Title: Histone variants shape the chromatin states in Arabidopsis

Authors: 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[1,6], and Frédéric Berger[1**]

Affiliations:

[1] Gregor Mendel Institute, Austrian Academy of Sciences, Vienna BioCenter, Dr. Bohr-Gasse 3, 1030 Vienna, Austria

[2] Vienna BioCenter PhD Program, Doctoral School of the University of Vienna and Medical University of Vienna, A-1030, Vienna, Austria

[3] Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan

[4] PRESTO, Japan Science and Technology Agency, Honcho, Kawaguchi, Saitama, 332-0012, Japan

[5] Present address: Institute of Molecular Pathology, Campus-Vienna-Biocenter 1, 1030 Vienna, Austria

[6] Present address: Department of Chromatin Regulation, Max Planck Institute of Immunobiology and Epigenetics, Freiburg, Germany

** Correspondence: Frederic.berger@gmi.oeaw.ac.at

Summary

How different intrinsic sequence variation or regulatory modifications of histones regulate nucleosome interactions with transcription remain unclear. To clarify this question, we examine how histone variants and histone modifications assemble in the *Arabidopsis thaliana* genome, and show that a limited number of chromatin states divide euchromatin and heterochromatin into biologically significant subdomains. We find that histone variants are as significant as histone modifications in determining the composition of chromatin states. Loss of function of the chromatin remodeler DECREASED IN DNA METHYLATION (DDM1) prevents the exchange of the histone variant H2A.Z to H2A.W and impacts the definition and distribution of chromatin states. In the mutant deprived of DDM1 transposons harbor chromatin states normally found only on proteins coding genes in the wild type. We propose that the dynamics of histone variant exchange control the organization of histone modifications into chromatin states, resulting in molecular landmarks that signify whether genes could be transcribed.

Keywords: Histone variant, H2A variants, Histone modification, Chromatin state, Transcription

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 are essential for genome organization and function in packaging DNA (Macadangdang *et al*, 2014), transcription (Allis *et al*, 1980; Costanzi & Pehrson, 1998; Giaimo *et al*, 2019) and genome maintenance (Rogakou *et al*, 1998). Nucleosomes contain two copies

each of the core histones H2A, H2B, H3, and H4. Histone variants have arisen through 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 & Berger, 2020; Talbert & Henikoff, 2021). Multicellular eukaryotes possess a centromeric H3 (CENP-A), a DNA replication–dependent H3.1 and a DNA replication-independent H3.3, as well as H2A.Z, H2A.X, and a DNA replication–dependent H2A variant. These variants regulate the properties of nucleosomes (Chakravarthy *et al*, 2004; Horikoshi *et al*, 2013; Koyama & Kurumizaka, 2018; Tachiwana *et al*, 2012; Tachiwana *et al*, 2011) and affect transcription (Rudnizky *et al*, 2016; Subramanian *et al*, 2015; Weber *et al*, 2014; Wollmann *et al*, 2017). In addition, unique H2A variants associated with transcriptional repression have evolved in animals and plants (Loppin & Berger, 2020; Talbert & Henikoff, 2021). In vascular plants, the H2A.W variants are distinguished by a C-terminal KSPK motif (Lei *et al*, 2021; Schmucker *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*, 2022; Osakabe *et al*, 2018; Schmucker *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 remain unknown. However, H2B variants might differ in their capacity to be ubiquitinated, an important modification for RNA polII elongation (Feng & 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 & Henikoff, 2021; Tessarz & Kouzarides, 2014). Applying a Hidden Markov Model [ChromHMM] (Ernst & Kellis, 2012, 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 & Kellis, 2012, 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 replicative 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 (Henikoff *et al*, 2004; Weber & 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 chromatin 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 the demonstration of the exchange of H2A variants by the remodeler DECREASE IN DNA METHYLATION 1 (DDM1) and its loss causing the mis-assembly of chromatin states.

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 H3.1 and H3.3 under the control of their native promoters (Jiang & 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. MS analysis revealed that circa 40% of nucleosomes contained both H3.1 and H3.3 (**Figure 1A**). Neither H3.1 nor H3.3 was preferentially associated with a specific H2A variant (**Figure 1B**). This was confirmed by MS analysis of nucleosomes pulled down with H2A variant-specific antibodies (**Figure 1 figure supplement 1D**). Therefore, H3 variants do not necessarily form homotypic nucleosomes, despite the presence of H3 variant-specific deposition mechanisms (Loppin & Berger, 2020; Nie *et al*, 2014; Probst *et al*, 2020), and they do not associate with a specific H2A variant.

We also explored the degree of co-occurrence of histone modifications with H3.1 and H3.3 variants. H3.1 and H3.3 bore the same sets and comparable levels of methylation and acetylation, except that H3.3 was more strongly associated with H3 acetylation and H3K36 methylation and H3.1 with H3K27me1 (**Figure 1C**). MS showed that the levels of H3K27 and H3K36 modifications on H3.1 and H3.3 were similar in endogenous and transgene-encoded H3.1 or H3.3 nucleosomes (**Figure 1D** and **Figure 1 figure supplement 1E–J**). As reported previously (Johnson *et al*, 2004), we observed some preferential association of H3.1 with H3K27me1 and H3.3 with H3K36 methylation, but this was not a strict association (**Figure 1D**

and **Figure 1 figure supplement 1E–J**). Thus, based on two different approaches, we conclude that there is no tight association between H3 modifications and the two H3 variants.

In contrast to H3 variants, homotypic nucleosomes containing either H2A.W, H2A.Z, H2A, or H2A.X showed marked enrichment of specific histone modifications (**Figure 1E**). Histone modifications in constitutive heterochromatin (H3K9me1, H3K9me2, and H3K27me1) were primarily associated with H2A.W. 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 and among themselves. 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.

Histone variants impact the determination of 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 replicative H2A (H2A.2 and H2A.13) in our analysis. We used profiles of enrichment obtained from ten days old seedlings. The algorithm ChromHMM (Ernst & Kellis, 2017) was used to define chromatin states in Arabidopsis (see Methods for details). We chose to analyze the most parsimonious model based on 26 chromatin states (**Figure 2 figure supplement 2A**). Chromatin states were clustered based on the emission probability for each modification and histone variant (**Figure 2A**). They showed distinct enrichment of various marks (**Figure 2A**), and distinct relative abundance on transposons, repeats, and elements of protein-coding genes (**Figure 2B**) and occupied different proportions

of the genome (**Figure 2 figure supplement 2F**). For each chromatin state, we measured the average level of transcriptional activity (**Figure 2C**), the degree of enrichment in CG methylation (**Figure 2D, figure supplement 2C** and **2D**), the level of accessibility by DNase I-seq (**Figure 2E**), and the nucleosome density determined by MNase-seq (**Figure 2 figure supplement 2E**). Chromatin states H1–H6 showed low accessibility, the highest levels of CG 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 2F**), 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 mark 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 (**Figures 1** and **2A**). 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 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 identified in our study with the 9 states identified 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 2B**). 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 identified in previous studies. In particular, states 2, 4, and 6 contained elements belonging to multiple newly defined chromatin 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, pointing to the importance of histone variants in the definition of chromatin states.

To test this idea, we calculated chromatin states after excluding either all histone modifications or all histone variants or single families of histone variants from our set used to calculate the 26 states model. A comparison between the resulting matrices of chromatin states was measured by the Jaccard index to evaluate the impact of removing either H2B variants, H3 variants, H2A variants, H3 modifications (no H3 PTMs) or all histone variants (no H2A/H3/H2B) on the definition of the 26 chromatin states (**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, histone modifications and histone variants are required to a comparable degree to define chromatin states. By combining computational and biochemical analyses, we 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 definition of chromatin states

We considered the idea that the dynamics of 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; Rougee *et al*, 2021). This suggested that the loss of H2A.W deposition had a profound impact on chromatin states. We thus obtained the profiles of H1, four H3 modifications and all H2A variants (11 profiles of enrichments) in leaves of *ddm1* and wild-type to identify the chromatin states in both genetic backgrounds.

Because we studied the impact of *ddm1* in leaves and defined the 26 reference chromatin states using profiles from seedlings we further 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 enrichment 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 3A)**. We observed an association between H2A variants and histone modifications comparable to the model describing the 26 chromatin states (**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 model describing the 26 chromatin states. The complexity of the different heterochromatin states was preserved in the concatenated model, supporting that the set of H2A variants and histone modifications used to study *ddm1* is sufficient to represent the domain of chromatin affected directly by *ddm1*. Each state occupied a similar proportion of the genome in seedling or leaves to the exception of state 5 present primarily in leaves and state 13 only present in seedlings (**Figure 3 figure supplement 3A, right column with green bars**). This indicated that neither tissue type nor developmental stage have 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 & Kellis, 2017). The concatenated model created a common set of 16 chromatin states that are shared between wild type and *ddm1* and identified the location of these chromatin states across the genome for each genotype individually (see Methods) (**Figure 3A**). The states were classified as constitutive heterochromatin (hI-hIII), facultative heterochromatin (fI-fVI), euchromatin (eI-eIIII), and intergenic (iI), based on their emission probability profiles 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 represent 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 euchromatin E 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 (hI-hIII) was reduced compared with the six heterochromatin in the wild type 26-state model. Overall, *ddm1* caused a perturbation of the chromatin states associated with constitutive heterochromatin and did not affect significantly 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 3C**).

To compare the chromatin state assignments between *ddm1* and wild-type genomes, we measured the Jaccard index (**Figure 3B**) 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 3F**). 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 not changed over protein coding genes, but were instead changed over TEs, including TE fragments and TE genes (assigned to TE families in TAIR10 annotation) (**Figure 3 figure supplement 3E**). 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, 3F** and **Figure 3 figure supplement 3E**). In addition we observed that *ddm1* caused mis-association of region of euchromatin and facultative heterochromatin with chromatin states, suggesting indirect effects of the disruption of constitutive heterochromatin caused directly by the loss of DDM1 (**Figure 3 figure supplement 3F**). We thus concluded that loss of DDM1 causes a profound perturbation of both the definition of chromatin states and their distribution.

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, Figure 3 figure supplement 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 3 figure supplement 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 a 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 3G** and **Figure 3 figure supplement 3H**). 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.

The DDM1 murine ortholog LSH binds and exchanges macroH2A and H2A (Ni & Muegge, 2021; Ni *et al*, 2020). The finding that DDM1 uses the same sites to bind both H2A.Z and H2A.W suggested that DDM1 controls the exchange between H2A.W and H2A.Z. To test whether this exchange occurs, we surveyed the enrichment of H2A variants and histone PTMs over TEs in *ddm1* and compared this with wild type. There was an overall enrichment of H2A.Z over TEs in *ddm1*, supporting the idea that DDM1 exchanges H2A.Z for H2A.W at TE genes in the wild type (**Figure 4A**). Remarkably, non-transcribed TE genes also accumulated H2A.Z (**Figure 4A**) supporting the conclusion that the exchange of H2A.Z for H2A.W 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 4A**). Transcription did not affect the degree of enrichment for H3K27me3 of facultative chromatin but only for H3K36me3, resulting in a chromatin profile resembling that observed in expressed protein coding genes (**Figure 4 figure supplement 4B**). Accordingly, strongly expressed TE genes showed an accumulation of euchromatic states, in contrast with non-expressed TEs that became covered with states typical of facultative heterochromatin, irrespective of the type of constitutive heterochromatin observed on these TEs in wild type (**Figure 4B, 4C** and **Figure 4 figure supplement 4E**). The change of chromatin states did not correlate with the length of the TE gene (**Figure 4 figure supplement 4C**) or the homogeneity of the chromatin landscape (**Figure 4 figure supplement 4D**). We conclude that the loss of DDM1 prevents the replacement of H2A.W by H2A.Z in a transcription-independent manner. This alteration in the dynamics of H2A.W and H2A.Z in *ddm1* causes a global perturbation of the allocation of chromatin states, in some cases permitting transcription of TEs, which is then accompanied by a further exchange of H2A.Z to H2A and H2A.X.

Discussion

Histone variants drive the overall organization of chromatin states

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. We do not exclude that minor variations of chromatin states occur between cell types or that rare cell types could harbor specific chromatin states that are

not reflected in our study. *In silico* and biochemical analyses demonstrate the strongest associations between nucleosomes homotypic for H2A variants and histone modifications while nucleosomes primarily heterotypic for H3 variants show a broader association with histone modifications.

The prominent role of H2A variants in the organization of chromatin states is supported by *in silico* simulation and the impact of the loss of the remodeler DDM1, which deposits H2A.W to maintain TE silencing. We further show that the DDM1 binding sites for H2A.W also bind H2A.Z, but not H2A or H2A.X. In *ddm1*, H2A.W is replaced by H2A.Z, whereas loss of H2A.W itself causes its replacement by H2A or H2A.X, but not H2A.Z (Bourguet *et al*, 2021) supporting the conclusion that DDM1 exchanges H2A.Z for H2A.W. This highlights further 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 & Muegge, 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 which remain silent in *ddm1* are occupied by a new state resembling facultative heterochromatin marked by 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 a driving force that shapes the deposition of histone modifications and thus participate in 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 ubiquitin ligase by the H2A.Z tail (Surface *et al*, 2016) explains 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 with

chromatin states as well as previous biochemical analyses (Gomez-Zambrano *et al*, 2019; Osakabe *et al.*, 2018). 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). These reports explain some of the relationships between histone variants and modifications. The importance of histone variants in the definition of chromatin states further suggests that the nature of the H2A variant directly controls the activity of the machinery that either deposits or erases the histone modifications.

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 (Reveron-Gomez *et al*, 2018). As plant cells divide, H3.1 sustains the deposition of H3K27me3 (Jiang & Berger, 2017) thus providing a forward feedback loop to sustain the polycomb repressive state in dividing cells. Recent data in mammalian cells show that the recycling of H2A-H2B after replication provides another positive feedback loop to maintain the patterns of post translational modifications linked to H2A variants and H2B (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 in constitutive heterochromatin was previously noted (Johnson *et al.*, 2004) 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

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.

For MNase digestion followed by immunoprecipitation, nuclei were isolated from 4 grams of 2-3 weeks old leaves and the procedure as described in (Lorkovic *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 1,800 × *g* at 4ºC nuclei were re-suspended in N buffer to a volume of 1 ml. Twenty microliters of MNase (0.1 u/µl) (SigmaAldrich) were added to each tube and incubated for 15 min at 37ºC and during the incubation nuclei were mixed 4 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 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 4ºC. 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 were kept for DNA extraction. Antibodies, including non-specific IgG from rabbit, were bound to protein A magnetic beads (GE Healthcare) and then incubated with MNase extracts overnight at 4ºC. 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 (ThermoFischer 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 1:5,000 dilution. Rat anti-HA antibody (Roche 3F10) was used at 1:2,000 dilution. Secondary antibodies were goat anti-rabbit IgG (BioRad) and goat-anti rat IgG (SigmaAldrich), both at a 1:10,000 dilution.

Mass spectrometry

For mass spectrometry immunoprecipitated nucleosomes were resolved on 4–20% gradient gels (Serva) and silver-stained. Histone H3 bands were excised, reduced, alkylated, in-gel trypsin and LysC digested, and processed for MS. The nano HPLC system used was a Dionex UltiMate 3000 HPLC RSLC nano system (ThermoFisher Scientific, Amsterdam, Netherlands) coupled to a Q Exactive mass spectrometer (ThermoFisher Scientific, Bremen, Germany), equipped with a Proxeon nanospray source (ThermoFisher Scientific, Odense, Denmark). Peptides were loaded onto a trap column (ThermoFisher Scientific, Amsterdam, Netherlands, PepMap C18, 5 mm 3 300 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 (ThermoFisher Scientific, Amsterdam, Netherlands, PepMap C18, 500mm 3 75 mm ID, 2 mm, 100 Å). Peptides were eluted using a flow rate of 230 nl/min and a binary 2-h 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 eneergy 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 state 1, 7, 8, >8) were put on a dynamic exclusion list for 40 s. Additionally, the minimum AGC target was set to 5E3 and 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 (7 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 in 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 1D**) were 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 threshold for modification probability set to 95% or higher. Relative modification levels were expressed as the number of modified peptides (**Figure 1 figure supplement 1F**) divided by the total number of peptides (**Figure 1 figure supplement 1E**) 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 1 figure supplement 1**). 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 1H and 1I**). 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 1J**) 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 grown on ½ MS sterilized plates in the growth chamber under long day (LD) conditions (21˚C; 16 h light/8 h dark). After 10 days seedling tissue was freshly harvested. For ChIP-seq from leaf tissue, Col-0 wild type (WT) and *ddm1* seeds were stratified at 4˚C in the dark for three days. Seeds were grown on soil in the growth chamber under long day (LD) conditions (21˚C; 16 h light/8 h dark) and leaf tissue was harvested 5 weeks after germination.

Chromatin immunoprecipitation (ChIP)

For ChIP nuclei isolation was performed using 10 day 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 were 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 hour 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 hours 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 hours and subsequently at 65˚C for 16 hours. 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*

$(Ingouff et al)) and then aligned to the TAIR10 *Arabidopsis* genome using Bowtie2 v2.4.1 (Langmead & Salzberg, 2012) with default settings, Picard v2.22.8 (broadinstitute.github.io/picard/) was used to remove duplicated reads. Deeptools v3.1.2 (Ramirez *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. log2 ratio (ChIP/ (Input or H3)) bigwig files.

***ChIP-seq data processing* for published data**

Publicly available ChIP-seq data for H2B variants was downloaded from GEO GSE151166 (Jiang *et al.*, 2020). ChIP-seq data for H3 variants was 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.

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 & Kellis, 2017) with default settings and window size 200 bp. We generated chromatin states from n=2 to n=50 (n, number of chromatin states). We chose to analyze the 26 chromatin state model in depth. We analyzed multiple states models within a window of n=13 to n=27 states. We looked at state association with genomic features. We observed that increasing state numbers from 26 to 27 gave rise to biologically redundant states. Hence we chose to analyze 26 state models in depth.

For the mixed seedling and leaf data, and for the mixed wild type and *ddm1* mutant data we used concatenated ChromHMM models (Ernst & 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, the separate files were joined. To this end, for each (200bp) bin of the genome, the information about which state had been assigned in wild type and ddm1 mutant respectively was combined. The state assignment datasets were then overlapped, using 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 and plots generated using R and the ggplot2 package from tidyverse.

Comparison of states from extensive and concatenated models

To compare the states from the two tissue model with the chromatin types defined in the extensive model, the states assigned to seedling by the two tissue model were overlapped with the chromatin types from the extensive model. For the comparison of the wildtype and *ddm1* mutant model states with the chromatin types from the extensive model, the states assigned to the wildtype were 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 3B** and **3C**).

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 state *s* 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 the function plotProfile from Deeptools v3.1.2 (Ramirez *et al.*, 2016).

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. 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 (github.com/nf-core/methylseq) as described in (doi: https://doi.org/10.1101/2022.12.04.519028). Cutadapt v2.10 was used to trim the adaptors (default parameters). Trimmed reads were then mapped to TAIR10 (Col-0) assembly using bismark v0.2.1 (Krueger & Andrews, 2011) allowing mismatches to 0.5. Methylation calls were performed using methylpy v1.3.7 (Schultz *et al*, 2015) on the aligned bam files. Methylation call bed files were used to calculate average methylation over chromatin state bed regions. Bedtools

map v2.3 (Quinlan, 2014) was used to calculate mean methylation levels across chromatin states. Box plots were generated using R v3.5 and the ggplot2 package from tidyverse to compare methylation levels across chromatin states.

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 with zero expression (tpm=0) in *ddm1* mutant were put in the “no expression” group. The remaining TE genes were divided into 4 quartiles based on the tpm values in the *ddm1* mutant.

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 GSE226469. Source data for all the main figures as well as supplementary figures have been provided with this manuscript. All other data supporting the conclusions of the study will be available from the corresponding author upon reasonable request.

Code availability

All the code used to analyze the genome-wide sequencing data presented in this study, as described in the Methods are available upon request.

Acknowledgements

FB acknowledges support from the next generation sequencing and PlantS facilities at the Vienna BioCenter Core Facilities (VBCF), the BioOptics facility and Molecular Biology Services from the Institute for Molecular Pathology (IMP), and 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 A.O.], the Austrian Science Fund (FWF): M2539-B21 [to A.O.], P32054, P30802, and P33380 [to F.B.], and DK1238 [to B.J.], the Austrian Academy of Sciences [to F.B., Z.J.L, E.A., S.A., R.P.and R.Y.], 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 A.O.].

Author contributions

Z.J.L. performed all the biochemical analysis of the composition of mononucleosomes. A.O. and A.L.K. performed biochemical experiments on DDM1 binding to H2A variants. S.A. prepared biological material. B.J. and A.O. performed the ChIPseq and RNAseq. R.Y. contributed the initial set of ChIP-seq data. B.J., E.A. and V.S. performed the genome-wide analysis. F.B. conceived, designed, and supervised the project. The manuscript was drafted by F.B., E.A, and Z.J.L., with comments from all co-authors.

Competing interests

All authors discussed the results and have no conflict of interest.

References

Allis CD, Glover CV, Bowen JK, Gorovsky MA (1980) Histone variants specific to the transcriptionally active, amitotically dividing macronucleus of the unicellular eucaryote, Tetrahymena thermophila. *Cell* 20: 609-617

Aslam M, Fakher B, Jakada BH, Cao S, Qin Y (2019) SWR1 Chromatin Remodeling Complex: A Key Transcriptional Regulator in Plants. *Cells* 8

Berger F, Muegge K, Richards EJ (2023) Seminars in cell and development biology on histone variants remodelers of H2A variants associated with heterochromatin. *Semin Cell Dev Biol* 135: 93-101

Bieluszewski T, Sura W, Dziegielewski W, Bieluszewska A, Lachance C, Kabza M, Szymanska-Lejman M, Abram M, Wlodzimierz P, De Winne N *et al* (2022) NuA4 and H2A.Z control environmental responses and autotrophic growth in Arabidopsis. *Nat Commun* 13: 277

Bourguet P, Picard CL, Yelagandula R, Pelissier T, Lorkovic ZJ, Feng S, Pouch-Pelissier MN, Schmucker A, Jacobsen SE, Berger F *et al* (2021) The histone variant H2A.W and linker histone H1 co-regulate heterochromatin accessibility and DNA methylation. *Nat Commun* 12: 2683

Bourguet P, Yelagandula R, To TK, Osakabe A, Alishe A, Lu RJ-H, Kakutani T, Chen P-Y, Berger F (2022) The histone variant H2A.W cooperates with chromatin modifications and linker histone H1 to maintain transcriptional silencing of transposons in Arabidopsis.

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. *Plant Cell* 30: 1337-1352

Chakravarthy S, Bao Y, Roberts VA, Tremethick D, Luger K (2004) Structural characterization of histone H2A variants. *Cold Spring Harb Symp Quant Biol* 69: 227-234

Costanzi C, Pehrson JR (1998) Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. *Nature* 393: 599-601

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. *J Proteome Res* 13: 3679-3684

Ernst J, Kellis M (2012) ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* 9: 215-216

Ernst J, Kellis M (2017) Chromatin-state discovery and genome annotation with ChromHMM. *Nat Protoc* 12: 2478-2492

Feng J, Shen WH (2014) Dynamic regulation and function of histone monoubiquitination in plants. *Front Plant Sci* 5: 83

Flury V, Reveron-Gomez 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 e1019

Giaimo BD, Ferrante F, Herchenrother A, Hake SB, Borggrefe T (2019) The histone variant H2A.Z in gene regulation. *Epigenetics Chromatin* 12: 37

Gomez-Zambrano A, Merini W, Calonje M (2019) The repressive role of Arabidopsis H2A.Z in transcriptional regulation depends on AtBMI1 activity. *Nat Commun* 10: 2828

Gu Z, Eils R, Schlesner M (2016) Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32: 2847-2849

Gu Z, Gu L, Eils R, Schlesner M, Brors B (2014) circlize Implements and enhances circular visualization in R. *Bioinformatics* 30: 2811-2812

Henikoff S, McKittrick E, Ahmad K (2004) Epigenetics, histone H3 variants, and the inheritance of chromatin states. *Cold Spring Harb Symp Quant Biol* 69: 235-243

Horikoshi N, Sato K, Shimada K, Arimura Y, Osakabe A, Tachiwana H, Hayashi-Takanaka Y, Iwasaki W, Kagawa W, Harata M *et al* (2013) Structural polymorphism in the L1 loop regions of human H2A.Z.1 and H2A.Z.2. *Acta Crystallogr D Biol Crystallogr* 69: 2431-2439

Ingouff M, Rademacher S, Holec S, Soljic L, Xin N, Readshaw A, Foo SH, Lahouze B, Sprunck S, Berger F (2010) Zygotic resetting of the HISTONE 3 variant repertoire participates in epigenetic reprogramming in Arabidopsis. *Curr Biol* 20: 2137-2143

Jacob Y, Bergamin E, Donoghue MT, Mongeon V, LeBlanc C, Voigt P, Underwood CJ, Brunzelle JS, Michaels SD, Reinberg D *et al* (2014) Selective methylation of histone H3 variant H3.1 regulates heterochromatin replication. *Science* 343: 1249-1253

Jeddeloh JA, Bender J, Richards EJ (1998) The DNA methylation locus DDM1 is required for maintenance of gene silencing in Arabidopsis. *Genes Dev* 12: 1714-1725

Jiang D, Berger F (2017) DNA replication-coupled histone modification maintains Polycomb gene silencing in plants. *Science* 357: 1146-1149

Jiang D, Borg M, Lorkovic 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 Genet* 16: e1008964

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 Res* 32: 6511-6518

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

Kharchenko PV, Alekseyenko AA, Schwartz YB, Minoda A, Riddle NC, Ernst J, Sabo PJ, Larschan E, Gorchakov AA, Gu T *et al* (2011) Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. *Nature* 471: 480-485

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

Kim YJ, Wang R, Gao L, Li D, Xu C, Mang H, Jeon J, Chen X, Zhong X, Kwak JM *et al* (2016) POWERDRESS and HDA9 interact and promote histone H3 deacetylation at specific genomic sites in Arabidopsis. *Proc Natl Acad Sci U S A* 113: 14858-14863

Koyama M, Kurumizaka H (2018) Structural diversity of the nucleosome. *J Biochem* 163: 85-95

Krueger F, Andrews SR (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27: 1571-1572

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9: 357-359

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. *Curr Biol* 31: 182-191 e185

Leng X, Thomas Q, Rasmussen SH, Marquardt S (2020) A G(enomic)P(ositioning)S(ystem) for Plant RNAPII Transcription. *Trends Plant Sci* 25: 744-764

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 Res* 46: D1157-D1167

Long J, Carter B, Johnson ET, Ogas J (2023) Contribution of the histone variant H2A.Z to expression of responsive genes in plants. *Semin Cell Dev Biol* 135: 85-92

Loppin B, Berger F (2020) Histone Variants: The Nexus of Developmental Decisions and Epigenetic Memory. *Annu Rev Genet* 54: 121-149

Lorkovic ZJ, Park C, Goiser M, Jiang D, Kurzbauer MT, Schlogelhofer P, Berger F (2017) Compartmentalization of DNA Damage Response between Heterochromatin and Euchromatin Is Mediated by Distinct H2A Histone Variants. *Curr Biol* 27: 1192-1199

Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* 389: 251-260

Macadangdang BR, Oberai A, Spektor T, Campos OA, Sheng F, Carey MF, Vogelauer M, Kurdistani SK (2014) Evolution of histone 2A for chromatin compaction in eukaryotes. *Elife* 3

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. *Nat Cell Biol* 10: 483-488

Ni K, Muegge K (2021) LSH catalyzes ATP-driven exchange of histone variants macroH2A1 and macroH2A2. *Nucleic Acids Res* 49: 8024-8036

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. *Nat Commun* 11: 5647

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. *Biol Open* 3: 794-802

Osakabe A, Jamge B, Axelsson E, Montgomery SA, Akimcheva S, Kuehn AL, Pisupati R, Lorkovic ZJ, Yelagandula R, Kakutani T *et al* (2021) The chromatin remodeler DDM1 prevents transposon mobility through deposition of histone variant H2A.W. *Nat Cell Biol* 23: 391-400

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 Res* 46: 7675-7685

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. *J Cell Sci* 126: 1323-1332

Probst AV, Desvoyes B, Gutierrez C (2020) Similar yet critically different: the distribution, dynamics and function of histone variants. *J Exp Bot* 71: 5191-5204

Quinlan AR (2014) BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics* 47: 11 12 11-34

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. *Mol Biol Evol* 39

Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dundar F, Manke T (2016) deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* 44: W160-165

Reveron-Gomez N, Gonzalez-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. *Mol Cell* 72: 239-249 e235

Riemondy KA, Sheridan RM, Gillen A, Yu Y, Bennett CG, Hesselberth JR (2017) valr: Reproducible genome interval analysis in R. *F1000Res* 6: 1025

Rogakou EP, Pilch DR, Orr AH, Ivanova VS, Bonner WM (1998) DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J Biol Chem* 273: 5858-5868

Roudier F, Ahmed I, Berard C, Sarazin A, Mary-Huard T, Cortijo S, Bouyer D, Caillieux E, Duvernois-Berthet E, Al-Shikhley L *et al* (2011) Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. *EMBO J* 30: 1928-1938

Rougee 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 Sci Alliance* 4

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. *Nat Commun* 7: 12958

Schmucker A, Lei B, Lorkovic ZJ, Capella M, Braun S, Bourguet P, Mathieu O, Mechtler K, Berger F (2021) Crosstalk between H2A variant-specific modifications impacts vital cell functions. *PLoS Genet* 17: e1009601

Schultz MD, He Y, Whitaker JW, Hariharan M, Mukamel EA, Leung D, Rajagopal N, Nery JR, Urich MA, Chen H *et al* (2015) Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature* 523: 212-216

Sequeira-Mendes J, Araguez I, Peiro 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. *Plant Cell* 26: 2351-2366

Stroud H, Greenberg MV, Feng S, Bernatavichute YV, Jacobsen SE (2013) Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. *Cell* 152: 352-364

Stroud H, Otero S, Desvoyes B, Ramirez-Parra E, Jacobsen SE, Gutierrez C (2012) Genome-wide analysis of histone H3.1 and H3.3 variants in Arabidopsis thaliana. *Proc Natl Acad Sci U S A* 109: 5370-5375

Subramanian V, Fields PA, Boyer LA (2015) H2A.Z: a molecular rheostat for transcriptional control. *F1000Prime Rep* 7: 01

Sullivan AM, Arsovski AA, Lempe J, Bubb KL, Weirauch MT, Sabo PJ, Sandstrom R, Thurman RE, Neph S, Reynolds AP *et al* (2014) Mapping and dynamics of regulatory DNA and transcription factor networks in A. thaliana. *Cell Rep* 8: 2015-2030

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 Rep* 14: 1142-1155

Tachiwana H, Kagawa W, Kurumizaka H (2012) Comparison between the CENP-A and histone H3 structures in nucleosomes. *Nucleus* 3: 6-11

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. *Proc Natl Acad Sci U S A* 107: 10454-10459

Tachiwana H, Kagawa W, Shiga T, Osakabe A, Miya Y, Saito K, Hayashi-Takanaka Y, Oda T, Sato M, Park SY *et al* (2011) Crystal structure of the human centromeric nucleosome containing CENP-A. *Nature* 476: 232-235

Talbert PB, Henikoff S (2021) Histone variants at a glance. *J Cell Sci* 134

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

Taus T, Kocher T, Pichler P, Paschke C, Schmidt A, Henrich C, Mechtler K (2011) Universal and confident phosphorylation site localization using phosphoRS. *J Proteome Res* 10: 5354-5362

Tessarz P, Kouzarides T (2014) Histone core modifications regulating nucleosome structure and dynamics. *Nat Rev Mol Cell Biol* 15: 703-708

Wang Y, Long H, Yu J, Dong L, Wassef M, Zhuo B, Li X, Zhao J, Wang M, Liu C *et al* (2018) Histone variants H2A.Z and H3.3 coordinately regulate PRC2-dependent H3K27me3 deposition and gene expression regulation in mES cells. *BMC Biol* 16: 107

Weber CM, Henikoff S (2014) Histone variants: dynamic punctuation in transcription. *Genes Dev* 28: 672-682

Weber CM, Ramachandran S, Henikoff S (2014) Nucleosomes are context-specific, H2A.Z-modulated barriers to RNA polymerase. *Mol Cell* 53: 819-830

Willige BC, Zander M, Yoo CY, Phan A, Garza RM, Wanamaker SA, He Y, Nery JR, Chen H, Chen M *et al* (2023) Author Correction: PHYTOCHROME-INTERACTING FACTORs trigger environmentally responsive chromatin dynamics in plants. *Nat Genet* 55: 355

Wollmann H, Holec S, Alden K, Clarke ND, Jacques PE, Berger F (2012) Dynamic deposition of histone variant H3.3 accompanies developmental remodeling of the Arabidopsis transcriptome. *PLoS Genet* 8: e1002658

Wollmann H, Stroud H, Yelagandula R, Tarutani Y, Jiang D, Jing L, Jamge B, Takeuchi H, Holec S, Nie X *et al* (2017) The histone H3 variant H3.3 regulates gene body DNA methylation in Arabidopsis thaliana. *Genome Biol* 18: 94

Yang Y, Zhang L, Xiong C, Chen J, Wang L, Wen Z, Yu J, Chen P, Xu Y, Jin J *et al* (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 Res* 50: 191-206

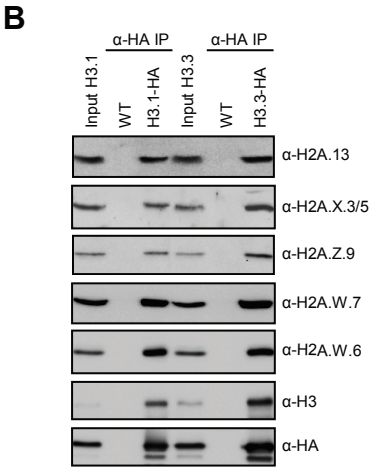
Yelagandula R, Stroud H, Holec S, Zhou K, Feng S, Zhong X, Muthurajan UM, Nie X, Kawashima T, Groth M *et al* (2014) The histone variant H2A.W defines heterochromatin and promotes chromatin condensation in Arabidopsis. *Cell* 158: 98-109

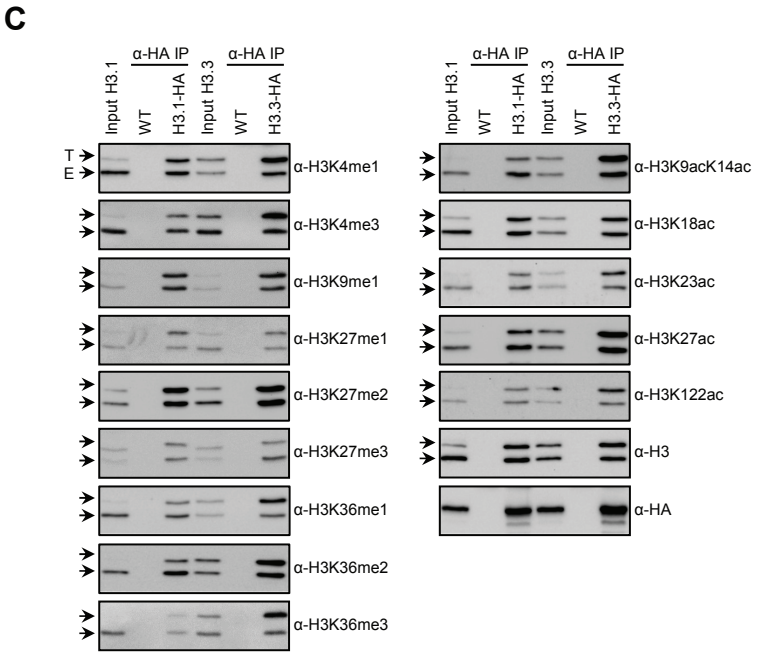
Zhao F, Xue M, Zhang H, Li H, Zhao T, Jiang D (2023) Coordinated histone variant H2A.Z eviction and H3.3 deposition control plant thermomorphogenesis. *New Phytol*

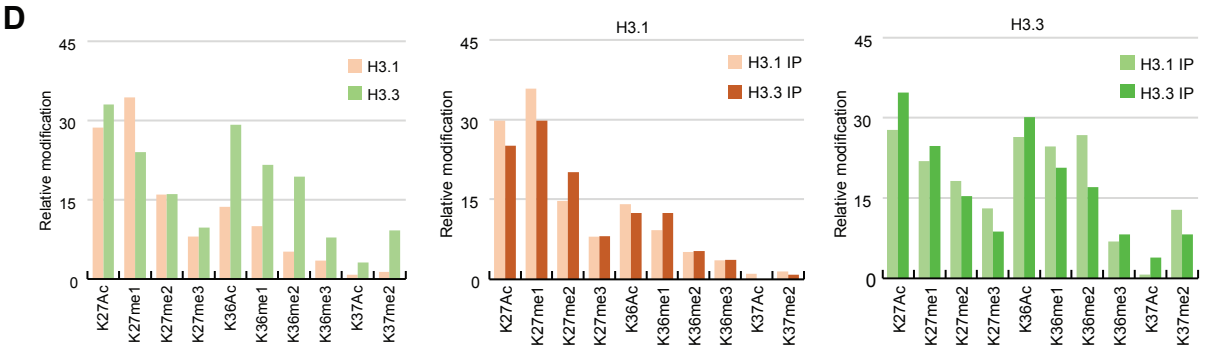
Zhou Y, Romero-Campero FJ, Gomez-Zambrano A, Turck F, Calonje M (2017) H2A monoubiquitination in Arabidopsis thaliana is generally independent of LHP1 and PRC2 activity. *Genome Biol* 18: 69

A

	H3.1 IP		H3.3 IP	
Peptides	T	E	T	E
H3.1	954	516	20	488
H3.3	16	560	1319	489
All H3	2577	2581	3493	2159







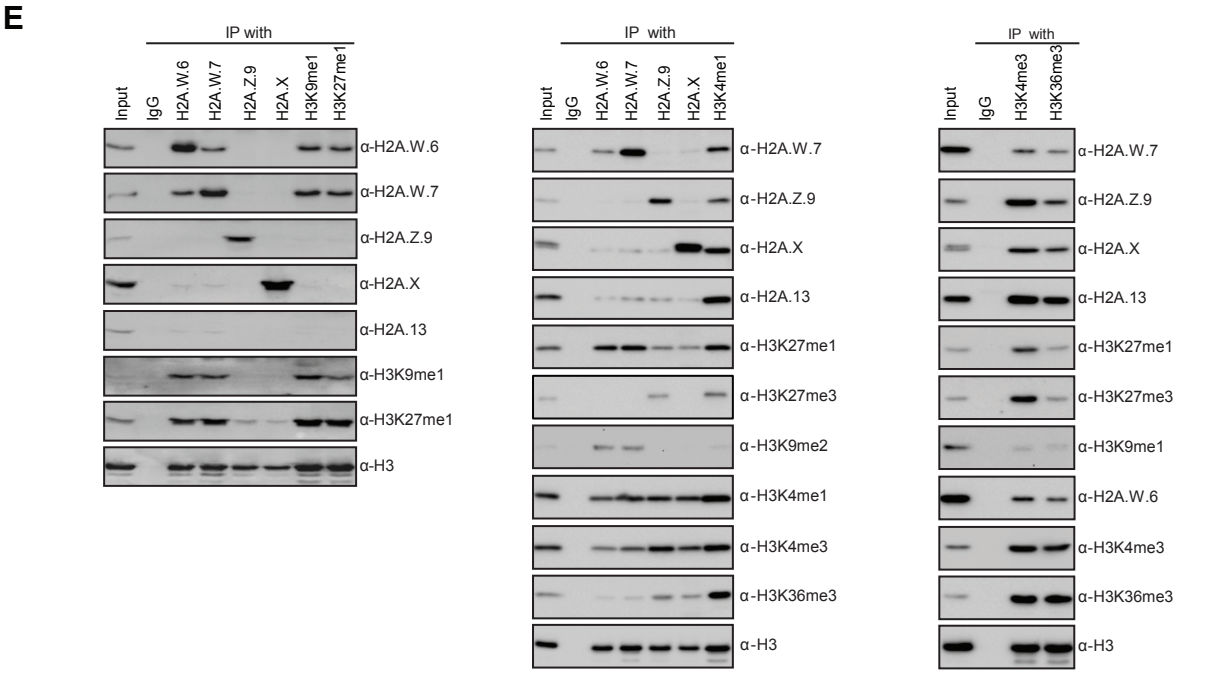


Figure 1: Biochemical analysis of the association between histone variants and histone marks.

(A) Histone H3.1 and H3.3 form homotypic and heterotypic nucleosomes. Spectral counts of H3.1- and H3.3-specific peptides in the respective immunoprecipitates (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 were immunoprecipitated with HA agarose and analyzed for the presence of H2A variants by immunoblotting. **(C)** Histone H3 marks are present on both H3.1 and H3.3. HA-tagged H3.1 and H3.3 were immunoprecipitated with HA agarose and analyzed for the presence of H3 marks by immunoblotting. Arrows indicate transgenic (T) and endogenous (E) H3. **(D)** MS analysis of cumulative H3.1 and H3.3 K27 and K36 modifications in H3.1 and H3.3 immunoprecipitates (left panel). Middle and right panels represent H3.1 and H3.3 K27 and K36 modifications, respectively, as analyzed separately in H3.1 and H3.3 immunoprecipitates. **(E)** 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 immunoblotting.

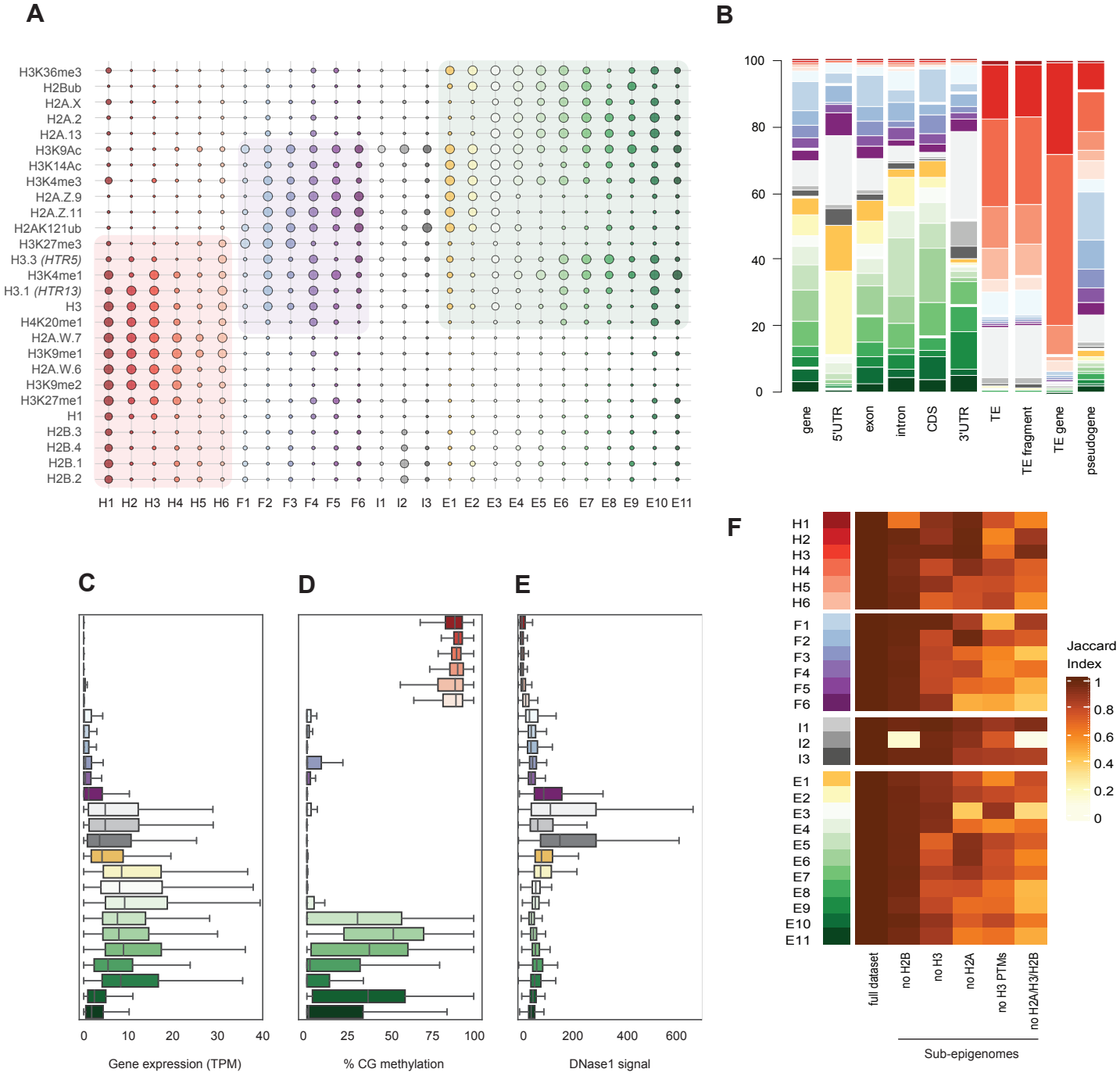


Figure 2: Histone variants define chromatin states in Arabidopsis thaliana.

(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) **(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. The X-axis from left to right (full dataset, no H2B variants, no H3 variants, no H2A variants, no histone modifications, and no histone variants (H2A/H3/H2B)).

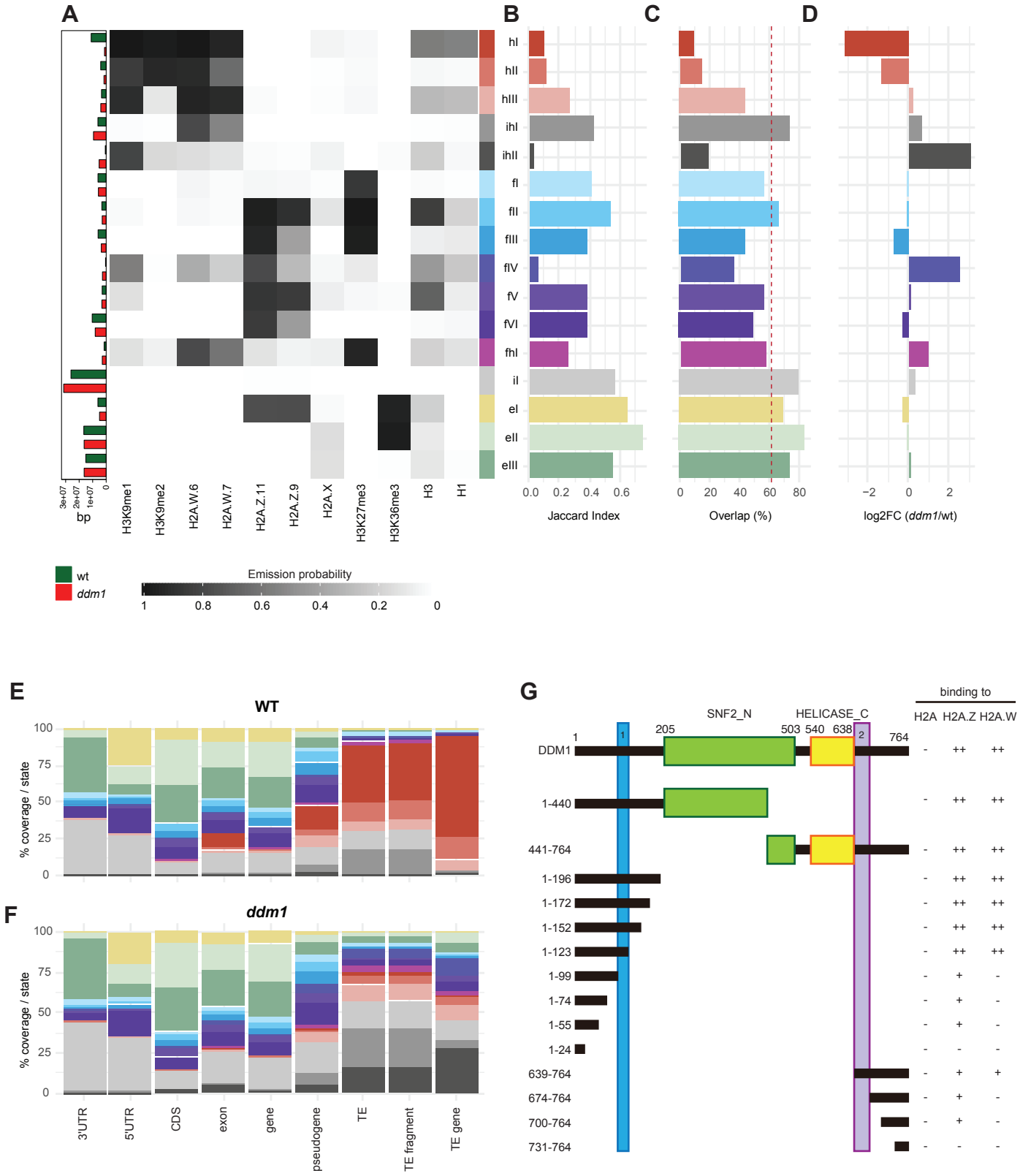


Figure 3: DDM1 loss of function disrupts chromatin states in Arabidopsis thaliana.

(A) Heatmap showing the emission probability for each mark/variant across the 16 chromatin states of the concatenated wild type and ddm1 mutant model. 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 dotted red vertical line represents the genome wide overlap (62.2%). **(D)** Bar plot showing the log2 fold changes of proportion of genome covered by each state across the ddm1 genome compared to 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. **(G)** Summary of the pull-down assay to identify binding regions in DDM1 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 gel pictures are shown in Figure 3 figure supplements 3G and 3H.

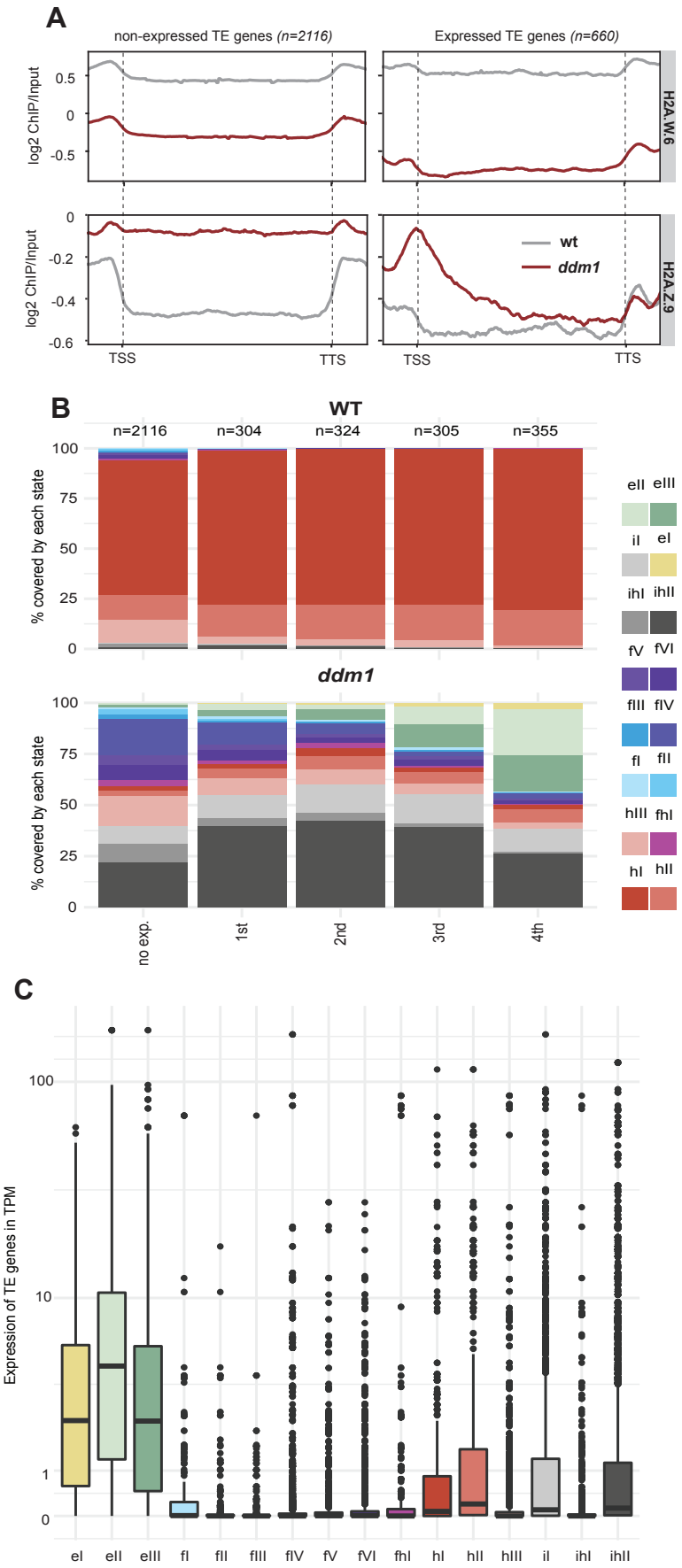


Figure 4: Impact of expression on chromatin states over TE genes in *ddm1*

(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. (The groups are: no expression and expression divided into four quartiles where the 4th is the group with the highest expression.) Expressed TEs belong to the 3rd and 4th quartiles. n represents the number of TE genes in each group. **(B-C)** Stacked bar plots of the proportion of states in wild type **(B)** and in ddm1 **(C)** overlapping TE genes grouped by expression in ddm1. **(D)** Box plot showing the expression of TE genes overlapping the 16 concatenated model states in ddm1.