155. Available from: https://www.biorxiv.org/content/10.1101/2025.03.30.646155v1

52. 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 Apr 12;18(1):69.

53. Chanvivattana Y, Bishopp A, Schubert D, Stock C, Moon YH, Sung ZR, et al. Interaction of Polycomb-group proteins controlling flowering in Arabidopsis. Development. 2004 Nov 1;131(21):5263–76.

54. 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 Feb 28;109(9):3576–81.

55. Brooks MD, Juang CL, Katari MS, Alvarez JM, Pasquino A, Shih HJ, et al. ConnecTF: A platform to integrate transcription factor-gene interactions and validate regulatory networks. Plant Physiol. 2021 Feb 25;185(1):49–66.

56. Romani F, Moreno JE. Molecular mechanisms involved in functional macroevolution of plant transcription factors. New Phytol. 2021;230(4):1345–53.

57. 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. 2014 Feb 11;111(6):2367–72.

58. Wong F, Gunawardena J. Gene Regulation in and out of Equilibrium. Annu Rev Biophys. 2020 May 6;49:199–226.

59. Guilfoyle TJ, Hagen G. Auxin response factors. Curr Opin Plant Biol. 2007 Oct;10(5):453–60.

60. Voichek Y, Hristova G, Mollá-Morales A, Weigel D, Nordborg M. Widespread position-dependent transcriptional regulatory sequences in plants. Nat Genet. 2024 Oct;56(10):2238–46.

61. Dietz KJ, 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 Sep;245(1–4):3–14.

62. Kizis D, Lumbreras V, Pagès M. Role of AP2/EREBP transcription factors in gene regulation during abiotic stress. FEBS Lett. 2001 Jun 8;498(2–3):187–9.

63. Sridhar VV, Surendrarao A, Liu Z. APETALA1 and SEPALLATA3 interact with SEUSS to mediate transcription repression during flower development. Development. 2006 Aug 15;133(16):3159–66.

64. Hu Y, Lai Y, Chen X, Zhou DX, Zhao Y. Distribution pattern of histone marks potentially determines their roles in transcription and RNA processing in rice. J Plant Physiol. 2020 Jun;249:153167.

65. Leng X, Thomas Q, Rasmussen SH, Marquardt S. A G(enomic)P(ositioning)S(ystem) for Plant RNAPII Transcription. Trends Plant Sci. 2020 Aug 1;25(8):744–64.

66. 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 Oct 23;33(20):4367-4380.e9.

67. 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]. p. 2025.03.17.643675. Available from: https://www.biorxiv.org/content/10.1101/2025.03.17.643675v1

68. Tamburri S, Conway E, Pasini D. Polycomb-dependent histone H2A ubiquitination links developmental disorders with cancer. Trends Genet TIG. 2022 Apr;38(4):333–52.

69. Loubiere V, Martinez AM, Cavalli G. Cell Fate and Developmental Regulation Dynamics by Polycomb Proteins and 3D Genome Architecture. BioEssays News Rev Mol Cell Dev Biol. 2019 Mar;41(3):e1800222.

70. 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 Nov 10;6(1):1144.

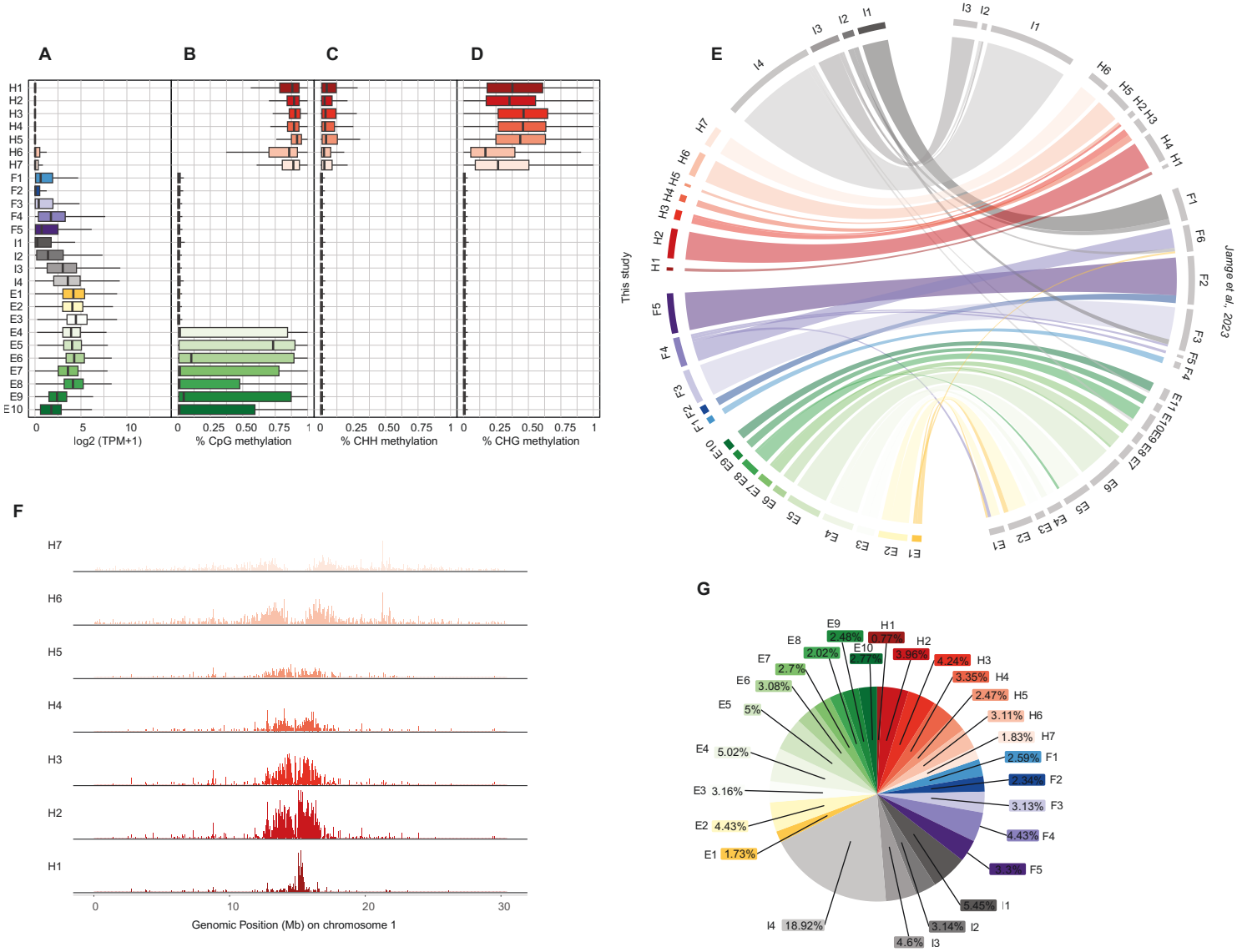
71. Wang Q, Shen WH. Chromatin modulation and gene regulation in plants: insight about PRC1 function. Biochem Soc Trans. 2018 Aug 20;46(4):957–66.

72. Luzete-Monteiro E, Zaret KS. Structures and consequences of pioneer factor binding to nucleosomes. Curr Opin Struct Biol. 2022 Aug;75:102425.

73. 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 Oct;88:102875.

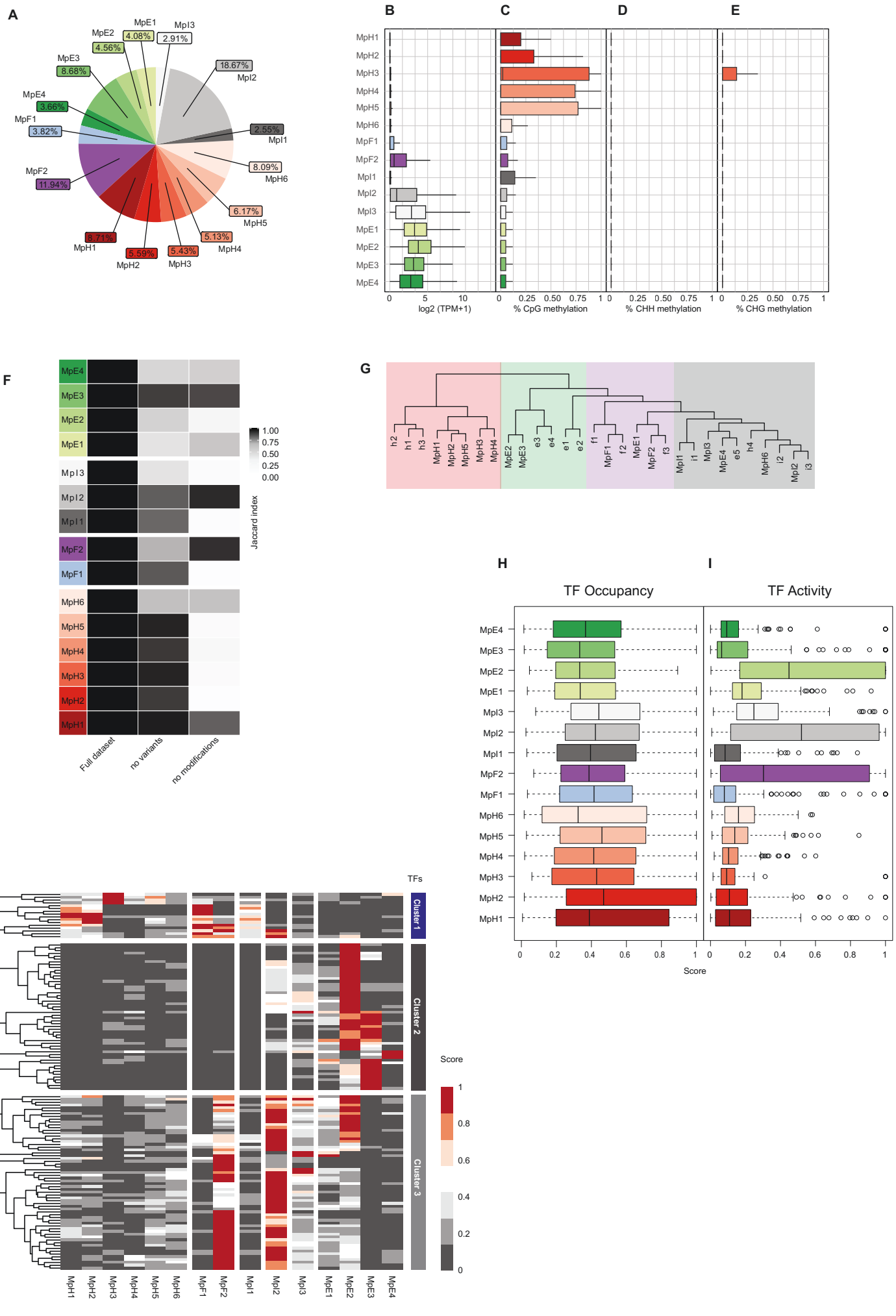
74. Jin R, Klasfeld S, Zhu Y, Fernandez Garcia M, Xiao J, Han SK, et al. LEAFY is a pioneer transcription factor and licenses cell reprogramming to floral fate. Nat Commun. 2021 Jan 27;12(1):626.

75. Strader L, Weijers D, Wagner D. Plant transcription factors - being in the right place with the right company. Curr Opin Plant Biol. 2022 Feb;65:102136.



Supplementary Figure S1

(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 colour code for the flow diagram represents the chromatin states in the current model. **(F)** Genomic distribution of heterochromatic states (H1–H7) [bottom to top] on Chr1 of Arabidopsis thaliana. **(G)** Pie chart showing the percentage of the genome covered by each state.



Supplementary Figure-S2

Supplementary Figure S2

(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).

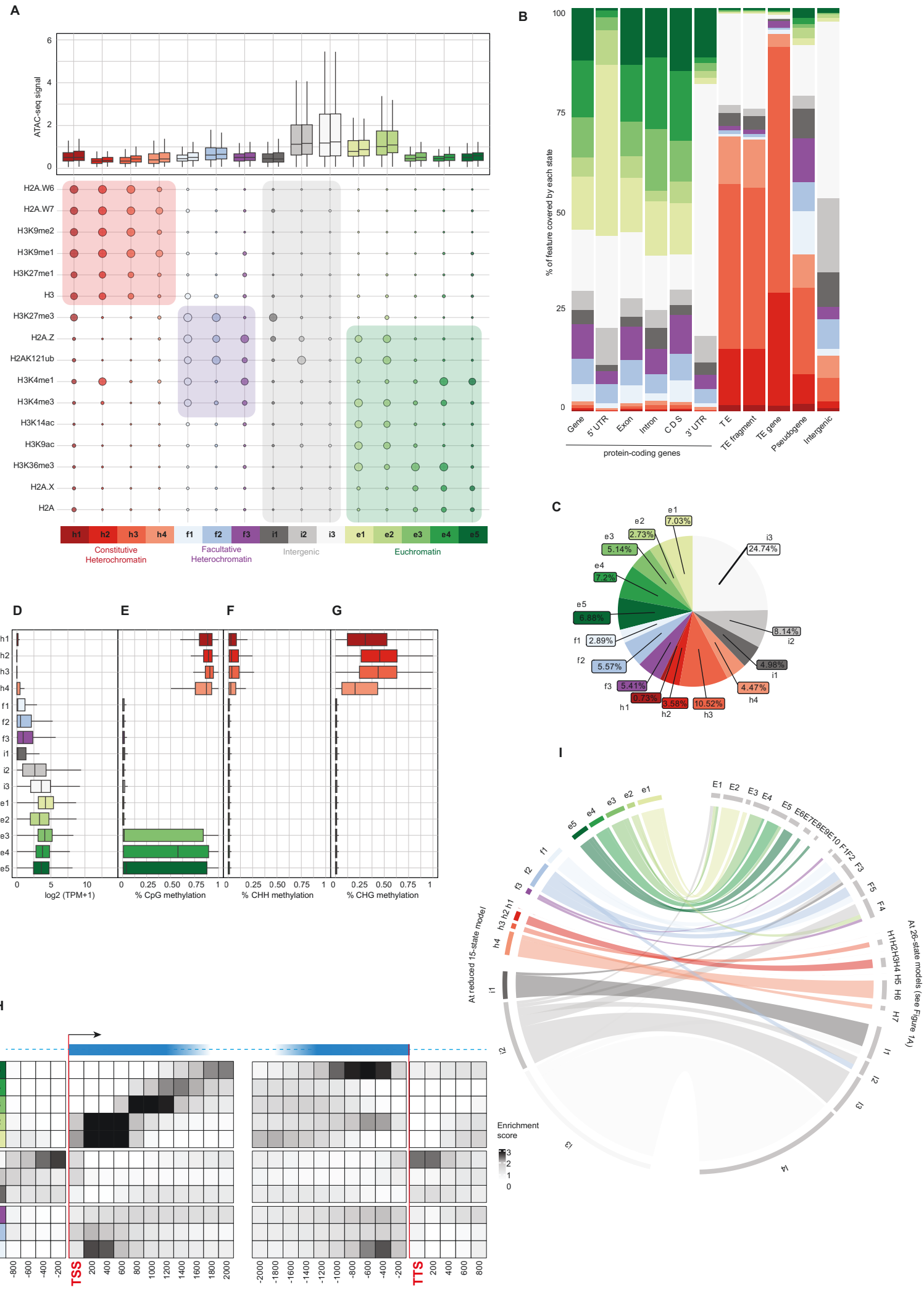
(C-E) Box plot showing levels of CpG **(C)**, CHH **(D)**, and CHG **(E)** methylation for all chromatin states of *Marchantia polymorpha*.

(F) 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.

(G) 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.

(H, I) Boxplot of TF occupancy **(H)** and TF activity **(I)** scores across different chromatin states. The later scores are calculated as the statistical enrichment of TF binding over putative targets (co-expressed genes).

(J) 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.



Supplementary Figure-S3

Supplementary Figure S3

(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*.

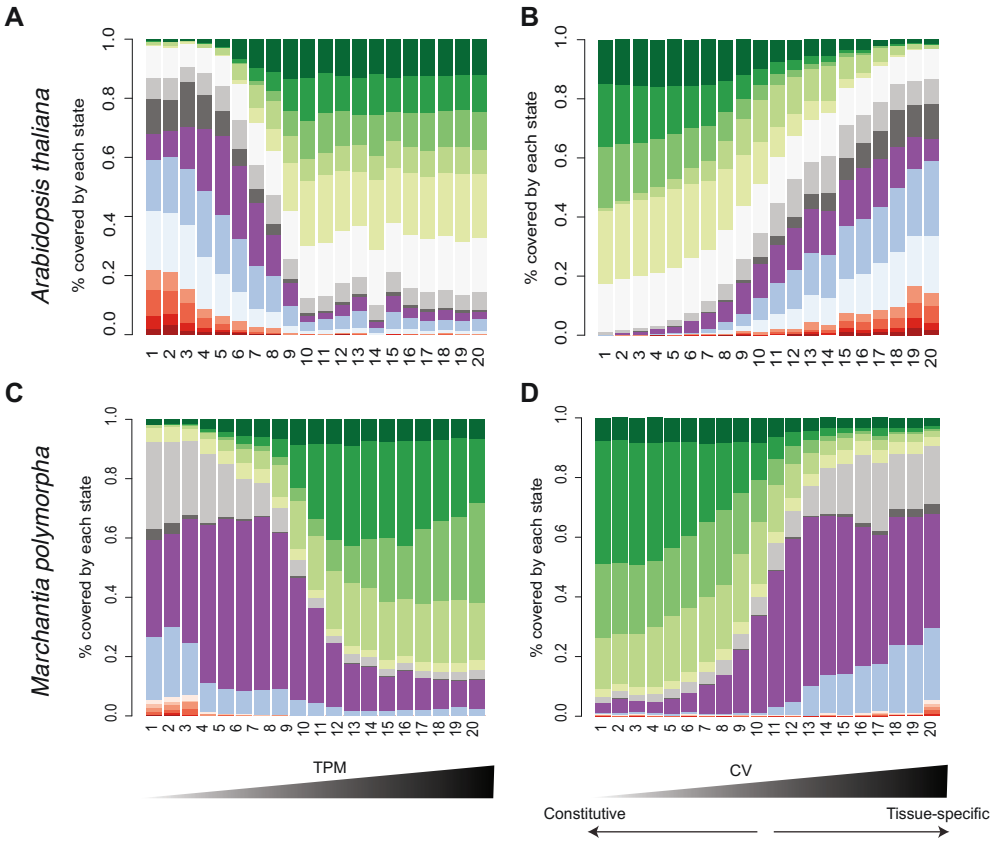
(C) Pie chart showing the percentage of the genome covered by each state of the 15-state model of *Arabidopsis thaliana*.

(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*.

(A) 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).

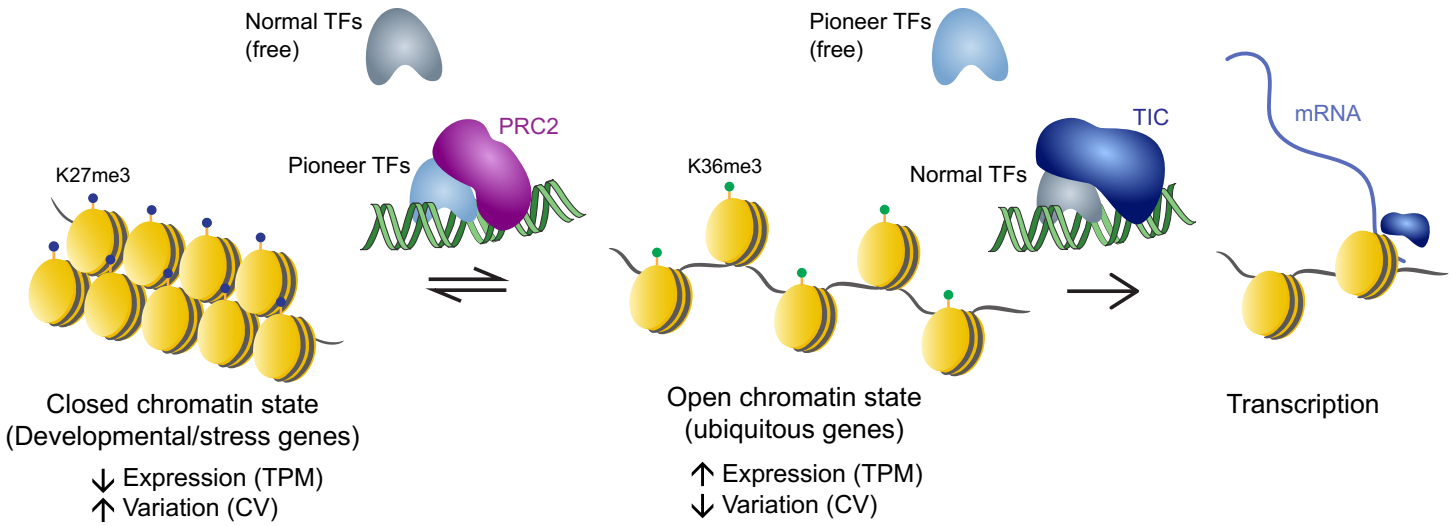
(B) Flow diagram showing the overlap (in bp) between the chromatin states of extensive 26-state model and the 15-state model *Arabidopsis thaliana*. The colour code for the flow diagram represents the colours for the states of the 15-state model.



Supplementary Figure-S4

(A-D) Bar plot of the proportion of chromatin states for 20 bins across coding regions of genes ordered by transcripts per million (TPM) Arabidopsis leaves **(A, B)** and Marchantia thallus **(C, D)**.

A



Supplemental Figure-S5
(A) Model of transcription factor interplay with facultative and open chromatin to control transcription rates.

Chapter 7

7. Discussion

Contents

7.1 H2A VARIANTS ARE THE MAJOR DETERMINANTS OF CHROMATIN STATE IDENTITY 143

7.2 DYNAMIC EXCHANGE OF HISTONE VARIANTS CONTROLS CHROMATIN ORGANIZATION 145

7.3 CONSERVATION OF CHROMATIN STATES ACROSS LAND PLANT EVOLUTION 145

7.4 CHROMATIN STATES DEFINE DISTINCT REGULATORY ENVIRONMENTS FOR GENE EXPRESSION 147

7.5 TRANSCRIPTION FACTOR-CHROMATIN STATE INTERACTIONS REVEAL REGULATORY STRATEGIES 148

7.6 CANDIDATE PIONEER TRANSCRIPTION FACTORS IN PLANTS 150

7.7 EVOLUTIONARY DIVERGENCE OF POLYCOMB REPRESSIVE PATHWAYS IN PLANTS 151

7.8 IMPLICATIONS FOR UNDERSTANDING PLANT GENE REGULATION 152

BIBLIOGRAPHY 154

7.1 H2A Variants are the Major Determinants of Chromatin State Identity

The organization of eukaryotic chromatin into functionally distinct domains has long been attributed to combinatorial patterns of histone post-translational modifications (PTMs). However, our comprehensive analysis of chromatin organization in *Arabidopsis thaliana* and *Marchantia polymorpha* reveals a striking and previously underappreciated role for H2A variants as primary determinants of chromatin state identity.

Through systematic chromatin state modeling using ChromHMM, we identified 26 distinct chromatin states in *Arabidopsis* that partition the genome into constitutive heterochromatin, facultative heterochromatin, euchromatin, and intergenic regions. Remarkably, each major chromatin domain is defined by the enrichment of specific H2A variants: H2A.W marks constitutive heterochromatin, H2A.Z defines facultative heterochromatin, while H2A and H2A.X are enriched in euchromatin (Jamge et al., 2023). This clear segregation contrasts sharply with

H3 variants, which show much weaker domain specificity.

Biochemical analysis of mononucleosomes provided mechanistic insights into these observations. While H3.1 and H3.3 form both homotypic and heterotypic nucleosomes, H2A variants predominantly assemble into homotypic nucleosomes (Jamge et al., 2023; Osakabe et al., 2018). Critically, immunoprecipitation experiments demonstrated that H2A variants show strong preferential associations with distinct sets of histone modifications: H2A.W nucleosomes are enriched for constitutive heterochromatin marks (H3K9me1/2, H3K27me1), H2A.Z nucleosomes exclusively carry H3K27me3, while H2A and H2A.X nucleosomes are enriched for euchromatic marks including H3K36me3 and H2Bub (Jamge et al., 2023).

To quantitatively assess the relative importance of histone variants, we performed systematic *in silico* removal experiments. Recalculating chromatin states after excluding different histone variants or H3 modifications revealed that H2A variants have the strongest impact on chromatin state definition, with effects comparable to or exceeding those observed upon removal of all H3 modifications. In contrast, H2B variants had minimal impact, and H3 variants showed intermediate effects (Jamge et al., 2023).

The functional importance of H2A variants extends beyond *Arabidopsis*. In *Marchantia polymorpha*, the orthologous H2A variant H2A.M.2 similarly defines constitutive heterochromatin, while H2A.Z marks facultative heterochromatin and H2A.X marks euchromatin. Computational analysis confirmed that H2A variants contribute as strongly as histone PTMs in defining chromatin states (Shukla et al., 2025).

These findings support a model in which H2A variants function as molecular landmarks that organize chromatin into functionally distinct domains by influencing the activity of chromatin-modifying enzymes. This is consistent with previous studies showing that H2A.Z inhibits H2B ubiquitination (Surface et al., 2016), that PRC2 preferentially methylates H3K27 on H2A.Z-containing nucleosomes (Wang et al., 2018), and that H2A.W synergizes with H3K9me2 in maintaining heterochromatic silencing (Bourguet et al., 2021; Osakabe et al., 2021). In contrast, H3 variants show less pronounced domain specificity, with most H3 modifications present on both H3.1 and H3.3, suggesting that to the notable exception of H3K27me1 deposition exclusive to H3.1 by ATXR5/6

(Bergamin et al., 2017), H3 variants play a secondary role compared to H2A variants in organizing chromatin states.

7.2 Dynamic Exchange of Histone Variants Controls Chromatin Organization

While H2A variants serve as major determinants of chromatin state identity, their genomic distribution must be actively maintained. Our analysis of the chromatin remodeler DDM1 (*Decreased DNA Methylation 1*) revealed that loss of DDM1 function causes genome-wide replacement of H2A.W with H2A.Z in constitutive heterochromatin, accompanied by corresponding shifts from H3K9me2 to H3K27me3 marks. This H2A variant redistribution results in conversion of constitutive heterochromatin to facultative heterochromatin states, leading to loss of DNA methylation and transcriptional reactivation of transposable elements (Jamge et al., 2023). These findings demonstrate that chromatin states are dynamic structures whose composition is actively maintained by chromatin remodeling activities, and that proper H2A variant distribution is essential for genome stability.

7.3 Conservation of Chromatin States Across Land Plant Evolution

A fundamental question in chromatin biology is whether the principles of chromatin organization are conserved across evolutionary time or whether they represent lineage-specific adaptations. To address this question, we performed comprehensive chromatin state analysis in *Marchantia polymorpha*, a liverwort that diverged from flowering plants approximately 450 million years ago, representing one of the earliest land plant lineages.

Using ChromHMM, we identified 15 distinct chromatin states in Marchantia based on the distribution of histone variants and modifications. Remarkably, despite the vast evolutionary distance, the fundamental organization of chromatin into constitutive heterochromatin, facultative heterochromatin, and euchromatin is conserved between *Marchantia* and *Arabidopsis*. Each major chromatin domain in *Marchantia* is defined by orthologous molecular signatures: H2A.M.2 (the *Marchantia* ortholog of H2A.W, sharing the characteristic C-terminal KSPK motif) marks constitutive heterochromatin, H2A.Z defines facultative heterochromatin, and H2A.X marks euchromatin (Shukla et al., 2025).

The conservation extends to histone modification patterns. Constitutive heterochromatin in both species is marked by H3K9me2 and H3K27me1, facultative heterochromatin by H3K27me3, and euchromatin by H3K36me3 and H2Bub. Furthermore, the functional properties of these chromatin states are conserved: constitutive heterochromatin is associated with transposable elements and DNA methylation, facultative heterochromatin with developmentally regulated genes, and euchromatin with actively transcribed genes (Shukla et al., 2025).

However, our analysis also revealed important species-specific differences. *Marchantia* lacks several chromatin states present in *Arabidopsis*, likely reflecting differences in genome complexity and regulatory requirements. Most notably, *Marchantia* shows a simpler organization of facultative heterochromatin, with fewer distinct H3K27me3-marked states compared to *Arabidopsis*. Additionally, while both species show strong enrichment of H2A.Z at facultative heterochromatin, the precise distribution patterns differ, with *Marchantia* showing broader H2A.Z domains around genes (Shukla et al., 2025).

Comparative analysis of transcription factor binding revealed both conserved and divergent features. Core transcription factors involved in basic cellular processes show similar chromatin state associations in both species, while transcription factors involved in lineage-specific developmental programs show distinct binding patterns.

Perhaps most strikingly, our analysis revealed differences in Polycomb repressive mechanisms between the two species. While both utilize H3K27me3 for facultative heterochromatin marking, Marchantia shows independence between H3K27me3 and H2AK121ub, with loss of PRC1 activity not affecting H3K27me3 levels. This contrasts with the interdependence observed in *Arabidopsis* (Derkacheva et al., 2013) and suggests evolutionary divergence in Polycomb pathway function (Shukla et al., 2025).

These findings demonstrate that the fundamental principles of chromatin organization: the use of specific H2A variants and histone modifications to define distinct functional domains, are deeply conserved across land plant evolution. This conservation suggests that the chromatin state framework represents an ancient and essential mechanism for organizing eukaryotic genomes. At the same time, the species-specific differences highlight how chromatin organization has been

adapted to meet the distinct regulatory requirements of different plant lineages, providing a flexible framework for evolutionary innovation in gene regulation.

7.4 Chromatin States Define Distinct Regulatory Environments for Gene Expression

A central goal of research in chromatin biology is to understand how chromatin organization influences gene regulation. Our comprehensive analysis reveals that chromatin states create distinct regulatory environments that correlate with specific patterns of transcriptional activity, DNA methylation, and chromatin accessibility.

In *Arabidopsis*, genes associated with different chromatin states show markedly different expression patterns. Euchromatic genes (marked by H2A/H2A.X, H3K36me3, and H2Bub) are highly expressed with broad tissue expression, consistent with housekeeping functions. Genes in facultative heterochromatin (H2A.Z and H3K27me3) show lower average expression but higher tissue specificity, indicating developmental regulation. Genes in constitutive heterochromatin (H2A.W and H3K9me2) are largely silenced (Jamge et al., 2023).

Chromatin accessibility measurements revealed that chromatin states predict accessibility patterns. Euchromatic states provide high accessibility at transcription start sites and gene bodies, facilitating transcription. Facultative heterochromatin associates with intermediate accessibility, primarily at regulatory elements. Constitutive heterochromatin prevents accessibility (Jamge et al., 2023).

DNA methylation patterns also correlate strongly with chromatin states. Constitutive heterochromatin shows high CG, CHG, and CHH methylation. Facultative heterochromatin shows selective depletion of DNA methylation despite being repressed, indicating that H3K27me3/H2A.Z mediated repression and DNA methylation represent alternative silencing mechanisms. The reduction of CG methylation is most likely the result of the presence of H2A.Z in facultative heterochromatin since H2A.Z covered genes in euchromatin are also depleted of gene body CG methylation in contrast to other euchromatic genes with CG gene body methylation and depletion of CHG and CHH methylation (Jamge et al., 2023).

Marchantia shows similar correlations between chromatin states and gene expression, though with a simpler methylation landscape more strictly confined to constitutive heterochromatin

(Shukla et al., 2025). Analysis of chromatin state dynamics revealed that genes can transition between states during development, particularly between euchromatin and facultative heterochromatin, with transitions correlating with expression changes. Genes undergoing transitions are enriched for developmental regulators, while housekeeping genes maintain stable euchromatic (Jamge et al., 2023).

These findings establish that chromatin states are functional units integrating multiple chromatin features to create distinct regulatory environments, providing a mechanistic framework for understanding how chromatin organization influences gene regulation across plant development and evolution.

7.5 Transcription Factor-Chromatin State Interactions Reveal Regulatory Strategies

Transcription factors (TFs) are the primary mediators of gene-specific regulation, but how they interact with chromatin organization to control gene expression had remained poorly understood in plants. By integrating chromatin state maps with genome-wide binding data for over 400 transcription factors in *Arabidopsis*, we uncovered systematic relationships between TF binding patterns and chromatin states that reveal fundamental regulatory strategies in plants.

Our analysis revealed that transcription factors show non-random associations with specific chromatin states. We identified three major groups of transcription factors based on their chromatin state preferences and binding locations relative to transcription start sites (TSS). The first group, comprising most TFs, binds primarily to euchromatic states near the TSS of actively transcribed genes. These TFs are enriched for basic helix-loop-helix (bHLH), AP2/ERF, NAC, and WRKY family members and typically function as activators of gene expression. Their association with accessible euchromatin is consistent with their role in fine-tuning expression of already active or poised genes (Shukla et al., 2025).

The second group shows preferential binding to intergenic chromatin states, often at large distances from annotated genes. These TFs tend to bind regions with intermediate chromatin accessibility and are enriched for families involved in environmental responses and stress signaling. Their distal binding suggests they may regulate genes through long-range chromatin interactions or control the expression of unannotated regulatory RNAs (Shukla et al., 2025).

Most intriguingly, the third group shows significant association with facultative

heterochromatin states marked by H2A.Z and H3K27me3. This group is highly enriched for MADS-box transcription factors and includes LEAFY (LFY), a master regulator of floral development. These TFs bind near the TSS of genes marked by repressive chromatin, suggesting they may function to activate or poise genes within repressive chromatin environments. The ability to bind H3K27me3-marked chromatin is a hallmark of pioneer transcription factors in animals, suggesting that plant MADS-box TFs and LEAFY may possess analogous pioneer activity (Shukla et al., 2025).

To validate these associations in an evolutionary context, we performed similar analysis in *Marchantia* polymorpha. Despite the evolutionary distance, we observed conservation of the general principle that different TF families associate with distinct chromatin states. However, the specific TF families showing facultative heterochromatin association differ between species, reflecting lineage-specific evolution of developmental regulatory networks. Notably, Marchantia MADS-box TFs also show association with H3K27me3-marked chromatin, suggesting that this regulatory strategy may be ancestral in land plants (Shukla et al., 2025).

Functional analysis of genes bound by facultative heterochromatin-associated TFs revealed enrichment for developmental regulators, particularly genes involved in meristem identity, floral organ specification, and phase transitions. Many of these genes are themselves transcription factors, suggesting a hierarchical regulatory network in which pioneer-like TFs activate master regulators within repressive chromatin domains. This organization may provide a mechanism for maintaining developmental genes in a repressed but poised state until appropriate developmental or environmental signals trigger their activation.

These findings reveal that chromatin states not only reflect the transcriptional status of genes but also define distinct regulatory environments that are targeted by specific classes of transcription factors. The association of TF families with specific chromatin states suggests a division of labor in gene regulation, with different TFs specialized for activating genes in different chromatin contexts. This framework provides new insights into how transcription factors and chromatin organization cooperate to control plant development and responses to environmental signals.

7.6 Candidate Pioneer Transcription Factors in Plants

Pioneer transcription factors are defined by their ability to bind nucleosomal DNA within closed chromatin and initiate chromatin remodeling to facilitate gene activation. While extensively characterized in animals, their existence in plants has remained unclear. Our systematic analysis of transcription factor binding patterns across chromatin states has identified potential candidate pioneer factors in plants, particularly among MADS-box transcription factors and LEAFY.

The defining characteristic of pioneer factors is their ability to bind target sites within repressive chromatin. Our analysis revealed that MADS-box TFs show significant enrichment at facultative heterochromatin states near transcription start sites, in stark contrast to most TF families that preferentially bind accessible euchromatic regions. MADS-box TFs bind near genes marked by H2A.Z and H3K27me3, often at sites with reduced chromatin accessibility compared to typical TF binding sites (Shukla et al., 2025).

LEAFY (LFY), a master regulator of floral identity, shows similar properties. LFY binding sites are significantly enriched at facultative heterochromatin states, and many LFY target genes carry H3K27me3 marks prior to floral induction. Previous studies showed that LFY can bind nucleosomal DNA and that its binding correlates with local chromatin opening and H3K27me3 loss (Sayou et al., 2016). Our genome-wide analysis demonstrates that facultative heterochromatin binding is a general property of LFY (Shukla et al., 2025).

Genes bound by MADS-box TFs in facultative heterochromatin are highly enriched for developmental regulators, including transcription factors involved in meristem identity, floral organ specification, and developmental phase transitions. Many show dynamic expression during development, transitioning from repressed to active states in specific contexts. This is consistent with a model in which MADS-box TFs bind repressed developmental genes and facilitate their activation in response to developmental or environmental cues.

Comparative analysis in *Marchantia polymorpha* revealed that MADS-box TFs also associate with H3K27me3-marked chromatin, suggesting this regulatory strategy is ancient and present in the common ancestor of land plants (Shukla et al., 2025). Our identification of potential candidate pioneer factors provides a framework for understanding how plants maintain developmental genes in a poised state—

repressed by H3K27me3 but accessible to specific transcription factors that can initiate their activation during developmental transitions.

7.7 Evolutionary Divergence of Polycomb Repressive Pathways in Plants

Polycomb Repressive Complexes (PRC1 and PRC2) are evolutionarily conserved chromatin-modifying complexes essential for developmental gene regulation. PRC2 catalyzes H3K27me3, while PRC1 catalyzes H2AK121ub. In animals, these complexes function hierarchically, with PRC2-deposited H3K27me3 recruiting PRC1. Our comparative analysis of chromatin states in *Arabidopsis thaliana* and *Marchantia polymorpha* reveals significant evolutionary divergence in Polycomb pathway function in plants.

In *Arabidopsis*, H3K27me3 and H2AK121ub show strong co-occurrence at facultative heterochromatin states, both predominantly on H2A.Z-containing nucleosomes. Genetic studies demonstrate interdependence between these marks, with loss of PRC2 reducing H2AK121ub levels and loss of PRC1 affecting H3K27me3 distribution, suggesting cooperative function (Shukla et al., 2025).

Strikingly, *Marchantia polymorpha* shows fundamentally different organization. While *Marchantia* possesses functional PRC2 that deposits H3K27me3 at facultative heterochromatin associated with H2A.Z, the relationship between H3K27me3 and H2AK121ub is uncoupled. Loss of PRC1 activity does not affect H3K27me3 levels or distribution, indicating PRC2 functions independently of PRC1. Furthermore, H2AK121ub shows broader genomic distribution and is not confined to H3K27me3-marked regions (Shukla et al., 2025).

This independence in *Marchantia* suggests that tight coupling observed in *Arabidopsis* and animals represents a derived state evolving after bryophyte-vascular plant divergence. Despite these mechanistic differences, functional output appears conserved: in both species, H3K27me3-marked facultative heterochromatin associates with developmental regulators showing tissue-specific expression, suggesting that using H3K27me3 to mark poised developmental genes is an ancient feature of land plant regulation (Shukla et al., 2025)

These findings reveal that while core Polycomb machinery is ancient and conserved, the regulatory logic governing its function has diverged significantly across plant evolution. The

independence of PRC1 and PRC2 in *Marchantia* versus their interdependence in *Arabidopsis* highlights the evolutionary plasticity of epigenetic regulatory systems and provides insights into how chromatin-based gene regulation has been adapted to meet distinct developmental requirements of different plant lineages.

7.8 Implications for Understanding Plant Gene Regulation

The comprehensive chromatin state maps and functional analyses presented in this thesis provide a new framework for understanding gene regulation in plants. By integrating histone variants, histone modifications, DNA methylation, chromatin accessibility, and transcription factor binding into a unified model, we can now view plant genomes as organized into discrete functional domains, each with characteristic regulatory properties and mechanisms.

A central insight from this work is that chromatin organization operates hierarchically. At the highest level, H2A variants partition the genome into major functional domains—constitutive heterochromatin, facultative heterochromatin, and euchromatin. Within these domains, specific combinations of histone modifications create distinct chromatin states that define more refined regulatory environments. This hierarchical organization provides both stability and flexibility: H2A variants establish stable chromatin domains that persist across cell divisions, while histone modifications enable dynamic regulation within these domains in response to developmental and environmental signals.

This framework resolves several longstanding questions in plant chromatin biology. First, it explains why H2A.Z has been implicated in seemingly contradictory functions—both gene activation and repression. Our data show that H2A.Z marks facultative heterochromatin, a domain characterized by conditional repression of developmental genes that can be activated in specific contexts. Thus, H2A.Z creates a chromatin environment that is repressive by default but accessible to specific transcription factors, particularly pioneer-like factors such as MADS-box TFs and LEAFY, that can initiate gene activation.

Second, our work clarifies the relationship between DNA methylation and Polycomb-mediated repression. These represent parallel silencing pathways that target different genomic regions: DNA methylation primarily silences constitutive heterochromatin (transposable elements and repeats),

while H3K27me3 marks facultative heterochromatin (developmental genes). The mutual exclusivity of these marks reflects their association with different H2A variants—H2A.W for DNA methylation and H2A.Z for H3K27me3—suggesting that H2A variant identity determines which silencing pathway operates at a given locus.

Third, our identification of candidate pioneer transcription factors provides a mechanistic explanation for how developmental genes transition from repressed to active states. The ability of MADS-box TFs and LEAFY to bind H3K27me3-marked chromatin suggests they can access developmental genes even when repressed, positioning them to rapidly activate these genes in response to appropriate signals. This may explain how plants coordinate complex developmental transitions, such as the vegetative-to-reproductive phase change, which requires precise temporal activation of multiple developmental regulators.

The evolutionary conservation of chromatin state organization between *Arabidopsis* and *Marchantia* demonstrates that these principles are fundamental to land plant biology. The 450-million-year conservation of the basic chromatin state framework, with specific H2A variants defining major domains and similar histone modifications marking these domains, suggests strong selective pressure to maintain this organizational logic. At the same time, the divergence in specific features, such as Polycomb pathway coordination and DNA methylation complexity, illustrates how chromatin regulatory systems can be modified to accommodate lineage-specific requirements.

From a practical perspective, the chromatin state framework provides a powerful tool for predicting gene regulatory behavior and designing strategies to manipulate gene expression. Understanding which chromatin state marks a gene can inform predictions about the potential of variability of its expression pattern, responsiveness to environmental signals, and the transcription factors likely to regulate it.

Looking forward, several important questions remain. How are H2A variants targeted to specific genomic regions? What are the molecular mechanisms by which H2A variants influence the activity of chromatin-modifying enzymes? How do chromatin states change during development, and what triggers these transitions? Do plant MADS-box TFs possess bona fide pioneer activity at the biochemical level? Addressing these questions will require integrating genomic approaches with biochemical

and structural studies to understand the molecular mechanisms underlying chromatin state establishment and maintenance.

In conclusion, this work establishes chromatin states as fundamental organizational units of plant genomes that integrate multiple layers of chromatin information to create distinct regulatory environments. This framework provides a conceptual foundation for understanding how chromatin organization influences gene regulation and offers new perspectives on the evolution and function of epigenetic regulatory systems in plants.

Bibliography

Bergamin, E., Sarvan, S., Malette, J., Eram, M. S., Yeung, S., Mongeon, V., Joshi, M., Brunzelle, J. S., Michaels, S. D., Blais, A., Vedadi, M., & Couture, J.-F. (2017). Molecular basis for the methylation specificity of ATXR5 for histone H3. *Nucleic Acids Research*, *45*(11), 6375–6387. https://doi.org/10.1093/nar/gkx224

Bourguet, P., Picard, C. L., Yelagandula, R., Pélissier, T., Lorković, Z. J., Feng, S., Pouch-Pélissier, M.-N., Schmücker, A., Jacobsen, S. E., Berger, F., & Mathieu, O. (2021). The histone variant H2A.W and linker histone H1 co-regulate heterochromatin accessibility and DNA methylation. *Nature Communications*, *12*(1), 2683. https://doi.org/10.1038/s41467-021-22993-5

Derkacheva, M., Steinbach, Y., Wildhaber, T., Mozgová, I., Mahrez, W., Nanni, P., Bischof, S., Gruissem, W., & Hennig, L. (2013). Arabidopsis MSI1 connects LHP1 to PRC2 complexes. *The EMBO Journal*, *32*(14), 2073–2085. https://doi.org/10.1038/emboj.2013.145

Jamge, B., Lorković, Z. J., Axelsson, E., Osakabe, A., Shukla, V., Yelagandula, R., Akimcheva, S., Kuehn, A. L., & Berger, F. (2023). Histone variants shape chromatin states in Arabidopsis. *eLife*, *12*, RP87714. https://doi.org/10.7554/eLife.87714

Osakabe, A., Jamge, B., Axelsson, E., Montgomery, S. A., Akimcheva, S., Kuehn, A. L., Pisupati, R., Lorković, Z. J., Yelagandula, R., Kakutani, T., & Berger, F. (2021). The chromatin remodeler DDM1 prevents transposon mobility through deposition of histone variant H2A.W. *Nature Cell Biology*, *23*(4), 391–400. https://doi.org/10.1038/s41556-021-00658-1

Osakabe, A., Lorkovic, Z. J., Kobayashi, W., Tachiwana, H., Yelagandula, R., Kurumizaka, H., & Berger, F. (2018). Histone H2A variants confer specific properties to nucleosomes and impact on chromatin accessibility. *Nucleic*

Acids Research, *46*(15), 7675–7685. https://doi.org/10.1093/nar/gky540

Sayou, C., Nanao, M. H., Jamin, M., Posé, D., Thévenon, E., Grégoire, L., Tichtinsky, G., Denay, G., Ott, F., Peirats Llobet, M., Schmid, M., Dumas, R., & Parcy, F. (2016). A SAM oligomerization domain shapes the genomic binding landscape of the LEAFY transcription factor. *Nature Communications*, *7*, 11222. https://doi.org/10.1038/ncomms11222

Shukla, V., Axelsson, E., Hisanaga, T., Haseloff, J., Berger, F., & Romani, F. (2025). *Conservation of chromatin states and their association with transcription factors in land plants* (p. 2025.07.12.664529). bioRxiv. https://doi.org/10.1101/2025.07.12.664529

Surface, L. E., Fields, P. A., Subramanian, V., Behmer, R., Udeshi, N., Peach, S. E., Carr, S. A., Jaffe, J. D., & Boyer, L. A. (2016). H2A.Z.1 Monoubiquitylation Antagonizes BRD2 to Maintain Poised Chromatin in ESCs. *Cell Reports*, *14*(5), 1142–1155. https://doi.org/10.1016/j.celrep.2015.12.100

Wang, Y., Long, H., Yu, J., Dong, L., Wassef, M., Zhuo, B., Li, X., Zhao, J., Wang, M., Liu, C., Wen, Z., Chang, L., Chen, P., Wang, Q., Xu, X., Margueron, R., & Li, G. (2018). Histone variants H2A.Z and H3.3 coordinately regulate PRC2-dependent H3K27me3 deposition and gene expression regulation in mES cells. *BMC Biology*, *16*(1), 107. https://doi.org/10.1186/s12915-018-0568-6

Chapter 8

Epilogue

The path that led to this thesis began with a sense of wonder about the invisible molecular choreography happening within every plant cell. How could the same DNA sequence give rise to a root cell in darkness and a leaf cell reaching toward the sun? This question became the thread I followed, leading me to discover that the secret lies in the sophisticated ways cells organize their genetic material.

What started as curiosity about cellular diversity evolved into a deeper appreciation for the elegant regulatory systems that plants have refined over hundreds of millions of years. The development of my chromatin state discovery pipeline using Nextflow opened new doors and suddenly, I could systematically map and analyze the complex patterns of chromatin organization in two plant species that had previously been invisible.

Through this work, I uncovered 26 distinct chromatin states in *Arabidopsis thaliana*, each defined by unique combinations of histone modifications and variants. Perhaps most remarkably, I discovered that H2A variants serve as master regulators of chromatin identity, with their dynamic exchange controlled by the chromatin remodeler DDM1. This finding challenged my initial assumptions about how chromatin states are established and maintained.

The evolutionary analysis revealed something truly extraordinary—chromatin organization has remained remarkably conserved across 450 million years of plant evolution. Comparing *Arabidopsis* and *Marchantia*, I found that the fundamental principles governing chromatin states transcend species boundaries, suggesting these regulatory mechanisms are deeply embedded in evolution.

Finally, by developing a transcription factor activity scoring system, I identified how different TFs navigate and interact with chromatin landscapes. Some transcription factors prefer open, accessible regions, while others (potential pioneers) can bind to more restrictive chromatin environments, reshaping the regulatory landscape as they go.

This journey has taught me that chromatin is not merely DNA packaging, but a sophisticated information system that plants use to orchestrate gene expression across time, space, and evolutionary history. What began as wonder about cellular diversity has become a deep appreciation for the molecular elegance underlying all plant life.