



universität
wien

DISSERTATION / DOCTORAL THESIS

Titel der Disseratation / Title of the Doctoral Thesis

„Histone H2A variants organize the chromatin landscape in
Arabidopsis thaliana“

verfasst von / submitted by

Bhagyshree Jamge

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

Doctor of Philosophy (PhD)

Wien, 2022 / Vienna, 2022

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

UA 794 620 490

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Molecular Biosciences

Betreut von / Supervisor:

Dr. Frédéric Berger PhD

Let everything happen to you: Beauty and Terror. Just keep going: no feeling is final.
Lass dir Alles geschehn: Schönheit und Schrecken. Man muss nur gehn: Kein Gefühl ist
das fernste.

Rainer Maria Rilke

Acknowledgements

Writing this final section of my thesis, I look back at the most exciting and the most formative years in my scientific career. I have nothing but deep gratitude to all the amazing people I met along my academic journey. I could not have made this journey on my own. Bitte entschuldigen Sie mich, Because it's a long one.

First, I would like to thank my supervisor, **Fred**. This has been an incredible journey. I would have not accomplished this without your constant unwavering support. You have taught me to be critical and rigorous in my research. Every meeting we had was inspiring with stimulating discussions and “Big picture” ideas. Thank-you for motivating me, even during the toughest times. Even while writing this thesis. I am truly indebted to you for your guidance and encouragement. Thank you from the bottom of my heart.

I'm extremely grateful to **Elin**, you have been my day-to-day mentor and go-to Bio-informatics “guru”. I admire your attention to detail. I learnt to put asterisks on my plots and it's not always about the p-values. **Akihisa Osakabe** San, I remember you purifying thousands of litres of culture for tiny amounts of protein in our project. This definitely made a huge contribution in the second part of my thesis. I admire your work ethics. It was an honour to work with you. **Ramesh**, thank you for training me during my initial years as a PhD student. The experimental skills I learnt from you laid the foundation for scientific career and projects. **Zdravko**, you are the third reviewer of the lab and we need you. I have always enjoyed our scientific discussions. Thank you for being patient and not giving up on that “States paper”. We still have a long journey to go.

Zdravko and **Sveta**, you literally run the lab! Thank-You for keeping things intact (Just like a nucleosome!).

Acknowledgements

Shout out to my fellow PhD mates, it has been fun to organise PhD Symposiums, lab beer hours and lab retreats with you both! **Sean**, thank you for sitting in the first row in the IMP lecture hall while I gave my Monday seminars. Thank you for all the constructive feedback on my research and literally proofreading every single manuscript in the lab. The lab is going to miss you. **Anna**, we started this journey together and faced our own challenges. It is inspiring to see your growth as a researcher. I enjoyed our time at GRS, Spain and I really wish to go to that GRC with you in the future!

Big thanks to current and past members of **Berger Office team** (Pierre,Tetsuya,Arie). The discussions that we have in our office are about anything and everything. Thanks for keeping the environment stimulating as well as fun. And **Mike** we still miss you. I would like to extend my thanks to all the current and past members of the **Berger Lab**.

Special thanks to **Ortrun**, it was fun to volunteer with you at multiple events such as Long night of research. We share a special bond because of Suraj.

I am also grateful to have excellent collaborators on multiple projects **Chang Liu** and **Facundo Romani**. It was a pleasure working with you. I would also like to extend my thanks to my **thesis committee**. **Alex Stark** and **Michael Nodine** for their insightful suggestions throughout the years. The **VBC PhD Programme** (Ines, Chris, Eva, Chiara) and all the facilities and people for creating a great research environment on campus.

Like any journey, this PhD also had it's roadblocks scientifically and emotionally. It would not have been possible without the support of my friends and family. Here I would like to take the opportunity to thank them.

Amrita, thank-you for those million calls and checking up on me. I really needed that, especially when I was in the final phase of my PhD. **Keerthi**, you taught me to be kind to myself. **Appu**, thanks for watching out for me and for those "Bollywood Dance classes" it helped me take my mind off things. **KJ**, thank you for a pep talk whenever I needed one. **Priya**, thank you for making all those homely dishes whenever I missed India. Thank you **Prabha** for your 5K runs and the "Biryani" sessions. Thank-You for all the love *My gal gang!*

When things did not work in the lab, it was great to cool it off with you guys. **Rev's** radioactive pasta to the rescue! **Roshan's** intense philosophical discussions and fun dance moves. **Sumanta**, thank you for being a sport and being on my team during those badminton games and helping me improve my game. **Chai**, thank-you for the soulful "sitar" melodies. Also a big shout out to you for formatting the "Marathi" acknowledgements last minute.

Thanks to **Raju**, **Prajakta** and VENT for "Bollywood Dance parties". All of you have left Vienna and I will be leaving too. Time Flies, but you made my PhD journey a fun one!

Suraj Bhaiya, thank you for introducing me to the VBC community. You set the stone rolling. You were my inspiration and constant motivation throughout my academic journey. In the family, you were the one who understood what it takes to be a PhD student. **Saily**, thank-you for making me try new new things and all those frequent visits you made to come see me. You and bhaiya always make me feel at home! Our small family in Europe.

Rahul, You have been with me on this journey every step of the way since the beginning. All the ups and downs you have seen it all. Thank-you for all your valuable feedback on my work and editing and formatting this thesis. For the gazillion coffees / teas we had discussing science, dancing on random songs and laughing over those insta-reels. Without your constant support this would have not been possible. Thanks is an understatement. You will be defending soon. And I will be there for you!.

Acknowledgements

Finally, I would like to say a few words thanking my family in "Marathi".

पप्पा,

मला पूर्ण जाणीव आहे कि, माझा हा पीएच डी चा प्रवास फक्त गेल्या ४-५ वर्षांचा नसून, तो त्या दिवशी चालू झाला कि ज्या दिवशी आपलं गाव सोडून ४०० कि.मी. चा प्रवास करून माझं नाव तुम्ही किमीन्स हायस्कुल मध्ये घातलं. तुमचं बोट धरून चालू केलेला माझा शैक्षणिक प्रवास आज या टप्प्यावर येऊन पोहोचलाय. "मुलींनी एवढं शिकून काय करायचंय? लग्न करून द्या म्हणजे झालं !" अश्या पद्धतीची विचारसरणी असलेल्या लोकांना न जुमानता तुम्ही माझ्या पाठीमागे कायमच समर्थपणे उभे राहिलात. तुमच्यासारख्या काळाच्या पुढ्या विचार करणाऱ्या व्यक्तीची मी मुलगी आहे याचा मला प्रचंड अभिमान आहे. मला आशा आहे कि तुम्ही दिलेली मूल्य घेऊन तुम्हाला अभिमान वाटेल असं आयुष्य मी जगेन.

मम्मी,

या सगळ्या प्रवासात माझा एक हात पप्पांनी धरला होता आणि दुसरा तू! तू उपसलेल्या कष्टांमुळेच आम्ही सगळी भावंडं आज या टप्प्यावर येऊन पोहोचलो आहोत. तू केलेल्या त्यागाला आणि कष्टांना तुलना नाही. या वाटेवरच्या प्रत्येक वळणावर तुझं मार्गदर्शन मिळालं नसतं तर आज आम्ही इथपर्यंत आलोच नसतो. मी तुला हे फारसं कधी बोलून दाखवत नाही, पण माझी खरी आदर्श तू आहेस. पाठीवरील तुझा हात आणि तुझी शिकवण हि शिंदोरी मला आयुष्यातील अडचणींना समर्थपणे झेलत पुढे जाण्यास पुरेशी आहे हा विश्वास मला आहे.

आई,

केवळ मला आवडतो म्हणून आजारी असतानादेखील धडपडत तो पदार्थ करून मला वाढण्याची एवढी ऊर्जा तुझ्यात येते तरी कुठून हा प्रश्न मला कायमच पडतो. तुझं निर्मळ प्रेम आणि भक्कम पाठिंबा आजपर्यंत आहे तसाच यापुढेही मिळत राहिल याबद्दल माझी पूर्ण खात्री आहे.

राजू भैया,

आम्ही इथे इतक्या दूर असताना आई आणि मम्मी-पप्पांची काळजी घ्यायला त्यांच्या जवळ तू आहेस हि गोष्ट खूप मोठी आहे. त्याच आधारावर आम्ही इथे निश्चिन्तपणे राहून काम करू शकतो. तुझे हे ऋण केवढे मोठे आहेत याची तुला कल्पनादेखील नसेल.

मंजू मावशी आणि काका,

या सगळ्या प्रवासातील दोन महत्वाच्या गोष्टी तुम्ही मला शिकवल्या. एक म्हणजे शिस्त. त्या शिस्तीशिवाय इथे पोहोचणं कठीण होतं. आणि दुसरी त्याहून महत्वाची गोष्ट म्हणजे पोळ्या. त्या कश्या करायच्या हे तुम्ही शिकवलं नसतं तर नुसता ब्रेड खाऊन इतकी वर्षे बाहेर राहणं अशक्य झालं असतं. श्रीनाथ आणि पार्थ, आपण एकत्र घालवलेल्या वर्षांच्या स्मृती कायमच मनाला आनंद देतात.

शेवटी, आजी-आजोबा, मामी-मामा , काकी-काका, मावशी-काका आणि सगळी भावंडं - तुमचे आशीर्वाद, प्रेम आणि शुभेच्छा पाठीशी आहेत म्हणूनच आतापर्यंतचा हा प्रवास इतका सुंदर झाला याची जाणीव मला आहे.

ऋणाईत

भाग्यश्री बाबासाहेब जामगे

Abstract

My research work focuses on understanding the contribution of Histone Variants in defining the chromatin state landscape in the model plant *Arabidopsis thaliana*. My results demonstrate that Histone H2A variants play a key role in dividing the *Arabidopsis* genome into distinct chromatin domains.

In my work, I characterised the landscapes of chromatin states in *Arabidopsis thaliana*. This work identified how histone variants are essential in defining the various chromatin states in *Arabidopsis*. We show that subtypes of H2A.W define distinct subtypes of constitutive heterochromatin. Surprisingly, distinct chromatin states are associated with specific transposon families and a novel class of genes expressed in gametes. H2A.Z is the histone variant associated with all developmentally repressed genes and it is not necessarily associated with H3K27me3. On expressed genes the chromatin states associate in a specific manner to form distinct chromatin landscapes. These landscapes define the mode of activity of RNA Pol II. Overall this extensive analysis introduces the concept of chromatin landscape as a stereotypical assembly of chromatin states and supports that chromatin landscapes instruct transcriptional regulation.

In a parallel project, I was interested to understand how chromatin remodelers maintain or influence the chromatin landscapes. Here, we uncovered the mechanism of action of the chromatin remodeler Decrease in DNA methylation 1 (DDM1). DDM1 is a SWI/SNF2 chromatin remodeler that was previously linked with loss of DNA methylation, and loss of H3K9me2 but the direct action of DDM1 was unknown. Through this work, we could show that DDM1 binds to Histone Variants H2A.W and deposits H2A.W at certain transposon loci to maintain transposons in a silenced chromatin state.

Zusammenfassung

Das ist eine deutsche Kurzfassung meiner in Englisch verfassten Masterarbeit.

Meine Forschungsarbeit konzentriert sich auf das Verständnis des Beitrags von Histon-Varianten bei der Definition der Chromatinzustandslandschaft in der Modellpflanze *Arabidopsis thaliana*. Meine Ergebnisse zeigen, dass Histon-H2A-Varianten eine Schlüsselrolle bei der Unterteilung des *Arabidopsis*-Genoms in verschiedene Chromatin-Domänen spielen.

In meiner Arbeit habe ich die Landschaften der Chromatinzustände in *Arabidopsis thaliana* charakterisiert. Dabei konnte ich feststellen, dass Histonvarianten bei der Definition der verschiedenen Chromatinzustände in *Arabidopsis* eine wesentliche Rolle spielen. Wir zeigen, dass Subtypen von H2A.W verschiedene Subtypen von konstitutivem Heterochromatin definieren. Überraschenderweise sind die verschiedenen Chromatinzustände mit bestimmten Transposon-Familien und einer neuen spezifischen Klasse von Genen verbunden, die in den Gameten exprimiert werden. H2A.Z ist die Histonvariante, die mit allen entwicklungsunterdrückten Genen assoziiert ist, und sie ist nicht unbedingt mit H3K27me3 verbunden. Bei exprimierten Genen assoziieren die Chromatinzustände auf spezifische Weise und bilden unterschiedliche Chromatinlandschaften. Diese Landschaften definieren die Art der Aktivität von RNA Pol II. Insgesamt führt diese umfassende Analyse das Konzept der Chromatinlandschaft als stereotype Anordnung von Chromatinzuständen ein und unterstützt, dass Chromatinlandschaften die Transkriptionsregulation steuern.

In einem parallelen Projekt war ich daran interessiert zu verstehen, wie Chromatin-Remodelers die Chromatinlandschaften aufrechterhalten oder beeinflussen. Hier haben wir den Wirkmechanismus des Chromatin-Remodelers Decrease in DNA methylation 1

Zusammenfassung

(DDM1) aufgedeckt. DDM1 ist ein SWI/SNF2-Chromatin-Remodeler, der zuvor mit dem Verlust von DNA-Methylierung und H3K9me2 in Verbindung gebracht wurde, dessen direkte Wirkung jedoch unbekannt war. Durch diese Arbeit konnten wir zeigen, dass DDM1 an Histonvarianten H2A.W bindet und H2A.W an bestimmten Transposon-Loci ablagert, um Transposons in einem silenced chromatin state zu halten.

Contents

Acknowledgements	iii
Abstract	vii
Zusammenfassung	ix
1 Introduction	1
1.1 General organisation of chromatin in eukaryotes	1
1.1.1 Nucleosomes – The Eukaryotic Wrappers in the house	1
1.1.2 Different “Scales” of chromatin organisation	1
1.1.3 Nucleosomes are dynamic units of the chromatin	2
1.2 Modifications of chromatin composition	4
1.2.1 Post-translational modifications of the histone tails	4
1.2.2 Isoforms of the core histones	8
1.3 Chromatin states	14
1.3.1 Transcription and chromatin	16
1.4 Plants as models for chromatin and epigenetic studies	17
1.4.1 <i>Arabidopsis thaliana</i> sets the perfect stage to study chromatin regulation in eukaryotes	18
2 Aims	19
2.1 Uncover the importance of Histone variants in shaping chromatin state landscape in <i>Arabidopsis</i> seedlings	19

Contents

2.2	Understand how chromatin remodelers maintain transposable elements in repressed chromatin states	20
3	Publications	21
3.1	List of Publications	21
3.2	Publication I	22
3.3	Publication II	77
4	Discussion	109
4.1	DDM1 deposits H2A.W to silence transposons	109
4.2	DDM1-H2A.W mediated silencing and other players	110
4.3	H2A variants are the major drivers in defining chromatin landscapes . . .	112
4.4	The new concept of chromatin landscape as assembly of chromatin states	113
4.5	The impact of chromatin landscapes on transcriptional activity	114
5	Epilogue	117
	Bibliography	119

1 Introduction

1.1 General organisation of chromatin in eukaryotes

1.1.1 Nucleosomes – The Eukaryotic Wrappers in the house

DNA encodes all the necessary information to “perform” the regulatory functions needed throughout the life cycle of eukaryotes and prokaryotes. But this information needs to be protected against damage, packaged in a compact space, and its access should be regulated. Eukaryotes evolved an efficient packaging/wrapping system of genetic information called nucleosomes. A nucleosome is composed of approximately 147 bp of DNA wrapped around an octamer of core histone proteins. Each octamer is composed of two H2A/H2B and two H3/H4 dimers [Luger et al., 1997]. Nucleosomes are separated by a linker DNA which varies from 20-80 bps long and is bound by the linker histone H1. This DNA-protein complex forms the 10 nm diameter fiber famously identified as “beads on a string” [Kornberg and Klug, 1981]. This basic packaging unit of chromatin is conserved across all eukaryotes. Therefore, nucleosomes enable eukaryotes to compact the genome inside the nucleus, while still enabling access to the genetic information during important processes such as transcription, DNA replication, and DNA repair.

1.1.2 Different “Scales” of chromatin organisation

History of chromatin

In the 1880s, long before the structure of nucleosomes was discovered, W. Flemming coined the term “chromatin”- a nuclear substance that stained strongly under a light microscope [Flemming, 1882, Wolffe and Hayes, 1999]. With the advancement in the field

1 Introduction

of cytogenetics and microscopy techniques Emil Heitz developed an inexpensive cytological method “kochmethode” boiling method to obtain chromosome spreads from moss and eventually applied it to a wide range of species (more than 115 plant species and several *Drosophila* and other Diptera) [Heitz, 1929]. This led him to make early observations of distinct cytological features of chromosomes. He interpreted them as euchromatin; which was lightly stained and heterochromatin, which was densely stained. Heitz also studied the changes in these cytological structures across the cell cycle stages and also distinguished the two types of chromatin during interphase. These broad classifications are still being used: i) Euchromatin which is loosely packed and promotes transcription, and ii) heterochromatin is densely packaged and restricts transcription.

Depending on the cell cycle stage the chromatin is also compacted at different scales. During mitosis, chromatin reaches its highest level of compaction, it is assembled as chromosomes replicated and then passed on to the new daughter nuclei. Conversely during interphase, chromatin is much less condensed and each chromosome occupies a spatially limited territory distributed around the nucleus [Stack et al., 1977].

Genome size varies hugely across eukaryotic species ranging from < 10 Mb to >100 Gb. Nevertheless, species with genome size over 100 Mb contain circa 25,000 protein-coding genes [Bowman and Poirier, 2015]. The difference in genome size is accounted for by increased number of regulatory regions, larger introns and most importantly by non coding sequences including repeats and transposons. Interestingly, the size of the nucleus does not vary with genome size but rather with different cell types in an individual.

1.1.3 Nucleosomes are dynamic units of the chromatin

Nucleosome assembly results from the deposition of pre-assembled tetramer of two H3 and two H4 followed by deposition of two heterodimers of H2A and H2B. Core histones contain a high number of positively charged residues that establish bonds with the negatively charged DNA. The nucleosome is further stabilised by non-covalent bonds between the core histones. Through their intimate DNA-protein interaction, nucleosomes limit access to the genome by other DNA binding proteins. Crystal structures have identified the most

1.1 General organisation of chromatin in eukaryotes

likely nucleosomal state as the fully wrapped nucleosome. When only a few DNA-histone bonds are dissociated, the nucleosomes acquire an unwrapped state. Previous studies have also shown nucleosomes are capable of spontaneous unwrapping and rewinding [Forties et al., 2011, Chereji and Morozov, 2014]. This dynamic nature of nucleosomes allows them to fine-tune the accessibility of the DNA to other chromatin interacting proteins [Struhl and Segal, 2013, Kornberg and Lorch, 2020].

Studies in multiple organisms have identified regions in the genome that are occupied by well-positioned nucleosomes and form phased nucleosome arrays [Baldi et al., 2018, Zhou et al., 2015a, Bernstein et al., 2004]. This nucleosome occupancy and positioning serve a critical biological function in protecting the DNA from endonucleases and serve as the initial barrier against DNA damage. This extends the primary role of nucleosomes from packaging the genome in a limited space during the interphase to the protection of the genome stability. This packaging constitutes a challenge for DNA replication and transcription that require a dynamic dissociation of the nucleosome to progress through the genomic sequences [Hartzog et al., 2002]. While the mechanisms that dissociate nucleosomes in front of the replication fork remain unknown, their re-assembly after the fork has been prescribed in great detail in the past years. We are beginning to understand how the transcriptional machinery goes through the nucleosome during elongation [Lai and Pugh, 2017, Venkatesh and Workman, 2015, Li et al., 2007, Hartzog et al., 2002]. Finally, how transcription is initiated relies on the action of specific transcription factors (TF). Pioneer TFs can bind to nucleosomal DNA and unwind the DNA to enable the assembly of the preinitiation complex (PIC) with RNA polymerase II (RNA Pol II) [Struhl and Segal, 2013, Ehara et al., 2019]. Therefore suggesting nucleosomes themselves can affect DNA - templated process.

However, variation in the composition of nucleosomes adds much complexity to the dynamic regulation of nucleosome activity. Core histones are subject to numerous and complex combinations of post-transcriptional modifications that affect the nucleosome stability [Strahl and Allis, 2000]. In addition, there are numerous isoforms of the core histones that may combine to produce diverse combinations of octamers

1 Introduction

[Marzluff, 2005, Lyons et al., 2016]. These two additional factors contribute to the diversity of the chromatin and its regulation as described below.

1.2 Modifications of chromatin composition

1.2.1 Post-translational modifications of the histone tails

Post-translational modifications (PTMs) occur on N- and C- terminal tails of core histones after their assembly into nucleosomes [Strahl and Allis, 2000]. Histone tails have three main properties. First, these tails are unstructured. Second, they protrude away from nucleosomes. Third, histone tails are enriched for basic amino acids such as Lysine (K), Arginine (R) and Serine (S) that undergo modifications. PTMs on histone tails occur during diverse regulatory processes such as cell cycle checkpoints, heterochromatic formation, centromere assembly, DNA replication, DNA repair and gene transcription [Strahl and Allis, 2000]. Plethora of PTMs have been identified using mass spectrometry of core histones. The most common modifications on histones are methylation, phosphorylation, acetylation, ubiquitination and SUMOylation. Most of these PTMs are conserved across eukaryotes but some are restricted to specific groups and are explained below. Mass spectrometry showed that some PTMs are exclusive of others, but these studies have a limited scope due to the short length of the peptides utilised in such analysis. Whereas, identification of antibodies specific for each PTM and genome-wide-mapping methods like ChIP-seq have allowed us to map the precise location of these PTMs in the genome [Barski et al., 2007, Johnson et al., 2007, Robertson et al., 2007]. Hence, we started documenting the genomic location for various histone PTMs across tissue types in different species [Feingold and Pachter, 2004, ENCODE Project Consortium, 2011, Park, 2009, Gerstein et al., 2010].

PTMs are enriched in specific genomic features and can be either positively or negatively correlated with the transcriptional status of the underlying gene. We still do not know the cause of these associations for most of the PTMs. Here, we will describe the most studied histone modifications and their implications in either promoting or inhibiting

transcription separately.

Post-transcriptional modifications of the histone tails that accompany transcription

Histone acetylation

Histone acetylation is linked to regulate transcription [Rothbart and Strahl, 2014, Zhou et al., 2011, Pokholok et al., 2005]. Acetylation occurs on the lysine residues, and this process neutralises the positive charge on the lysine residue resulting in reduced interaction with negatively charged DNA. Hence, acetylation promotes nucleosome dissociation, which is required during transcription initiation [Allfrey et al., 1964, Turner, 2000]. Consequently, histone acetylation (H3K9Ac and H3K14Ac) are usually enriched at promoters, enhancers and linked with accessible regions of the chromatin. Histone acetylation is tightly regulated by the action of histone acetyltransferases (HAT) and histone deacetylases (HDAC) [Turner, 2000]. The acetylation/deacetylation of histones is a very dynamic process and makes it a good candidate to regulate gene expression.

H3K4 methylation

Addition of 3-methyl groups on the lysine 4 on histone H3 protein (H3K4me3) is a PTM, associated with transcribed genes [Shilatifard, 2012, Dou et al., 2006, Pokholok et al., 2005]. This PTM is deposited by Trithorax Group proteins and marks transcription start site (TSS) and promoters of actively transcribed genes [Petruk et al., 2001]. H3K4me3 is capable of recruiting the transcriptional initiation factor TFIID subunit 4 and hence facilitates transcription [Lauberth et al., 2013]. However, studies in other model systems indicate that H3K4me3 is not required for transcription [Pavri et al., 2006]. So even though the H3K4me3 is conserved across species, the precise involvement of H3K4me3 in transcription still remains to be clarified.

In mammals, H3K4me1 marks tissue-specific active enhancers together with H3K27Ac [Herz et al., 2012, Creighton et al., 2010]. In the compact genomes like *Arabidopsis* — where very few enhancer elements have been identified — H3K4me1 is present over gene bodies [Zhang et al., 2009, Weber et al., 2016]. It is associated positively with

1 Introduction

transcriptional activity. However H3K4me1 is both the product of demethylation of H3K4me3 and monomethylation of H3K4 and its role in the context of regulatory elements in *Arabidopsis* remains to be investigated.

H3K36 methylation

H3K36me3 is associated with actively transcribed regions [Bannister et al., 2005]. The Ser2 phosphorylated C terminal domain (CTD) of the elongating form of RNA polymerase II recruits the H3K36 methyltransferase that deposits H3K36me3 [Kizer et al., 2005]. Increased transcription of genes can lead to hyperacetylation of gene bodies, allowing cryptic transcription. H3K36me3 serves as the docking site for histone deacetylases like Rpd3S, to reestablish the repressive chromatin after RNA pol-II has passed through the gene body [Li et al., 2009]. Therefore, inhibiting cryptic transcription [Li et al., 2009, Wagner and Carpenter, 2012]. H3K36me3 regulates splicing [Kim et al., 2011] and inhibits PRC2 (Polycomb Repressive Complex 2) activity in vitro [Schmitges et al., 2011]. Hence H3K36me3 plays an important role during transcription elongation. In addition, demarcates active chromatin from repressive chromatin.

Post-transcriptional modifications of the histone tails that inhibit transcription

PTMs in facultative heterochromatin

H3K27me3 and H2A ubiquitination of lysine 121 (H2Aub) are associated with facultative heterochromatin; enriched over protein-coding genes that are developmentally regulated. H3K27me3 and H2AK121Aub are deposited by Polycomb Repressive Complexes 2 and 1 (PRC2 and PRC1), respectively [Schuettengruber and Cavalli, 2009, Di Croce and Helin, 2013, Simon and Kingston, 2013]. H3K27me3 is crucial for silencing developmental genes. In contrast, how H2Aub actually represses genes is yet to be understood. There have been studies reporting the role of PRC1 in chromatin repression independent of the H2Aub [Pengelly et al., 2015]. PRC1 is also suggested to promote liquid-liquid phase separation although the definite proof of this form of regulation in vivo is still missing [Plys et al., 2019, Tatavosian et al., 2019]. Although there is a degree of

1.2 Modifications of chromatin composition

overlap of H2Aub and H3K27me3 recent studies have identified the role of H2Aub independent of H3K27me3 in gene repression. And finally, H2Aub is not only associated with repression but also found on genes that are actively transcribed. Hence the non-obligate relationship between PRC1 and PRC2 marks suggest there exists complex chromatin landscapes and these marks cannot be viewed in isolation.

PTMs associated with constitutive heterochromatin

Unlike facultative heterochromatin, constitutive repressed chromatin covers transposons that remain silenced irrespective of the developmental stage. H3K9 methylation is found in constitutive heterochromatin in eukaryotes. In yeast and animals H3K9me3 is bound by heterochromatic protein HP1 which promotes compaction of heterochromatin and could phase separate by recruiting other heterochromatin interacting proteins. Removal of H3K9me3 also leads to derepression of transposons. In plants, H3K9me2 does not bind to HP1, but rather recruits the plant-specific DNA methyltransferases CHROMO-METHYLASE (CMT) 2 and 3 [Du et al., 2012]. CMT2/3 deposits DNA methylation specifically on cytosines in CHG context [Hume Stroud et al., 2014]. Presence of CHG methylation recruits SER domain proteins, which in turn deposits H3K9me. This process maintains heterochromatin, forming a feed-back loop with histones and DNA methylation.

Apart from H3K9me, there are PTMs such as H3K27me1, H4K20me1 and H3K23me3 associated with heterochromatic regions such as repeats and transposable elements. In flowering plants, H3K27me1 is deposited by specific histone methyltransferases ATXR5 and 6. We do not completely understand the mechanisms depositing H4K20me1 and H3K23me3 in constitutive heterochromatin. Furthermore, we do not understand the link between DNA methylation and these PTMs. Given the variety of repressive marks associated with repeats and transposons, we can hypothesise that there are specific PTMs associated with distinct transposon families and they could be targeted by distinct silencing mechanisms.

1 Introduction

1.2.2 Isoforms of the core histones

In addition to PTMs, histones in the nucleosome can be replaced by their isoforms. Replication coupled histones, also called “Canonical” histones are mainly expressed during the S-phase of the cell cycle. These canonical histones are incorporated into nucleosomes during DNA replication. In animals, the canonical histones are encoded by multiple gene copies in the genome. Genes which encode the canonical histone lack introns and 3’ polyadenylation tails (polyA) [Marzluff, 2005, Lyons et al., 2016].

In contrast to the replication-coupled histones, histone variants are non-allelic isoforms, encoded by single-copy genes that have both introns and polyA tails. They can be expressed throughout the cell cycle stage and are incorporated into the chromatin independent of the cell cycle. Some histone variants are expressed in specialised cell types, whereas others are uniformly expressed across tissue types. Histone variants have been shown to occupy certain genomic regions, suggesting the role of specific chaperones or chromatin remodelers in their deposition. Extensive studies across various organisms have identified the evolution of many histone variants in fungi, plants, and metazoans. Histone variants are usually characterised by peculiar amino acid differences in their tails and the linked regions between the very conserved alpha helices [Talbert and Henikoff, 2010, Talbert and Henikoff, 2014, Talbert and Henikoff, 2017, Talbert and Henikoff, 2021]. In the following sub-section, we will explain the histone variants present across multiple species with a focus on plants.

H3 Variants

Multicellular eukaryotes have evolved a major replication independent variant H3.3, which is distinguished from the replication-coupled H3.1 only by a few amino acid differences. H3.3 is important for development and can be deposited throughout the cell cycle and in non-dividing cells. H3.3 variants are enriched over gene bodies and have been positively correlated with their transcription in plants [Mito et al., 2005, Mito et al., 2007, Wollmann et al., 2012, Wirbelauer et al., 2005] and mammals. In animals H3.3 is also deposited to silent loci such as telomeres and H3.3 is a crucial factor in setting up correct

1.2 Modifications of chromatin composition

heterochromatin boundaries in mouse embryos. Hence there is evidence suggesting the dual role of H3.3 in transcription and as well as establishing heterochromatin.

H3.3 and H3.1 have evolved independently across plants, animals, and fungi. Similar to metazoans, Arabidopsis H3.3 and H3.1 differ by only four amino acid residues. Arabidopsis H3.3 is enriched towards 3'end of gene bodies which is also associated with CG methylation of gene bodies. This enrichment of H3.3 is positively correlated with the level of transcription [Wollmann et al., 2012, Wirbelauer et al., 2005]. It is unclear why H3.3 is associated with transcription. Unlike H3.3, H3.1 covers the gene bodies evenly and negatively correlated with gene expression. The association of H3.1 and repression of transposons is explained in part with the fact that H3.1 is the unique substrate for the deposition of H3K27me1 by ATXR5/6.

In all eukaryotes with point centromeres, CenH3 (CENPA) is the H3 variant that is found in centromeric nucleosomes. CenH3 is important for kinetochore assembly and as a consequence crucial for chromosome segregation and progression through mitosis [De Wulf and Earnshaw, 2009]. CenH3 only shares 50-60% identity with H3.1 or H3.3 [Dawe and Henikoff, 2006, Henikoff et al., 2001, Malik and Henikoff, 2003]. CenH3 has an extended L1-loop and a rapidly evolving N-terminal tail, which has played a key role in speciation [Dawe and Henikoff, 2006, Henikoff et al., 2001].

Furthermore, both mammals and flowering plants evolved specialised H3 variants that are expressed in reproductive cell types. In Arabidopsis, H3.10 is specifically expressed in sperm and cannot be modified by PRC2 [Borg et al., 2020]. Hence, H3.10 plays an important role in removing H3K27me3 from the paternal genome. Sperm-specific H3 variants are also found in mammals, that alter nucleosome properties and are fundamental in sperm maturation [Borg et al., 2020]. The similarities between histone variants and their association with certain genome function is conserved across phyla, suggesting a convergent evolution of H3 variants and thereby to achieve similar forms of transcriptional regulation.

Deposition of H3 variants

H3 chaperones assemble nucleosomes in either replication dependent (RD) or independ-

1 Introduction

ent (RI) manner [Franklin and Zweidler, 1977, Hake and Allis, 2006, Ahmad and Henikoff, 2002, Urban and Zweidler, 1983]. CAF1 (Chromatin Assembly Factor1) loads newly synthesised H3.1/H3.2 : H4 dimers on DNA in RD manner [Natsume et al., 2007]. Whereas , HIRA (histone cell cycle regulator) is H3.3 specific chaperone deposits newly synthesised H3.3 in RI manner [Tang et al., 2006, English et al., 2006]. HIRA was initially linked with H3.3 deposition at active sites of transcription. DAXX (death domain associated protein) , H3.3 specific chaperone was first identified in mouse embryonic stem cells (mESCs)[Elsässer et al., 2015]. DAXX together with ATRX deposit H3.3 at repressive loci such as repeats,telomeres and pericentromeric DNA [Voon et al., 2015, Goldberg et al., 2010, Elsässer et al., 2015]. In Arabidopsis, all homologues for H3 chaperones have been identified except for DAXX [Elsässer et al., 2012, Lewis et al., 2010]. Recent studies in Arabidopsis have also linked loss of CAF1 with loss of H3.1, H2A.W and H3K9me2 [Benoit et al., 2019]. Suggesting a potential link between H3 and H2A depositing mechanism in a replication dependent manner. Specific chaperones for H2A variants are yet to be discovered.

H2A Variants

The H2A family contains the largest number of variants identified across species. The most common histone variants in all eukaryotes are H2A.Z, H2A.X and the cell cycle dependent H2A.

H2A.Z

H2A.Z is the most widely studied H2A variant. It is characterised by its peculiar docking domain with an extended acidic patch [Ausió and Abbott, 2002, Marzluff et al., 2002, Osakabe et al., 2018]. H2A.Z is essential in flies, frogs, and mice[Faast et al., 2001, ?]. Whereas in Arabidopsis and yeast the absence of H2A.Z causes several growth defects [Jackson and Gorovsky, 2000, Jarillo and Piñeiro, 2015]. Several studies across species show that H2A.Z containing nucleosomes are usually found in the proximity of nucleosome depleted regions (NDR) as well as transcription start sites (TSS) of genes that are highly transcribed. H2A.Z occupies the +1 nucleosome of the TSS and is rather depleted at

1.2 Modifications of chromatin composition

-1 in expressed genes for most eukaryotes. In Arabidopsis, H2A.Z also occupies the entire gene body of stress-responsive genes together with H3K27me3 [Sura et al., 2017, Dai et al., 2017]. The pattern of H2A.Z over genes is anticorrelated with the pattern of CG DNA methylation[Zilberman et al., 2008]. How distinct PTMs associate with H2A.Z nucleosomes in relation with the dual function of H2A.Z remains to be explained.

H2A.X

H2A.X can be distinguished by their C terminal motif SQ[E/D] , where stands for a hydrophobic amino acid. Apart from this motif, canonical H2A and H2A.X are very similar. H2A.X is found in all eukaryotes with the exception of nematodes although the motif is associated together with another type of variant such as H2A.Z (zebrafish, drosophila) or replication-coupled H2A (yeast)[Pinto and Flaus, 2010, Talbert et al., 2002]. In mammals, H2A.X mRNA transcripts produced during the S-phase lack poly-A tails, whereas during the rest of the cell cycle they are polyadenylated suggesting H2A.X can be expressed in a replication coupled as well replication-independent manner [Mannironi et al., 1989]. H2A.X also plays a crucial role during DNA damage response. Upon DNA double-strand break (DSB) H2A.X is phosphorylated at the Serine 139 position resulting in H2A.X [Downs et al., 2007, Li et al., 2010]. H2A.X occupies regions adjacent to DSB. This accumulation of H2A.X results in loosening up of chromatin and enabling access to DNA repair machinery.

H2A.W and macroH2A mark heterochromatin

Plants and animals both have evolved specific H2A variants for transcriptional repression. [Yelagandula et al., 2014, Talbert and Henikoff, 2010, Bourguet et al., 2021]. macroH2A is the largest core histone that evolved in metazoans and has a macro domain 25KDa at its C-terminal [Talbert and Henikoff, 2010]. This macro domain is not unique to histones and is found in different non-histone proteins across various species. The macro domain has been associated with many functions such as DNA repair, gene regulation and chromatin remodelling [Han et al., 2011]. The linker region that separates the core histone fold from the macro-domain is enriched with

1 Introduction

Lysine (K) residues and has been associated with compaction of nucleosome arrays in vitro [Muthurajan et al., 2011]. Land plants also evolved a fourth type of H2A variants which is characterised by a longer C terminal tails enriched for lysine rich. These variants are found both in bryophytes (H2A.M) and in vascular plants (H2A.W) [Kawashima et al., 2015, Yelagandula et al., 2014, Bourguet et al., 2021]. H2A.W variants share a characteristic KSPK motif. Similar motifs are also found in macroH2A as well as linker histone H1. Though macroH2A occupies both constitutive as well as facultative heterochromatin. H2A.W is exclusively enriched at repeats and transposons that are constitutively repressed.

Deposition of H2A variants

In Arabidopsis, through biochemical analysis of mononucleosomes, it was shown that H2A variants associate exclusively to form homotypic nucleosomes [Osakabe et al., 2018]. The exclusive presence of homotypic nucleosomes suggests a dedicated deposition mechanism for each variant at least in Arabidopsis. ChIP-seq has established that H2A variants occupy distinct regions of the Arabidopsis genome suggesting further that certain mechanisms dock the molecular machines that deposit H2A variants to specific regions of the genome[Yelagandula et al., 2014, Lorković and Berger, 2017]. Histone PTMs might provide cues for these suspected docking mechanisms and thus setting up the chromatin landscapes and thereby regulating transcription. Chromatin has to be precisely duplicated during DNA replication, reassembled after replication, and or repackaged after DNA repair. Chromatin remodelling factors play a fundamental role during these processes. These chromatin remodelling factors include Histone chaperones and ATP-dependent chromatin remodelers. Histone chaperones facilitate nucleosome assembly and disassembly. During replication, chaperones like Facilitates Chromatin Transcription (FACT) and Nucleosome assembly protein 1 (Nap1) complex facilitate H2A-H2B deposition irrespective of the H2A variant [D’Arcy et al., 2013, Zhou et al., 2015b]. No chaperones have yet been reported that deposit specific H2A variants at the replication fork. Suggesting, there are undiscovered mechanisms that deposit specific H2A histone variants such as H2A.W,

1.2 Modifications of chromatin composition

macroH2A to certain regions in the genome.

The other factors that can dissociate or relocate nucleosomes are ATP-dependent chromatin remodelers. ATP dependent chromatin remodelers can be subdivided in four groups based on the domains they contain. 1) SWI/SNF (switch/sucrose-non-fermenting) contain bromodomain, 2) ISWI (imitation switch) contain SANT-SLIDE domain, 3) CHD (chromodomain-helicase-DNA binding) contains tandem chromodomains and 4) INO80(inositol requiring 80) family have Helicase SANT domain. These domains are critical for remodeler recruitment and or facilitating interaction with other PTMs or important for regulating the ATPase activity of the remodeler.

SWI2/SNF2 complex was first identified in yeast, SWR1 facilitates the removal of the H2A-H2B dimer and exchanges it for an H2A.Z-H2B dimer[Luk et al., 2010]. INO80 facilitates the exchange of H2A.Z with H2A. Plants also have several such chromatin remodelling complexes and their roles appear to be conserved. Genetic screens in Arabidopsis identified the SWI/SNF chromatin remodeler DDM1 (Decrease in DNA methylation 1) [Kato et al., 2003, Vongs et al., 1993]. Loss of DDM1 accompanies loss of DNA methylation and H3K9me2, transposons are de-repressed, and several genes are misregulated [Kakutani, 1997, Vongs et al., 1993, Kato et al., 2003] . LSH (Lymphoid-Specific Helicase) is the mammalian ortholog of DDM1. LSH is also important for CpG methylation and development [Lyons and Zilberman, 2017]. DDM1 lacks DNA methyltransferase activity. Hence there is a missing link between the function of these chromatin remodelers and must be investigated in the context of heterochromatin variants such as H2A.W and macroH2A.

Other histone variants

H2B variants

Early studies have identified H2B variants in mammals but were only restricted to their function in spermatogenesis. Ten H2Bs were identified in Arabidopsis [Bergmüller et al., 2007], and these variants have specific expression patterns in different reproductive tissues and developmental stages. For example, H2B.S is a novel class of seed specific H2B variants.

1 Introduction

H2B variants are shown to be associated with distinct genomic regions using genome wide profiles in somatic tissue. Amongst the four H2B variants which were studied in depth, H2B.3 is relatively depleted from heterochromatin while H2B.2 is relatively enriched over repeats and transposons [Jiang et al., 2020]. This peculiar distribution of H2B suggests a functional specialisation of these variants and also specific deposition mechanisms. How does the localization of H2Bs affect transcription remains unknown.

H4 variants

Functional variants for H4 were only discovered recently. H4.G in humans is one such example. It has truncated C-terminus and shares only 85% identity with the replication coupled H4 histone [Long et al., 2019]. The gene H4C7 encodes for H4.G variant, forms a polyA transcript that is expressed at low levels in certain cancer cell lines. The H4.G variant is also localised to the nucleolus and forms an unstable nucleosome-like structure, therefore affecting the compaction of rDNA clusters and eventually promoting rRNA transcription [Long et al., 2019]. H4 variants have also been described in parasitic species but their presence in other organisms are yet to be discovered.

1.3 Chromatin states

With the advancement of genome wide sequencing technologies and development of specific antibodies against chromatin proteins, it became possible to map the localization of histone PTMs in great detail. Given the plethora of histone PTMs that occur on histone tails, nucleosomes could be assembled in 2^n combinations (where n is the number of histone PTMs). In addition, the histones can be exchanged with histone variants, thus giving the theoretical possibility of a huge number of distinct nucleosomes. However, the correlations established between histone PTMs, histone variants and different domains of chromatin suggested a limitation of the number of combinations of histone PTMs and variants per nucleosome. So, it is important to determine if and how PTMs and histone variants combine and how these combinations associate with specific elements of the genome. Computational methods like the Hidden Markov model (ChromHMM)

have been widely used to define “chromatin states”, which reflect the combinations of PTMs that could exist in the genome [Ernst and Kellis, 2012, Ernst and Kellis, 2010, Ernst and Kellis, 2017]. As expected, only a limited number of chromatin states have been identified in multiple organisms [Ernst and Kellis, 2010, Ernst and Kellis, 2012, Roudier et al., 2011, Sequeira-Mendes et al., 2014]. In drosophila and mammalian cells, combinatorial analyses showed that specific chromatin states associate with regulatory elements like promoters, enhancers as well as transcribed regions vs non transcribed elements of the genome. Chromatin states have also been previously identified in Arabidopsis, in two seminal studies described below [Roudier et al., 2011, Sequeira-Mendes et al., 2014].

In Arabidopsis, the first attempt to study the combinatorial patterns of chromatin marks was performed by Roudier et al. [Roudier et al., 2011]. In this study 12 different PTMs were profiled to identify four distinct chromatin types namely active genes, repressed genes, repeats and intergenic regions. H3K4me3 and H3K27me3 were diagnostic of two distinct types of genes. 10% of genes were not assigned to any particular chromatin type, indicating that there were other epigenetic missing states. This study served as the first sketch of the Arabidopsis epigenome until 2013 when additional PTMs and histone variants were studied in detail.

In 2013, a second study describing the Arabidopsis genome into nine chromatin states was published [Sequeira-Mendes et al., 2014]. Sequeira-Mendes2014 Sequeira-Mendes et al. utilised 16 features including, CG methylation, multiple histone PTMs and two H3 variants and H2A.Z. Inclusion of additional features led to extend the previous simplified chromatin state model to nine chromatin states. To summarise, this study identified two chromatin states associated with open chromatin on expressed genes (euchromatin) with a chromatin state covering promoters and a chromatin state for gene bodies. Two chromatin states were associated with repressed genes (facultative heterochromatin) regulated by PRC2 but the contribution of PRC1 was not documented. intergenic regions. Two chromatin states over repeats and transposons divided constitutive heterochromatin between the pericentromeres and the small islands of heterochromatin on the chromosome arms. The remaining chromatin states were primarily but not only present in intergenic

1 Introduction

regions, although intergenic regions were also covered by other chromatin states, suggesting a lack of information to clearly differentiate the chromatin over coding and non-coding regions. This study provide a valuable resource to ascribe the localization of the three major types of chromatin domains and some of their subdivisions but did not provide higher-order degree of integration of the genomic information to reveal new rules of organisation of chromatin. This limitation might originate from the lack of comprehensive integration of the profile of all histone variants and/or some histone PTMs.

1.3.1 Transcription and chromatin

Transcription can be divided into four major steps: the recruitment of a preinitiation complex (PIC) that primes RNA Pol II activity, pausing at the 5' end, active elongation and termination. Histone PTMs have been associated with each step. For example, H3 and H4 acetylation reduce the bonds between histones and DNA and are important for the association of RNA Pol II PIC. H3K36me3 is deposited with active elongation. In yeast, a single SET domain protein (Set2p) deposits H3K36me3 but in mammals and Arabidopsis there are many more “writers” of H3K36me. The role of these writers during RNA-Pol II elongation still needs to be investigated. We know that transcriptional activity influences the deposition of PTMs. But the role of PTMs together with histone variants and how they could influence transcriptional activity still remains unclear.

]There is also evidence for an impact of histone variants on transcription. In yeast, H2A.Z is associated with TSS of actively transcribing genes and it was proposed that H2A.Z would lower the energy barrier opposed to RNA Pol II interaction with the nucleosome. However, there is contrasting evidence supporting that H2A.Z is also found on genes that are expressed at low level and on genes repressed by H3K27me3. This dual nature of H2A.Z and its association with distinct PTMs in regulating transcription needs to be further investigated. There is evidence that H3.3 is enriched at the 3' end of the genes but whether this plays a role in termination of transcription is unclear and we do not know if specific histone PTMs are involved in this process.

1.4 Plants as models for chromatin and epigenetic studies

Chromatin research in plants has provided various pivotal discoveries that led the field of chromatin and epigenetics in a new direction. To point out a few, the early observations of distinct cytological structures in chromosomes (as mentioned earlier) were initially described in plants, now interpreted as Euchromatin/Heterochromatin [Heitz, 1929]. Laws of Mendelian inheritance and the first evidence of non-Mendelian inheritance (epialleles) were also reported originally through studies done on crops like peas, maize, and tomato [Pikaard and Scheid, 2014]. Evidence of genomic imprinting, i.e., biased expression of specific genes from maternal or paternal allele was also first described in maize [Pikaard and Scheid, 2014] and later found in various organisms. Last but not least, in the 1940s, the remarkable discovery of jumping genes – “transposable elements” for long considered “Junk DNA”. This pioneering work led by McClintock and her colleagues was identified through her careful observations of Maize kernels [Reviewed in [Lisch et al., 2009]]. We are now discovering that transposons play a crucial role in regulating the expression of genes where they are inserted and serve as critical regulatory elements in gene and genome evolution in several species [Pikaard and Scheid, 2014].

So, even though plants and metazoans have a large evolutionary distance between them, there are striking similarities of conserved epigenetic mechanisms. Mechanisms that control the genome and chromatin organisation are highly conserved such as. 1) Features like the ratio of euchromatin vs heterochromatin, genome size vs complexity or the number of genes is comparable within plants and mammals. 2) the histone proteins responsible for chromatin packaging. 3) the DNA modifying enzymes and histone-modifying enzymes. 4) The chromatin remodeler and chaperones have similar functions wrt. nucleosome spacing. Hence the epigenetic pathways identified in plants or mammals are conserved across eukaryotes and plant models are suitable for understanding their functions [Pikaard and Scheid, 2014].

1.4.1 *Arabidopsis thaliana* sets the perfect stage to study chromatin regulation in eukaryotes

With the advancement of molecular biology techniques and transgenics, *Arabidopsis* became a popular model for plant researchers in the 1980s. *Arabidopsis thaliana* is a small weed that belongs to the Brassicaceae family. It mainly grows across the northern hemisphere and serves as an excellent model for studying natural variation. The key points that made it a popular research model were the availability of genetic tools, short generation time, and a relatively small compact genome (135 Mb). In 2000, *Arabidopsis* was the first flowering plant to have a complete sequenced reference genome published. *Arabidopsis* has a complex repertoire of partially redundant genes encoding chromatin modifiers, erasers, reader and histone variants [Pikaard and Scheid, 2014]. The functional redundancy of these gene families turns out to be an asset because the mutants are not lethal in contrast to mammals. In addition, the relatively compact genome of *Arabidopsis* provides a good model for study of chromatin and epigenetics.

2 Aims

The goal of my thesis work is to characterise the importance of histone variants in establishing the chromatin landscape in the model plant *Arabidopsis thaliana*. First, I described how histone variants and PTMs combine in a new set of chromatin states in *Arabidopsis* seedlings. Second, I discovered that DDM1(Decrease in DNA methylation 1) a chromatin remodeler, binds and deposits the variant H2A.W on transposons and participates in the silencing of transposons.

2.1 Uncover the importance of Histone variants in shaping chromatin state landscape in *Arabidopsis* seedlings

Chromatin composition can be affected by two major players mainly post translational modifications (PTM) and histone variants. In *Arabidopsis*, previous studies have established chromatin states utilising PTMs and few histone variants. But we still lacked a complete overview of how all histone variants contribute in shaping the chromatin state landscape. Given that histone variants occupy specific regions in the genome. It was important to understand how this peculiar distribution is associated with PTMs. Furthermore, we wanted to understand how chromatin landscapes could instruct transcription. To further understand this, we described chromatin state landscape using histone variants and PTMs. This allowed us to discover novel features of the *Arabidopsis* genome.

2.2 Understand how chromatin remodelers maintain transposable elements in repressed chromatin states

Decrease in DNA methylation 1 (DDM1) is a SWI/SNF2 chromatin remodeler that was previously linked with loss of DNA methylation, loss of H3K9me2 and thus leading to de-repression of transposons. But the direct action of DDM1 was unknown. Here we uncovered the mechanism of action of DDM1. Through this work, we could show that DDM1 binds histone variant H2A.W and deposits H2A.W at certain transposon loci to maintain transposons in a silenced chromatin state.

3 Publications

3.1 List of Publications

Publication I

Jamge, B., Lorković, Z., Axelsson, E., Yelagandula, R., Akimcheva, S., Berger, F. (2022). Transcriptional activity is shaped by the chromatin landscapes in Arabidopsis. bioRxiv.

bioRxiv. <https://doi.org/10.1101/2022.06.02.494419>

DOI :10.1101/2022.06.02.494419

Contribution: study design, experimental work, data analysis, and manuscript writing

Publication II

Osakabe, A.*, **Jamge, B.***, Axelsson, E., Montgomery, S.A., Akimcheva, S., Kuehn, A.L., Pisupati, R., Lorković, Z.J., Yelagandula, R., Kakutani, T. and Berger, F., 2021. The chromatin remodeler DDM1 prevents transposon mobility through deposition of histone variant H2A. W. Nature Cell Biology, 23(4), pp.391-400.

Nat Cell Biol 23, 391–400 (2021). Nat Cell Biol 23, 391–400 (2021). <https://doi.org/10.1038/s41556-021-00658-1>

DOI : 10.1038/s41556-021-00658-1

Contribution: experimental work, data analysis, and manuscript writing.

3.2 Publication I

Transcriptional activity is shaped by the chromatin landscapes in Arabidopsis

Jamge, B., Lorković, Z., Axelsson, E., Yelagandula, R., Akimcheva, S., Berger, F.
(2022). bioRxiv.

Submitted to BioRxiv

bioRxiv. <https://doi.org/10.1101/2022.06.02.494419>

DOI :10.1101/2022.06.02.494419

Title: Transcriptional activity is shaped by the chromatin landscapes in *Arabidopsis*

Authors: Bhagyshree Jamge^{1,2}, Zdravko J. Lorković¹, Elin Axelsson¹, Ramesh Yelagandula^{1,3}, Svetlana Akimcheva¹, 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 Present address: Institute of Molecular Biotechnology of the OeAW, Austrian Academy of Sciences, Vienna BioCenter, Dr. Bohr-Gasse 3, 1030 Vienna, Austria

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

One Sentence Summary: Histone variants and histone modifications build a limited number of distinct chromatin landscapes that instruct the transcriptional regulation of genes and transposons in *Arabidopsis*.

Summary

How histone variants and histone modifications shape nucleosome-mediated transcriptional repression, and how transcriptional activity shapes the enrichment of histone modifications and variants remain unclear. To clarify these relationships, we examined chromatin organization in the *Arabidopsis thaliana* genome, identifying a limited number of chromatin landscapes that distinguish transposon families and distinct groups of genes based on their transcriptional regulation. Unexpectedly, H2A variants are strong determinants of the landscape architecture. Combinations of H2A.W and four histone modifications define six domains that are occupied by specific transposon families and organized concentrically around the centromere. Moreover, H2A.Z defines transcriptional gene repression in specific domains. Expressed genes occupy four chromatin landscapes with specific RNA Polymerase II profiles. Although the composition of each chromatin landscape is invariant, they cover genes with a wide range of

expression levels. Therefore, chromatin landscapes control the range of transcriptional activity, but transcriptional activity has little effect on chromatin composition.

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

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 (Macadangang et al. 2014), transcription (Allis et al. 1980; Costanzi and Pehrson 1998; Giaimo et al. 2019; Meneghini, Wu, and Madhani 2003; Yelagandula et al. 2014), and genome maintenance (Rogakou et al. 1998). Notably, nucleosomes contain two copies each of histones H2A, H2B, H3, and H4. The histones associated with arrays of nucleosomes, which wrap around genes and transposons, are subject to a range of post-translational modifications (Talbert and Henikoff 2021; Tessarz and Kouzarides 2014), and chromatin states can be defined based on the characteristic combinations of histone modifications they exhibit (Ernst and Kellis 2012, 2017). Indeed, chromatin states distinguish the major domains of chromatin in the genome, comprising euchromatin (with active genes), facultative heterochromatin (with repressed genes) and constitutive heterochromatin (with transposons and repeats) (Roudier et al. 2011). They also highlight specific features such as promoters and enhancers (Ernst and Kellis 2010). How chromatin states assemble at different scales of genome organization and whether transcriptional activity shapes chromatin states, or chromatin states shape the transcriptional regulation of genes and transposons, remain unclear.

Histone variants have arisen from divergence—both within and between organisms—of their intrinsically disordered loop and tail regions, with dozens of variants in the H2A, H2B, and H3 families identified (Talbert and Henikoff 2021). All multicellular eukaryotes possess a centromeric H3 (CENPA), a DNA replication-dependent H3.1 (which is incorporated in the S phase), and a DNA replication-independent H3.3 as well as H2A.Z, H2A.X, and a DNA replication-dependent H2A variant. These variants can have important functions in chromatin state establishment. For example, nucleosomes cause RNA Polymerase II (Pol II) to stall, and

the transcriptional complex must overcome the energy barrier imposed by nucleosomes to initiate transcription (Bintu et al. 2012; Bondarenko et al. 2006; Kireeva et al. 2005). This energy barrier is reduced when the first nucleosome encountered by RNA Pol II contains the histone variant H2A.Z, highlighting the potential for variants to influence transcription (Rudnizky et al. 2016; Subramanian, Fields, and Boyer 2015; Weber and Henikoff 2014).

This idea is further supported by the following observation. In fission yeast, H2A.Z participates in the initiation and termination of transcription (Albert et al. 2007; Raisner et al. 2005; Venters et al. 2011). In plants, opposing roles for H3.3 and H2A.Z have been suggested (Wollmann et al. 2017). In addition, unique types of H2A variants that are associated with transcriptional repression have evolved in animals and plants (Loppin and Berger 2020; Talbert and Henikoff 2021). In mammals, the so-called macroH2A variants have a C-terminal macro domain attached to the core by a linker enriched in lysine residues (Douet et al. 2017; Hsu et al. 2021; Sun and Bernstein 2019). In vascular plants, the H2A.W variants are further distinguished by a C-terminal KSPK motif (Lei and Berger 2020; Schmücker et al. 2021; Yelagandula et al. 2014) which favors heterochromatic silencing by directly altering H2A.W interaction with DNA (Osakabe et al. 2018; Schmücker et al. 2021; Yelagandula et al. 2014). Furthermore, a diversity of H2B variants was recently recognized in both plants (Jiang et al. 2020) and mammals (Raman et al. 2022), although the roles of these variants in transcription remains unknown. Histone variants thus represent complex, diverse nucleosome components that could regulate and organize the functional activities of chromatin.

Applying Hidden Markoff Model [ChromHMM] (Ernst and 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 and Kellis 2010, 2012, 2017; Kharchenko et al. 2011; Liu et al. 2018; Roudier et al. 2011; Sequeira-Mendes et al. 2014). Conserved chromatin states include H3 and H4 acetylation, which is deposited by the preinitiation complex (PIC) associated with the SAGA complex and the general transcription factor TFIID (Watanabe and Kokubo 2017). 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). Trimethylation of lysine 4 of H3 (H3K4me3) deposited by the complex COMPASS is also correlated with the PIC, but its direct roles remain unclear (Fromm and Avramova 2014). After the initiation of transcription, RNA Pol II begins the early elongation process, which involves

pausing sites (Nielsen et al. 2019). These sites are marked by the histone tail acetylation levels, H3K4me3, and H3K36me3 which are positively correlated with transcript levels (Gates, Foulds, and O'Malley 2017). Active elongation is marked by monoubiquitination of H2B (H2Bub), which requires specific ubiquitin ligases: HUB1 and Hub2 in *Arabidopsis* (Feng and Shen 2014) and UBE2A:RNF20/40 in human (Kim et al. 2009). This modification is associated with transcribed regions in mammals (Minsky et al. 2008) and plants (Roudier et al. 2011). With the exception of H3K9me2, which marks the 3' ends of genes in mammals but not plants (Gates, Foulds, and O'Malley 2017; Leng et al. 2020), the histone marks or variants that control the termination of transcriptional elongation are not understood.

Compared with the wealth of data regarding the relationships of histone modifications with transcription, the role of histone variants in this process remains to be explored. In *Arabidopsis*, there is a clear correlation between transcript levels and the enrichment of the variant H3.3 (Stroud et al. 2012; Wollmann et al. 2012, 2017) and replicative H2A and H2A.X (Yelagandula et al. 2014)(Yelagandula et al. 2014) on gene bodies marked by H3K36me2 (Leng et al. 2020). By contrast, H2A.Z is enriched on the bodies of 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). The complex, intimate associations between histone variants and modifications have not been fully characterized in plants or mammals. In addition, it remains unclear to what extent transcriptional activity generates these associations or whether feedback loops generate a dialog between the transcription and deposition of chromatin elements that shape chromatin organization in the genome.

Here, we analyzed how all 13 histone variants expressed in vegetative tissues associate with the 12 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 among euchromatin, facultative heterochromatin, and constitutive heterochromatin. Unsupervised clustering of chromatin states showed that chromatin is organized into stereotypical Chromatin Landscapes (CLs) that distinguish distinct classes of transposons and genes with different modes of transcriptional regulation. Each CL was associated with a specific mode of transcriptional repression or activity. Remarkably, the levels of transcription did not affect the organization of the CLs, suggesting that CLs define specific modes of transcription, whereas transcription has little influence on the composition of the CL.

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, whether H3 variants also assemble in a homotypic manner and whether H2A variants preferably assemble with a specific H3 variant or histone modification have been unclear. To answer these questions, we immunoprecipitated mononucleosomes (**Fig S1A–C**) from transgenic *Arabidopsis* lines expressing HA-tagged H3.1 and H3.3 under the control of their native promoters (Jiang and Berger, 2017). Because tagged H3 ran slower than endogenous H3 on SDS-PAGE (**Fig S1B**), 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 ~40% of nucleosomes contained both H3.1 and H3.3 (**Fig 1A**). Neither H3.1 nor H3.3 was preferentially associated with a specific H2A variant (**Fig 1B**). This was confirmed by MS analysis of nucleosomes pulled down with H2A variant-specific antibodies (**Fig S1D**). Therefore, H3 variants do not necessarily form homotypic nucleosomes, despite the presence of H3 variant-specific deposition mechanisms (Nie et al. 2014), and they do not associate with specific H2A variants.

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 (**Fig 1C**). We also analyzed H3.1 and H3.3 modifications by MS using information from both tagged H3 and endogenous H3 in the respective immunoprecipitates. MS unambiguously distinguishes H3K27 methylation and H3K36 methylation in H3.1 and H3.3 variants (**Fig S1E**). Here, we focused on H3K27 and H3K36 modifications because they are diagnostic of constitutive (H3K27me1) versus facultative (H3K27me3, silenced by Polycomb repressive complex 2 [PRC2]) heterochromatin and actively transcribed (H3K36me) regions of the genome, respectively. The levels of H3K27 and H3K36 modifications on H3.1 and H3.3 displayed the same trends regardless of whether they were precipitated with H3.1 or H3.3 nucleosomes; the same trend was also observed by comparing endogenous and transgene-encoded H3.1 and H3.3 (**Fig 1D** and **S1E–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 (**Fig 1D** and **S1E–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. H2A.W variants were strongly associated with histone modifications of constitutive heterochromatin (H3K9me1, H3K9me2, and H3K27me1) (**Fig 1E**). The modification H3K27me3, which is associated with facultative heterochromatin, was detected only in H2A.Z nucleosomes, and H3K36me3 (marking euchromatin) was primarily detected in H2A, H2A.X, and H2A.Z nucleosomes (**Fig 1E**). Other modifications (such as H3K4me1 and H3K4me3) displayed complex patterns of association with H2A variants and associations among themselves. Therefore, H2A but not H3 variants form homotypic nucleosomes that preferentially carry specific histone modifications that are associated with the transcriptional status of protein-coding genes or transposons.

Identification of chromatin states using histone variants in Arabidopsis

We performed ChIP-seq using Arabidopsis seedlings to identify the combinations of all histone variants present in somatic cells and their associations with histone modifications. We included the most abundant isoforms of H2A.W (H2A.W.6 and H2A.W.7), H2A.Z (H2A.Z9 and H2A.Z.11), and replicative H2A (H2A.2 and H2A.13) in our analysis. The algorithm ChromHMM (Ernst and Kellis 2017) was used to describe chromatin states in Arabidopsis (see Methods for details). We chose to analyze the most parsimonious model based on 26 chromatin states (**Fig S2A**). Chromatin states were clustered based on the emission probability for each modification and histone variants (**Fig 2A**) and showed distinct enrichment of various marks and occupied different proportions in the genome (**Fig 2A, 2B**). For each chromatin state, we measured the average level of transcriptional activity (**Fig 2C**), the degree of enrichment in CG methylation (**Fig 2D, S2C and S2D**), the level of accessibility by DNase I-seq (**Fig 2E**) and the relative abundance of transposons, repeats, and elements of protein-coding genes (**Fig 2F**). Altogether, these parameters defined four groups: constitutive heterochromatin (H), facultative heterochromatin (F), euchromatin (E), and intergenic regions (I).

States H1–H6 shared the highest levels of CG methylation, with the near absence of transcriptional activity, and was primarily occupied by transposons and repeats, thus representing constitutive heterochromatin. States F1–F6 were associated with low transcriptional activity and were present over genes and half of pseudogenes, indicating that they represented facultative heterochromatin. States E1–E11 occupied expressed genes and thus comprised euchromatin. Three states (I1–I3) occupying 15% of the genome comprised

non-coding regions such as transposons, repeats in regions dense with protein-coding genes, and untranslated regions of genes (3'UTR and 5'UTR). Hence, I1–I3 were classified as intergenic states. As defined by state occupancy, constitutive and facultative heterochromatin occupied 17% and 20% of the genome (**Fig S2F**), respectively, which corresponded to previous estimates (Roudier et al. 2011). Chromatin accessibility was (to some degree) inversely correlated with the degree of nucleosome density determined by MNase-seq (**Fig S2E**). Chromatin accessibility differed among the four groups, from the less accessible constitutive heterochromatic states H1–H6 to the most accessible intergenic states I1–I3, and states E1, E2, and F6 were present in the 5' and 3'UTRs of genes (**Fig 2E**). Thus, chromatin accessibility, reflecting the potential for transcriptional activity, was broadly associated with chromatin states.

Overall, the chromatin states recapitulated the preferential associations between H2A variants and histone modifications observed in mononucleosomes (**Fig 2A**). Constitutive heterochromatin was enriched in H3K9me1/2, H3K27me1, and H4K20me1, but only H3K9me1 and H2A.W.7 truly represented all six heterochromatic states. The histone modifications H2AK121ub and H3K27me3, which are deposited by PRC1 and PRC2, respectively, were associated with chromatin states F1–F6, and H2A.Z was present in all facultative heterochromatin states. H3K27me3 covered chromatin states F1–F3 and was also observed in H5, while H2AK121Ub marked several euchromatin states. Similarly, H2Bub and H3K36me3 was associated with H2A and H2A.X in all euchromatic states (E1–E11), but all other marks that are usually associated with euchromatin were also distributed over a larger spectrum of states. Contrasting with the strong association between H2A variants and the three major domains of chromatin, much less prominent associations were observed between histone modifications and H3 variants or H2B variants; this finding is in agreement with the results of our biochemical analysis of mononucleosome composition and previous studies (Osakabe et al. 2018; Jiang et al. 2020).

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). The blocks of heterochromatic states H1–H6 corresponded to the previously identified states 8 and 9, which were also defined as heterochromatin in (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 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 (**Fig S2B**). Altogether, the 26-states model provides a more refined and coherent classification of elements of chromatin than states defined primarily based on chromatin modifications, pointing to the importance of histone variants in the definition of chromatin states.

To test this idea directly, we calculated chromatin states after excluding either all histone modifications or all histone variants or single families of histone variants. A comparison between the resulting matrices of chromatin states (measured by the Jaccard index) showed that the loss of histone modifications or histone variants affected the definition of the 26 chromatin states to comparable degrees (**Fig 2G**). Removing H2B variants did not affect most chromatin states, except for H1 and I2, which showed high emissions probabilities for H2B variants (**Fig 2A**). By contrast, H3 variants contributed more effectively towards defining specific states, and removing all H2A variants (no H2A) caused the loss of several states marked by a Jaccard index < 0.7 (**Fig 2G**). In summary, histone modifications and histone variants are required to a comparable degree to define biological chromatin states. Among histone variants, H2A variants have the strongest effect on the association with histone modifications to define chromatin states.

Transposons share distinct CLs that are distinguished by the identity of transposons and their modes of silencing

Constitutive heterochromatin comprised six states that were all enriched to varying degrees in H2A.W.6 and H2A.W.7. These states were further distinguished by their relative enrichment in four histone modifications and H3.1/H3.3 (**Fig 2A**), distinct levels of CHG and CHH DNA methylation (**Fig S3 D,E**), nucleosome density (**Fig S2E**), length (**Fig S2H**), and position relative to the peri-centromere and chromosome arms (**Fig 3A, S3A, Table 1**). The relative proximity of heterochromatic states H1–H4 versus H5 and H6 was also revealed by analyzing the closest neighbor relationships between chromatin states (**Fig 2H**). States H2–H4 (occupying pericentromeres) were enriched in transposon genes (TEGs), whereas states H5 and H6 occupied chromosome arms and were primarily associated with transposable element (TE)

fragments (**Fig 2F**, and **3A**). TE fragments were methylated by RNA-directed DNA methylation (RdDM), a process involving RNA Pol V (Matzke et al. 2009). Accordingly, we observed enrichment in Pol V ChIP peaks over chromatin states H5 and H6 (**Fig 3B**). These chromatin states were covered by H2A.W.7 (**Fig 2A**) and were typically present in chromosome arms (**Fig 3A**), which are targeted by RdDM (Lorković et al. 2017a; Matzke et al. 2009).

We extended our analysis to determine whether there were preferential associations between heterochromatin states H1–H6 and the various mechanisms that deposit DNA methylation (**Fig 3C**) or silence transposons (**Fig 3D**). This confirmed the strong association between RdDM and chromatin states H5 and H6. By contrast, H2 and H3 were not methylated by RdDM but were instead methylated by CHROMOMETHYLASES 2 AND 3 (CMT2 and CMT3), which deposit CHH and CHG methylation (**Fig 3C** and **S3D** and **S3E**). We also observed clear preferential associations between transposon silencing mechanisms and chromatin states (**Fig 3D**), suggesting that chromatin states provide cues that guide silencing mechanisms to their targets.

To establish whether chromatin states associate with each other in a preferential manner, we performed unsupervised hierarchical clustering analysis of the associations between the chromatin states across all transposons. We included only TEs >500 bp long to eliminate fragments of TEs with low mapping ability. We identified seven clusters of TEs, which we named TE1 to TE7 (**Fig S3B**). In each cluster, TEs were primarily enriched in a single chromatin state. The states of constitutive heterochromatin were organized into five clusters (TE1–TE5) over transposons, except for state H1, which did not cover TEs but instead primarily covered repeats at the centers of pericentromeres and were thus likely close to the centromere (**Fig 2F**, **S2H**, **S3A**, **S3B** and **Table 1**). Transposons with no CG methylation were associated with state F1 and were classified as Polycomb group transposons TE6 (**Fig S3B**). Lastly, group TE7 comprised transposons in regions covered with intergenic state I1 (**Fig S3B**).

We also observed marked differences between the seven clusters of TEs in terms of their association with chromatin modifications and histone variants (**Fig 3E**, **S3F**, **S3G**), specific TE families (**Fig 3G**), relative proportions amongst all TEs (**Fig S3C**), lengths (**Fig 3F**), and DNA methylation patterns (**Fig 3I** and **S3D–E**). The near exclusive association between TEs from clusters 1–5 and the constitutive heterochromatin states marked by H3K9 methylation

confirmed the specific associations among TEs in each cluster, DNA methylation pathways, and enrichment in specific families of TEs. Clusters 1–3, associated with LTR Copia and Gypsy transposons in pericentromeric heterochromatin, were primarily covered by H3.1, H2A.W.6, and H2A.W.7, while TEs from clusters 4 and 5 were located in the arms of chromosomes, covered by H2A.W.7, and methylated by the RdDM pathway. Cluster 6 and 7 primarily contained short Helitrons covered by H2A.W.7 and histone modifications present on protein-coding genes. TE6 transposons were covered with H3K27me3 but no CG methylation, while CG methylation was associated with H3K9Ac on TE7 transposons. TEs from these two clusters were the shortest and located in proximity of protein coding genes (**Fig 2H**).

In conclusion, there were striking associations between TE families and each chromatin state. Such an association was described for the guidance of RdDM by non-coding RNAs on TEs from clusters 4 and 5 (Matzke et al. 2009). Our results generalize this idea and suggest that transposons contain information that guides the deposition of the chromatin marks responsible for their silencing. In addition, we identified a new, large set of associations between TEs and facultative heterochromatin marked by H3K27me3 or intergenic states in euchromatin.

H2A.Z marks repressed genes in facultative heterochromatin

We applied unsupervised clustering of the relative enrichment of chromatin states to all protein-coding genes of Arabidopsis and observed ten distinct clusters. Unlike TEs, protein-coding genes were occupied by an assembly of chromatin states (**Fig S4A, S4C, FigS5A**). Clustering chromatin states over protein-coding genes revealed that the facultative heterochromatin states F1–F6 formed only three types of associations that represented stereotypical chromatin landscapes (CLs) (**Fig S4A**). Each CL involved the association of one or two chromatin states in the gene body (e.g. F2 and F3 with CL1) with another identical chromatin state upstream and downstream of the gene body (e.g. F1 with CL1) (**Fig 4A, Fig S4 A, C**). These associations were also supported by neighborhood analysis of chromatin states (**Fig 2H**). Each CL was characterized by a specific combination of chromatin modifications (**Fig 4B**), but CL1–CL3 shared enrichment of H2A.Z on the gene bodies, which represents the hallmark of facultative heterochromatin (**Fig 4B**). The histone modification H2AK121Ub deposited by PRC1 was present in CL1–CL3, although to varying degrees (**Fig 4B**). In CL1, H2AK121Ub associated with H3K27me3 deposited by PRC2 (**Fig 4B**). CL1 comprised ~70% genes in facultative heterochromatin (**Fig 4A**). These genes shared the lowest expression in the wild type (**Fig 4E**), and many of these genes were expressed in mutants deprived of PRC1 or PRC2 activity (**Fig**

S4B, Table 2). In conclusion, genes from CL1 are the genes typically regulated by Polycomb group proteins.

CL2 and CL3 comprised genes covered by H2A.Z marked by H2AK121Ub and H3.3 (Fig. 4B, S4E and Table 2). Whereas genes from CL1 were bordered by regions enriched in H3K27me3, genes within CL2 also carried marks present in constitutive heterochromatin and were occupied and surrounded by regions enriched in H3K4me1 and H3K27me1 (**Fig 4B**). Genes within CL3 were shorter than those in CL1 and CL2 (**Fig 4D**) and were enriched in H3.3 compared to other genes in constitutive heterochromatin (**Fig S4E**).

In addition to the 9564 repressed genes covered by CL1–CL3 typical of facultative heterochromatin (dominated by the presence of H2A.Z), 1115 protein-coding genes were not expressed and were covered primarily by heterochromatic states (**Fig 4A, C, E, and S4C–D, Table 2**). These genes were covered by CL4–CL6, which were distinguished by constitutive heterochromatin states present at their gene bodies (**Fig 4A and S4C**) and were preferentially enriched for H2A.W.7 versus H2A.W6 and H3.1 versus H3.3 in their gene bodies (**Fig S4D–E**). These genes were preferentially associated with pericentric heterochromatin (CL4 and CL5) and were surrounded by H2A.W.7 and H3K9me (**Fig 4C**) or with heterochromatin islands in chromosome arms (CL6).

We conclude that constitutive heterochromatin hosts genes with three types of chromatin landscapes dominated by H3.1 in association with H2A.W and H3K9me. By contrast, H2AK121Ub is the main marker of facultative heterochromatin, which comprises three types of chromatin landscapes across genes controlled by Polycomb repressive complexes.

Active genes share one of four stereotypical chromatin landscapes

All 16,776 expressed protein-coding genes shared some of the euchromatin states (E1–E11), which assembled into four chromatin landscapes (CL7–CL10; **Fig S5A, Table 2**). These associations were also supported by neighborhood analysis of chromatin states (**Fig 2H**). The E1 state enriched in H2A.Z (**Fig 2A**), the only state common to all euchromatin genes, marked the 5'UTR, except CL9, where E1 also marked the entire gene body (**Fig 5A and S5A**). CL9 was most similar to CL1–CL3 and occupied genes with gene bodies covered by H2AK121Ub, but it was associated with euchromatin marks H3K4me3, H3K36me3, and H3 acetylation. All other CLs comprised an assembly of euchromatic states arranged in the 5' to 3' direction from

E1–E9 for cluster 1, E1–E3, E8 for cluster 2, and E1 followed by E10/E11 for CL10 (**Fig 5A** and **S5A**). All euchromatic CLs were characterized by gene bodies surrounded by the intergenic state I1 (except for I3, which was located upstream of the 5' end for CL8) enriched in H3K9Ac and H3K14Ac and with the highest chromatin accessibility (**Fig 2A, 2E**). In genes covered by CL7–CL10, the transcription end site (TES) was also routinely followed by a 3'UTR marked by the state I2 characterized by enrichment in histone variant H2B.1 and H2B.2 (**Fig 2A, 2F, 5A, S5A, S6B**).

Genes from all euchromatic CLs shared similar profiles of H3K4me3 just after the TSS overlaid with a broad asymmetrical enrichment of H3K36me3 declining towards the 3' end (**Fig S5B**). The H3K4me3 profile overlaid a similar peak of H2A.Z, H3K9Ac, and H3K14Ac at the 5' end, which mark the presence of the preinitiation complex (Leng et al., 2020). One exception was genes with CL9, which showed a symmetrical enrichment in these marks parallel to their symmetrical organization of chromatin. Overall, each euchromatic CL exhibited distinctive features relative to the genomic profiles of H2A.Z, H2A.X, and replicative H2A (**Fig 5B**). Genes with CL7 and CL10 showed a strong peak of H2A.Z at the 5' end, while this peak was extended towards the 3' end in genes with CL8 and CL9. In opposition to the enrichment of H2A.Z, CL7 and CL10 shared the strong presence of CG methylation towards the 3' of the gene body, while this enrichment was reduced in CL8 and CL9 (**Fig 5C**). In addition, we observed a gradual enrichment of H3.3 towards the 3' ends of all euchromatic genes, as reported previously (Wollmann et al. 2017). In addition, in general, the variant H2B.3 was located over gene bodies, while H2B.1 and H2B.2 marked the beginning of the 3'UTRs (**Fig S6B**).

The characteristics of CL8–CL10 are detailed in Table 2. These CLs lacked the states E5, E6, and E7 and were present in genes with lower expression levels, and except for CL10, in genes that were shorter than those covered with CL7 (**Fig 5A, 5F, and 5G**). Two-thirds of euchromatin genes were covered with CL7 and were (on average) the longest and most strongly transcribed (**Fig 5A, 5F, and 5G**). These genes covered with CL7 were marked by strong enrichment of H2Bub from the 5' to 3' end, which was not observed in other euchromatic genes (**Fig 5E**). The transition to the gene bodies was marked by the replacement of H2A.Z by H2A and H2A.X associated with H2Bub and H3K36me3. The loss of the latter histone modifications marked the transition to states E8 and E9, with a return of H3K9Ac that marked the terminators.

In conclusion, all CLs of euchromatin showed an asymmetrical modular organization that primarily reflected the distribution of H2A and H2A.X in association with H3.3, H2B.3, and H3K36me3. This organization contrasts with the symmetric arrangement of chromatin over CLs associated with gene repression.

Chromatin landscapes determine a range of transcriptional activity

CL7 and CL10 showed a peak of S5P Pol II at the 5' ends of genes and a peak of S2P towards the ends of genes, although these features were much less pronounced in CL10 (**Fig 5C**). By contrast, these peaks were absent in CL8 and CL9, with a much higher and uniform profile of enrichment of Pol II covering the entire transcriptional unit from the 5' to 3' end (**Fig 5C**). Surprisingly, the coupling of CL7 and CL10 in opposition to CL8 and CL9 was not reflected in the range of gene expression, which was strongest in CL7 genes and diminished from CL8 to CL10 genes (**Fig 5F**). Similarly, compared to the high levels of expression related to CL7 and the lowest levels of expression related to CL10, the transcript levels were correlated with the enrichment of H2A.X, H2A, and several modifications over the gene bodies (**Fig 5F and S6**).

To identify features that were directly correlated with the level of transcription, we compared the enrichment of each mark across gene quartiles based on their transcript levels. In each CL, we observed a clear positive correlation between transcript levels, chromatin accessibility revealed by ATAC-seq (**Fig S7**) and the levels of enrichment in RNA Pol II and its phosphorylated forms S5P (paused) and S2P (elongating) (**Fig S8**). Each CL was associated with a distinct profile of RNA Pol II and its phosphorylated forms, with patterns of peaks and troughs unique for each CL. The overall shape of the pattern did not vary between quartiles which correlated positively with increased enrichment of RNA Pol II. Remarkably, we did not observe such a marked correlation between transcript levels and the enrichment of H2A and H3 variants or histone H3 and H4 modifications or gene body methylation, except for weak correlations for CL8 (**Fig S7, S9–S10**). The exception was H2Bub enrichment, which was positively associated with the levels of transcripts in each CL (**Fig S8–S10**). In conclusion, within the set of genes associated with a particular CL, the chromatin composition did not vary with transcript levels. Moreover, each CL was associated with a specific range of transcript levels and a distinct profile of RNA Pol II, suggesting that euchromatic CLs are set up in a stereotypical manner that constrains RNA Pol II activity to a range of levels.

Discussion

Histone variants drive the overall organization of chromatin into CLs

We demonstrated that histone modifications and histone variants subdivide the intergenic space, the non-protein-coding space, and the protein-coding genic space of the genome into 26 chromatin states. This refines the previous definition of nine chromatin states in *Arabidopsis* (Sequeira-Mendes et al. 2014). Our analysis indicated that histone variants are as impactful as histone modifications in defining chromatin states. *In silico* analyses showed that H2A variants are most important in defining the broad groups of chromatin states. These chromatin states include constitutive heterochromatin defined by H2A.W and H3K9me1; facultative heterochromatin defined by H2A.Z carrying K121Ub, with a strong association with H3K27me3; and euchromatin, primarily occupied by H2A and H2A.X in conjunction with H3K4me3, H3K36me3, and histone acetylation. These strong associations between H2A variants and histone modifications are supported by the results of biochemical analysis (our analysis and (Osakabe et al. 2018)).

The mechanisms that drive the stereotypical associations between H2A variants and specific modifications remain unclear in the case of H2A.W, H2A.X, and replicative H2A, but we have some clues about the role of H2A.Z and the Polycomb writers PRC1 and PRC2. H2A.Z is the major carrier of H2AK121Ub deposited by PRC1, as indicated by their co-occurrence with chromatin states as well as previous biochemical analyses (Osakabe et al. 2018). *In vitro* biochemical assays also showed that PRC2 strongly prefers arrays of H2A.Z as a substrate (Y. Wang et al. 2018), thus supporting the link between H2A.Z and H3K27me3 (Carter et al. 2018). Overall, our observations suggest that chromatin writers and erasers likely operate with different affinities and efficacies based on the histone variants in the nucleosome. This is also illustrated by the anticorrelation between H2Bub and H2A.Z resulting in the inhibition of the ubiquitin ligase by the H2A.Z tail (Surface et al. 2016) and by the link between H3K27 monomethylation of its preferred substrate 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. Although a general enrichment of H3.1 in constitutive heterochromatin was also noted previously, H3.1 and H3.3 do not differentiate the three main compartments of chromatin but tend to differentiate gene bodies among the different CLs. H2B variants define specific states: H1 close to the centromere, and I2 marking the 3'UTRs of expressed genes. However, their role in defining

these states is minor compared to roles of specific H2A variants, whose presence or absence provides global discrimination of the three major domains of the genome and their subdivisions.

Chromatin states are arranged into stereotypical CLs on transposons and genes

A major advance using our approach resulted from clustering chromatin states over transposons and genes. As summarized in Tables 1 and 2, the Arabidopsis genome is occupied by only a limited number of distinct CLs. In addition to these we expect two additional CLs including the centromere, which was recently described (Naish et al. 2021), and the ribosomal DNA arrays that were not studied here due to lack of unique mapping ability of these regions (Sims et al. 2021). We described seven CLs that occupy transposons, while protein-coding genes are occupied by ten contrasting CLs. The symmetrical organization of chromatin states over CLs invariably marks low or no expression, while modular directional assemblies of chromatin states over CLs are associated with expressed genes.

Of the 11640 annotated TEs, 7043 are covered by CLs in constitutive heterochromatin. Among these, most TEGs are present in CLs in the pericentromeres and are dominated by H2A.W and H3K9me, which act in synergy to silence transposons in the pericentromeric regions (Bourguet et al. n.d.) Broadly, we observed strong associations between the identity of transposons and specific silencing mechanisms. This provides concrete evidence for the selective adaptation of transposons to distinct silencing machineries (Bourguet et al. n.d.). Our analysis supports and extends the idea that different mechanisms act in distinct territories of heterochromatin. Our results suggest that TEs segregate in distinct territories that recruit different actors of silencing. We also determined that 4597 TEs (clusters 2 and 3), mostly transposon fragments and short TEs, are present in intergenic regions. Hence, a large fraction of annotated TEs that are associated with CLs typically associate with protein-coding genes, suggesting that TEs play a pervasive role in transcriptional regulation; this represents one of only a few examples of this concept reported to date (Quadrana 2020).

In Arabidopsis seedlings, circa 40% of protein-coding genes are repressed and covered by three types of CLs typical of facultative heterochromatin; 70% of these genes are present in CLs dominated by domains targeted by both PRC1 and PRC2 and represent Polycomb complex-controlled genes. The remaining 30% of genes are isolated and covered only by the marks typical of constitutive heterochromatin in addition to H2AK121Ub (CL2 and CL3) or are

present in constitutive heterochromatin (CL4 to CL6). The other protein-coding genes are expressed to varying degrees and are covered with four distinct CLs in which the first few nucleosomes are enriched in H2A.Z and H3K4me3. CL9 occupies short genes covered by H2A.Z that are expressed at low levels, while most genes with CL7 or CL8 are largely occupied by replicative H2A and H2A.X. The central module of CL7 hosts gene body methylation and is surrounded by a module of four states that structures the 5' end, while a module of two states marks the 3' ends of genes. CL8 is similar to CL7 but lacks the central module and thus shows very low gene body methylation. By contrast, approximately 1000 genes with the lowest expression levels are occupied by CL10, with a specific core state occupied by high gene body methylation and constitutive heterochromatin marks H3K27me1 and H4K20me1.

We conclude that CLs define distinct environments associated with transcriptional silencing, repression, or activity. These CLs are differentiated by different patterns of enrichment of H2A variants, H3 variants, and histone modifications. The limited number of the CLs indicates that they are produced by via the orchestration of dedicated mechanisms, which depend on the coordinated activity of chromatin writers/erasers as well as proteins that regulate the deposition and eviction of histones. While in plants chaperones specific for H3 deposition are well studied, chaperones for each type of H2A variant have not been identified (Nie et al. 2014), several remodelers are involved in the exchange of H3.1 with H3.3 (H. Wang et al. 2018), H2A.Z with replicative H2A (Aslam et al. 2019), the eviction of H2A.Z (Aslam et al. 2019), or the deposition of H2A.W (Osakabe et al. 2021). We propose that these remodelers participate in determining the architecture of the different CLs together with the chaperones that control the specific deposition of H3.1 or H3.3.

CLs determine the range of transcriptional activities of protein-coding genes

The well-characterized genome-wide association between transcript levels and the enrichment of several histone modifications and histone variants suggests that epigenetic modifications and/or the deposition of histone variants either modify transcription or are a consequence of transcription (Leng et al. 2020). Histone H3 and H4 acetylation neutralize the charge of lysine residues, destabilize nucleosomes, and facilitate the promotion of transcription (Leng et al. 2020), but the role of H3K4me3 and H3K36me3 in promoting transcription is less clear (Talbert and Henikoff 2021). For each CL associated with expressed genes, we discovered the surprising lack of correlation between transcript levels and the levels of histone variants, gene body DNA methylation, and H3 and H4 modifications. The range of transcription in each CL

is not narrow and declines from CL7 to CL10, and there is some overlap between the ranges of these four CLs. We believe that this explains the correlations observed between transcript levels and the enrichment of chromatin marks when the entire pool of protein-coding genes was included. However, it is clear that each CL is associated with a distinct profile of RNA Pol II. We thus conclude that each CL (comprising a fixed pattern of enrichment of chromatin modifications and histone variants) enables a specific regime of RNA Pol II activity, resulting in a specific range of transcripts produced. H2bUb is the only histone modification studied here that was consistently affected in all CLs by transcript levels. This could be explained by the fact that H2Bub is a hub for the recruitment of factors involved in transcriptional elongation through interactions with H3K4me3 (Leng et al. 2020) and associations with H3K36me3 writers (McGinty et al. 2008; Zhao et al. 2019). The association of HUB1 with transcription elongation factors highlights the importance of this modification for productive transcriptional elongation (Antosz et al. 2017; Pfab, Breindl, and Grasser 2018). The ubiquitination of H2B is strongly inhibited by the H2A.Z tail and the specific acidic patch of H2A.Z in mammals (Wojcik et al. 2018). The strong opposition of H2A.Z and H2Bub in Arabidopsis supports the likely conservation of the inhibition of ubiquitination of H2B by the H2A.Z tail and the specific acidic patch shown in mammals (Wojcik et al. 2018). H2Bub stimulates the deposition of H3K36me3 and productive transcription elongation and thus appears to function as a relay between the CL and the activity of RNA Pol II. The effect of transcription on H2Bub observed in our study also suggests a feedback loop that could involve H3K4me3 (Leng et al. 2020).

Our findings argue in favor of a deterministic nature of CLs that would inform the transcriptional activity associated with genes or transposons in Arabidopsis. Although the organization of the transcriptional units of plants is somewhat different from that of yeast or mammals, chromatin modifications and histone variants are strongly conserved across a large range of eukaryotes (Grau-Bové and Sebé-Pedrós 2021), suggesting that comparable CLs will be identified in a broad range of organisms. If so, the parsimonious nature of CLs and their capacity to instruct the orientation and level of transcriptional activity imply the existence of a genetic code that shapes the organization of CLs. Like the preferred sequences that direct chromatin remodelers to define the structure of the TSS (Raisner et al. 2005; J. Wang et al. 2019) and the regular spacing of nucleosomes (Li et al. 2014), the sequences of genes that share a common CL share common features that should now be deciphered. Finding the cis-code that guides the complex machinery that builds the CL will be the next step in decoding the relationship between the genetic code and the code provided by the CLs of histone variants and

modifications. Perhaps this code could ultimately be used to engineer chromatin organization in synthetic eukaryotic systems.

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 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 (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 MgCl₂, 1 mM CaCl₂, 250 mM sucrose, 1 mM DTT, 10 mM β-glycerophosphate) supplemented with protease inhibitors (Roche). After spinning for 5 min at $1,800 \times g$ at 4°C nuclei were re-suspended in N buffer to a volume of 1 ml. Twenty microliters of MNase (0.1 u/μl) (SigmaAldrich) were added to each tube and incubated for 15 min at 37°C 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 $20,000 \times g$ 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 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 over night 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 $1 \times$ Laemmli loading buffer in $0.2 \times$ 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 Fischer 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 (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 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 (Thermo Fisher 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 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 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 *A. 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 on 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 **Fig S1D**) 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 number of modified peptides (**Fig S1F**) divided by total number of peptides (**Fig S1E**) that were measured for each lysine position resulting in total modification levels for H3.1 and H3.3 (see **Fig 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 (**Fig S1**). 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 (**Fig S1H and S1I**). 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 (**Fig S1J**) as without differentiation between H3.1 and H3.3.

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.

Chromatin immunoprecipitation (ChIP)

For ChIP nuclei isolation was performed using 10-day seedlings from WT Col-0. 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 MgCl₂, 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). Nuclei were 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 anti-H3, anti-H2A.X.3/5, anti-H2A , anti-H2A.Z.9/11, anti-H2A.W.6, anti H2A.W.7, anti-H1, H3K36me3 (Abcam, ab9050), anti-H3K27me3 (Millipore, 07-449), anti-H3K4me3, anti-H3K4me1, anti-H3K27me1, anti-H4K20me1, anti-H3K9me1 or anti-H3K9me2 (Abcam, ab1220) [J1] 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 NaHCO₃), 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 analysis

For ChIP-seq, libraries were prepared with an Nugen Ovation Ultralow V2 DNA-Seq library prep kit (NuGen) 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 was used to convert the raw BAM files to fastq. FastQC v0.11.8 (Andrews 2010) (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) was used to generate quality reports for all sequencing data. Reads were filtered and then aligned to the TAIR10 *Arabidopsis* genome using Bowtie2 (Langmead and Steven L Salzberg 2013) with default settings. 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. log2 ratio (ChIP/ (Input or H3)) bigwig files were then used to generate metaplots and heatmaps across the different features in the genome (protein coding genes, Transposable elements and TEGenes (https://www.arabidopsis.org/download/index-auto.jsp?dir=%2Fdownload_files%2FGenes%2FTAIR10_genome_release%2FTAIR10_gff3)).

Defining chromatin states

Aligned ChIP-seq BAM files for 27 chromatin features were generated as described above. Aligned BAM files were then converted to BED format using . These genome wide BED files were then used for learning chromatin state by a multivariate Hidden Markov Model. We used the LearnModel program from ChromHMM (Ernst and Kellis 2017) with default settings and window size 200bp. 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 states 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.

Analysis of TE groups

To identify Transposable elements that were associated with each chromatin state. TEs with length greater than 500 bp were considered for further analysis. Using Deeptools we performed hierarchical clustering on TEs. The State region file was used to create a signal bigwig file. Presence of state was assigned score “1” and absence of state was assigned “0”. This scored signal file for each state was then used in Deeptools to generate TE groups based on state enrichment. Not all the states overlapped with TEs above 500bp, Hence only heterochromatic states H1-H6 and I1 and F1 were used. Clustering for TE groups was performed using $n=2$ to $n=10$ clusters. Clusters with similar profiles were merged and we chose to analyze 7 unique groups of TEs.

Analysis of chromatin landscape across protein coding genes

To identify Chromatin landscapes across protein coding genes. Using Deeptools we performed hierarchical clustering on all protein coding genes annotated in TAIR10. The State region file was used to create a signal bigwig file. Presence of state was assigned score “1” and absence of state was assigned “0”. This scored signal file for each state was then used in Deeptools to generate CL groups based on state enrichment across these genes. Clustering for protein coding genes was performed using $n=2$ to $n=26$ clusters. Clusters with similar profiles were merged and we chose to analyze 10 unique groups of genes each with a specific chromatin landscape.

Analysis of gene quantiles in each chromatin landscape

For each euchromatic chromatin landscape (CL7-10), genes in each CL were divided in four quantiles (q1=lowest expression and q4=highest expression) based on the expression in TPM (Transcript per million). Deeptools v3.1.2(Ramírez et al. 2016) was used to generate heatmaps and average profiles for various ChIP-seq, ATAC-seq, Pol-II phosphorylation forms as well as CG methylation profiles across these gene quantiles(q1-q4) in each CL.

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 XXXX.

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.

Figure Legends

Fig. 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.

Fig. 2: Histone variants define chromatin states in *Arabidopsis thaliana*.

(A) Heatmap showing the emission probabilities for each mark/variant across the 26 chromatin states.

(B) Bar plot representing the proportion of the genome occupied by each state (% Genome).

(C) Boxplot showing the expression of protein-coding genes overlapping with each chromatin state in Transcripts per Million (TPM).

- (D) Boxplot showing CG methylation levels across chromatin states.
- (E) Boxplot comparing DNase I-seq read coverage across states.
- (F) Bar plot showing the overlap between annotated genomic features and chromatin states.
- (G) Heatmap showing the Jaccard similarity index between the states generated using the whole model and states using a subset of marks, i.e. excluding a set of marks the X-axis from left to right (complete data, no H2B variants, no H3 variants, no H2A variants, no histone modifications, and no variants (H2A/H3/H2B)).
- (H) Network diagram representing the transition matrix of all chromatin states. The thickness of the arrows is proportional to the probability of the state transition. [The thicker the connection between the states, the more likely they are nearby on a linear scale and transition from one state to other.] (The probability for self-transition in the same state was omitted from the plots to simplify the diagram.)

Fig. 3: H2A.W.6 and H2A.W.7 variants together describe heterochromatin and repeats.

Heterochromatic states can be divided into H2AW6/H2AW7 states (H1–H4) and H2A.W.7-only states (H5–H6).

- (A) Genomic distribution of heterochromatic states (H1–H6) [bottom to top] along Chr1 in *Arabidopsis thaliana*.
- (B) Aggregate heatmap showing the signal of heterochromatic states plotted over Pol V ChIP-seq peaks.
- (C) Heatmap showing the overlap enrichment scores across TEs regulated by different DNA methylation pathways and their association with heterochromatic states.
- (D) Heatmap showing overlap enrichment scores across TE mis-regulated in different heterochromatic mutants and their association with heterochromatic states.
- (E) Aggregate ChIP-seq profiles and heatmaps for various histone modifications across the TE groups.
- (F) Boxplot showing the lengths of TEs in each TE group.
- (G, H) Stacked bar plot showing the proportion of each TE superfamily and TE pathway respectively in each TE group.
- (I) Boxplot showing the CG methylation levels in each TE group based on chromatin states.

Fig. 4: Most H2A.Z in the genome is associated with developmentally repressed genes.

There are two types of facultative chromatin states: H2AZ with H3K27me3 (F1–F3) and H2AZ with other repressive marks (F4–F6).

- (A) Schematic representation of six types of repressed genes the chromatin landscapes (CL). [CL1–CL3: Facultative repressed genes and CL4–CL6: Constitutively repressed genes.
- (B, C) Aggregate profile plots and heatmaps showing ChIP-seq signals for histone marks and variants relevant to each CL.
- (D) Boxplot comparing gene lengths across repressed CLs.
- (E) Boxplot comparing gene expression in TPM (transcripts per million) across repressed CLs.
- (F) Line plot showing % Genes expressed (TPM>1) within each CL in different tissue types across various developmental stages in Arabidopsis.

Fig. 5: Transcribed chromatin states assemble in a stereotypical fashion across active genes. Euchromatic states assemble in a stereotypic fashion across expressed protein-coding genes.

- (A) Schematic representation of states occupying the four types of euchromatic gene clusters CL7–CL10 [CL7: Regular active genes; CL8: Low CG methylated genes; CL9: Short H2AZ genes; and CL10: Bivalent genes].
- (B–D) Aggregation profile plots across the four types of active gene CLs of histone H2A variants (B), Pol II occupancy (C), and DNA methylation levels (D).
- (E) Aggregation profile and heatmap showing log₂(ChIP/Input) for chromatin marks that differentiate across the four gene CLs.
- (F, G) Boxplots showing gene expression (F) and gene length (G) across active gene clusters.

Supplementary Figure Legends

Fig. S1: Biochemical analysis of the association between histone variants and histone marks.

- (A) DNA in input samples used for H3.1 and H3.3 immunoprecipitation.
- (B) Silver-stained gel of immunoprecipitated H3.1 and H3.3 mononucleosomes.
- (C) DNA extracted from immunoprecipitated H3.1 and H3.3 mononucleosomes.
- (D) Spectral counts of H3.1- and H3.3-specific peptides in immunoprecipitates with H2A variants.
- (E) H3.1- and H3.3-specific peptides used for the analysis of H3 K27 and K36 modifications.
- (F) Number of MS spectra measured that covered the indicated residues.

- (G) Number of measured MS spectra containing the indicated modifications.
- (H) Sequences of peptides used to evaluate H3 acetylation by MS.
- (I) Relative acetylation levels of the indicated H3 residues in H3.1 and H3.3 immunoprecipitates.
- (J) Relative acetylation levels of the indicated H3 residues on H3.1 and H3.3 revealed from transgenic copies of the respective H3 variants.

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). (E) Number of MS spectra measured covering the indicated residues.

Fig. S2: Comparison between published chromatin states and chromatin states and properties associated with states identified in the current study.

- (A) Heatmap showing the correlation across states from different numbers of chromatin states. States model with 50 states was used to calculate correlation with other models with lower numbers of states (n=49: n=2).
- (B) Flow diagram showing the overlap (in bp) between published states (Sequeria-Mendes 2014 et al) and the current state model. Color code for the flow diagram represents the chromatin states in the current model.
- (C, D) Boxplots showing methylation levels for CHG and CHH across states.
- (E) Boxplot showing nucleosome occupancy (MNase-seq read density) across different states.
- (F) Pie plot showing the genome % of each state.
- (G) Overlap between non-protein-coding gene features and states.

Fig. S3: Transposable elements clusters and their properties

- (A) Genomic distribution of heterochromatic states (H1–H6) [bottom to top] along all chromosomes in *Arabidopsis thaliana*.
- (B) Heatmap showing the occupancy of heterochromatic chromatin states (H1–H6) across TE groups TE1–TE7.
- (C) Bar plot showing the number of TEs in each TE group.

- (D, E) Boxplot showing the methylation levels across TE groups from left to right (CHG and CHH, respectively).
- (F) Aggregate ChIP-seq profiles and heatmaps for H3 variants across the TE groups.
- (G) Aggregate ChIP-seq profiles and heatmaps for H2A variants across the TE groups.

Fig. S4: Chromatin landscapes across developmentally repressed genes

- (A) Heatmap displaying the occupancy of facultative chromatin states (F1–F6) across the CL1–CL3 genes.
- (B) Boxplot showing the expression levels of CL1–CL3 genes in PRC mutants.
- (C) Heatmap displaying the occupancy of heterochromatic chromatin states (H1–H6) across gene clusters CL4–CL6.
- (D) Aggregate profile plots and heatmaps showing ChIP-seq signals for H2A variants across CL4–CL6.
- (E) Aggregate profile plots showing ChIP-seq signals for H3.1 and H3.3 along repressed genes (CL1–CL6).

Fig. S5: Expressed genes chromatin landscapes

- (A) Heatmap showing the intergenic and euchromatic states across active protein-coding gene clusters (CL7–CL10).
- (B) Heatmap showing $\log_2(\text{ChIP}/\text{Input})$ for active chromatin marks across four types of gene clusters.

Fig. S6: Expressed genes chromatin landscapes. Aggregation profiles across CL7–CL10 genes showing $\log_2(\text{ChIP}/\text{Input})$ for chromatin marks. (A) H2A variants (B) H2B variants (C, D) Histone PTMs, and (E) H3 variants.

Fig. S7: Expression quantiles within chromatin landscapes. (A) Aggregation profiles across expression quantiles (q1= lowest; q4=highest) in CL7–CL10: **I**) Accessibility (ATAC-seq signal) across gene quantiles, **II**) CG methylation levels, and **III**) $\log_2(\text{ChIP}/\text{Input})$ signal for H3 variants across gene quantiles in each CL type.

Fig. S8: Expression quantiles within CLs and Pol II. (A) Aggregation profiles for Pol II occupancy across expression quantiles (q1= lowest; q4=highest) in each CL type (CL7–CL10).

Fig. S9: Expression quantiles within CLs and H2A and H2B variants. Aggregation profiles for log2 ChIP/Input signal for (A) H2A variants and (B) H2B variants across expression quantiles (q1= lowest; q4=highest) in each CL type (CL7–CL10).

Fig. S10: Expression quantiles within CLs and H3 modifications. Aggregation profiles for log2 ChIP/Input signal for (A) H2AZ and multiple H3 modifications across expression quantiles (q1= lowest; q4=highest) in each CL type (CL7–CL10).

Main Figures

Figure 1

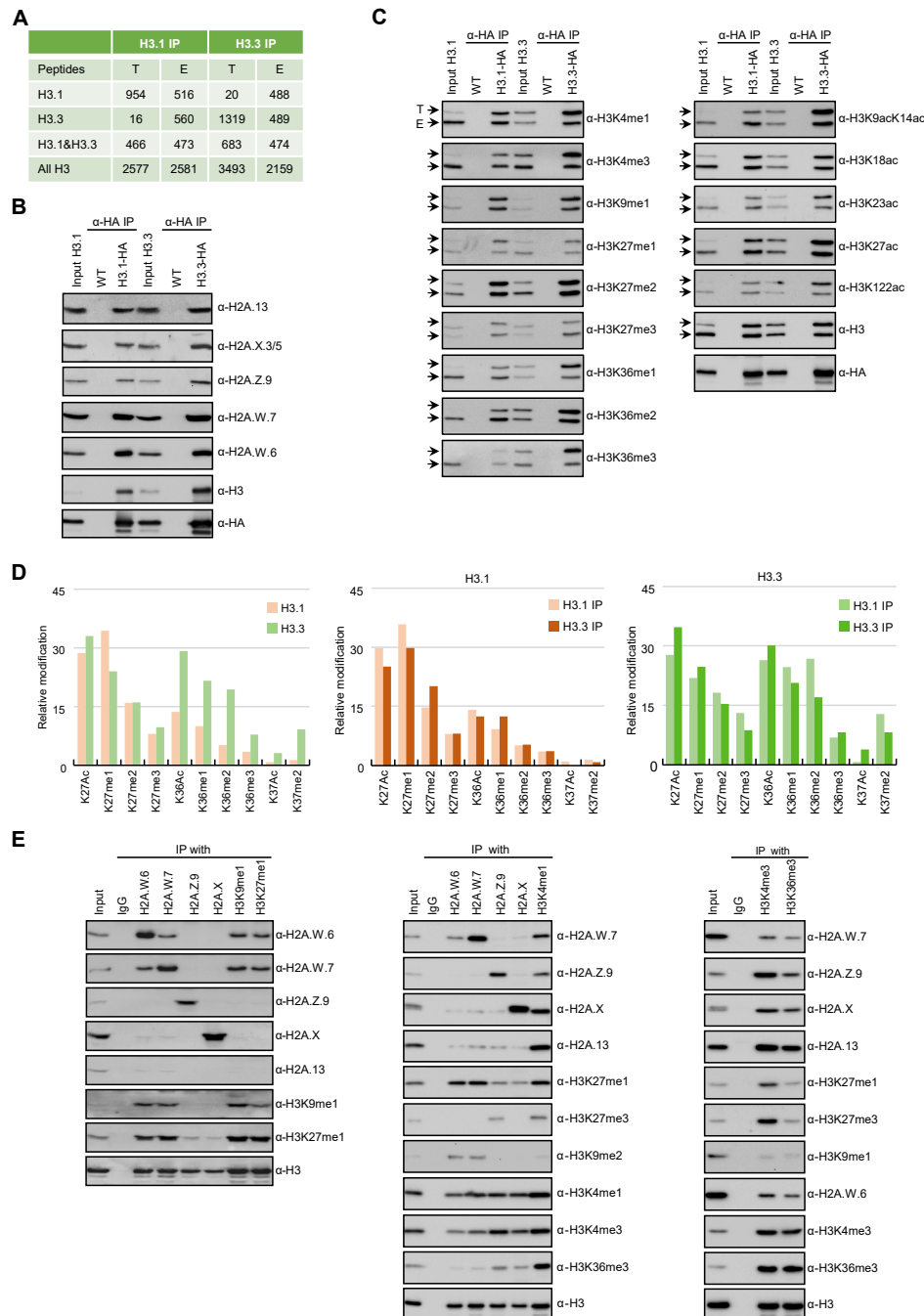


Figure 1

Figure 2

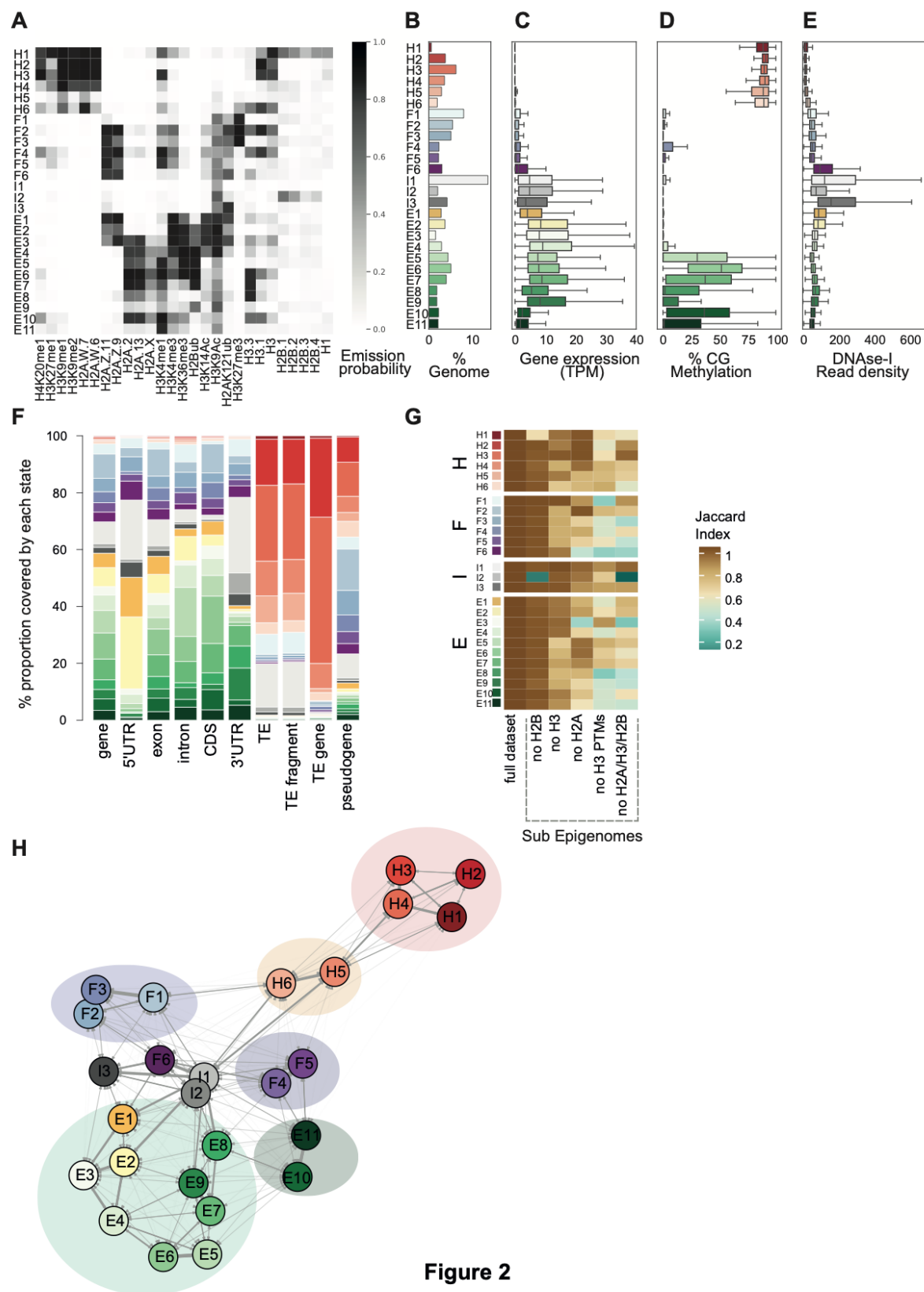


Figure 2

Figure 3

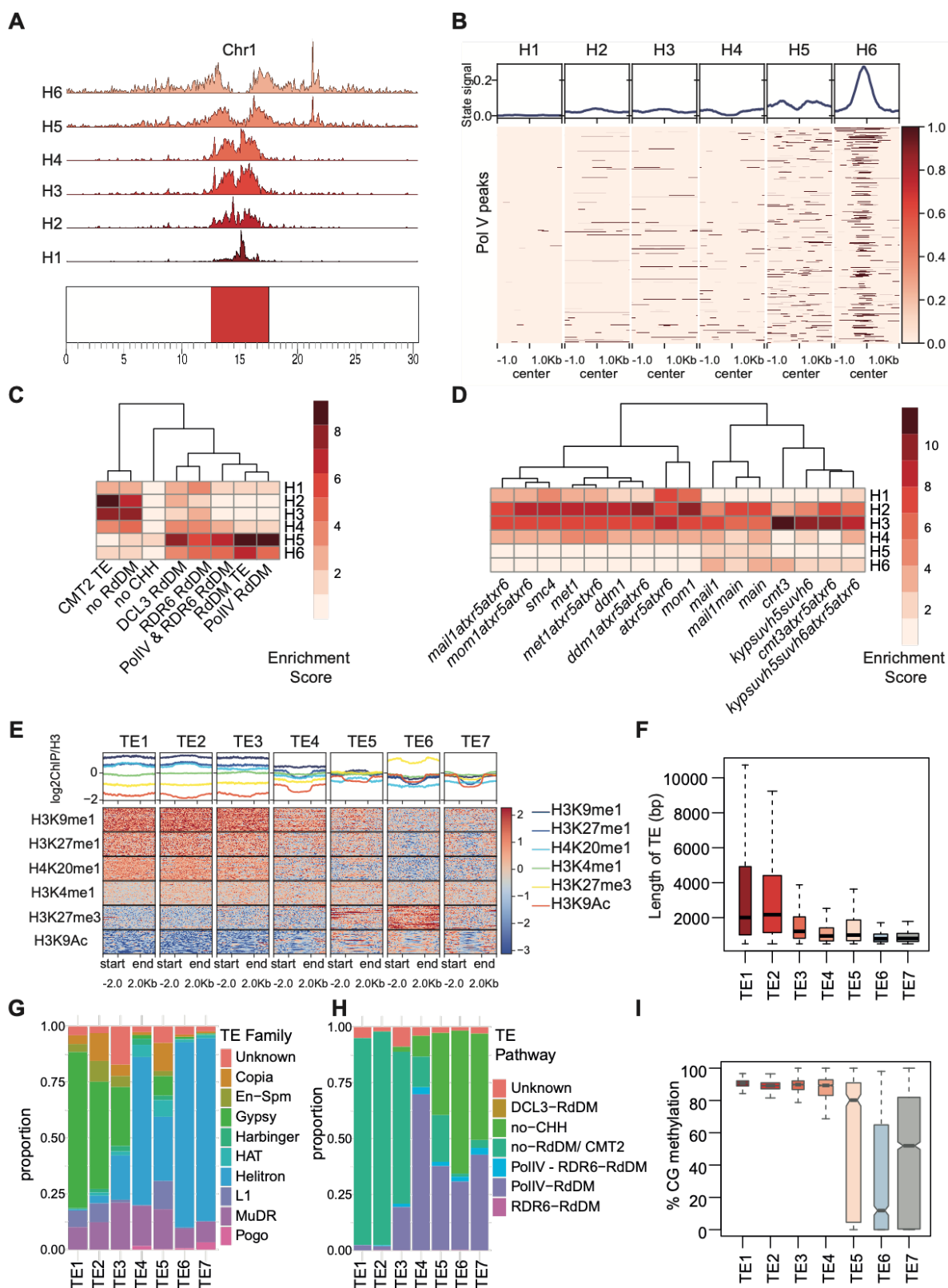


Figure 3

Figure 4

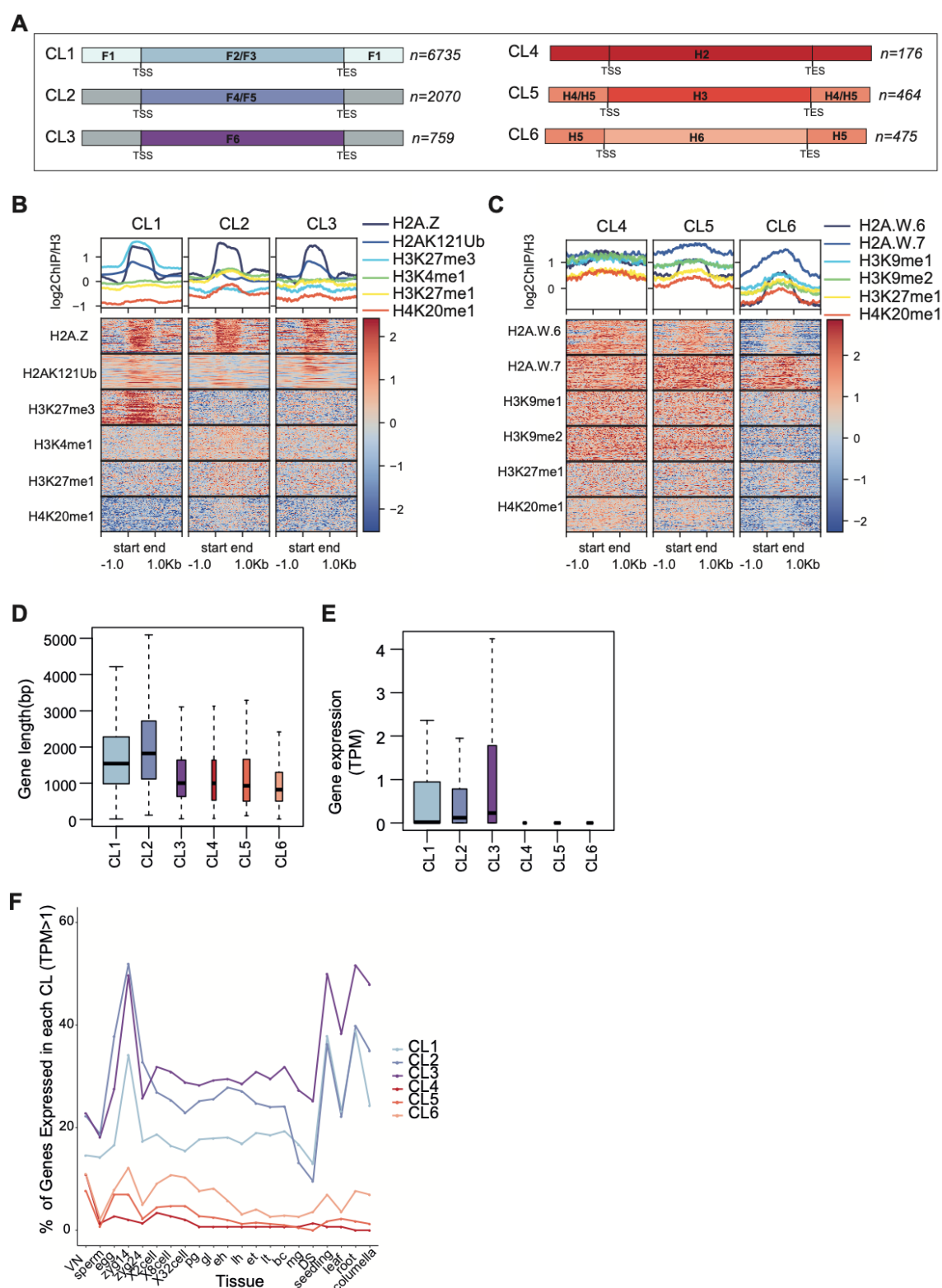


Figure 4

Figure 5

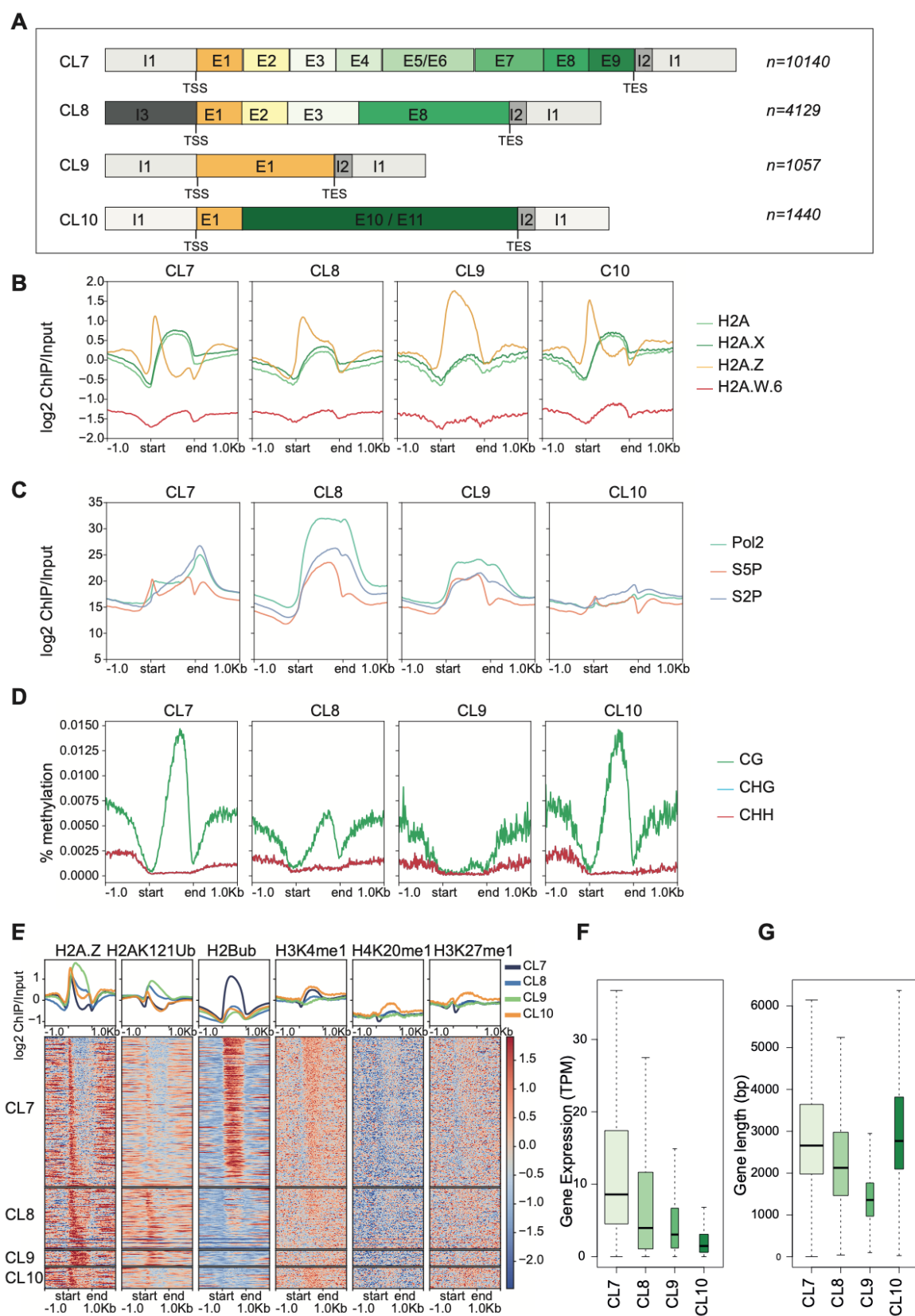


Figure 5

Supplementary Figures

Fig S1

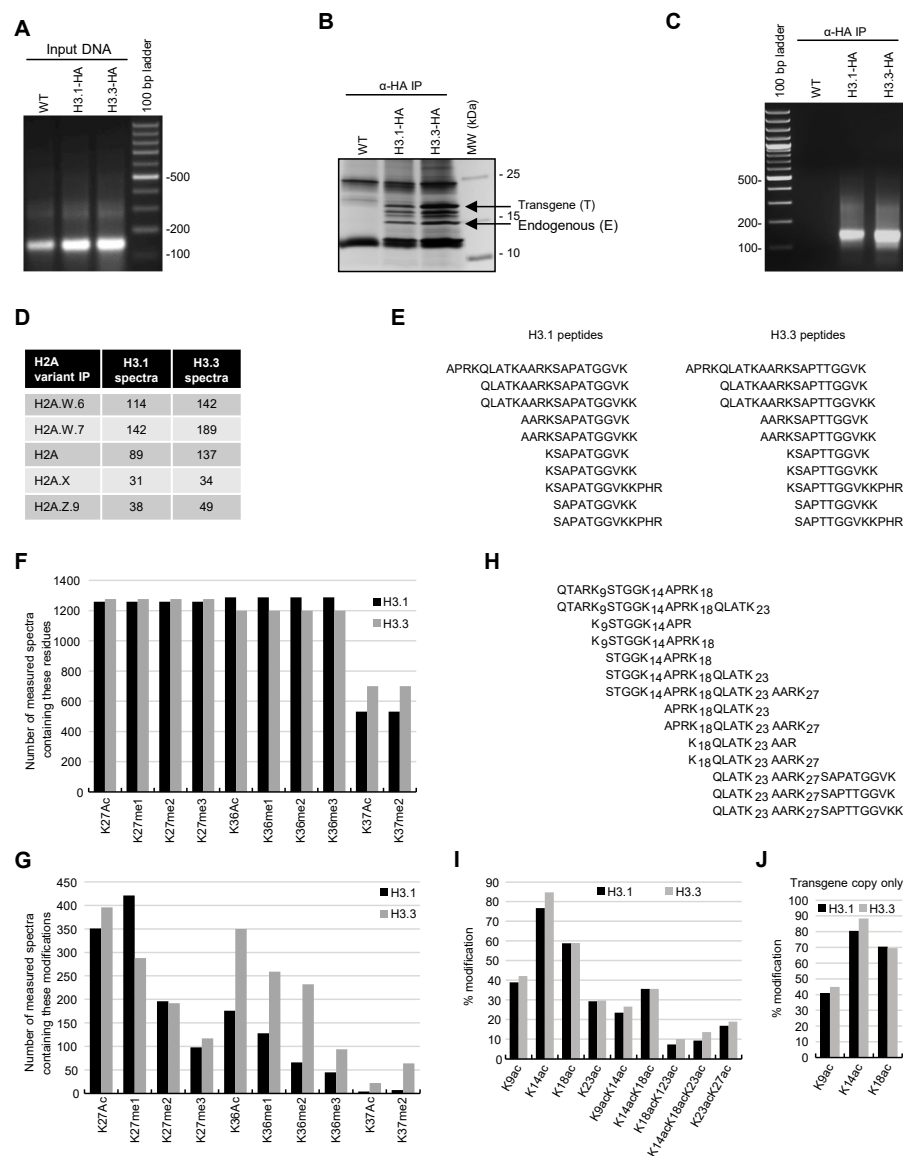


Figure S1

Fig S3

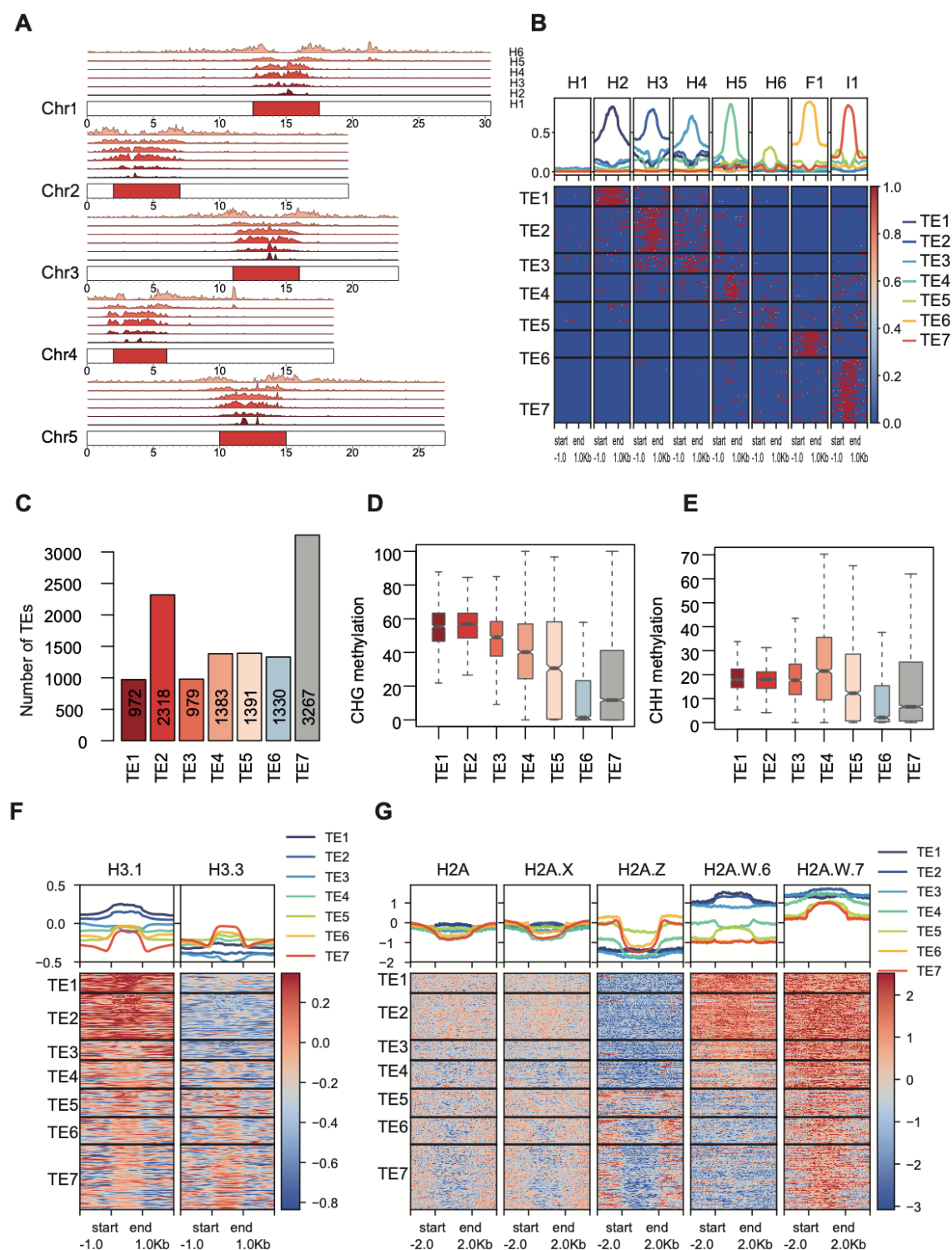


Figure S3

Fig S4

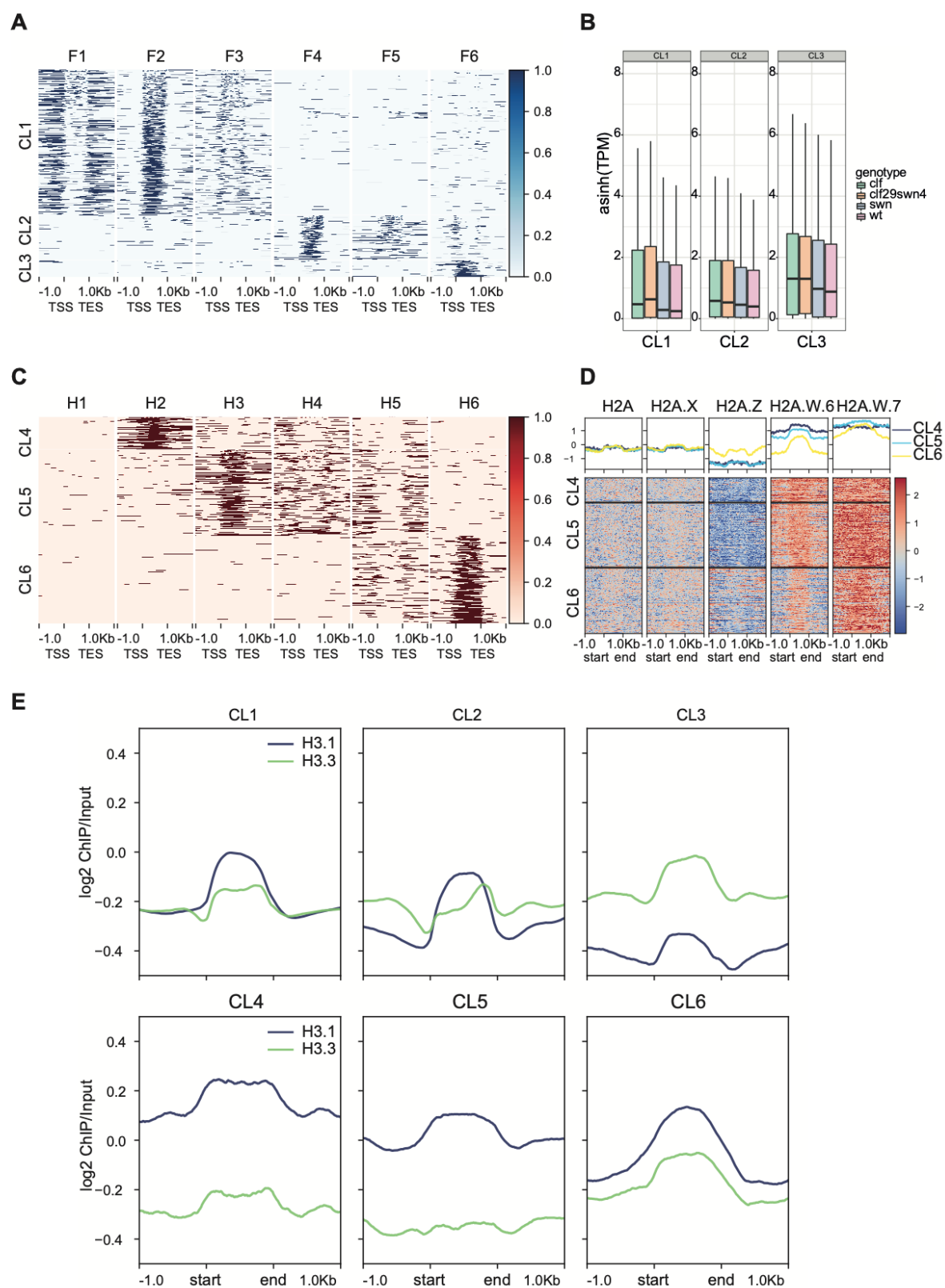
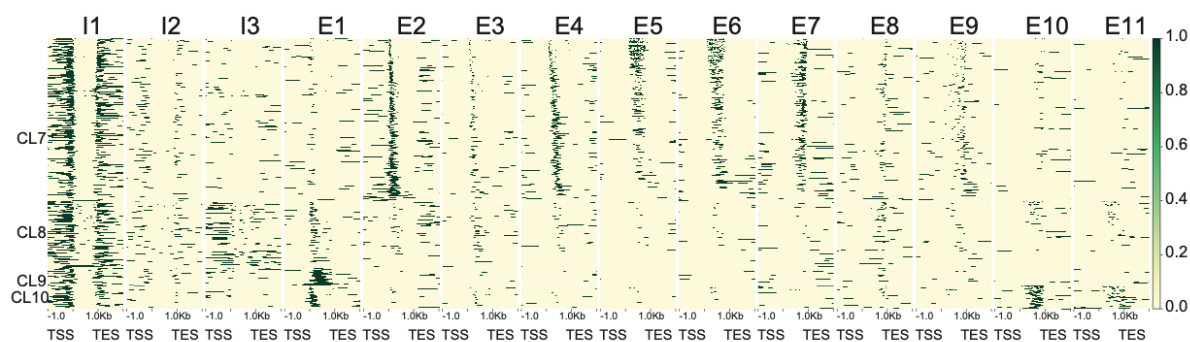


Figure S4

Fig S5

A



B

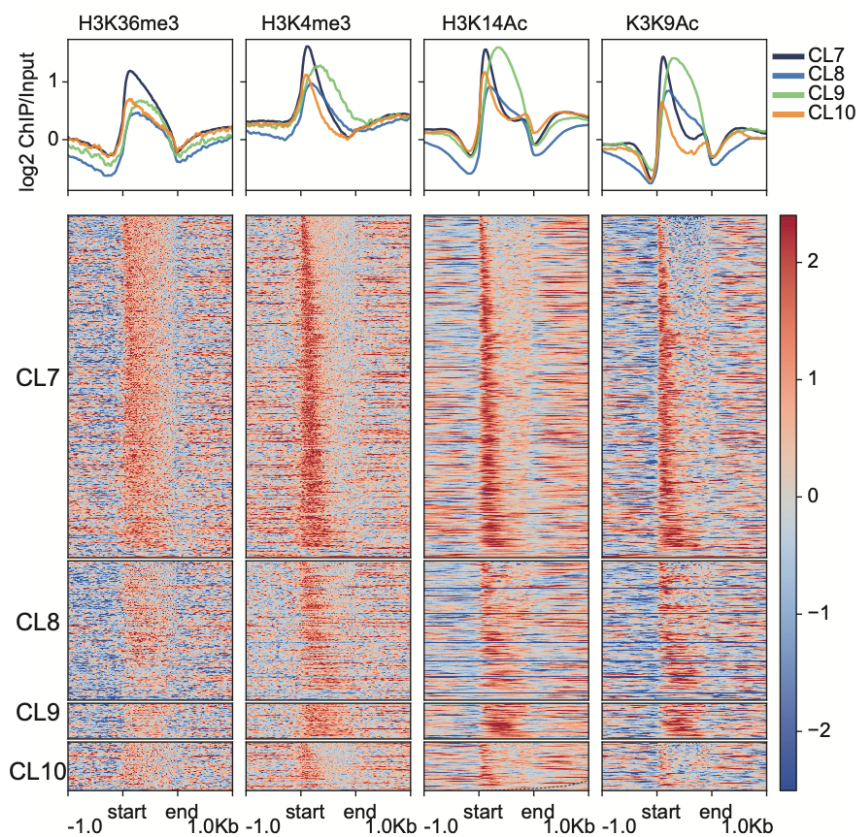


Figure S5

Fig S6

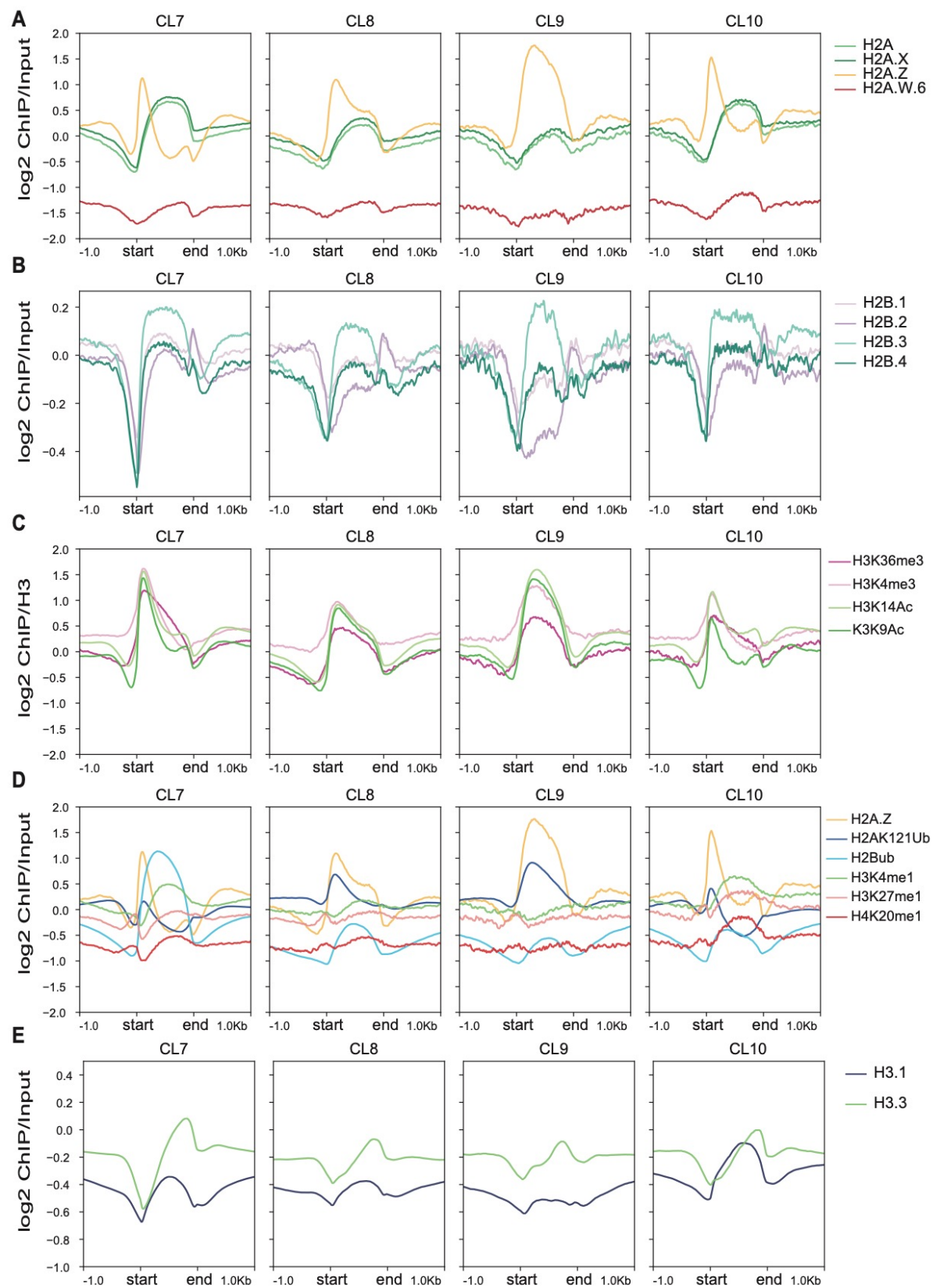


Figure S6

Fig S7

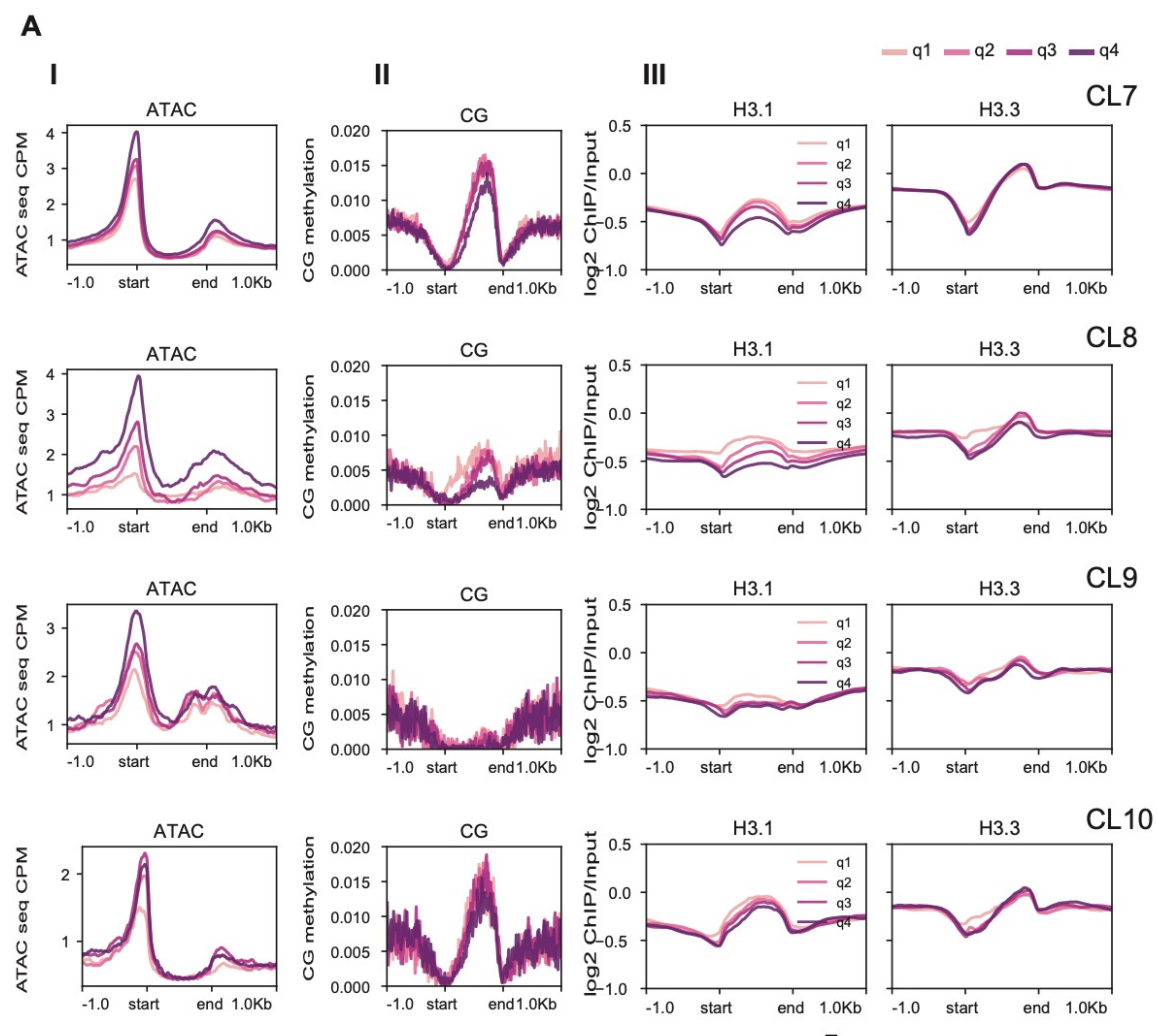


Figure S7

Fig S8

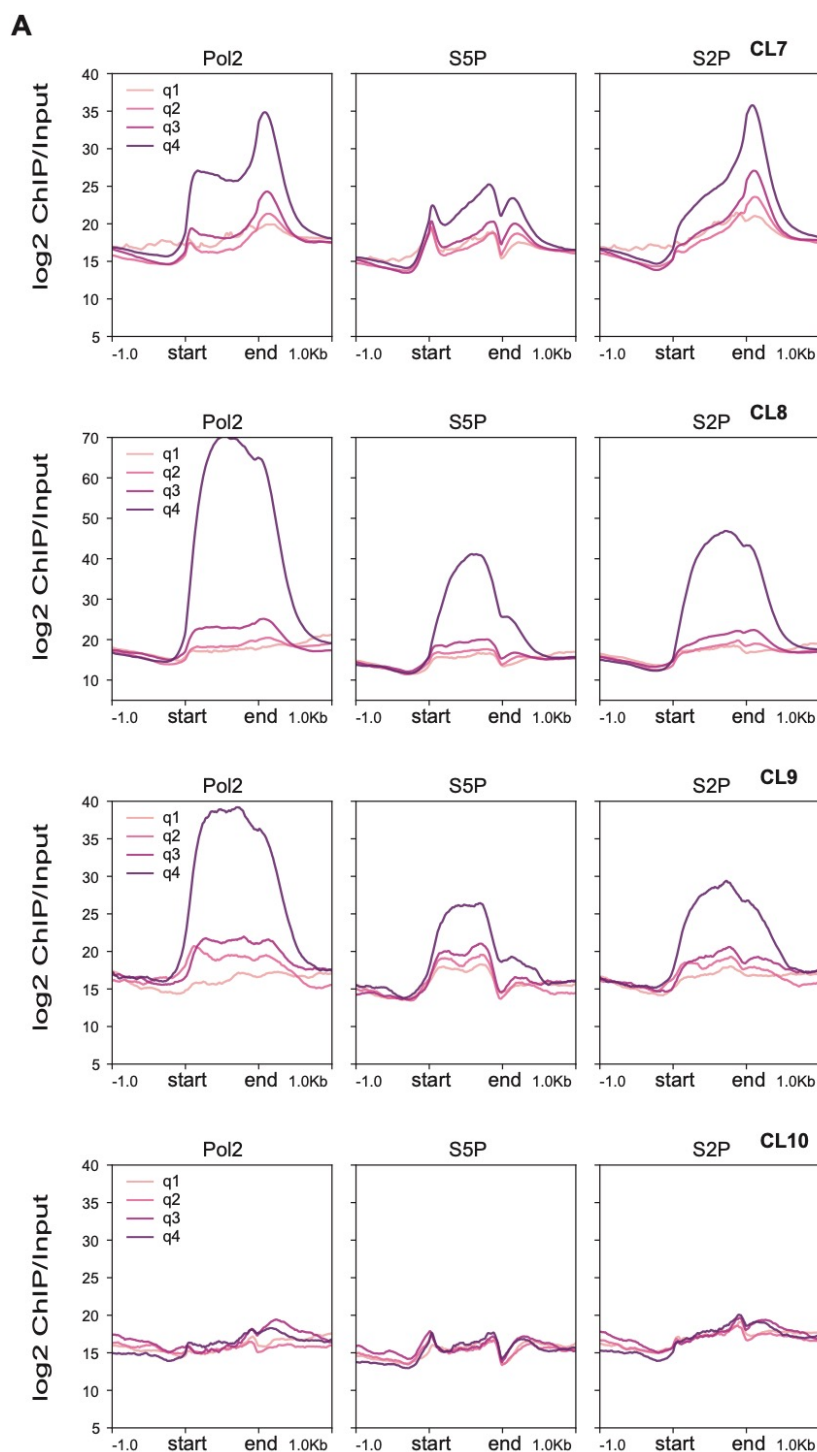


Figure S8

Fig S9

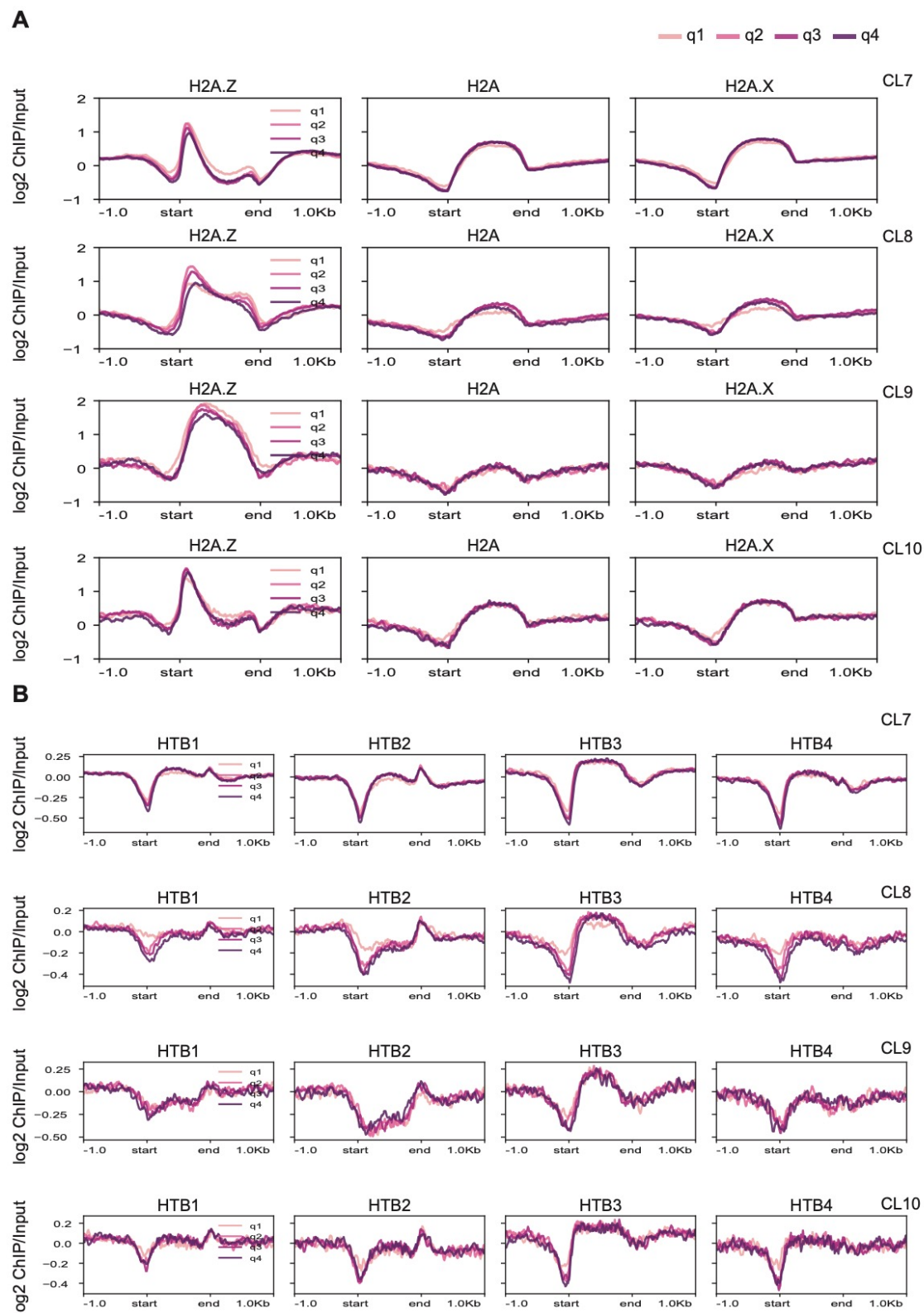


Figure S9

Fig S10

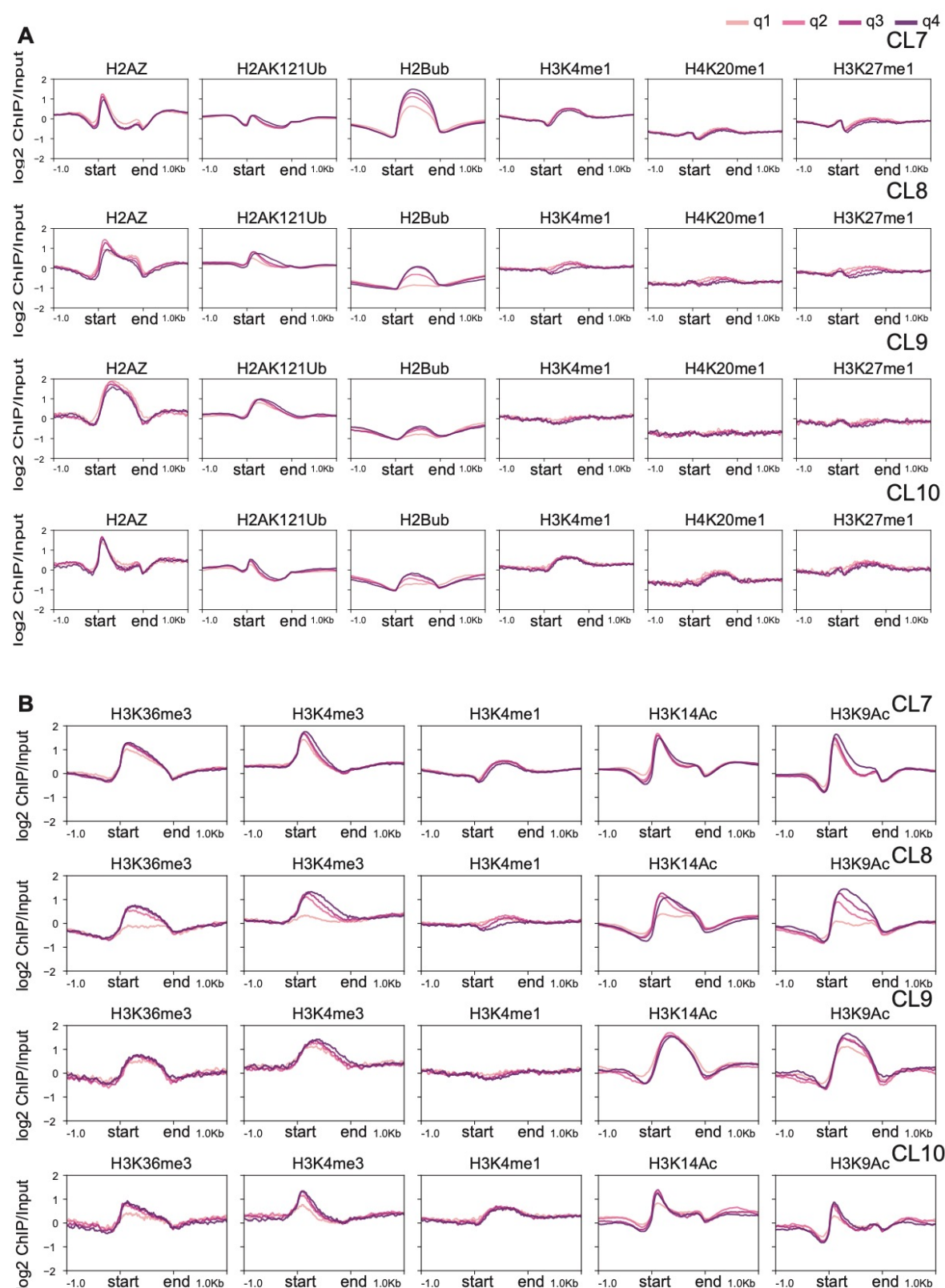


Figure S10

Tables

Table 1 Transposable Elements clusters and their chromatin properties.

#TE clusters	% TEs (TEs >500bp n= 11640)	State enriched	Variant	PTMs	Notes	Location
TE1	8.35% (n=972)	H2	H2A.W.6/7 H3.1	H3K9me1/2 H4K20me1	Long TE, CMT2, regulated, LTR Gypsy	Pericentromeric
TE2	19.9 % (n= 2318)	H3	H2A.W.6/7 H3.1	H3K9me1/2 H3K27me1 H4K20me1 H3K4me1	Long TE, CMT2 regulated, LTR Gypsy, Copia	Pericentromeric
TE3	9.4% (n=979)	H4	H2A.W.6/7 H3.1	H3K9me1/2 H3K27me1 H3K4me1	Long TE, CMT2 regulated, LTR Gypsy	Pericentromeric
TE4	11.8% (n=1383)	H5	H2A.W.7	H3K9me1	Short TE, RdDM, RC Helitrons	Chromosome arms
TE5	11.9% (n=1391)	H6	H2A.W.7	H3K9me1/2 H4K20me1 H3K4me1	Short TE, RdDM, L1, Copia, RC Helitrons	Chromosome arms
TE6	11.4% (n=1330)	F1	H2A.W.7	H3K27me3	Short TE, RC Helitrons	Chromosome arms and gene rich regions covered by Polycomb-regulated genes
TE7	28% (n=3267)	I1	H2A.W.7	Neighboring regions are covered with H3K9Ac	Short TE, RC Helitrons	Chromosome arms

Table 2 Specific chromatin landscape of each gene cluster. The proportion of genes, the main histone variant and modifications (PTMs), and specific traits are indicated.

#Chromatin Landscape	% Protein coding genes	Variant	PTMs	Notes	Gene Ontology terms
CL1	24.5% (n=6735)	H2A.Z H3.1	H3K27me3, H2AK121Ub	Strongly repressed by Polycomb, inserted in facultative heterochromatin. Encode developmental genes, miRNAs, lncRNAs	Development, process, and regulation and response related genes
CL2	7.5% (n=2070)	H2A.Z H3.1 covering gene body H3.3 only peaks at 3'	H4K20me1, H3K27me1	Repressed genes independent from Polycomb, inserted in euchromatin. Also include genes encoding snRNAs and snoRNAs	Cell cycle, protein modification related genes
CL3	2.7% (n=759)	H2A.Z H3.3	H2AK121Ub	Short genes weakly repressed by Polycomb and inserted in euchromatin	Defense and response genes
CL4	0.7% (n=176)	H2A.W H3.1	H3K9me1/2	Genes in pericentric heterochromatin, which are expressed during reproduction.	Regulation of fertilization
CL5	1.7% (n=464)	H2A.W H3.1	H3K9me1/2	Genes in pericentric heterochromatin	Fertilization
CL6	1.8% (n=475)	H2A.W.7 H3.1 H3.3	H3K27me1 H4K20me1	Genes in chromosome arms	Pollination
CL7	37.0% (n=10140)	H2A/X H2A.Z in 5'	H2Bub H3K36me3 H3K4me1 H3K4me3	Highly expressed ubiquitous housekeeping genes, high levels of gene body methylation. Longest genes	Metabolism, housekeeping genes
CL8	15.0% (n=4129)	H2A/X H2A.Z in 5'	H3K4me3 H3K36me3	Extended enrichment of H2A.Z inside the gene body and low DNA methylation. Shorter and expressed at lower levels than CL1 genes	Transport, signaling, and response
CL9	3.9% (n=1057)	H2A.Z	H2AK121Ub	Short genes devoid of intron and gene body methylation. Low levels of RNA Pol II. Resemble genes in clusters CL5 and CL7	Response to karrikin Only one GO terms found for this group
CL10	5.2% (n=1440)	H2A/X	H3K4me1 high H3K4me3 low H3K36me3 low H4K20me1 H3K27me1	Lowest levels of RNA Pol II, transcripts, and chromatin accessibility and high levels of gene body methylation. Similar to CL6 genes.	Cell cycle, chromosome segregation

References

- Albert, Istvan, Travis N. Mavrich, Lynn P. Tomsho, Ji Qi, Sara J. Zanton, Stephan C. Schuster, and B. Franklin Pugh. 2007. "Translational and Rotational Settings of H2A.Z Nucleosomes across the *Saccharomyces Cerevisiae* Genome." *Nature* 446 (7135): 572–76. <https://doi.org/10.1038/NATURE05632>.
- Allis, C. David, Claiborne V.C. Glover, Josephine K. Bowen, and Martin A. Gorovsky. 1980. "Histone Variants Specific to the Transcriptionally Active, Amitotically Dividing Macronucleus of the Unicellular Eucaryote, *Tetrahymena Thermophila*." *Cell* 20 (3): 609–17. [https://doi.org/10.1016/0092-8674\(80\)90307-4](https://doi.org/10.1016/0092-8674(80)90307-4).
- Andrews, Simon. 2010. "FastQC." *Babraham Bioinformatics*. <https://doi.org/citeulike-article-id:11583827>.
- Antosz, Wojciech, Alexander Pfab, Hans F. Ehrnsberger, Philipp Holzinger, Karin Köllen, Simon A. Mortensen, Astrid Bruckmann, et al. 2017. "The Composition of the Arabidopsis RNA Polymerase II Transcript Elongation Complex Reveals the Interplay between Elongation and MRNA Processing Factors." *The Plant Cell* 29 (4): 854–70. <https://doi.org/10.1105/TPC.16.00735>.
- Aslam, Mohammad, Beenish Fakher, Bello Hassan Jakada, Shijiang Cao, and Yuan Qin. 2019. "SWR1 Chromatin Remodeling Complex: A Key Transcriptional Regulator in Plants." *Cells* 8 (12). <https://doi.org/10.3390/CELLS8121621>.
- Bintu, Lacramioara, Toyotaka Ishibashi, Manchuta Dangkulwanich, Yueh Yi Wu, Lucyna Lubkowska, Mikhail Kashlev, and Carlos Bustamante. 2012. "Nucleosomal Elements That Control the Topography of the Barrier to Transcription." *Cell* 151 (4): 738–49. <https://doi.org/10.1016/J.CELL.2012.10.009>.
- Bondarenko, Vladimir A., Louise M. Steele, Andrea Újvári, Daria A. Gaykalova, Olga I. Kulaeva, Yury S. Polikanov, Donal S. Luse, and Vasily M. Studitsky. 2006. "Nucleosomes Can Form a Polar Barrier to Transcript Elongation by RNA Polymerase II." *Molecular Cell* 24 (3): 469–79. <https://doi.org/10.1016/j.molcel.2006.09.009>.
- Bourguet, Pierre, Ramesh Yelagandula, Taiko Kim To, Akihisa Osakabe, Archana Alishe, Rita Jui-Hsien Lu, Tetsuji Kakutani, Pao-Yang Chen, and Frédéric Berger. n.d. "The Histone Variant H2A.W Cooperates with Chromatin Modifications and Linker Histone H1 to Maintain Transcriptional Silencing of Transposons in Arabidopsis." Accessed June 2, 2022. <https://doi.org/10.1101/2022.05.31.493688>.
- Carter, Benjamin, Brett Bishop, Kwok Ki Ho, Ru Huang, Wei Jia, Heng Zhang, Pete E.

- Pascuzzi, Roger B. Deal, and Joe Ogas. 2018. “The Chromatin Remodelers PKL and PIE1 Act in an Epigenetic Pathway That Determines H3K27me3 Homeostasis in Arabidopsis.” *The Plant Cell*. <https://doi.org/10.1105/tpc.17.00867>.
- Costanzi, C., and J. R. Pehrson. 1998. “Histone MacroH2A1 Is Concentrated in the Inactive X Chromosome of Female Mammals.” *Nature* 393 (6685): 599–601. <https://doi.org/10.1038/31275>.
- Douet, Julien, David Corujo, Roberto Malinverni, Justine Renauld, Viola Sansoni, Melanija Posavec Marjanović, Neus Cantariño, et al. 2017. “MacroH2A Histone Variants Maintain Nuclear Organization and Heterochromatin Architecture.” *Journal of Cell Science* 130 (9): 1570–82. <https://doi.org/10.1242/jcs.199216>.
- Ernst, Jason, and Manolis Kellis. 2010. “Discovery and Characterization of Chromatin States for Systematic Annotation of the Human Genome.” *Nature Biotechnology* 28 (8): 817–25. <https://doi.org/10.1038/NBT.1662>.
- . 2012. “ChromHMM: Automating Chromatin-State Discovery and Characterization.” *Nature Methods*. <https://doi.org/10.1038/nmeth.1906>.
- . 2017. “Chromatin-State Discovery and Genome Annotation with ChromHMM.” *Nature Protocols* 12 (12): 2478–92. <https://doi.org/10.1038/NPROT.2017.124>.
- Feng, Jing, and Wen Hui Shen. 2014. “Dynamic Regulation and Function of Histone Monoubiquitination in Plants.” *Frontiers in Plant Science* 5 (MAR). <https://doi.org/10.3389/FPLS.2014.00083>.
- Fromm, Michael, and Zoya Avramova. 2014. “ATX1/AtCOMPASS and the H3K4me3 Marks: How Do They Activate Arabidopsis Genes?” *Current Opinion in Plant Biology* 21: 75–82. <https://doi.org/10.1016/J.PBI.2014.07.004>.
- Gates, Leah A., Charles E. Foulds, and Bert W. O’Malley. 2017. “Histone Marks in the ‘Driver’s Seat’: Functional Roles in Steering the Transcription Cycle.” *Trends in Biochemical Sciences* 42 (12): 977–89. <https://doi.org/10.1016/J.TIBS.2017.10.004>.
- Giaimo, Benedetto Daniele, Francesca Ferrante, Andreas Herchenröther, Sandra B. Hake, and Tilman Borggreffe. 2019. “The Histone Variant H2A.Z in Gene Regulation.” *Epigenetics & Chromatin* 12 (1). <https://doi.org/10.1186/S13072-019-0274-9>.
- Grau-Bové, Xavier, and Arnau Sebé-Pedrós. 2021. “Orthology Clusters from Gene Trees with Possvm.” *Molecular Biology and Evolution* 38 (11): 5204–8. <https://doi.org/10.1093/MOLBEV/MSAB234>.
- Hsu, Chen Jen, Oliver Meers, Marcus Buschbeck, and Florian H. Heidel. 2021. “The Role of MacroH2A Histone Variants in Cancer.” *Cancers* 13 (12).

- <https://doi.org/10.3390/CANCERS13123003>.
- Jacob, Yannick, Elisa Bergamin, Mark T.A. Donoghue, Vanessa Mongeon, Chantal LeBlanc, Philipp Voigt, Charles J. Underwood, et al. 2014. “Selective Methylation of Histone H3 Variant H3.1 Regulates Heterochromatin Replication.” *Science (New York, N.Y.)* 343 (6176): 1249. <https://doi.org/10.1126/SCIENCE.1248357>.
- Jiang, Danhua, Michael Borg, Zdravko J. Lorković, Sean A. Montgomery, Akihisa Osakabe, Ramesh Yelagandula, Elin Axelsson, and Frédéric Berger. 2020. “The Evolution and Functional Divergence of the Histone H2B Family in Plants.” *PLoS Genetics* 16 (7). <https://doi.org/10.1371/JOURNAL.PGEN.1008964>.
- Kharchenko, Peter V., Artyom A. Alekseyenko, Yuri B. Schwartz, Aki Minoda, Nicole C. Riddle, Jason Ernst, Peter J. Sabo, et al. 2011. “Comprehensive Analysis of the Chromatin Landscape in *Drosophila Melanogaster*.” *Nature* 471 (7339): 480–86. <https://doi.org/10.1038/NATURE09725>.
- Kim, Jaehoon, Mohamed Guermah, Robert K. McGinty, Jung Shin Lee, Zhanyun Tang, Thomas A. Milne, Ali Shilatifard, Tom W. Muir, and Robert G. Roeder. 2009. “RAD6-Mediated Transcription-Coupled H2B Ubiquitylation Directly Stimulates H3K4 Methylation in Human Cells.” *Cell* 137 (3): 459–71. <https://doi.org/10.1016/J.CELL.2009.02.027>.
- Kireeva, Maria L., Brynne Hancock, Gina H. Cremona, Wendy Walter, Vasily M. Studitsky, and Mikhail Kashlev. 2005. “Nature of the Nucleosomal Barrier to RNA Polymerase II.” *Molecular Cell* 18 (1): 97–108. <https://doi.org/10.1016/J.MOLCEL.2005.02.027>.
- Langmead and Steven L Salzberg. 2013. “Bowtie2.” *Nature Methods*. <https://doi.org/10.1038/nmeth.1923>.Fast.
- Lei, Bingkun, and Frédéric Berger. 2020. “H2A Variants in Arabidopsis: Versatile Regulators of Genome Activity.” *Plant Communications* 1 (1): 100015. <https://doi.org/10.1016/j.xplc.2019.100015>.
- Leng, Xueyuan, Quentin Thomas, Simon Horskjær Rasmussen, and Sebastian Marquardt. 2020. “A G(Enomic)P(ositioning)S(ystem) for Plant RNAPII Transcription.” *Trends in Plant Science* 25 (8): 744–64. <https://doi.org/10.1016/J.TPLANTS.2020.03.005>. <https://doi.org/10.1016/J.TPLANTS.2020.03.005>.
- Li, Guang, Shujing Liu, Jiawei Wang, Jianfeng He, Hai Huang, Yijing Zhang, and Lin Xu. 2014. “ISWI Proteins Participate in the Genome-Wide Nucleosome Distribution in Arabidopsis.” *The Plant Journal : For Cell and Molecular Biology* 78 (4): 706–14. <https://doi.org/10.1111/TPJ.12499>.

- Liu, Yue, Tian Tian, Kang Zhang, Qi You, Hengyu Yan, Nannan Zhao, Xin Yi, Wenying Xu, and Zhen Su. 2018. “PCSD: A Plant Chromatin State Database.” *Nucleic Acids Research* 46 (D1): D1157–67. <https://doi.org/10.1093/NAR/GKX919>.
- Loppin, Benjamin, and Frédéric Berger. 2020. “Histone Variants: The Nexus of Developmental Decisions and Epigenetic Memory.” *Annual Review of Genetics* 54 (November): 121–49. <https://doi.org/10.1146/ANNUREV-GENET-022620-100039>.
- Lorković, Zdravko J., Chulmin Park, Malgorzata Goiser, Danhua Jiang, Marie Therese Kurzbaue, Peter Schlögelhofer, and Frédéric Berger. 2017a. “Compartmentalization of DNA Damage Response between Heterochromatin and Euchromatin Is Mediated by Distinct H2A Histone Variants.” *Current Biology : CB* 27 (8): 1192–99. <https://doi.org/10.1016/J.CUB.2017.03.002>.
- Luger, Karolin, Armin W Mä Der, Robin K Richmond, David F Sargent, and Timothy J Richmond. 1997. “Crystal Structure of the Nucleosome Core Particle at 2.8 Å Resolution.” *Nature © Macmillan Publishers Ltd.* Vol. 389.
- Macadangdang, Benjamin R., Amit Oberai, Tanya Spektor, Oscar A. Campos, Fang Sheng, Michael F. Carey, Maria Vogelauer, and Siavash K. Kurdastani. 2014. “Evolution of Histone 2A for Chromatin Compaction in Eukaryotes.” *ELife* 2014 (3). <https://doi.org/10.7554/ELIFE.02792>.
- Matzke, Marjori, Tatsuo Kanno, Lucia Daxinger, Bruno Huettel, and Antonius JM Matzke. 2009. “RNA-Mediated Chromatin-Based Silencing in Plants.” *Current Opinion in Cell Biology* 21 (3): 367–76. <https://doi.org/10.1016/J.CEB.2009.01.025>.
- McGinty, Robert K., Jaehoon Kim, Champak Chatterjee, Robert G. Roeder, and Tom W. Muir. 2008. “Chemically Ubiquitylated Histone H2B Stimulates HDot1L-Mediated Intranucleosomal Methylation.” *Nature* 2008 453:7196 453 (7196): 812–16. <https://doi.org/10.1038/nature06906>.
- Meneghini, Marc D., Michelle Wu, and Hiten D. Madhani. 2003. “Conserved Histone Variant H2A.Z Protects Euchromatin from the Ectopic Spread of Silent Heterochromatin.” *Cell* 112 (5): 725–36. [https://doi.org/10.1016/S0092-8674\(03\)00123-5](https://doi.org/10.1016/S0092-8674(03)00123-5).
- Minsky, Neri, Efrat Shema, Yair Field, Meromit Schuster, Eran Segal, and Moshe Oren. 2008. “Monoubiquitinated H2B Is Associated with the Transcribed Region of Highly Expressed Genes in Human Cells.” *Nature Cell Biology* 10 (4): 483–88. <https://doi.org/10.1038/NCB1712>.
- Naish, Matthew, Michael Alonge, Piotr Wlodzimierz, Andrew J. Tock, Bradley W.

- Abramson, Anna Schmücker, Terezie Mandáková, et al. 2021. “The Genetic and Epigenetic Landscape of the Arabidopsis Centromeres.” *Science* 374 (6569).
https://doi.org/10.1126/SCIENCE.ABI7489/SUPPL_FILE/SCIENCE.ABI7489_MДАР_REPRODUCIBILITY_CHECKLIST.PDF.
- Nie, Xin, Haifeng Wang, Jing Li, Sarah Holec, and Frédéric Berger. 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/-/DC1>.
- Nielsen, Mathias, Ryan Ard, Xueyuan Leng, Maxim Ivanov, Peter Kindgren, Vicent Pelechano, and Sebastian Marquardt. 2019. “Transcription-Driven Chromatin Repression of Intragenic Transcription Start Sites.” *PLoS Genetics* 15 (2).
<https://doi.org/10.1371/journal.pgen.1007969>.
- Osakabe, Akihisa, Bhagyshree Jamge, Elin Axelsson, Sean A. Montgomery, Svetlana Akimcheva, Annika Luisa Kuehn, Rahul Pisupati, et al. 2021a. “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, Akihisa, Zdravko J. Lorković, Wataru Kobayashi, Hiroaki Tachiwana, Ramesh Yelagandula, Hitoshi Kurumizaka, and Frédéric Berger. 2018. “Histone H2A Variants Confer Specific Properties to Nucleosomes and Impact on Chromatin Accessibility.” *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gky540>.
- Pfab, Alexander, Matthias Breindl, and Klaus D. Grasser. 2018. “The Arabidopsis Histone Chaperone FACT Is Required for Stress-Induced Expression of Anthocyanin Biosynthetic Genes.” *Plant Molecular Biology* 96 (4–5): 367–74.
<https://doi.org/10.1007/s11103-018-0701-5>.
- Quadrana, Leandro. 2020. “The Contribution of Transposable Elements to Transcriptional Novelty in Plants: The FLC Affair.” *Transcription* 11 (3–4): 192.
<https://doi.org/10.1080/21541264.2020.1803031>.
- Raisner, Ryan M., Paul D. Hartley, Marc D. Meneghini, Marie Z. Bao, Chih Long Liu, Stuart L. Schreiber, Oliver J. Rando, and Hiten D. Madhani. 2005. “Histone Variant H2A.Z Marks the 5’ Ends of Both Active and Inactive Genes in Euchromatin.” *Cell* 123 (2): 233–48. <https://doi.org/10.1016/J.CELL.2005.10.002>.
- Raman, Pravurutha, Mary C. Rominger, Janet M. Young, Antoine Molaro, Toshio Tsukiyama, and Harmit S. Malik. 2022. “Novel Classes and Evolutionary Turnover of Histone H2B

- Variants in the Mammalian Germline.” *Molecular Biology and Evolution* 39 (2).
<https://doi.org/10.1093/molbev/msac019>.
- Ramírez, Fidel, Devon P. Ryan, Björn Grüning, Vivek Bhardwaj, Fabian Kilpert, Andreas S. Richter, Steffen Heyne, Friederike Dündar, and Thomas Manke. 2016. “DeepTools2: A next Generation Web Server for Deep-Sequencing Data Analysis.” *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkw257>.
- Rogakou, Emmy P., Duane R. Pilch, Ann H. Orr, Vessela S. Ivanova, and William M. Bonner. 1998. “DNA Double-Stranded Breaks Induce Histone H2AX Phosphorylation on Serine 139.” *The Journal of Biological Chemistry* 273 (10): 5858–68.
<https://doi.org/10.1074/JBC.273.10.5858>.
- Roudier, François, Ikhlaq Ahmed, Caroline Bérard, Alexis Sarazin, Tristan Mary-Huard, Sandra Cortijo, Daniel Bouyer, et al. 2011. “Integrative Epigenomic Mapping Defines Four Main Chromatin States in Arabidopsis.” *EMBO Journal*.
<https://doi.org/10.1038/emboj.2011.103>.
- Rudnizky, Sergei, Adaiah Bavly, Omri Malik, Lilach Pnueli, Philippa Melamed, and Ariel Kaplan. 2016. “H2A.Z Controls the Stability and Mobility of Nucleosomes to Regulate Expression of the LH Genes.” *Nature Communications* 7 (September).
<https://doi.org/10.1038/NCOMMS12958>.
- Schmücker, Anna, Bingkun Lei, Zdravko J. Lorković, Matías Capella, Sigurd Braun, Pierre Bourguet, Olivier Mathieu, Karl Mechtler, and Frédéric Berger. 2021. “Crosstalk between H2A Variant-Specific Modifications Impacts Vital Cell Functions.” *PLoS Genetics* 17 (6). <https://doi.org/10.1371/journal.pgen.1009601>.
- Sequeira-Mendes, J., I. Araguez, R. Peiro, R. Mendez-Giraldez, X. Zhang, S. E. Jacobsen, U. Bastolla, and C. Gutierrez. 2014. “The Functional Topography of the Arabidopsis Genome Is Organized in a Reduced Number of Linear Motifs of Chromatin States.” *The Plant Cell*. <https://doi.org/10.1105/tpc.114.124578>.
- Sims, Jason, Giovanni Sestini, Christiane Elgert, Arndt von Haeseler, and Peter Schlögelhofer. 2021. “Sequencing of the Arabidopsis NOR2 Reveals Its Distinct Organization and Tissue-Specific rRNA Ribosomal Variants.” *Nature Communications* 12 (1). <https://doi.org/10.1038/S41467-020-20728-6>.
- Stroud, Hume, Sofia Otero, Bénédicte Desvoves, Elena Ramírez-Parra, Steven E. Jacobsen, and Crisanto Gutierrez. 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): 5370–75.

- <https://doi.org/10.1073/PNAS.1203145109>.
- Subramanian, Vidya, Paul A. Fields, and Laurie A. Boyer. 2015. "H2A.Z: A Molecular Rheostat for Transcriptional Control." *F1000prime Reports* 7 (January).
<https://doi.org/10.12703/P7-01>.
- Sun, Zhen, and Emily Bernstein. 2019. "Histone Variant MacroH2A: From Chromatin Deposition to Molecular Function." *Essays in Biochemistry* 63 (1): 59–74.
<https://doi.org/10.1042/EBC20180062>.
- Surface, Lauren E., Paul A. Fields, Vidya Subramanian, Russell Behmer, Namrata Udeshi, Sally E. Peach, Steven A. Carr, Jacob D. Jaffe, and Laurie A. Boyer. 2016. "H2A.Z.1 Monoubiquitylation Antagonizes BRD2 to Maintain Poised Chromatin in ESCs." *Cell Reports* 14 (5): 1142–55. <https://doi.org/10.1016/J.CELREP.2015.12.100>.
- Talbert, Paul B., and Steven Henikoff. 2021. "Histone Variants at a Glance." *Journal of Cell Science* 134 (6). <https://doi.org/10.1242/JCS.244749>.
- Tessarz, Peter, and Tony Kouzarides. 2014. "Histone Core Modifications Regulating Nucleosome Structure and Dynamics." *Nature Reviews. Molecular Cell Biology* 15 (11): 703–8. <https://doi.org/10.1038/NRM3890>.
- Venters, Bryan J., Shinichiro Wachi, Travis N. Mavrich, Barbara E. Andersen, Peony Jena, Andrew J. Sinnamon, Priyanka Jain, et al. 2011. "A Comprehensive Genomic Binding Map of Gene and Chromatin Regulatory Proteins in *Saccharomyces*." *Molecular Cell* 41 (4): 480–92. <https://doi.org/10.1016/J.MOLCEL.2011.01.015>.
- Wang, Haifeng, Danhua Jiang, Elin Axelsson, Zdravko J. Lorković, Sean Montgomery, Sarah Holec, Bas J.G.E. Pieters, Abbas H.K. Al Temimi, Jasmin Mecinović, and Frédéric Berger. 2018. "LHP1 Interacts with ATRX through Plant-Specific Domains at Specific Loci Targeted by PRC2." *Molecular Plant* 11 (8): 1038–52.
<https://doi.org/10.1016/J.MOLP.2018.05.004>.
- Wang, Jianhao, Sujuan Gao, Xiuling Peng, Keqiang Wu, and Songguang Yang. 2019. "Roles of the INO80 and SWR1 Chromatin Remodeling Complexes in Plants." *International Journal of Molecular Sciences* 20 (18). <https://doi.org/10.3390/IJMS20184591>.
- Wang, Yan, Haizhen Long, Juan Yu, Liping Dong, Michel Wassef, Baowen Zhuo, Xia Li, et al. 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). <https://doi.org/10.1186/S12915-018-0568-6>.
- Watanabe, Kiyoshi, and Tetsuro Kokubo. 2017. "SAGA Mediates Transcription from the TATA-like Element Independently of Taf1p/TFIID but Dependent on Core Promoter

- Structures in *Saccharomyces Cerevisiae*.” *PLoS ONE* 12 (11).
<https://doi.org/10.1371/journal.pone.0188435>.
- Weber, Christopher M., and Steven Henikoff. 2014. “Histone Variants: Dynamic Punctuation in Transcription.” *Genes and Development* 28 (7): 672–82.
<https://doi.org/10.1101/GAD.238873.114>.
- Wojcik, Felix, Geoffrey P. Dann, Leslie Y. Beh, Galia T. Debelouchina, Raphael Hofmann, and Tom W. Muir. 2018. “Functional Crosstalk between Histone H2B Ubiquitylation and H2A Modifications and Variants.” *Nature Communications* 9 (1).
<https://doi.org/10.1038/S41467-018-03895-5>.
- Wollmann, Heike, Sarah Holec, Keith Alden, Neil D. Clarke, Pierre Étienne Jacques, and Frédéric Berger. 2012. “Dynamic Deposition of Histone Variant H3.3 Accompanies Developmental Remodeling of the Arabidopsis Transcriptome.” *PLoS Genetics* 8 (5).
<https://doi.org/10.1371/JOURNAL.PGEN.1002658>.
- Wollmann, Heike, Hume Stroud, Ramesh Yelagandula, Yoshiaki Tarutani, Danhua Jiang, Li Jing, Bhagyshree Jamge, et al. 2017. “The Histone H3 Variant H3.3 Regulates Gene Body DNA Methylation in Arabidopsis Thaliana.” *Genome Biology* 18 (1).
<https://doi.org/10.1186/S13059-017-1221-3>.
- Yelagandula, Ramesh, Hume Stroud, Sarah Holec, Keda Zhou, Suhua Feng, Xuehua Zhong, Uma M. Muthurajan, et al. 2014. “The Histone Variant H2A.W Defines Heterochromatin and Promotes Chromatin Condensation in Arabidopsis.” *Cell*.
<https://doi.org/10.1016/j.cell.2014.06.006>.
- Zhao, Wei, Pia Neyt, Mieke Van Lijsebettens, Wen Hui Shen, and Alexandre Berr. 2019. “Interactive and Noninteractive Roles of Histone H2B Monoubiquitination and H3K36 Methylation in the Regulation of Active Gene Transcription and Control of Plant Growth and Development.” *The New Phytologist* 221 (2): 1101–16.
<https://doi.org/10.1111/NPH.15418>.

3.3 Publication II

The chromatin remodeler DDM1 prevents transposon mobility through deposition of histone variant H2A.W.

Osakabe, A.* , Jamge, B.* , Axelsson, E., Montgomery, S.A., Akimcheva, S., Kuehn, A.L., Pisupati, R., Lorković, Z.J., Yelagandula, R., Kakutani, T. and Berger, F., 2021.

Published in Nature Cell Biology

Nat Cell Biol 23, 391–400 (2021). <https://doi.org/10.1038/s41556-021-00658-1>

DOI : 10.1038/s41556-021-00658-1



The chromatin remodeler DDM1 prevents transposon mobility through deposition of histone variant H2A.W

Akihisa Osakabe^{1,5,8}, Bhagyshree Jamge^{1,8}, Elin Axelsson¹, Sean A. Montgomery¹, Svetlana Akimcheva¹, Annika Luisa Kuehn^{1,6}, Rahul Pisupati¹, Zdravko J. Lorković¹, Ramesh Yelagandula^{1,7}, Tetsuji Kakutani^{1,2,3,4} and Frédéric Berger¹✉

Mobile transposable elements (TEs) not only participate in genome evolution but also threaten genome integrity. In healthy cells, TEs that encode all of the components that are necessary for their mobility are specifically silenced, yet the precise mechanism remains unknown. Here, we characterize the mechanism used by a conserved class of chromatin remodelers that prevent TE mobility. In the *Arabidopsis* chromatin remodeler DECREASE IN DNA METHYLATION 1 (DDM1), we identify two conserved binding domains for the histone variant H2A.W, which marks plant heterochromatin. DDM1 is necessary and sufficient for the deposition of H2A.W onto potentially mobile TEs, yet does not act on TE fragments or host protein-coding genes. DDM1-mediated H2A.W deposition changes the properties of chromatin, resulting in the silencing of TEs and, therefore, prevents their mobility. This distinct mechanism provides insights into the interplay between TEs and their host in the contexts of evolution and disease, and potentiates innovative strategies for targeted gene silencing.

Transposable elements (TEs) are mobile DNA sequences that self-propagate and comprise a substantial proportion of eukaryotic genomes^{1–3}. TE mobilization, which is also known as transposition, carries regulatory elements that rewire gene regulatory networks^{4,5}. As a result, transposition provides rich substrates for evolution^{6–9}. At the scale of an individual organism, TE expression threatens genome stability due to their potentially disruptive genomic insertions, deletions and rearrangements¹⁰. Thus, specific mechanisms must limit transposition in healthy somatic cells, and deviation from this state leads to diseases¹¹. In invertebrates, fungi and plants, non-coding RNAs target H3K9 methylation to form repressive heterochromatin at TEs^{1,2}. In *Drosophila*, Piwi-interacting RNA (piRNAs) are small RNAs that are produced by long precursors expressed in the germline^{12–14}. piRNAs drive the deposition of the heterochromatin mark H3K9me3 that recruits heterochromatin protein 1 (HP1), which compacts chromatin leading to transcriptional silencing^{15–21}. The role of piRNAs is conserved across metazoans but is not general to all eukaryotes. Although an HP1 orthologue exists in plants, it has no impact on transposons¹⁵. Plant short interfering RNAs (siRNAs) produced by the RNA-dependent DNA methylation (RdDM) pathway guide the deposition of DNA methylation on transposons but, on its own, this pathway is ineffective at maintaining the silencing of transposon-coding regions and does not prevent transposition^{17–19}. Furthermore, the loss of DNA methylation in mutants deprived of the maintenance DNA METHYLTRANSFERASE 1 (MET1) or the plant-specific methyltransferases CHROMOMETHYLASES 2 and 3 (CMT2 and CMT3) does not cause transposition^{22,23}. Only the joint loss of the activity of

several DNA methyltransferases causes a strong derepression of TEs and transposition^{22–24}, but whether this effect is direct or indirect is unknown.

In flowering plants, single mutations in the SWI/SNF2 chromatin remodelling factor DECREASE IN DNA METHYLATION 1 (DDM1) are characterized by massive transposition events leading to genome instability^{22,25–31}. Loss of DDM1 also causes a global reduction in DNA methylation and H3K9 methylation^{32,33}, which was assumed to be responsible for the loss of TE silencing^{28,33,34}. In plants, DNA methylation is deposited both in the CG context by the DNA methyltransferase MET1, and the non-CG context by the methyltransferases CMT2 and CMT3 and de novo methyltransferases from the RdDM pathway¹⁷. Most of the TEs that are upregulated in *ddm1* mutants are methylated in the non-CG context by CMT2, but not by the RdDM pathway^{35,36}. Other pathways that are associated with the deposition of the heterochromatic modifications H3K9me1/2 and H3K27me1 do not impact silencing of TEs with an amplitude that is comparable to DDM1 (ref. ²⁴). These observations suggest that the mechanism of DDM1 does not rely directly on the feedback loops involving H3K9 methylation, non-coding RNAs and DNA methylation. This idea is also supported by the fact that the loss of the DDM1 orthologue in the fungus *Neurospora crassa* does not affect DNA methylation³⁷. The mammalian DDM1 orthologues LSH and HELLS do silence transposons³⁸, suggesting that DDM1 orthologues share a common mechanism. Similar to the loss of DDM1 in plants, the loss of LSH causes DNA demethylation^{39,40}, but recent evidence shows that this effect is indirect⁴¹. Furthermore, it has been proposed that DDM1 and its orthologue LSH provide

¹Gregor Mendel Institute (GMI), Austrian Academy of Sciences, Vienna Biocenter (VBC), Vienna, Austria. ²National Institute of Genetics, Mishima, Japan.

³Department of Genetics, School of Life science, The Graduate University of Advanced Studies (SOKENDAI), Mishima, Japan. ⁴Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Tokyo, Japan. ⁵Present address: Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Tokyo, Japan. ⁶Present address: Department of Chromatin Regulation, Max Planck Institute of Immunobiology and Epigenetics, Freiburg, Germany. ⁷Present address: Institute of Molecular Biotechnology of the Austrian Academy of Science (IMBA), Vienna BioCenter (VBC), Vienna, Austria. ⁸These authors contributed equally: Akihisa Osakabe, Bhagyshree Jamge. ✉e-mail: frederic.berger@gmi.oeaw.ac.at

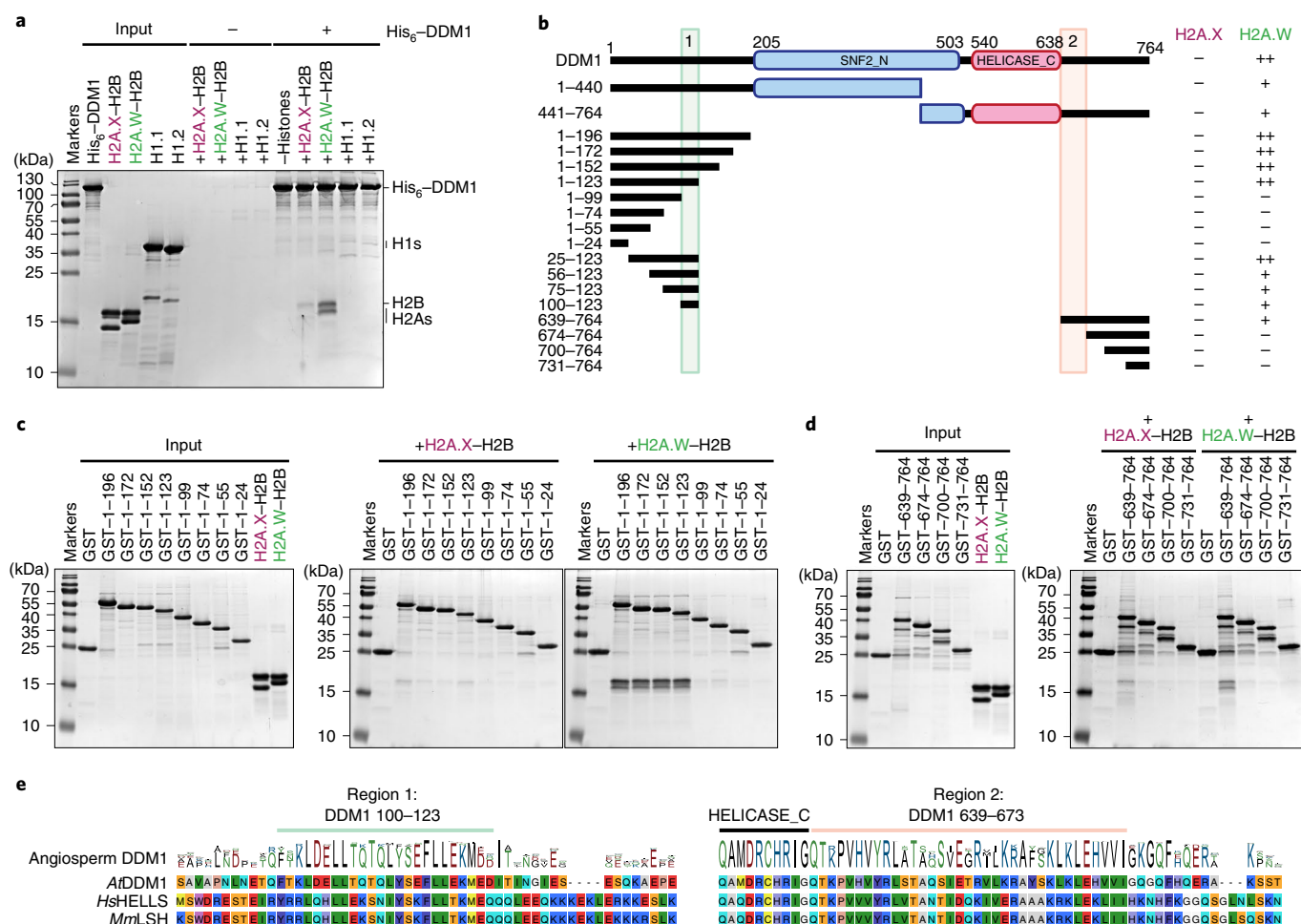


Fig. 1 | DDM1 binds directly to H2A.W through two conserved regions. **a**, SDS-polyacrylamide gel electrophoresis (SDS-PAGE) analysis of pull-down assays. His₆-tagged DDM1 was incubated with histone dimers, containing either H2A.X or H2A.W, or the linker histones H1.1 or H1.2. The samples were analysed using 15% SDS-PAGE and visualized using Coomassie Brilliant Blue. One out of three representative gels obtained is provided as source data. **b**, Summary of the H2A.W-binding sites in DDM1 identified by the pull-down assays shown in **c** and **d** and Extended Data Fig. 1g,i. **c,d**, SDS-PAGE analysis of pull-down assays. GST-tagged DDM1 fragments (N-terminal regions in **c** and C-terminal regions in **d**) were incubated with histone dimers containing H2A.X or H2A.W. After pull-down, the samples were analysed using 15% SDS-PAGE and were visualized using Coomassie Brilliant Blue. One out of two representative gels obtained is provided as source data. Pull-down assays were repeated twice with similar results. **e**, Sequence logo of alignments of H2A.W-binding regions 1 and 2 in DDM1 from flowering plants (angiosperms) with sequences from *Arabidopsis* DDM1, human (Hs) HELLS and mouse (Mm) LSH. Full alignments are shown in Supplementary Fig. 1. The list of species used is provided in Supplementary Table 1. Source data are available online.

DNA methyltransferases access to heterochromatin containing the linker histone H1 (refs. ^{36,42}). However, the direct molecular target of DDM1 and the molecular mechanisms through which DDM1 prevents expression and transposition of TEs remain unknown.

In this Article, we focus on the relationship between DDM1 and the histone variant H2A.W that evolved in vascular plants^{43,44}. Specific residues in the L1 loop and a long C-terminal tail with a KSPK motif distinguish H2A.W from the other H2A variants and confer specific properties to the nucleosome^{43,45}. H2A.W is confined to heterochromatin, but evidence of the mechanisms that are responsible for its deposition is lacking. Here we show that DDM1 binds to and deposits H2A.W into heterochromatin and, through the deposition of H2A.W, DDM1 participates in the silencing of TE protein-coding regions that are responsible for mobility.

Results

Conserved domains of DDM1 bind to the histone variant H2A.W. DDM1 is proposed to regulate the dynamics of linker histone H1, which is enriched in constitutive heterochromatin^{36,42},

and we considered that DDM1 could control its deposition. Other chromatin remodelers that are closely related to DDM1, such as SWR1 and INO80, control H2A variant dynamics^{46,47}, leading us to consider a possible link between DDM1 and another major component of heterochromatin, the histone variant H2A.W⁴⁴. To test whether DDM1 binds either to H1 or H2A.W, we expressed and purified DDM1 as a recombinant protein (Extended Data Fig. 1a). To validate DDM1 activity in vitro, we showed that purified DDM1 has DNA-dependent ATPase activity (Extended Data Fig. 1b,c), as reported previously⁴⁸. This activity was abolished by substitution of one amino acid, which was predicted to be essential for catalytic activity on the basis of structural homology with the DEAD-box RNA helicase Vasa⁴⁹ (Extended Data Fig. 1c–e). Thus, we identified the conserved catalytic site of DDM1 and obtained active recombinant DDM1, which could be used to investigate whether DDM1 has histone-binding activity.

We performed a nickel-nitrilotriacetic acid (Ni-NTA) agarose pull-down assay with recombinant histone heterodimers composed of H2B with either H2A.W or H2A.X. We used H2A.X as a control

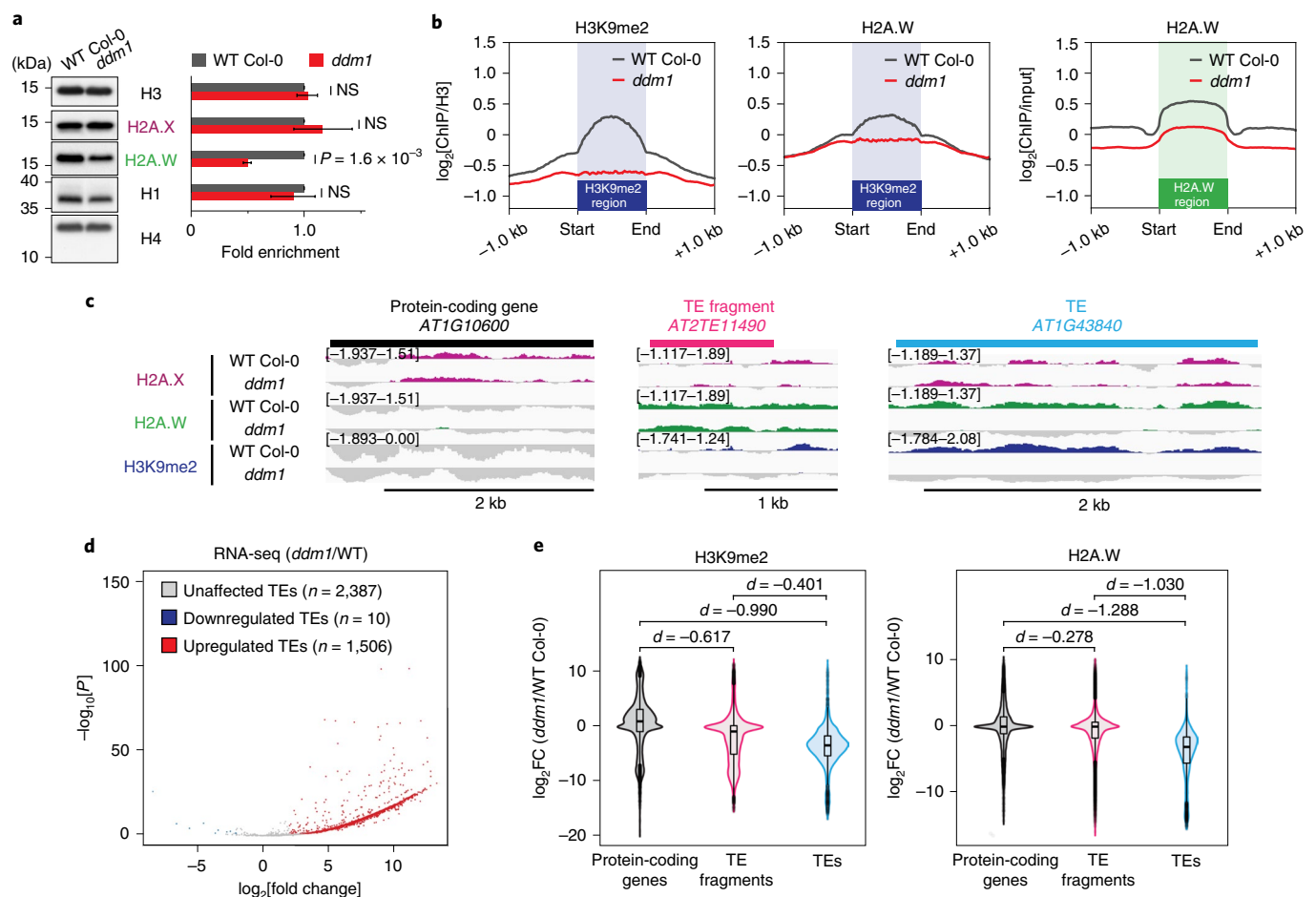


Fig. 2 | DDM1 is necessary for H2A.W deposition on TEs. **a**, Western blot analyses of purified chromatin in WT Col-0 and *ddm1* plants (left). Right, the fold enrichment of each histone in the WT (grey) and *ddm1* mutants (red). The intensity of each signal was normalized to H4. The ratios in the WT were set to 1 and the fold enrichment of each protein was calculated relative to the WT. Data are mean $\pm$ s.d. of $n = 3$ independent biological replicates. NS, not significant. **b**, Aggregate profile plots of the profiles of H3K9me2 (left) and H2A.W (middle) over heterochromatic regions defined by H3K9me2 enrichment, and H2A.W (right) over heterochromatic regions defined by H2A.W enrichment in the WT (grey) and *ddm1* mutants (red). **c**, IGV genome browser snapshot of the distribution of H2A.X (purple), H2A.W (green) and H3K9me2 (blue) in the WT and *ddm1* mutants. Coloured and grey shading indicate an enriched or depleted signal, respectively. Annotated protein-coding genes, TE fragments and TEs are shown as black, pink and light blue boxes, respectively. Values are shown as the $\log_2$ ratio of immunoprecipitated (ChIP) DNA relative to H3. **d**, The expression of TEs in *ddm1* mutants compared to the WT. The numbers in grey, blue and red indicate unaffected, downregulated and upregulated TEs, respectively. **e**, The enrichment of H3K9me2 (left) and H2A.W (right) on protein-coding genes ($n = 27,655$), TE fragments ($n = 25,695$, not including TEs) and TEs ($n = 3,903$) in *ddm1* mutants compared with the WT. The box plots show the minimum and maximum values as well as the 25th, 50th and 75th quartiles. Statistical significance was determined using effect size (Cohen's d), which was computed for each comparison. $|d| < 0.2$, negligible; $|d| < 0.5$, small; $|d| < 0.8$, medium; otherwise, large, as reported previously⁶¹. FC, fold change. Source data are available online.

because it is a close relative of H2A.W among the family of H2A variants⁵⁰. We also tested whether DDM1 could bind to recombinant linker histones H1.1 or H1.2. We found that His₆-tagged DDM1 co-precipitated with heterodimers containing H2A.W, but not H2A.X or linker histone H1 (Fig. 1a), and concluded that DDM1 binds to H2A.W but not to linker histone H1. To identify conserved motifs that are involved in H2A.W binding, we prepared a series of DDM1 fragments fused to a His₆ or a glutathione S-transferase (GST) tag. Our pull-down assay identified fragments containing residues 100–123 or 639–673 that were able to bind to H2A.W (Fig. 1b–d and Extended Data Fig. 1f–i). Sequence alignment of DDM1 orthologues from a representative array of flowering plants showed a strong conservation of residues in these domains, supporting the importance of these two H2A.W binding sites (Fig. 1e, Supplementary Table 1 and Supplementary Fig. 1). In conclusion, we have identified the H2A.W-binding regions of DDM1.

The conservation of this binding region between DDM1 and its mammalian orthologues HELLS and LSH (Fig. 1e) suggests that these three remodelers use a comparable mechanism to prevent transposon activity and form a specific class that we named DHL (DDM1, HELLS and LSH).

DDM1 is required for H2A.W deposition on transposons. DDM1 binds to H2A.W but does not accumulate specifically in heterochromatin⁵¹ and it is therefore improbable that H2A.W is involved in the recruitment of DDM1. Alternatively, binding of H2A.W to DDM1 suggested that DDM1 is involved in the deposition of H2A.W into heterochromatin. To address this point, we used first-generation *Arabidopsis ddm1* homozygous mutants, which lost a large fraction of DNA methylation in all contexts (Extended Data Fig. 2). Western blot analysis showed that H2A.W levels decreased substantially in *ddm1* mutants while global levels of H3, H2A.X and H1 remained

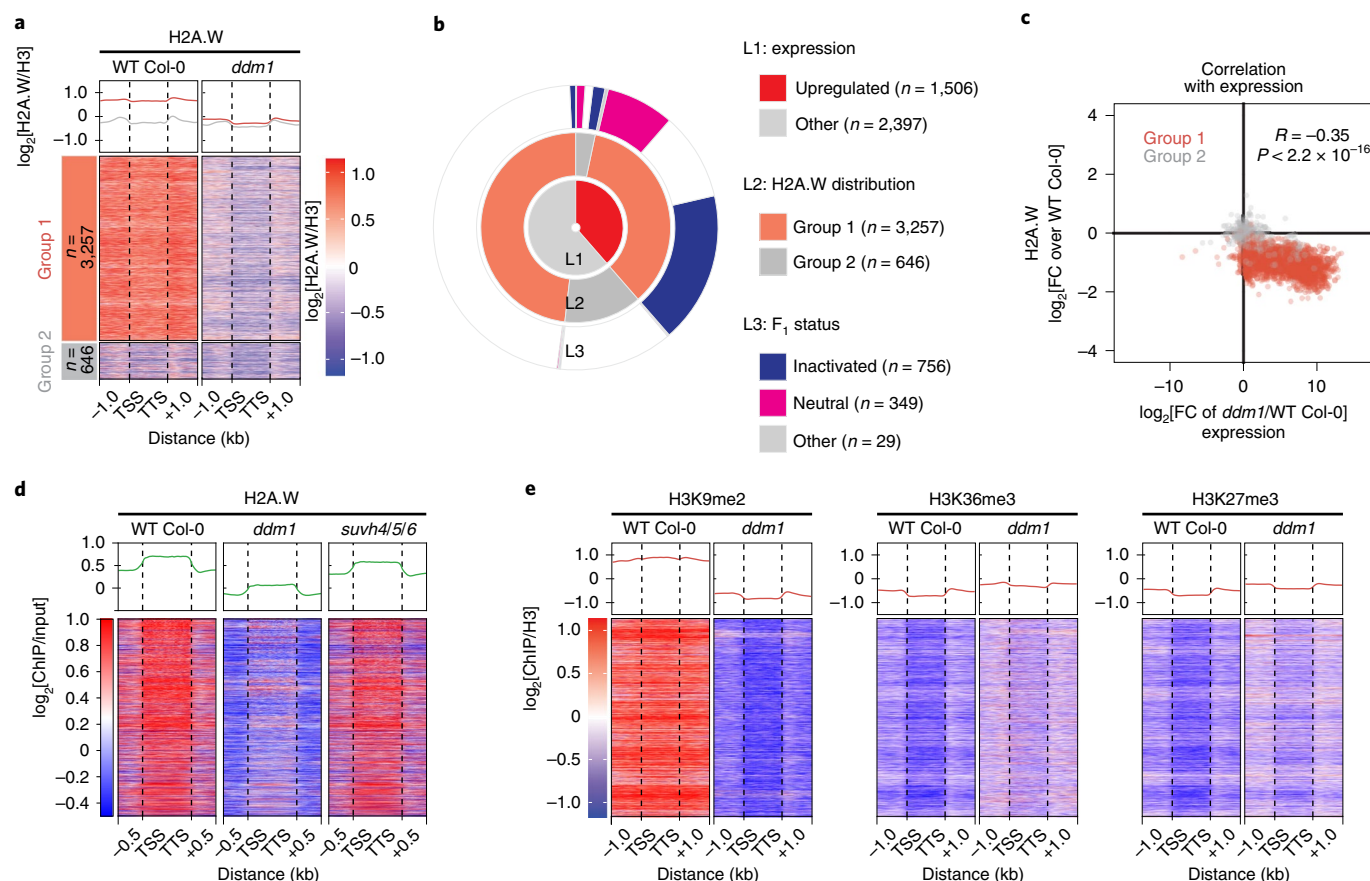


Fig. 3 | DDM1 changes the chromatin state of TEs. **a**, Aggregate profile plots (top) and heat maps (bottom) showing the distribution of H2A.W over TEs in the WT and *ddm1* mutants. TEs are clustered according to their enrichment in H2A.W in the WT (enriched in group 1; not enriched in group 2). **b**, Summary of the overlap across different groups of TEs. Layers from the inside to outside: L1, expression in *ddm1* mutants (shown in Fig. 2d); L2, TE clusters based on the distribution of H2A.W in WT Col-0 (shown in **a**); L3, TEs status in *F₁* *DDM1/ddm1* plants (shown in Fig. 5d). **c**, Comparison between TE expression and the loss of H2A.W in *ddm1* mutants. The y axis represents the log₂-transformed fold change in H2A.W (*ddm1*/WT). The x axis represents the expression level of TEs (log₂[fold change of *ddm1*/WT]). The *R* value corresponds to the Spearman's rank correlation coefficient for group 1 TEs. The statistical significance (*P*) was determined using a *t*-test. **d**, Aggregation plots (top) and heat maps (bottom) of the distribution of H2A.W over TEs in the WT and *ddm1* and *suvh4/5/6* mutants. **e**, Aggregation plots (top) and heat maps (bottom) of the distribution of H3K9me2, H3K36me3 and H3K27me3 over TEs in which H2A.W is lost in *ddm1* mutants (group 1). Source data are available online.

unchanged compared to the wild type (WT) (Fig. 2a). To identify the sites at which DDM1 mediates the deposition of H2A.W, we performed chromatin immunoprecipitation coupled with DNA sequencing (ChIP-seq) and compared the profiles of H2A.W with H2A.X and H1 in *ddm1* mutants and WT plants. In the *ddm1* mutant, a global loss of H2A.W was observed at H3K9me2-enriched heterochromatic regions and at H2A.W-enriched regions, while the levels of H2A.X and H1 remained unchanged (Fig. 2b,c and Extended Data Fig. 3a,b). This supports the notion that DDM1 specifically mediates H2A.W deposition but does not have an impact on H2A.X nor H1 enrichment in heterochromatin. Together with the lack of binding of H1 by DDM1 (Fig. 1a), these results led us to conclude that DDM1 silences TEs through a mechanism that is independent of H1, and that DDM1 was involved in deposition of H2A.W.

The loss of DDM1 caused the expression of about 40% of all TE genes, (1,506 out of 3,903; Fig. 2d), a result that is comparable to previous reports¹⁹. We focused our study on TEs with protein-coding potential as defined in TAIR10. Hereafter, we refer to them collectively as TEs and distinguish them from TE fragments, which are remnants of coding and non-coding regions of transposons that are unable to gain mobility on their own. The loss of H2A.W in *ddm1* mutants was more accentuated on TEs than on TE fragments

(Fig. 2c,e, effect size (Cohen's *d*) was -0.278 for TE fragments over protein-coding genes versus -1.288 for TEs over protein-coding genes; Extended Data Fig. 3b,c). This observation suggested that DDM1 mediates H2A.W deposition preferentially on the coding regions of transposons. We assessed how widespread this mechanism is and found that 83% of all TEs were enriched with H2A.W in the WT (3,257 out of 3,903 TEs; cluster group 1 in Fig. 3a,b) and, more specifically, in pericentromeric regions (Extended Data Fig. 3d). A small group of TEs that were not covered with H2A.W in the WT (646 out of 3,903 TEs; cluster group 2 in Fig. 3a,b) were mainly distributed outside of pericentric heterochromatin (Extended Data Fig. 3d) and were not upregulated in *ddm1* mutants (Fig. 3b,c). All TEs from group 1 lost H2A.W in *ddm1* mutants (Fig. 3a), leading to the conclusion that DDM1 is necessary for the deposition of H2A.W on the majority of TEs.

Although DDM1 binds to and mediates the deposition of H2A.W, other factors, such as the loss of H3K9me2 and nucleosome turnover associated with transcription⁵², might also explain the loss of H2A.W in *ddm1* mutants. We observed that, in the triple mutant *suvh4/5/6*, which lacks H3K9me2 (ref. ⁵³), H2A.W is not lost over TEs to the extent observed in *ddm1* mutants (Fig. 3d), and we concluded that the loss of H2A.W in *ddm1* mutants does not depend

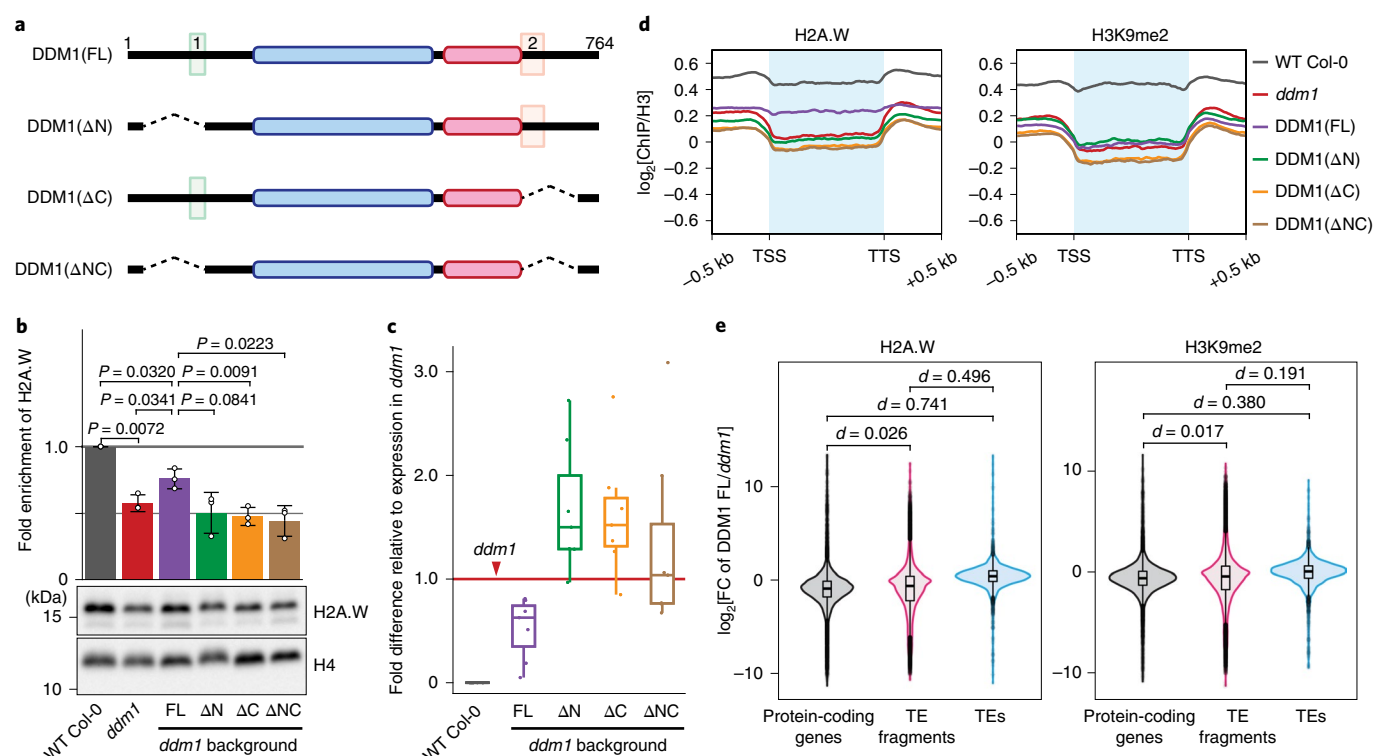


Fig. 4 | DDM1 binding to H2A.W is required for TE silencing. **a**, Schematic representation of DDM1 deletions devoid of H2A.W-binding regions. Boxes 1 and 2 correspond to the H2A.W-binding regions shown in Fig. 1. Forms of DDM1 that lack the N-terminal region (residues 25–123) and C-terminal region (residues 639–730) were defined as ΔN and ΔC, respectively. **b**, Western blot analyses of purified chromatin in WT Col-0, *ddm1* mutants and transgenic lines expressing DDM1 mutants. Top, the fold enrichment of H2A.W. The intensity of the H2A.W signal was normalized to H4. The ratios in the WT were set to 1 and the fold enrichment was calculated relative to the WT. Data are mean $\pm$ s.d. of $n = 3$ independent biological replicates. **c**, Fold changes in the expression levels of TEs in WT Col-0, *ddm1* mutants and transgenic lines expressing DDM1 mutants obtained using RT-qPCR (data for each TE are shown in Extended Data Fig. 4b). The value in the *ddm1* type was set to 1. Each dot corresponds to one TE. Data are the mean values of $n = 3$ independent biological replicates. The box plots show the minimum and maximum values as well as the 25th, 50th and 75th quartiles. **d**, Aggregate profile plots of profiles of H2A.W (left) and H3K9me2 (right) over TEs in WT (grey), *ddm1* mutants (red), and *ddm1* mutants expressing DDM1(FL) (purple), DDM1(ΔN) (green), DDM1(ΔC) (orange) and DDM1(ΔNC) (brown). **e**, The enrichment of H2A.W (left) and H3K9me2 (right) on protein-coding genes ($n = 27,655$), TE fragments ($n = 25,695$, not including TEs) and TEs ($n = 3,903$) in *ddm1* mutants expressing DDM1(FL) compared to the *ddm1* mutant. The box plots show the minimum and maximum values as well as the 25th, 50th and 75th quartiles. Statistical significance was determined using effect size d (Cohen's d), which was computed for each comparison. $|d| < 0.2$, negligible; $|d| < 0.5$, small; $|d| < 0.8$, medium; otherwise, large, as reported previously⁶¹. Source data are available online.

on H3K9me2. In agreement with this conclusion, we observed that, in *ddm1* mutants, H2A.W was depleted over TEs but not over TE fragments, whereas H3K9me2 was lost from both TE fragments and TEs (Fig. 2c,e, effect size d was -0.617 for TE fragments over protein-coding genes versus -0.990 for TEs over protein-coding genes; Extended Data Fig. 3b,c). We also considered that the loss of H2A.W in *ddm1* mutants at TEs might be due to increased transcription. However, we did not observe a significant correlation between the levels of H2A.W depletion and the levels of TE transcripts in *ddm1* versus the WT (Fig. 3c, cluster group 1). We further investigated whether the loss of DDM1 affects the post-translational modifications that associate positively (H3K36me3) and negatively (H3K27me3) with transcriptional regulation of protein-coding genes. We observed that, as H3K9me2 was depleted from TEs in *ddm1* mutants, H3K36me3 and H3K27me3 accumulated to some degree over TEs (Fig. 3e and Extended Data Fig. 3e,f). It was reported that H3K27me3 replaces H3K9me2 over heterochromatic regions that lose DNA methylation in *met1* mutants (ref. ⁵⁴) and the accumulation of H3K27me3 in *ddm1* plants is probably related to the loss of CG methylation and independent on TE expression. Thus, DDM1 activity causes a global maintenance of H3K9me2

and prevents the deposition of post-translational modifications associated with protein-coding genes on TEs. Regardless of whether they generate mRNAs in *ddm1* mutants, TEs globally lost H2A.W in *ddm1* mutants (Fig. 3c). We conclude that DDM1 deposits H2A.W and, as a result, installs a chromatin state that is incompatible with the transcription of the coding regions of TEs.

DDM1-mediated H2A.W deposition silences transposons. To investigate the impact of H2A.W deposition by DDM1 on TE silencing, we used genetic complementation experiments using the WT and mutant forms of DDM1 devoid of H2A.W-binding sites (Fig. 4a). Western blot and quantitative PCR with reverse transcription (RT-qPCR) analyses showed that the expression of native DDM1 protein in *ddm1* plants restored levels of H2A.W (Fig. 4b) and restored TE silencing (Fig. 4c and Extended Data Fig. 4). By contrast, expression of DDM1 mutant forms lacking either one or both H2A.W-binding regions did not restore H2A.W levels or TE silencing (Fig. 4b,c and Extended Data Fig. 4). We obtained genomic profiles of H2A.W in the *ddm1* mutants complemented with WT or mutant forms of DDM1 and observed that only complementation with WT DDM1 restored enrichment of H2A.W on TEs (Fig. 4d,e, effect size d was

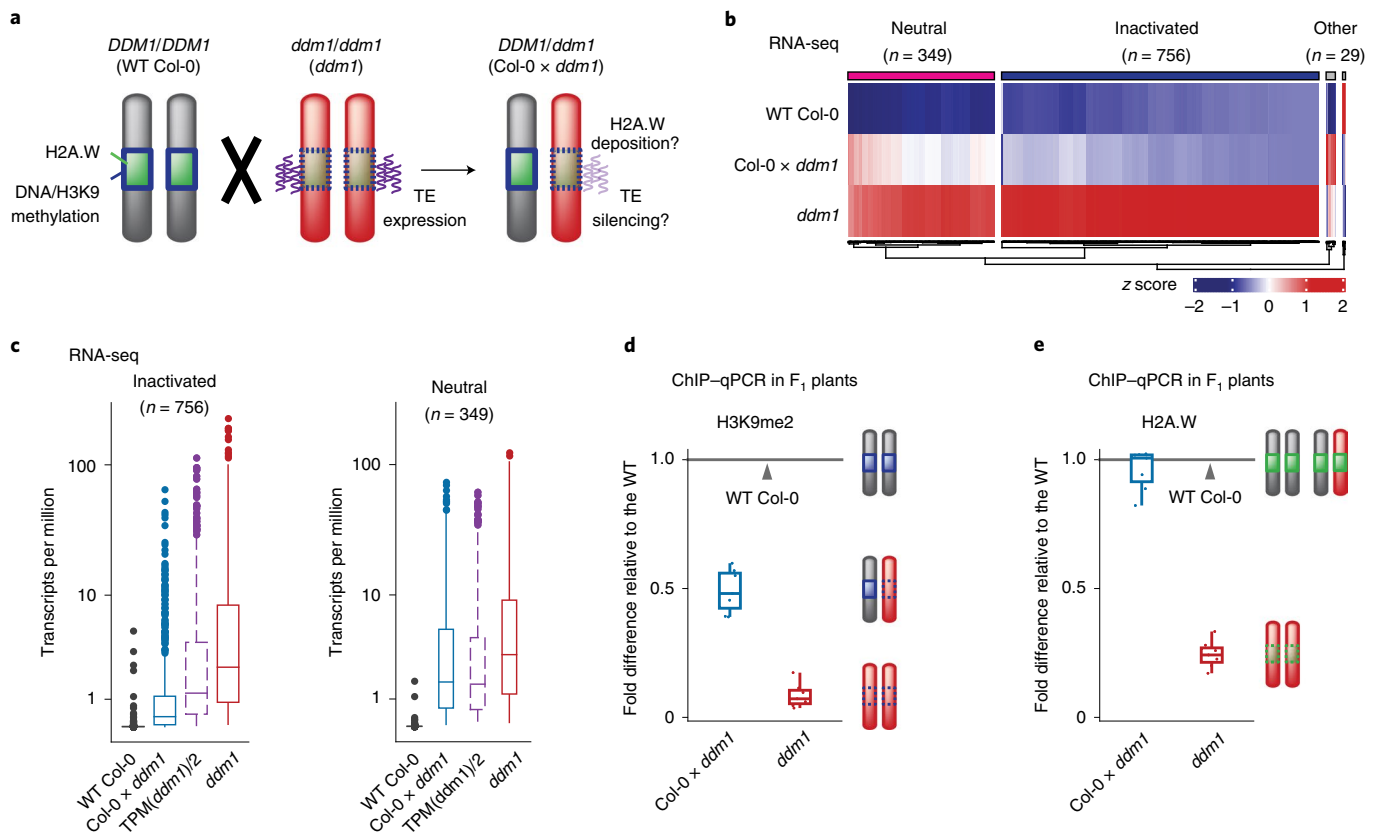


Fig. 5 | DDM1 deposits H2A.W over TEs to mediate silencing. **a**, Schematic of F_1 plants generated by crossing WT Col-0 and *ddm1* mutants. F_1 plants inherit chromosomes from the WT marked by H3K9me2 and DNA methylation (blue) and from *ddm1*-mutant chromosomes devoid of these marks. We tested the inheritance of H2A.W (green). **b**, Heat map of the z-transformed expression levels of differentially expressed TEs. TEs that remained expressed were classified as neutral ($n = 349$) whereas TE that were resiled were classified as inactivated ($n = 756$). **c**, The expression levels (transcripts per million (TPM)) of neutral (left) or inactivated (right) TEs from WT Col-0 (grey), F_1 DDM1/*ddm1* (blue) or *ddm1* mutants (red). As a reference, the 50% expression levels of TEs in *ddm1* mutants were calculated and are shown in the dashed purple boxes. The box plots show the minimum and maximum values as well as the 25th, 50th and 75th quartiles. Data are the mean values of $n = 3$ independent biological replicates. **d,e**, Fold changes in H3K9me2 (**d**) and H2A.W (**e**) at seven TEs in F_1 and *ddm1* plants obtained using ChIP-qPCR (Extended Data Fig. 6a,b). The value in the WT was set to 1. Each dot corresponds to one TE. The box plots show the minimum and maximum values as well as the 25th, 50th and 75th quartiles. The mean values of $n = 3$ independent biological replicates are shown. Source data are available online.

0.496 for TEs over TE fragments; Extended Data Fig. 5a–c). This showed the functionality of the H2A.W-binding domains of DDM1. Furthermore, the levels of H3K9me2 were similar in *ddm1* mutants and in complemented *ddm1* plants, further supporting that the deposition of H2A.W does not depend on H3K9me2 and that, in the absence of H3K9me2, H2A.W was sufficient to silence TEs (Fig. 4d,e and Extended Data Fig. 5a,b). The restoration of DDM1 function did not affect the levels of H2A.W on protein-coding genes or TE fragments (Fig. 4e and Extended Data Fig. 5d,e). Thus, we concluded that DDM1 binding to H2A.W is essential for H2A.W deposition to the coding regions of TEs and for TE silencing.

We hypothesized that, in DDM1/*ddm1* F_1 plants, DDM1 might be sufficient to cause silencing of TEs expressed on the chromosomes from the *ddm1* parent. Indeed, *k*-means clustering of TE expression showed that 70% of the TEs expressed in the *ddm1* mutant parent were resiled in F_1 plants (756 out of 1,134 TEs; Figs. 3b and 5a,b, inactivated cluster). We considered that, if DDM1 was not able to silence TEs from the alleles from *ddm1*-inherited alleles, the level of expression of TEs measured in the F_1 plants would be the half of that measured in *ddm1* mutants (Fig. 5c, dashed box). By contrast, in F_1 plants, we observed a much lower value (Fig. 5c, blue box), which would be predicted if DDM1 produced by the WT-inherited

allele silences in *trans* TEs from the *ddm1*-inherited allele. These results led us to conclude that DDM1 is not only required but also sufficient to repress TEs in F_1 plants.

F_1 DDM1/*ddm1* plants inherit a set of H3K9me2-marked chromosomes from the WT parent and a set of chromosomes devoid of H3K9me2 from the *ddm1* parent³³ (Fig. 5a). In F_1 plants, we used ChIP-qPCR analysis to measure the levels of H2A.W enrichment on a panel of individual TEs. We also measured the levels of H3K9me2 in the WT and *ddm1* mutants as a control. As expected, the averaged levels of H3K9me2 over TEs in F_1 plants were almost half of the levels of the WT (Fig. 5d and Extended Data Fig. 6a). By contrast, the levels of H2A.W in F_1 plants were restored to WT values (Fig. 5e and Extended Data Fig. 6b) irrespective of the absence of H3K9me2 marks. These results support that DDM1 is sufficient to mediate H2A.W deposition on TEs independently of H3K9me2.

We tested whether the maintenance of silencing of at least a subset of transposons is mediated by DDM1 and not by the other main silencing pathways that are associated with constitutive heterochromatin, CG methylation by MET1, and H3K9 methylation by SUVH4, SUVH5 and SUVH6. A comparison of the transcriptomes from mutants in each pathway showed that most TEs expressed either in *met1* or *suvh4/5/6* mutants also lost silencing in *ddm1*

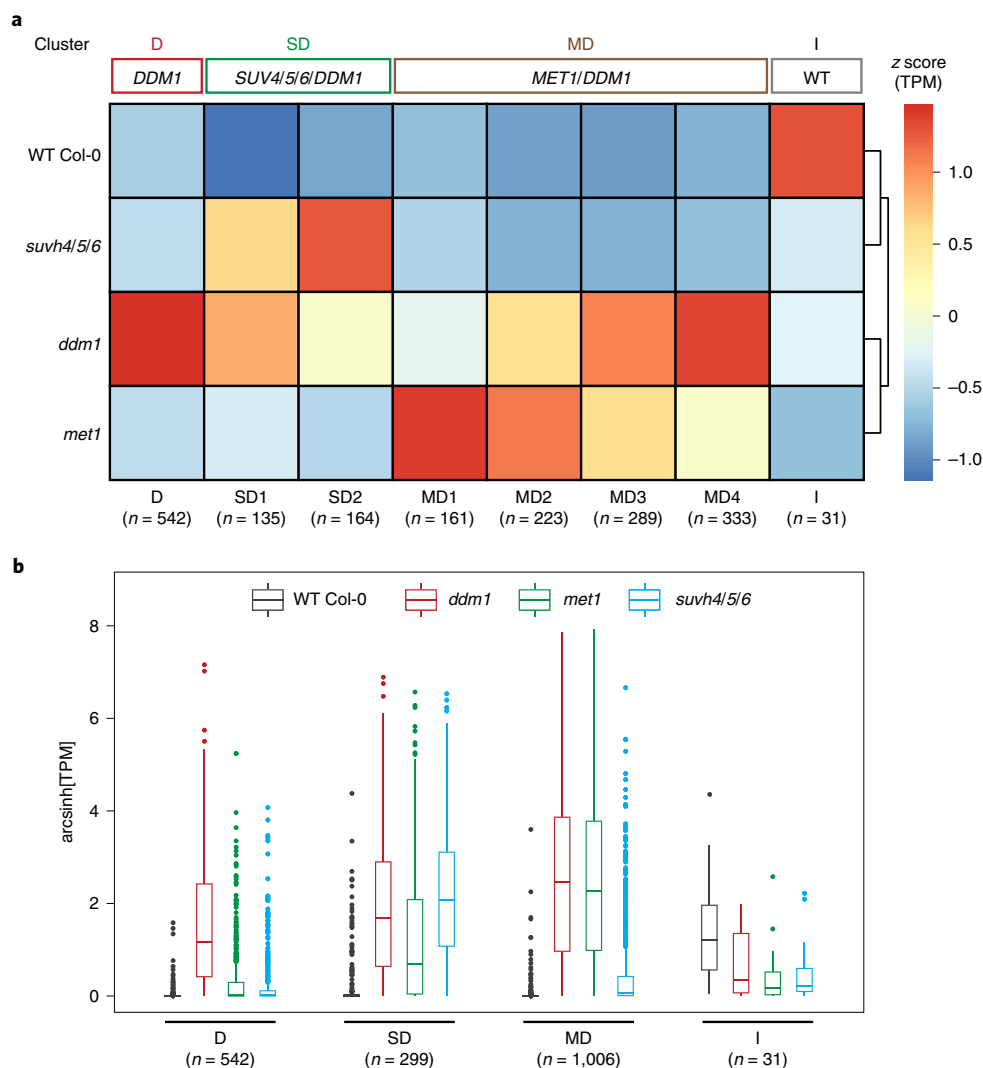


Fig. 6 | Expression of TEs in *ddm1* and other mutants. a, *k*-means clustering of TEs expressed in WT Col-0 or *suvh4/5/6*, *ddm1* or *met1* mutants. **b**, The expression of each TE cluster. The box plots of normalized expression levels (arcsinh[transcripts per million]) of each cluster of TEs from WT Col-0 (grey), and *ddm1* (red), *met1* (green) and *suvh4/5/6* (light blue) mutants. Sample sizes (where *n* is the number of TEs) are noted on each box plot. The box plots show the minimum and maximum values as well as the 25th, 50th and 75th quartiles. Source data are available online.

mutants (classes SD and MD in Fig. 6a and Extended Data Fig. 7). By contrast, around 30% of TEs lost silencing in *ddm1* mutants but not in *met1* or *suvh4/5/6* mutants (class D in Fig. 6a and Extended Data Fig. 7). The levels of expression of these TEs in *ddm1* mutants were comparable to the levels of expression in TEs that lost silencing in the two other mutants (Fig. 6b and Extended Data Fig. 7). This observation suggested that DDM1 is able to maintain some TEs silenced independently of deposition of CG and H3K9 methylation. We next considered whether DDM1-mediated silencing of TEs in *F₁* plants is associated with a gain of DNA methylation, and obtained genomic profiles of DNA methylation in the CG, CHG and CHH contexts in WT, *ddm1* and *F₁* DDM1/*ddm1* plants. Globally, we observed that the levels of DNA methylation over resiled TEs in *F₁* plants were either similar or higher than the calculated mean of the levels between the WT and *ddm1* mutants (Fig. 7a–c). With a few exceptions, we observed no or only a mild increase in CG methylation at TEs silenced by DDM1 in *F₁* plants (inactivated; Fig. 7a and Extended Data Fig. 8a). Furthermore, the mild increase in CG methylation was also observed to a similar magnitude in TEs that are not resiled (neutral), further supporting that CG

methylation does not have a major role in DDM1-mediated resiliencing of TEs (Fig. 7d). A small proportion of TEs gained CHG or CHH methylation at similar levels to the WT in resiled TEs but not in neutral TEs (Fig. 7b,c,e,f and Extended Data Fig. 8b,c). Thus, DDM1 probably cooperates with the pathways that methylate DNA in the CHG and CHH contexts to silence a small fraction of TEs. Such a cooperation between DDM1 and RdDM was reported previously³⁶. However, note that the gain of DNA methylation in *F₁* plants was not correlated with the levels of expression of TEs (Fig. 7d–f) and that a high proportion of resiled TEs did not gain DNA methylation, supporting the idea that resiliencing of a large fraction of TEs by DDM1 in *F₁* plants is not primarily related to DNA methylation.

We next examined whether DDM1 affects families of TEs that show high mobility in *Arabidopsis*. In addition to the majority of TEs that were resiled in *F₁* plants (Fig. 5b and Extended Data Fig. 6c), a smaller group of TEs remained expressed in *F₁* plants (neutral cluster in Figs. 3b and 5b,c and Extended Data Fig. 6d) and this cluster included five out of six TEs that were examined in a previous study³⁴ (Extended Data Fig. 6e). Compared with neutral

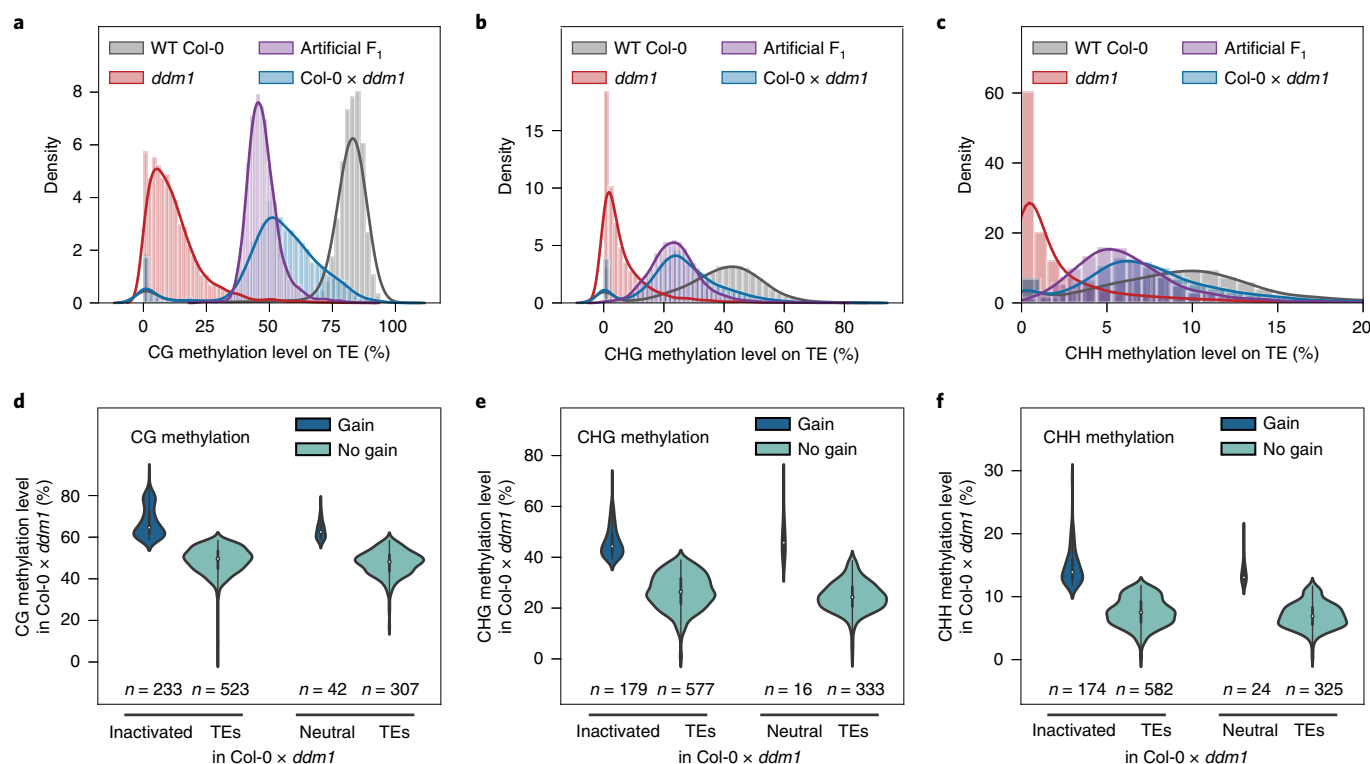


Fig. 7 | DNA methylation of TEs in *ddm1* and F₁ Col-0 × *ddm1* plants. **a–c**, Density plot of the distribution for weighted DNA methylation levels in TEs in WT Col-0 (grey), *ddm1* mutants (red), artificial F₁ (purple) and F₁ cross Col-0 × *ddm1* (blue) plants for CG methylation (**a**), CHG methylation (**b**) and CHH methylation (**c**). **d–f**, DNA methylation levels for inactivated (n = 756) and neutral (n = 349) TEs. TEs were classified on the basis of gain (dark blue) and no gain (greenish blue) in CG (**d**), CHG (**e**) and CHH (**f**) methylation in each context. Source data are available online.

TEs, TEs that were resiled by DDM1 in F₁ plants belonged preferentially to several large retrotransposon families, with a preference for Gypsy elements over Copia elements, which are able to transpose in *Arabidopsis*⁵⁵ (Fig. 8a,b). We also noticed that DDM1 suppresses a class of endogenous retroviruses that are active in *Arabidopsis*⁵⁶ (Fig. 8c). Furthermore, 125 out of the 142 mobile TEs identified in *Arabidopsis*⁵⁵ were expressed in *ddm1* mutants and 53% (66 out of 125) were resiled in F₁ plants (WT × *ddm1*; Fig. 8d). Thus, DDM1 mediates the deposition of H2A.W to silence TEs that threaten the stability of the genome.

Discussion

In conclusion, the *Arabidopsis* chromatin remodeler DDM1 is necessary and sufficient to inactivate potentially mobile TEs. We identified two broadly conserved domains of DDM1 that bind to H2A.W. These domains are required for the deposition of H2A.W and to initiate and maintain transcriptional silencing of genes that enable transposon mobility. DDM1-mediated remodelling of nucleosomes prevents the acquisition of chromatin marks that are normally found on protein-coding genes. Thus, DDM1 promotes distinct chromatin properties opposed to transcriptional activity at coding regions of active transposons.

Although DDM1 is required for the maintenance of DNA methylation and H3K9 methylation, the mechanism involved remains unclear. The linker histone H1 ensures sufficient chromatin accessibility for the action of the DNA methyltransferases⁴², but we report that DDM1 does not bind to H1 and has no major impact on H1 deposition. It is therefore improbable that H1 is directly involved in the loss of DNA methylation in *ddm1*. An alternative mechanism might involve the chromatin environment enriched in H3K27me3 that takes the place of H2A.W/H3K9me2 in *ddm1*,

as suggested by the antagonism between H3K27me3 and DNA methylation^{54,57}.

In mammals, the DDM1 orthologues LSH and HELLS⁵⁸ also maintain silencing of a large fraction of TEs in somatic cells³⁸. The conservation of the H2A.W-binding regions between the DHL remodelers DDM1 and LSH/HELLS (Fig. 1e) and a recent report that links deposition of macroH2A to LSH⁵⁹ suggests that the mechanism of H2A variant deposition described in this study also pertains to the deposition of heterochromatic macroH2A variants⁶⁰ by DDM1 orthologues in mammals. However, in mice, the loss of macroH2A deposition in *lsh* mutants also affects gene repression⁵⁹, suggesting that DDM1 orthologues have roles that are more diverse than suppressing TE transposition. Uncovering the molecular mechanism of this special class of chromatin remodelers will yield a clearer understanding of how chromatin dynamics and genome evolution are governed. Mobility of transposons and endogenous retroviruses threatens genome integrity and is associated with diseases¹¹, and we anticipate that the mechanism uncovered in this study will provide the foundation to study health issues associated with mobile elements and find innovative strategies to silence genes.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41556-021-00658-1>.

Received: 12 September 2020; Accepted: 1 March 2021;
Published online: 8 April 2021

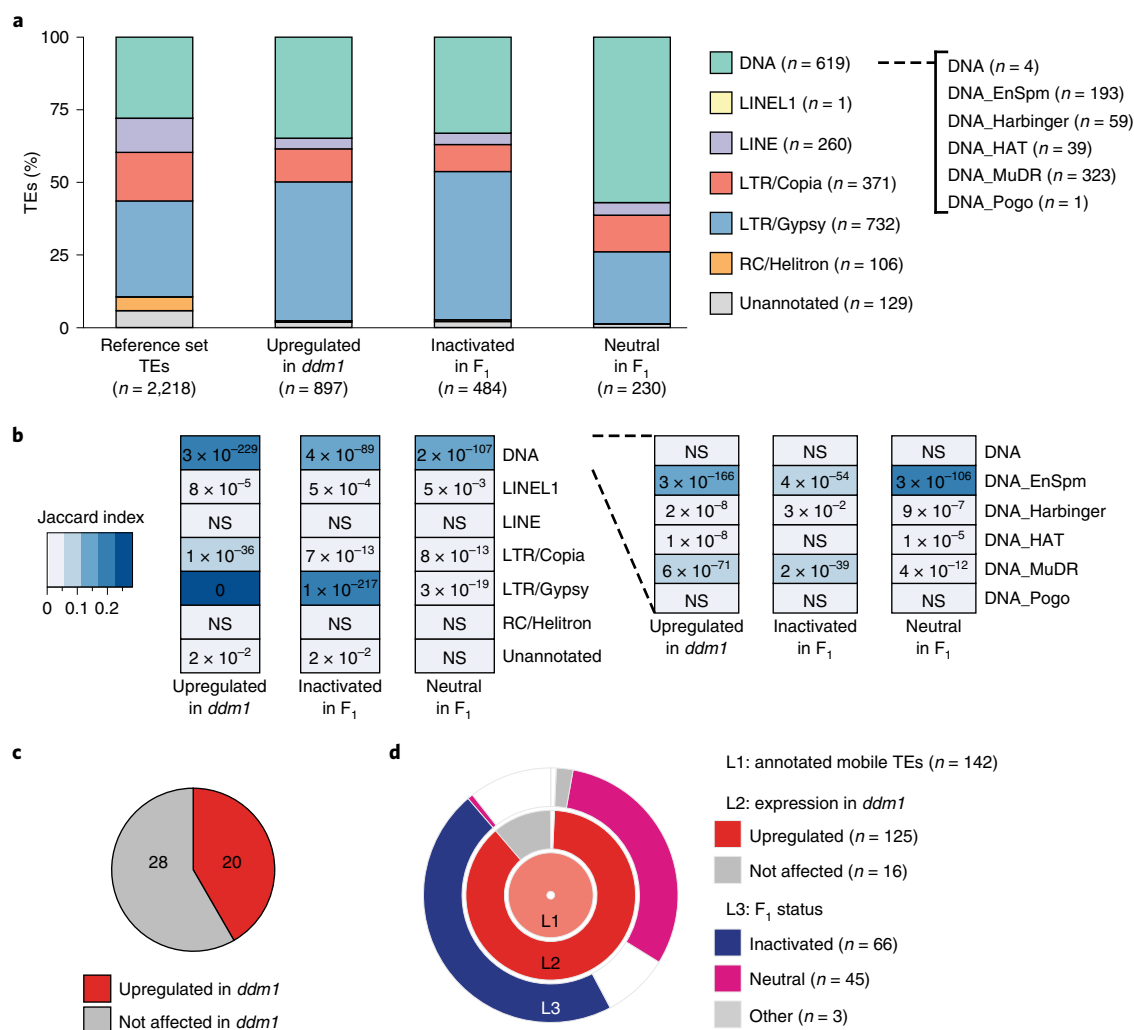


Fig. 8 | DDM1 silences mobile TEs. **a**, The proportion of TE families affected by DDM1. The proportion of TE gene super families (the total number (*n*) of TE genes assigned to the TE family in the TAIR10 annotation is indicated on the plot; y axis) that is upregulated in *ddm1* mutants or inactivated or neutral in the *F₁* DDM1/*ddm1* plants is shown. **b**, The association between TE super families and TE families affected by DDM1. Fisher's exact test was used to test the association between TE super families and TEs upregulated in *ddm1* mutants or inactivated or neutral in the *F₁* DDM1/*ddm1* plants. The number in each box is the corresponding adjusted *P* value for each comparison and the colour key shows the Jaccard index, which measures the similarity between the two sets (0 = no similarity, 1 = identical). **c**, Differentially expressed *Athila4* TE genes in *ddm1* mutants. **d**, Summary of the overlap across different groups of mobile TEs. Layers from the inside to the outside: L1, all annotated mobile TEs; L2, expression in *ddm1* mutants; L3, TE status in *F₁* DDM1/*ddm1* plants. Source data are available online.

References

- Cosby, R. L., Chang, N. C. & Feschotte, C. Host-transposon interactions: conflict, cooperation, and cooption. *Genes Dev.* **33**, 1098–1116 (2019).
- Bourque, G. et al. Ten things you should know about transposable elements. *Genome Biol.* **19**, 199 (2018).
- Wicker, T. et al. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8**, 973–982 (2007).
- Gagnier, L., Belancio, V. P. & Mager, D. L. Mouse germ line mutations due to retrotransposon insertions. *Mob. DNA* **10**, 15 (2019).
- Kidwell, M. G. & Lisch, D. Transposable elements as sources of variation in animals and plants. *Proc. Natl Acad. Sci. USA* **94**, 7704–7711 (1997).
- Chuong, E. B., Elde, N. C. & Feschotte, C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* **351**, 1083–1087 (2016).
- Jangam, D., Feschotte, C. & Betran, E. Transposable element domestication as an adaptation to evolutionary conflicts. *Trends Genet.* **33**, 817–831 (2017).
- Velanis, C. N. et al. The domesticated transposase ALP2 mediates formation of a novel Polycomb protein complex by direct interaction with MSI1, a core subunit of Polycomb Repressive Complex 2 (PRC2). *PLoS Genet.* **16**, e1008681 (2020).
- Rishishwar, L. et al. Evidence for positive selection on recent human transposable element insertions. *Gene* **675**, 69–79 (2018).
- Boissinot, S., Davis, J., Entezam, A., Petrov, D. & Furano, A. V. Fitness cost of LINE-1 (L1) activity in humans. *Proc. Natl Acad. Sci. USA* **103**, 9590–9594 (2006).
- Payer, L. M. & Burns, K. H. Transposable elements in human genetic disease. *Nat. Rev. Genet.* **20**, 760–772 (2019).
- Czech, B. et al. piRNA-guided genome defense: from biogenesis to silencing. *Annu. Rev. Genet.* **52**, 131–157 (2018).
- Handler, D. et al. The genetic makeup of the *Drosophila* piRNA pathway. *Mol. Cell* **50**, 762–777 (2013).
- Ozata, D. M., Gainetdinov, I., Zoch, A., O'Carroll, D. & Zamore, P. D. PIWI-interacting RNAs: small RNAs with big functions. *Nat. Rev. Genet.* **20**, 89–108 (2019).
- Derkacheva, M. et al. *Arabidopsis* MSI1 connects LHP1 to PRC2 complexes. *EMBO J.* **32**, 2073–2085 (2013).
- Grewal, S. I. & Jia, S. Heterochromatin revisited. *Nat. Rev. Genet.* **8**, 35–46 (2007).
- Law, J. A. & Jacobsen, S. E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* **11**, 204–220 (2010).

18. Nuthikattu, S. et al. The initiation of epigenetic silencing of active transposable elements is triggered by RDR6 and 21–22 nucleotide small interfering RNAs. *Plant Physiol.* **162**, 116–131 (2013).
19. Oberlin, S., Sarazin, A., Chevalier, C., Voinnet, O. & Mari-Ordonez, A. A genome-wide transcriptome and translome analysis of *Arabidopsis* transposons identifies a unique and conserved genome expression strategy for Ty1/Copia retroelements. *Genome Res.* **27**, 1549–1562 (2017).
20. Senti, K. A. & Brennecke, J. The piRNA pathway: a fly's perspective on the guardian of the genome. *Trends Genet.* **26**, 499–509 (2010).
21. Yu, R., Wang, X. & Moazed, D. Epigenetic inheritance mediated by coupling of RNAi and histone H3K9 methylation. *Nature* **558**, 615–619 (2018).
22. Kato, M., Miura, A., Bender, J., Jacobsen, S. E. & Kakutani, T. Role of CG and non-CG methylation in immobilization of transposons in *Arabidopsis*. *Curr. Biol.* **13**, 421–426 (2003).
23. Mirouze, M. et al. Selective epigenetic control of retrotransposition in *Arabidopsis*. *Nature* **461**, 427–430 (2009).
24. Stroud, H. et al. DNA methyltransferases are required to induce heterochromatic re-replication in *Arabidopsis*. *PLoS Genet.* **8**, e1002808 (2012).
25. Corem, S. et al. Redistribution of CHH methylation and small interfering RNAs across the genome of tomato *ddm1* mutants. *Plant Cell* **30**, 1628–1644 (2018).
26. Higo, H. et al. *DDM1* (decrease in DNA methylation) genes in rice (*Oryza sativa*). *Mol. Genet. Genomics* **287**, 785–792 (2012).
27. Hirochika, H., Okamoto, H. & Kakutani, T. Silencing of retrotransposons in *Arabidopsis* and reactivation by the *ddm1* mutation. *Plant Cell* **12**, 357–369 (2000).
28. Jeddelloh, J. A., Stokes, T. L. & Richards, E. J. Maintenance of genomic methylation requires a SWI2/SNF2-like protein. *Nat. Genet.* **22**, 94–97 (1999).
29. Mittelsten Scheid, O., Afsar, K. & Paszkowski, J. Release of epigenetic gene silencing by trans-acting mutations in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **95**, 632–637 (1998).
30. Singer, T., Yordan, C. & Martienssen, R. A. Robertson's mutator transposons in *A. thaliana* are regulated by the chromatin-remodeling gene *Decrease in DNA Methylation (DDM1)*. *Genes Dev.* **15**, 591–602 (2001).
31. Tan, F. et al. *DDM1* represses noncoding RNA expression and RNA-directed DNA methylation in heterochromatin. *Plant Physiol.* **177**, 1187–1197 (2018).
32. Gendrel, A. V., Lippman, Z., Yordan, C., Colot, V. & Martienssen, R. A. Dependence of heterochromatic histone H3 methylation patterns on the *Arabidopsis* gene *DDM1*. *Science* **297**, 1871–1873 (2002).
33. Soppe, W. J. et al. DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in *Arabidopsis*. *EMBO J.* **21**, 6549–6559 (2002).
34. Lippman, Z., May, B., Yordan, C., Singer, T. & Martienssen, R. Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. *PLoS Biol.* **1**, E67 (2003).
35. Stroud, H. et al. Non-CG methylation patterns shape the epigenetic landscape in *Arabidopsis*. *Nat. Struct. Mol. Biol.* **21**, 64–72 (2014).
36. Zemach, A. et al. The *Arabidopsis* nucleosome remodeler *DDM1* allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* **153**, 193–205 (2013).
37. Basenko, E. Y., Kamei, M., Ji, L., Schmitz, R. J. & Lewis, Z. A. The LSH/*DDM1* homolog *MUS-30* is required for genome stability, but not for DNA methylation in *Neurospora crassa*. *PLoS Genet.* **12**, e1005790 (2016).
38. Dunican, D. S. et al. Lsh regulates LTR retrotransposon repression independently of *Dnmt3b* function. *Genome Biol.* **14**, R146 (2013).
39. Dennis, K., Fan, T., Geiman, T., Yan, Q. & Muegge, K. Lsh, a member of the SNF2 family, is required for genome-wide methylation. *Genes Dev.* **15**, 2940–2944 (2001).
40. Zhu, H. et al. Lsh is involved in de novo methylation of DNA. *EMBO J.* **25**, 335–345 (2006).
41. Ren, J., Finney, R., Ni, K., Cam, M. & Muegge, K. The chromatin remodeling protein Lsh alters nucleosome occupancy at putative enhancers and modulates binding of lineage specific transcription factors. *Epigenetics* **14**, 277–293 (2019).
42. Lyons, D. B. & Zilberman, D. *DDM1* and Lsh remodelers allow methylation of DNA wrapped in nucleosomes. *eLife* **6**, e30674 (2017).
43. Lei, B. et al. A Synthetic approach to reconstruct the evolutionary and functional innovations of the plant histone variant H2A.W. *Curr. Biol.* **31**, 182–191.e5 (2021).
44. Yelagandula, R. et al. The histone variant H2A.W defines heterochromatin and promotes chromatin condensation in *Arabidopsis*. *Cell* **158**, 98–109 (2014).
45. Osakabe, A. et al. Histone H2A variants confer specific properties to nucleosomes and impact on chromatin accessibility. *Nucleic Acids Res.* **46**, 7675–7685 (2018).
46. Mizuguchi, G. et al. ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science* **303**, 343–348 (2004).
47. Papamichos-Chronakis, M., Watanabe, S., Rando, O. J. & Peterson, C. L. Global regulation of H2A.Z localization by the INO80 chromatin-remodeling enzyme is essential for genome integrity. *Cell* **144**, 200–213 (2011).
48. Brzeski, J. & Jerzmanowski, A. Deficient in DNA methylation 1 (*DDM1*) defines a novel family of chromatin-remodeling factors. *J. Biol. Chem.* **278**, 823–828 (2003).
49. Sengoku, T., Nureki, O., Nakamura, A., Kobayashi, S. & Yokoyama, S. Structural basis for RNA unwinding by the DEAD-box protein *Drosophila Vasa*. *Cell* **125**, 287–300 (2006).
50. Talbert, P. B. et al. A unified phylogeny-based nomenclature for histone variants. *Epigenetics Chromatin* **5**, 7 (2012).
51. Zemach, A. et al. *DDM1* binds *Arabidopsis* methyl-CpG binding domain proteins and affects their subnuclear localization. *Plant Cell* **17**, 1549–1558 (2005).
52. Workman, J. L. Nucleosome displacement in transcription. *Genes Dev.* **20**, 2009–2017 (2006).
53. Li, X. et al. Mechanistic insights into plant SUVH family H3K9 methyltransferases and their binding to context-biased non-CG DNA methylation. *Proc. Natl Acad. Sci. USA* **115**, E8793–E8802 (2018).
54. Deleris, A. et al. Loss of the DNA methyltransferase MET1 Induces H3K9 hypermethylation at PcG target genes and redistribution of H3K27 trimethylation to transposons in *Arabidopsis thaliana*. *PLoS Genet.* **8**, e1003062 (2012).
55. Lanciano, S. et al. Sequencing the extrachromosomal circular mobilome reveals retrotransposon activity in plants. *PLoS Genet.* **13**, e1006630 (2017).
56. Slotkin, R. K. et al. Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* **136**, 461–472 (2009).
57. Rougée, M. et al. Polycomb mutant partially suppresses DNA hypomethylation-associated phenotypes in *Arabidopsis*. *Life Sci. Alliance* **4**, e202000848 (2020).
58. Thijssen, P. E. et al. Mutations in *CDCA7* and *HELLS* cause immunodeficiency-centromeric instability-facial anomalies syndrome. *Nat. Commun.* **6**, 7870 (2015).
59. Ni, K. et al. LSH mediates gene repression through macroH2A deposition. *Nat. Commun.* **11**, 5647 (2020).
60. Douet, J. et al. MacroH2A histone variants maintain nuclear organization and heterochromatin architecture. *J. Cell Sci.* **130**, 1570–1582 (2017).
61. Cohen, J. Statistical power analysis. *Curr. Dir. Psychol. Sci.* **1**, 98–101 (1992).

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

© The Author(s), under exclusive licence to Springer Nature Limited 2021, corrected publication 2022

Methods

Plant material. The WT *Arabidopsis thaliana* plants used were accession Col-0. Homozygous first-generation *ddm1-2* mutant plants were obtained from the self-pollinated progeny of the heterozygote, and F₁ *DDM1/ddm1* plants were obtained from crosses between WT and first-generation homozygous *ddm1-2* mutants. We also used *met1-3* first-generation plants from *met1-3/+* and *suva4/5/6* triple mutants (SALK_041474, Gabi_263C05, SAIL_1244_F04).

Seeds were sown on soil and then stratified at 4°C in the dark for 3 d. Plants were grown in the growth chamber under long-day conditions (21°C; 16 h light–8 h dark). Genotyping for *ddm1-2* was performed using the cleaved amplified polymorphic sequence method with the primers listed in Supplementary Table 2. This was followed by RsaI (Thermo Fisher Scientific) restriction enzyme digestion.

DNA methylation analysis. Genomic DNA was purified using the DNeasy plant mini kit (Qiagen) from 4–5-week-old leaves according to the manufacturer's instructions. After purification of genomic DNA, genomic DNA (200 ng) was sheared by sonication using a Covaris S2 Focused Ultrasonicator for 45 s at 4°C (duty cycle, 10%; intensity, 5.0; cycles per burst, 200) in 55 µl Covaris microTUBES (AFA Fiber Snap-Cap). After sonication, end repair/dA-tailing, adaptor ligation, oxidation of 5mC and 5hmC, deamination and library preparation were performed using the NEBNext Enzymatic Methyl-seq kit (New England Biolabs) according to the manufacturer's instructions. Libraries were sequenced using Illumina Novaseq SP to generate paired-end 250 bp reads. The samples were prepared from two independent biological replicates. The sequenced EM-seq treated reads were mapped to the TAIR 10 assembly (<https://www.arabidopsis.org/index.jsp>) along with chloroplast genome using methylseq, a Nextflow pipeline (<https://github.com/nf-core/methylseq>). Conversion rates were calculated on the basis of methylation levels present on reads mapped to the chloroplast genome using methylpy^{62,63}. We calculated weighted methylation levels across protein-coding genes, TE fragments and TE genes, hereafter referred to as TEs annotated from TAIR 10 using custom python scripts.

Purification of recombinant *Arabidopsis* DDM1. The DNA fragment encoding the *Arabidopsis* DDM1 coding sequence was cloned into a pET-15b vector (Novagen). Freshly transformed *Escherichia coli* strain BL21 (DE3) containing minor tRNA expression vector (Codon (+) RIL, Stratagene) was grown on LB plates containing ampicillin (50 µg ml⁻¹) at 37°C for 16 h. After incubation, a single colony was inoculated into LB medium containing ampicillin (50 µg ml⁻¹) and cultured at 25°C. When cell density reached an absorbance at 600 nm (*A*₆₀₀) of 0.4, 0.5 mM isopropyl-β-D-thiogalactopyranoside was added to induce DDM1 expression and cultured cells were incubated at 25°C for an additional 12 h. Cells expressing DDM1 were collected and lysed with sonication in buffer A (50 mM Tris-HCl pH 7.5, 10% glycerol, 1 mM phenylmethylsulfonyl fluoride, 0.05% IGEPAL CA-630 (Sigma-Aldrich) and 2 mM 2-mercaptoethanol) containing 500 mM NaCl. Cell debris was removed by centrifugation and the lysate was mixed with Ni-NTA agarose resin (Qiagen) at 4°C for 1 h. After binding, the resin was washed with buffer B (50 mM Tris-HCl pH 7.5, 500 mM NaCl, 10% glycerol and 2 mM 2-mercaptoethanol) containing 5 mM imidazole. The His₆-tagged DDM1 protein was eluted by addition of buffer B containing 300 mM imidazole. The His₆ tag was cleaved from the DDM1 fragment by the addition of thrombin protease (15 U mg⁻¹ of protein; GE Healthcare) and then immediately dialysed against buffer C (20 mM Tris-HCl pH 7.5, 10% glycerol and 2 mM 2-mercaptoethanol) containing 150 mM NaCl. After removing the His₆ tag, DDM1 was processed using the RESOURCE S anion exchange column (GE Healthcare) and proteins were collected from the flow-through fractions. The DDM1 protein solution was diluted with buffer C to a final concentration of 50 mM NaCl and then processed using the RESOURCE S cation exchange column (GE Healthcare) and purified by elution with a linear gradient of 50–300 mM NaCl in buffer C. The peak fractions containing DDM1 were collected, and DDM1 was further purified by Superdex 200 (GE Healthcare) size-exclusion column chromatography with buffer C containing 50 mM NaCl. Purified DDM1 was collected and concentrated using the Amicon Ultra 30K filter (Millipore) with buffer exchange into buffer C containing 150 mM NaCl. The vector encoding the DDM1^{E334N} mutant was generated by site-directed mutagenesis using the Q5 Site-Directed Mutagenesis Kit (New England Biolabs) and the DDM1^{E334N} protein was purified as described above.

DNA fragments encoding DDM1 truncations, DDM1(1–440) or DDM1(441–764), were cloned into pET-15b. His₆-DDM1(1–440) and His₆-DDM1(441–764) were purified as described above, with the exclusion of the His₆ tag cleavage step. However, His₆-DDM1(441–764) was collected in the flow-through fractions from the RESOURCE S cation exchange column.

DNA fragments encoding the 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), DDM1(25–123), DDM1(56–123), DDM1(75–123), DDM1(100–123), DDM1(639–764), DDM1(674–764), DDM1(700–764) and DDM1(731–764) were inserted into the EcoRI and XhoI sites of pGEX-4T-1 (GE Healthcare). Expression of DDM1 truncations was performed as described above. Cells expressing DDM1 truncations were collected and lysed by sonication in buffer D (20 mM Tris-HCl pH 7.5, 1 M NaCl, 0.2 mM EDTA, 1% Triton X-100 and 1 mM dithiothreitol). Cell debris was removed by centrifugation and the lysate was mixed

with glutathione Sepharose 4B resin (GE Healthcare) at 4°C for 1 h. After binding, resin was washed with buffer D, rinsed with 20 mM Tris-HCl pH 8.0 and then GST-tagged DDM1 truncations were eluted by buffer E (20 mM Tris-HCl pH 8.0 and 20 mM glutathione). Eluted proteins were immediately dialysed against buffer C containing 150 mM NaCl.

ATPase assay. DDM1 or DDM1^{E334N} (1 µM) was mixed with 0.5 µM, 1 µM, 2 µM, 3.75 µM, 5 µM, 7.5 µM or 10 µM 193-base-pair Widom 601 DNA^{64,65} (DNA concentration expressed as moles of nucleotides) in a 100 µl total reaction volume containing 53 mM Tris-HCl pH 7.5, 15 mM NaCl, 2.5 mM MgCl₂, 0.5 mM ATP, 1% glycerol, 0.2 mM 2-mercaptoethanol and 0.01 mM EDTA, and incubated at 30°C for 1 h. ATPase assays were performed using a colorimetric kit (Innova Biosciences) and activity was calculated from values of optical density at 600 nm.

Purification and reconstitution of *Arabidopsis* histone dimers. Recombinant *Arabidopsis* histones H2A.X.3, H2A.W.6 and H2B.9 were expressed and purified as previously described^{45,66,67}. The H2A–H2B dimers were reconstituted and purified as previously described⁶⁸.

Purification of *Arabidopsis* linker histones. DNA fragments encoding H1.1 and H1.2 were cloned into pET-28a (Novagen), which has an additional gene encoding a SUMO tag between H1 and the His₆ tag. Recombinant H1 proteins were expressed in *E. coli* BL21 (DE3) containing the minor tRNA expression vector (Codon (+) RIL) and grown on LB plates containing kanamycin (50 µg ml⁻¹) at 37°C for 16 h. After incubation, a single colony was inoculated into LB medium containing kanamycin (50 µg ml⁻¹) and cultured at 25°C. When the cell density reached *A*₆₀₀ = 0.4, 0.5 mM isopropyl-β-D-thiogalactopyranoside was added to induce of H1 expression and cells were incubated at 25°C for an additional 12 h. Cells expressing H1 were collected and lysed by sonication in buffer A containing 1 M NaCl. Cell debris was removed by centrifugation and the lysate was mixed with Ni-NTA agarose resin (Qiagen) at 4°C for 1 h. After binding, resin was washed with buffer B containing 5 mM imidazole. The His₆-tagged H1 protein was eluted by the addition of buffer B containing 300 mM imidazole. The SUMO–His₆ tag was cleaved from H1 by the addition of GST-fused human rhinovirus (HRV) 3C protease (0.01 mg per mg of H1) and then immediately dialysed against buffer C (20 mM Tris-HCl pH 7.5, 10% glycerol and 2 mM 2-mercaptoethanol) containing 150 mM NaCl. After removing the SUMO–His₆ tag, H1 was loaded onto a RESOURCE S cation exchange column and purified by elution with a linear gradient of 150–1,000 mM NaCl in buffer C. Peak fractions containing H1 were collected and dialysed against buffer C containing 150 mM NaCl.

Ni-NTA or GS4B bead pull-down assays. Purified His₆-tagged DDM1 or truncations (10 µM) were mixed with H2A–H2B dimers or H1 (20 µM) in a total volume of 10 µl containing 32 mM Tris-HCl pH 7.5, 300 mM NaCl, 4% glycerol, 40 mM imidazole, 0.4 mM EDTA, 0.1 µg ml⁻¹ BSA and 2.4 mM 2-mercaptoethanol, and incubated at 30°C for 1 h. Ni-NTA agarose beads (5 µl, 50% slurry) were added to the reaction mixtures and incubated at 30°C for 30 min with shaking. After incubation, beads were washed three times with buffer containing 20 mM Tris-HCl pH 7.5 buffer, 500 mM NaCl, 40 mM imidazole and 0.25% Triton X-100.

Purified GST-tagged DDM1 truncations (2 µM) were mixed with H2A–H2B dimers (4 µM) in a total volume of 50 µl reaction containing 38 mM Tris-HCl pH 7.5, 150 mM NaCl, 7% glycerol, 1.4 mM EDTA, 3 mM 2-mercaptoethanol, 1 mM dithiothreitol and 0.1 µg ml⁻¹ BSA, and incubated at 30°C for 1 h. Glutathione Sepharose 4B beads (5 µl, 50% slurry) were added to reaction mixtures and incubated at 30°C for 30 min with shaking. After incubation, the beads were washed three times with buffer containing 20 mM Tris-HCl pH 7.5, 250 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol and 0.1% Triton X-100.

Proteins that precipitated with beads were analysed using 15% SDS–PAGE, with Coomassie Brilliant Blue staining.

Amino acid sequence alignment. The *A. thaliana* DDM1 protein sequence was used in a BLAST search of flowering plant genomes. These sequences were aligned, and sequence logos were generated using CLC Genomics Workbench 11.0 (<https://digitalinsights.qiagen.com>). A list of all sequences and alignments is provided in Supplementary Table 1.

Western blot. Nuclei from 4–5-week-old leaves of WT Col-0, first-generation *ddm1* mutants, transgenic lines expressing DDM1 mutants and second-generation *ddm1* mutants were isolated in NIB buffer⁶⁹. Proteins were separated using 15% SDS–PAGE and then transferred to nitrocellulose membrane (GE Healthcare). Western blot analysis was performed using antibodies against H4 (0.5 µg ml⁻¹; Abcam, ab10158), H3 (1 µg ml⁻¹; Abcam, ab1791), H2A.X.3/5 (1 µg ml⁻¹; affinity-purified rabbit polyclonal antibodies against the peptide CKVGKNGKDIIGSASQ⁴³), H2A.W.6 (1 µg ml⁻¹; affinity-purified rabbit polyclonal antibodies against the peptide GGRKPPGAPKTKSV⁴⁴) and H1.1/1.2 (1 µg ml⁻¹; affinity-purified rabbit polyclonal antibodies against the peptide RTGSSQYAIQKFIEEK) (Supplementary Fig. 2). As secondary antibodies, goat anti-rabbit IgG (BioRad) was used at a dilution of 1:10,000. Blots were developed using an enhanced chemiluminescence kit (Thermo Fisher Scientific) and signals

were detected using the ChemiDoc instrument (BioRad). Signal quantification was performed using Image Lab (BioRad).

ChIP analysis. Nuclei isolation was performed using 4–5 week-old leaves from WT Col-0, first-generation *ddm1* mutants, *F₁ DDM1/ddm1* plants, transgenic lines expressing DDM1 mutants and second-generation *ddm1* mutants as described previously^{70–72}. 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. Cross-linked tissues were frozen and ground with a pestle and mortar 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 a 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 MgCl₂, 0.5% Triton X-100, 10 mM 2-mercaptoethanol and protease inhibitor), and then 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 sonication buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.1% SDS and protease inhibitor). Nuclei were sonicated 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. 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 1 g of tissue; Thermo Fisher Scientific) were added to sheared chromatin and incubated at 4°C for 1 h with rotation. Pre-cleared samples were collected and incubated with 5 µg of anti-H3, anti-H2A.X.3/5, anti-H2A.W.6, anti-H1.1/1.2, anti-H3K36me3 (Abcam, ab9050), anti-H3K27me3 (Millipore, 07-449) 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 h 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 NaHCO₃), incubated at 65°C for 15 min and 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)). The reaction mixture was then incubated at 45°C for 3 h and subsequently 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).

For ChIP-qPCR, purified DNA was used for qPCR assays using FastStart Essential DNA Green Master (Roche). The PCR reaction and quantitative measurement to obtain *C_q* values were performed on a LightCycler 96 (Roche). qPCR assays were performed with three independent technical replicates and three biological replicates for each sample. For each qPCR experiment, technical replicates were averaged. For each biological replicate, ΔC_q values were obtained using the matched H3 values as reference. Mean and s.d. values for the ΔC_q values were calculated using the biological replicates. Fold changes were calculated as $2^{-\Delta\Delta C_q}$. The primers used for ChIP-qPCR were the same as for RT-qPCR. qPCR analysis and figure generation were performed using R v.3.5.1 (<https://www.R-project.org/>) and the package ggplot2 v.3.1.0.

For ChIP-seq, sequencing libraries were prepared using the NEBNext Ultra II DNA library prep kit for Illumina (New England Biolabs) according to the manufacturer's instructions. The libraries were sequenced using an Illumina HiSeq 2500 to generate single-end 50 bp reads. Samples were prepared from two independent biological replicates. For alignment and quality check of the sequenced samples, bedtools v.2.27.1 was used to convert the raw BAM files to fastq. FastQC v.0.11.8 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) was used to generate quality reports for all sequencing data. Reads were filtered and then aligned to the TAIR10 *Arabidopsis* genome using Bowtie2 (ref.⁷³) using the default settings. DeepTools (v.3.1.2)⁷⁴ was used to examine correlations between the ChIP samples. The bamCompare function from DeepTools was used to normalize ChIP samples to the input or H3. Bigwig files of the log₂-transformed ratio of ChIP/input or H3 were then used to generate metaplots and heat maps across the different features in the genome (protein-coding genes, TE fragments and TEs (https://www.arabidopsis.org/download/index-auto.jsp?dir=%2Fdownload_files%2FGenes%2FTAIR10_genome_release%2FTAIR10_gf3), and H3K9me2 regions). Bedtools map was used to calculate the average log₂[ChIP/H3] enrichment over TEs in both *ddm1* mutants, WT Col-0, transgenic lines expressing DDM1 mutants and second-generation *ddm1* mutants. The log₂-transformed fold change of *ddm1*/Col-0 and DDM1(FL)/*ddm1* for each ChIP analysis was then calculated in R v.3.5.1 (<https://www.R-project.org/>). These data were used to generate scatter plots and calculate the correlations between TE

expression and log₂[fold change of *ddm1*/Col-0] for ChIP (H2A.W). DANPOS2 (v.2.2.2)⁷⁵ was used to perform differential binding analysis. Each sample was normalized to its respective input. ChIP samples for each histone mark/variant were compared between *ddm1* mutants and Col-0 using the dtriple function from DANPOS2. Wig files generated by dtriple were then used for calculating the ChIP enrichment across various genomic features, such as protein-coding genes, TE fragments and TEs. Bedtools map was next used to calculate the coverage of each ChIP. The log₂-transformed fold change for the enrichment between *ddm1* mutants and WT Col-0 for each ChIP was then calculated in R v.3.5.1 (<https://www.R-project.org/>). These data were then used to generate box plots across different ChIPs for different regions (protein-coding genes, TE fragments and TEs). We performed *t*-tests in R v.3.5.1 (<https://www.R-project.org/>) to determine the significance of the log₂[fold change of *ddm1*/Col-0] across different regions. To estimate the magnitude of the log₂[fold change of *ddm1*/Col-0] over different regions, we also calculated effect size (Cohen's *d*) for each comparison in the box plot using the effsize package v.0.7.6 (<https://github.com/mtorchiano/effsize/>) in R with the following thresholds: $|d| < 0.2$, negligible; $|d| < 0.5$, small; $|d| < 0.8$, medium; otherwise, large, as reported previously⁶¹.

ChIP for H2A.W in *suvh4/5/6* plants. Nuclei isolation was performed using seedlings (aged 10 d) from WT Col-0, *ddm1* mutants and *suvh4/5/6* triple mutants as described above. For ChIP-seq, sequencing libraries were prepared using the Nugen Ovation Ultralow V2 DNA-Seq library prep kit (NuGen) according to the manufacturer's instructions. The libraries were sequenced using the Illumina HiSeq 2000 system to generate single-end 50 bp reads. The data were analysed as described above.

RNA extraction. Total RNA was extracted using the Spectrum plant total RNA kit (Sigma-Aldrich) from 4–5-week-old leaves of WT Col-0, *F₁ DDM1/ddm1*, *ddm1* first-generation transgenic lines expressing DDM1 mutants, *met1* mutants, *suvh4/5/6* triple mutants and *ddm1* second-generation mutants. Contaminating DNA was removed using a DNA-free DNA removal kit (Thermo Fisher Scientific).

RNA-seq analysis. Ribo-Zero libraries. For RNA-seq analyses in WT Col-0, *F₁ DDM1/ddm1*, *ddm1* first-generation mutants, transgenic lines expressing DDM1 mutants and *ddm1* second-generation mutants, rRNA was removed using the Ribo-Zero rRNA Removal kit (Illumina). After depletion of rRNA, RNA-seq libraries were generated using the NEBNext UltraII directional RNA library prep kit for Illumina (New England Biolabs) according to the manufacturer's instructions. The libraries were sequenced using the Illumina HiSeq 2500 system to generate single-end 50 bp reads. Samples were prepared from three independent biological replicates.

Poly(A) libraries. For RNA-seq analyses of *met1*, *ddm1* and *suvh4/5/6* mutants, RNA-seq libraries were generated using the NEBNext UltraII directional RNA library prep kit for Illumina (New England Biolabs) according to the manufacturer's instructions. The libraries were sequenced using the Illumina HiSeq 2500 system to generate single-end 50 bp reads. Samples were prepared from three independent biological replicates.

For alignment and visualization, reads were mapped to the TAIR10 *Arabidopsis* genome using the STAR aligner⁷⁶ and bigwig files were generated using the bamCoverage function in DeepTools⁷⁴. Kallisto (v.0.46.0)⁷⁷ was used to estimate transcript abundance and read counts for all transcripts in TAIR10. The resulting values per transcript were mapped to gene level using R v.3.6.1 (<https://www.R-project.org/>) and the packages tximport (v.1.12.3)⁷⁸ and TxDb.Athaliana.BioMart.plantsmart28 v.3.2.2 (<https://bioconductor.org/packages/release/data/annotation/html/TxDb.Athaliana.BioMart.plantsmart28.html>). Statistical analysis was performed with R v.3.5.1 (<https://www.R-project.org/>) using the DESeq2 (v.1.22.2) package⁷⁹.

The pairwise comparison between WT Col-0 and *ddm1* mutants was performed using the default DESeq2 settings. Furthermore, the shrunken log₂-transformed fold changes were calculated using the default settings. Genes with a false-discovery rate (FDR)-adjusted $P < 0.05$ and an absolute shrunken log₂-transformed fold change of > 2 were classified as differentially expressed.

For analyses in *F₁ DDM1/ddm1* plants, genes with less than 10 reads per sample were excluded. To find genes that showed expression changes across any of the three genotypes (WT Col-0, *F₁ DDM1/ddm1* and *ddm1* (second generation)), a likelihood ratio test was performed. A full model including the genotype of each sample was compared to a reduced model without genotype information. TEs with a FDR-adjusted $P < 0.05$ and for which the log₂-transformed fold change between the most and least expressed genotype was larger than 1 were classified as differentially expressed. Details on how the log₂-transformed fold change was calculated are provided below. The transcript per million values of the differentially expressed TEs were *z*-transformed across the genotypes and the resulting *z* scores were clustered using *k*-means. On the basis of the within-cluster sum of square (WSS) values, the number of clusters was set to 6. On the basis of their profiles, the resulting clusters were classified as either neutral, inactivated or other (see below for details). Visualizations of the clustering results were generated using ComplexHeatmap (v.2.3.2)⁸⁰ and ggplot2 (v.3.1.0)⁸¹.

For the analyses of *met1*, *ddm1* and *svuh4/5/6* mutants, these files were processed as described above. To find genes differentially expressed across any of the three genotypes (*met1*, *ddm1* and *svuh4/5/6*), a likelihood ratio test was performed as described above, and the $\log_2$ -transformed fold change was calculated as described below. The transcript per million values of the differentially expressed TEs were z -transformed across the genotypes, and the resulting z scores were clustered using k -means. On the basis of the WSS values, the number of clusters was set to 8. On the basis of their profiles, the resulting clusters were classified as follows: D, DDM1 specific; SD, SUVH4/5/6/DDM1; MD, MET1/DDM1 specific; I, expressed in WT Col-0 (see below for details).

Calculation of $\log_2$ -transformed fold change in F_1 DDM1/*ddm1* and *met1*, *svuh4/5/6* and ddm1 datasets. The counts per gene dataset was $\log_2$ -transformed using the DESeq function `rlog()` with the default settings. Replicates were averaged and, for each gene, the fold change between the genotypes with the largest and smallest $\log_2$ -transformed value was calculated.

Analysis of F_1 DDM1/*ddm1* TE clusters. To find an appropriate number of clusters for the k -means clustering, we performed k -means clustering on the data for $k=2, 3, \dots, 15$. For each clustering, the WSS was calculated. $k=6$ was picked as a trade-off between a low WSS and a low number of clusters. Two clusters for which the TEs showed intermediate expression levels in F_1 DDM1/*ddm1* compared with the *ddm1* mutant were defined as neutral ($n=349$). Two clusters for which TE expression levels were closer to Col-0 than the *ddm1* mutant were defined as inactivated ($n=756$). One cluster with the highest expression in Col-0 ($n=22$) and another with the highest TE expression in F_1 DDM1/*ddm1* ($n=7$) were defined as other.

Analysis of *met1*, *svuh4/5/6* and *ddm1* TE clusters. To find an appropriate number of clusters for the k -means clustering, we performed k -means clustering on the data for $k=2, 3, \dots, 15$. For each clustering, the WSS was calculated. $k=8$ was picked as a trade-off between a low WSS and a low number of clusters. One cluster with the highest expression in *ddm1* ($n=542$) was defined as D. Two clusters for which the TE showed similar expression levels in *svuh4/5/6* and *ddm1* mutants were defined as SD ($n=299$). Four clusters for which the TE showed similar expression levels in *met1* and *ddm1* mutants were defined as MD ($n=1,006$). Another cluster with the highest TE expression in WT Col-0 ($n=31$) was defined as I.

RT-qPCR. For RT-qPCR, cDNA was synthesized from DNase-treated RNA (1 μ g) using the RevertAid RT Reverse Transcription Kit (Thermo Fisher Scientific) and 1.25 ng of synthesized cDNA was used for the qPCR assay using FastStart Essential DNA Green Master (Roche). The PCR reaction and quantitative measurement to obtain C_q values were performed using the LightCycler 96 (Roche) instrument. qPCR assays were performed with two independent technical replicates and three biological replicates for each sample. For each qPCR experiment, technical replicates were averaged. For each biological replicate ΔC_q values were obtained using the matched *ACT2* gene values as a reference. Mean and s.d. values for the ΔC_q values were calculated using the biological replicates. The fold changes were calculated as $2^{-\Delta C_q}$. The $\log_2$ -transformed fold changes between F_1 DDM1/*ddm1* and *ddm1* plants were calculated as $\Delta C_{q,ddm1} - \Delta C_{q,F_1}$. A list of the primers used for RT-qPCR is provided in Supplementary Table 2. qPCR analysis and figure generation were performed using R v.3.5.1 (<https://www.R-project.org/>) and package ggplot2 (v.3.1.0)⁸¹.

Fisher's exact test. We performed Fisher's exact tests using the GeneOverlap Package (<https://github.com/shenlab-sinai/geneoverlap>) in R to test the significance of the association between different TE superfamilies and the TE categories defined in this paper (upregulated in *ddm1* ($n=1,506$), inactivated ($n=756$) and neutral ($n=349$)). The adjusted P values and Jaccard index were plotted for each comparison.

H2A.W enriched regions. Regions enriched in H2A.W.6 were defined using deepTools2.0, R3.2.2 and bedtools; the input DNA was used as a control. For plotting H2A.W.6 over H2A.W.6-rich regions, we defined the $\log_2$ -transformed ratio of ChIP-H2A.W.6/input coverage with a bin size of 10 bp. We next selected all of the regions with scores of >0 and, finally, from the resulting regions, we selected only those with enrichment scores of ≥ 1 and a width of ≥ 200 (both had to be true).

DNA methylation analysis in Col-0, *ddm1* and Col-0 $\times$ *ddm1* (F_1 stage) plants. We compared DNA methylation in F_1 plants to an in silico cross (average of DNA methylation in Col-0 and *ddm1* plants) and calculated the z score for each TE. We categorized each TE into gain and no gain on the basis of its normalized DNA methylation values ($z > 1.65$ as gain, assuming a standard normal and P -value significance of <0.05). We repeated this for each methylation context (CG, CHG and CHH) independently.

Generation of transgenic plants. To construct the plasmid containing DDM1-3 $\times$ Flag and the promoter sequence of DDM1 corresponding to the 2,438 bp upstream plus 99 bp 5' untranslated region of the DDM1 gene, the DDM1 coding sequence and 3 $\times$ Flag sequence were amplified by PCR and cloned into the

entry vector pDONR P4-P1R, pDONR 221 or pDONR P2R-P3 by a BP reaction (Invitrogen). The fragments were then subcloned into a pAlligatorG43-derived plasmid⁸² using an LR reaction in MultiSite Gateway technology (Invitrogen) to obtain the *pDDM1:DDM1-3 $\times$ Flag* fusion construct. To construct the plasmid containing DDM1(Δ N)(Δ 25–123)-3 $\times$ Flag, DDM1(Δ C)(Δ 639–730)-3 $\times$ Flag or DDM1(Δ NC)(Δ 25–123; Δ 639–730)-3 $\times$ Flag, the plasmid vector *pDDM1:DDM1-3 $\times$ Flag* was used as a template for the mutagenesis. The *ddm1-2* heterozygous (*ddm1*^{+/+}) plants were transformed using the floral dipping method, and T_1 plants harbouring transgene in the *ddm1*^{+/+} background were selected. The T_2 -generation *ddm1* homozygous plants with the transgene were selected, and then the T_3 -generation *ddm1* homozygous plants with the transgene were used for further analyses. A list of the primers used to check the genotyping of *ddm1-2* and transgene plants is provided in Supplementary Table 2.

Statistics and reproducibility. All statistical procedures performed are indicated in the figure legends. ChIP-seq and EM-seq were performed using two replicates. RNA-seq analyses were performed using three replicates for each genotype. Western blot analyses were repeated three times, pull-down assays were repeated twice and ATPase assays were performed once. The numbers of biological replicates for each experiment are also provided in the captions. Differential gene expression analysis was performed using DESeq2. Genes with an FDR-adjusted $P < 0.05$ and an absolute shrunken $\log_2$ -transformed fold change of >2 were classified as differentially expressed. To estimate the magnitude of $\log_2$ [fold change of *ddm1*/Col-0] (ChIP-seq) over different regions, for each comparison in the box plot (Figs. 2e and 4e), the effect size (Cohen's d) was calculated using the effsize package in R with the following thresholds: $|d| < 0.2$, negligible; $|d| < 0.5$, small; $|d| < 0.8$, medium; otherwise, large. Furthermore, to test the significance of the association between different TE super-families and the TE categories defined in this paper (upregulated in *ddm1*, inactivated and neutral), we performed Fisher's exact tests using the GeneOverlap package in R. All bar plots (Figs. 2a and 4b) are labelled with precise P values, and the violin plots (Figs. 2e and 4e) are provided with effect sizes (Cohen's d). The P values and Cohen's d shown in the plots (Figs. 2a,e and 4b,e) and data used to generate the box plots (Figs. 4c, 5c–e and 6b) are provided as source data.

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

Data availability

The deep-sequencing (ChIP-seq, RNA-seq and EM-seq) data that support the findings of this study have been deposited to the NCBI Gene Expression Omnibus (GEO) under the accession number GSE150436 and the SRA (BioProject ID: PRJNA689609). All other data supporting the findings of this study are available from the corresponding author on reasonable request. Source data are provided with this paper.

Code availability

Custom code used to process and analyse the deep-sequencing data, as described in the Methods, is available from the corresponding author on request.

References

- Lister, R. et al. Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* **133**, 523–536 (2008).
- Kawakatsu, T. et al. Epigenomic diversity in a global collection of *Arabidopsis thaliana* accessions. *Cell* **166**, 492–505 (2016).
- Lowary, P. T. & Widom, J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.* **276**, 19–42 (1998).
- Arimura, Y., Tachiwana, H., Oda, T., Sato, M. & Kurumizaka, H. Structural analysis of the hexasome, lacking one histone H2A/H2B dimer from the conventional nucleosome. *Biochemistry* **51**, 3302–3309 (2012).
- Tanaka, Y. et al. Expression and purification of recombinant human histones. *Methods* **33**, 3–11 (2004).
- Tachiwana, H. et al. Structural basis of instability of the nucleosome containing a testis-specific histone variant, human H3T. *Proc. Natl Acad. Sci. USA* **107**, 10454–10459 (2010).
- Osakabe, A. et al. Vertebrate Spt2 is a novel nucleolar histone chaperone that assists in ribosomal DNA transcription. *J. Cell Sci.* **126**, 1323–1332 (2013).
- Lorkovic, Z. J. et al. Compartmentalization of DNA damage response between heterochromatin and euchromatin is mediated by distinct H2A histone variants. *Curr. Biol.* **27**, 1192–1199 (2017).
- Ito, T., Takahashi, N., Shimura, Y. & Okada, K. A serine/threonine protein kinase gene isolated by an in vivo binding procedure using the *Arabidopsis* floral homeotic gene product, AGAMOUS. *Plant Cell Physiol.* **38**, 248–258 (1997).
- Wollmann, H. et al. Dynamic deposition of histone variant H3.3 accompanies developmental remodeling of the *Arabidopsis* transcriptome. *PLoS Genet.* **8**, e1002658 (2012).

72. 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.* **1610**, 93–106 (2017).
73. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
74. Ramirez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
75. Chen, K. et al. DANPOS: dynamic analysis of nucleosome position and occupancy by sequencing. *Genome Res.* **23**, 341–351 (2013).
76. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
77. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
78. Sonesson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* **4**, 1521 (2015).
79. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
80. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
81. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2016).
82. Kawashima, T. et al. Dynamic F-actin movement is essential for fertilization in *Arabidopsis thaliana*. *eLife* **3**, e04501 (2014).

Acknowledgements

F.B. acknowledges support from the staff at 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 and H. Wang for sharing material; E. Sasaki, Z. Harvey, M. Borg,

A. Marí-Ordóñez, P. Refsing Andersen and J. M. Watson for discussions and reading of the manuscript. This research 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.), P26887, P28320, P32054, P30802 and P33380 (to F.B.), and DK1238 (to B.J. and S.A.M.), the Austrian Academy of Sciences (to F.B., Z.J.L., E.A., S.A., R.P. and R.Y.), MEXT/JSPS 15H05963 and 19H00995 (to T.K.) and JST CREST JPMJCR15O1 (to T.K.).

Author contributions

A.O., B.J. and E.A. performed the genome-wide analysis. A.O., B.J. and R.P. performed DNA methylation analysis in *ddm1* mutants. A.O., E.A. and S.A. performed RT-qPCR analysis of TEs in *F₁ DDM1/ddm1* plants and transgenic lines. A.O. and A.L.K. performed biochemical experiments. S.A.M. performed the phylogenetic analysis. R.Y. contributed the initial set of ChIP-seq data. Z.J.L. contributed the anti-H1 antibodies and T.K. contributed the design of experiments based on *F₁* plants. F.B. and A.O. conceived, designed and supervised the project. The manuscript was drafted by F.B., A.O. and B.J. with comments from all of the coauthors.

Competing interests

The authors declare no competing interests.

Additional information

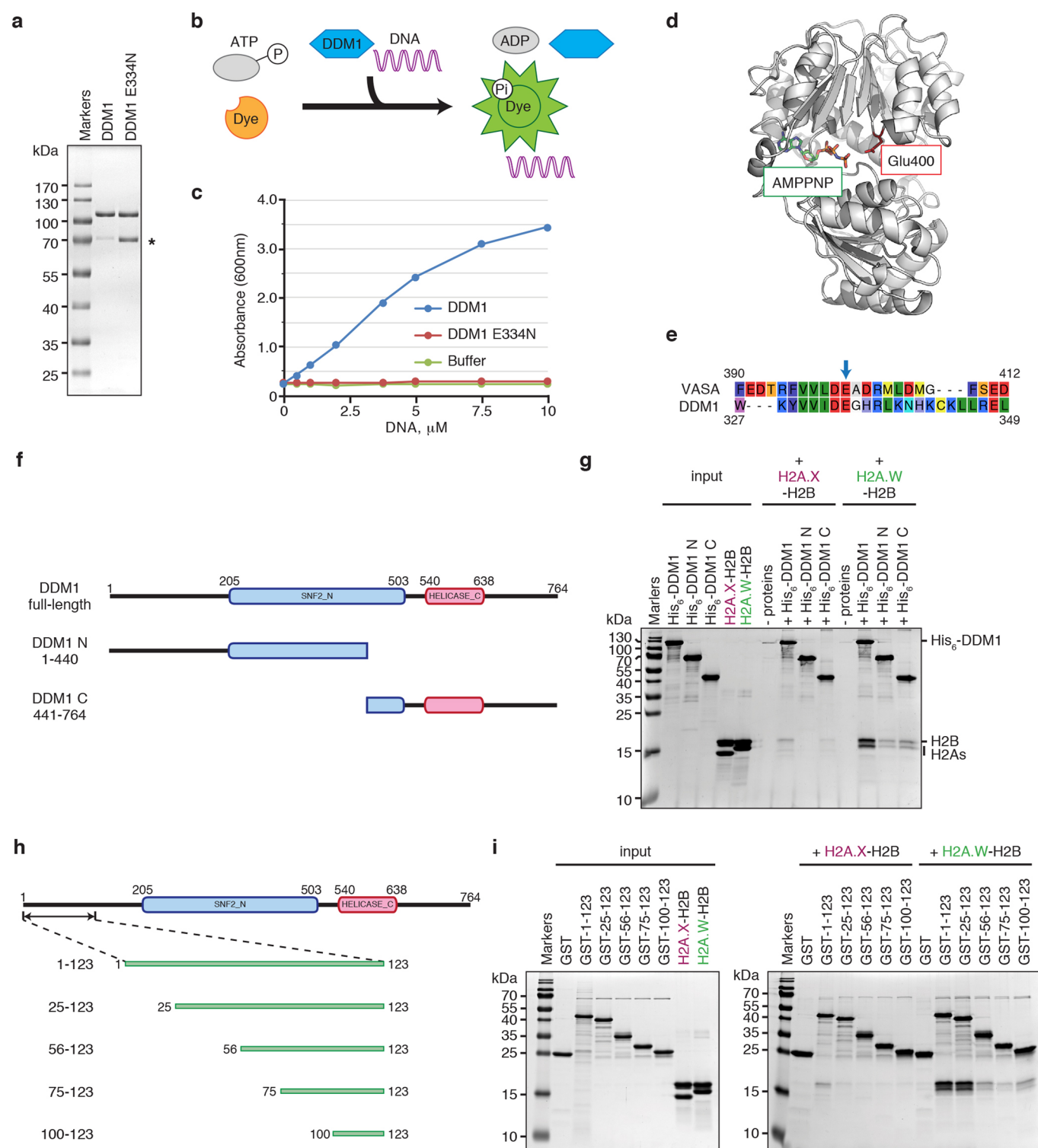
Extended data is available for this paper at <https://doi.org/10.1038/s41556-021-00658-1>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41556-021-00658-1>.

Correspondence and requests for materials should be addressed to Frédéric Berger.

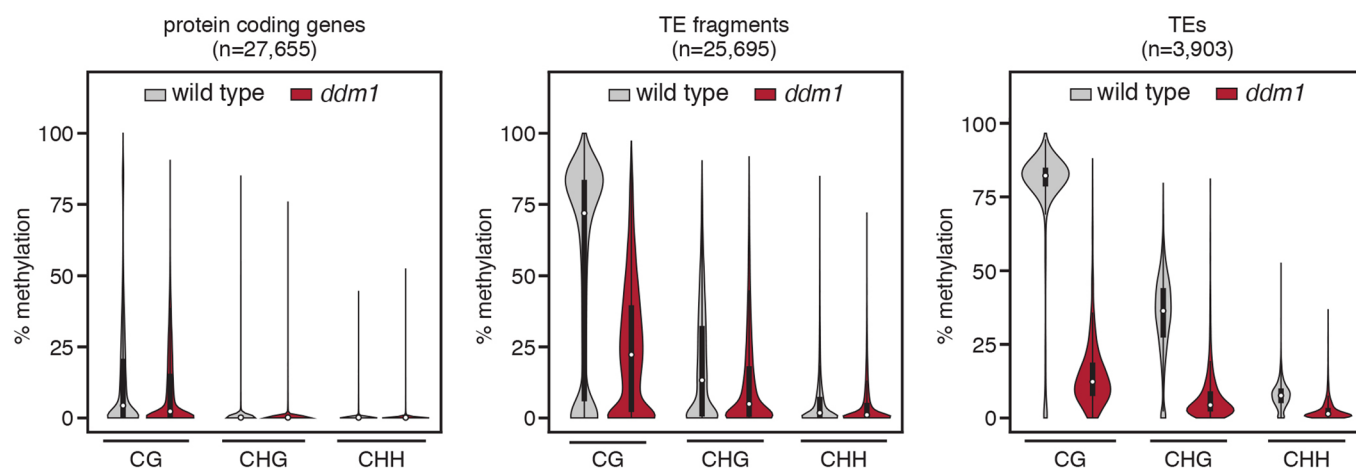
Peer review information *Nature Cell Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

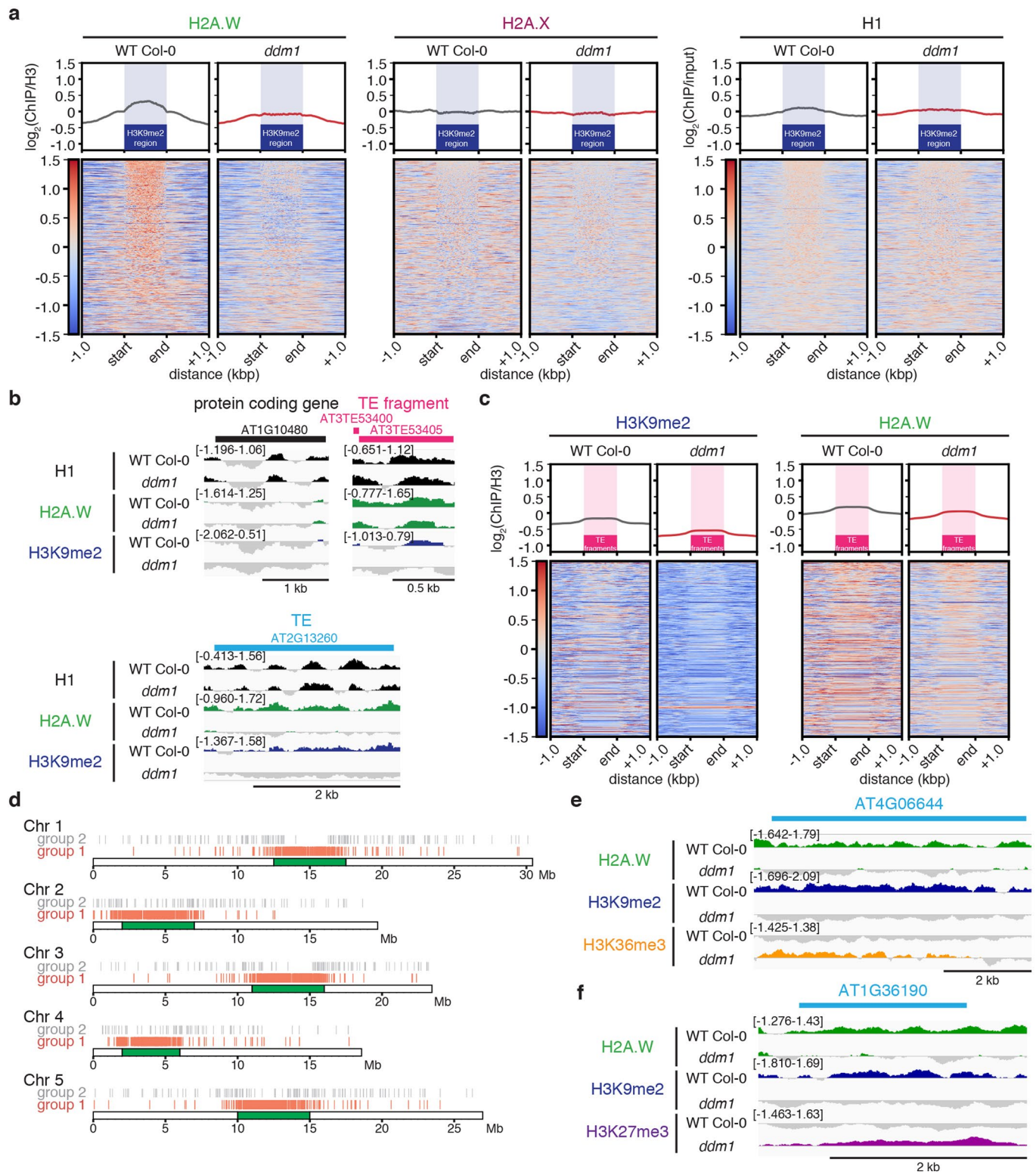


Extended Data Fig. 1 | See next page for caption.

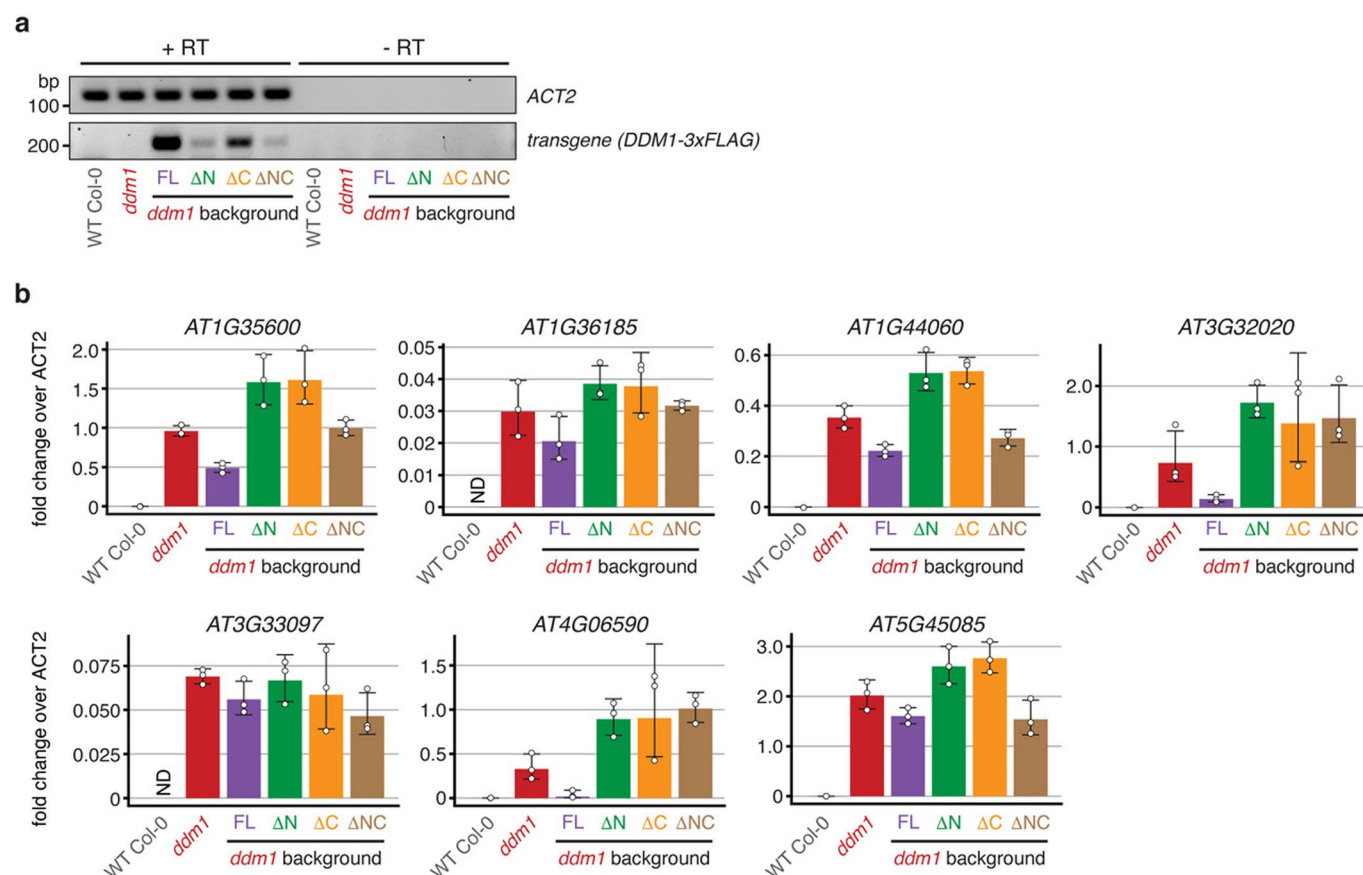
Extended Data Fig. 1 | Validation of recombinant DDM1 protein and H2A.X or H2A.W pulldown assay with DDM1 fragments. Related to Fig. 1. **a**, SDS-PAGE analysis of purified recombinant DDM1 and DDM1 E334N. Proteins were separated by 10% SDS-PAGE and visualized by Coomassie Brilliant Blue staining. Asterisk indicates the degraded products that could not be removed after purification. (Source data shows one of three representative gels obtained). **b**, Schematic representation of the ATPase assay. Malachite green solution was used to detect free phosphate liberated from ATP by enzyme-catalyzed hydrolysis. **c**, Graphical representation of the results of ATPase assays with buffer (green), wild type DDM1 (blue), or DDM1 E334N (red), respectively. DNA concentration is expressed as moles of nucleotides. **d**, The crystal structure of *Drosophila* RNA helicase protein VASA⁴⁹ together with the ATP analog (AMPPNP) (PDB ID: 2DB3). The glutamic acid residue at position 400, which coordinates a catalytic water molecule for stabilizing the interaction with ATP, is indicated. **e**, Sequence alignment between the Glu400 between *Drosophila* VASA and *Arabidopsis* DDM1 centered on GLU400, indicated by the blue arrow. Glu400 in VASA corresponds to Glu334 in *Arabidopsis*. **f, h** Schematic representation of the DDM1 fragments used in the pulldown assay. **g, i** SDS-PAGE analysis of the pulldown assay. His₆-tagged DDM1 and its truncations (panel g) or GST-tagged DDM1 fragments (panel i) were incubated with histone dimers containing either H2A.X or H2A.W. After pulldown, samples were analyzed by 15% SDS-PAGE and visualized by Coomassie Brilliant Blue. Data in a, g and i represent three independent experiments. Source data are provided in Source Data Extended Data Fig. 1.



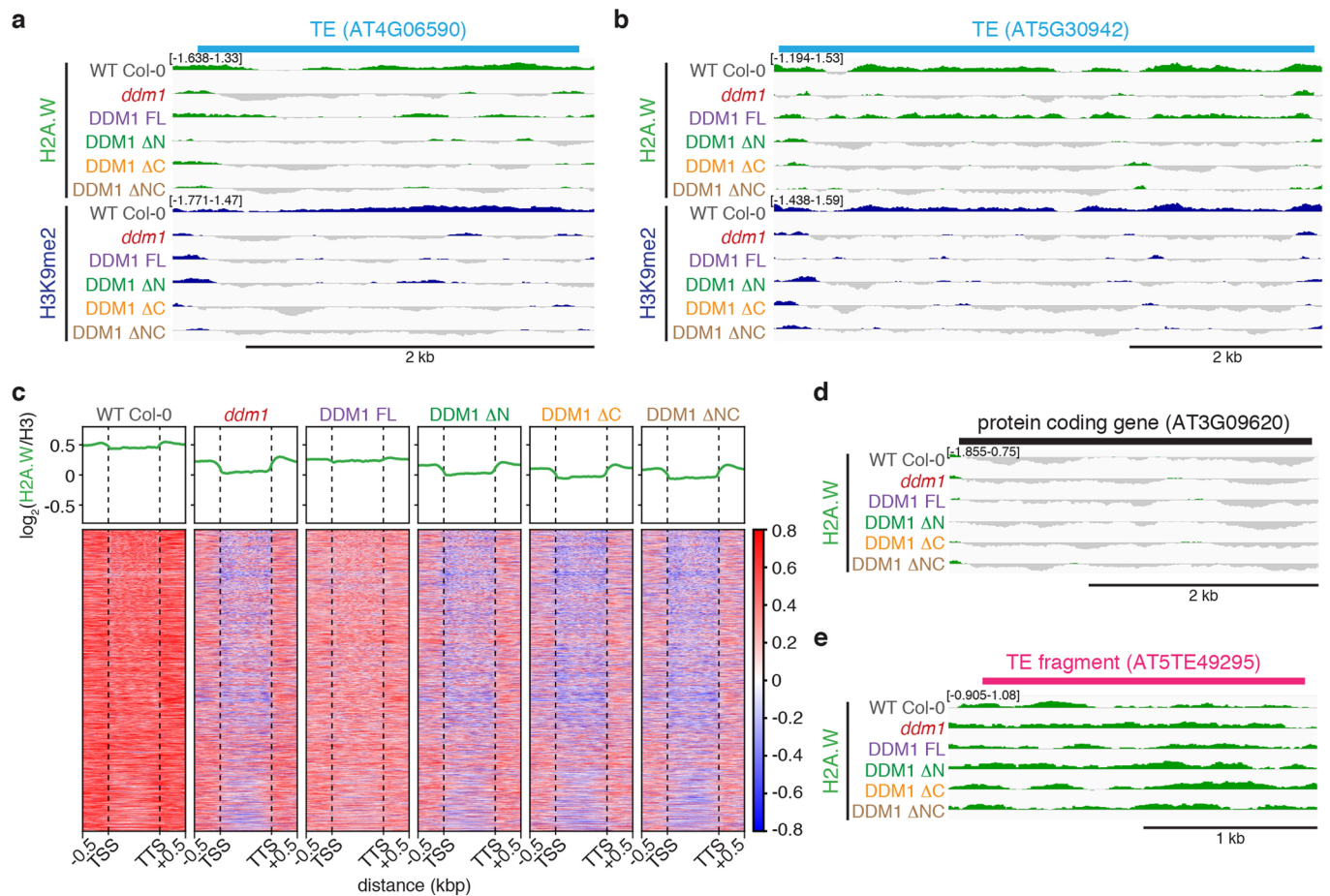
Extended Data Fig. 2 | DNA methylation in *ddm1*. Related to Fig. 2. Violin plots for the proportion of DNA methylation in CG, CHG, and CHH contexts on protein coding genes (left, $n=27,655$), TE fragments not including TEs (middle, $n=25,695$), and TEs (right, $n=3,903$) in wild type Col-0 (grey) and *ddm1* (red).



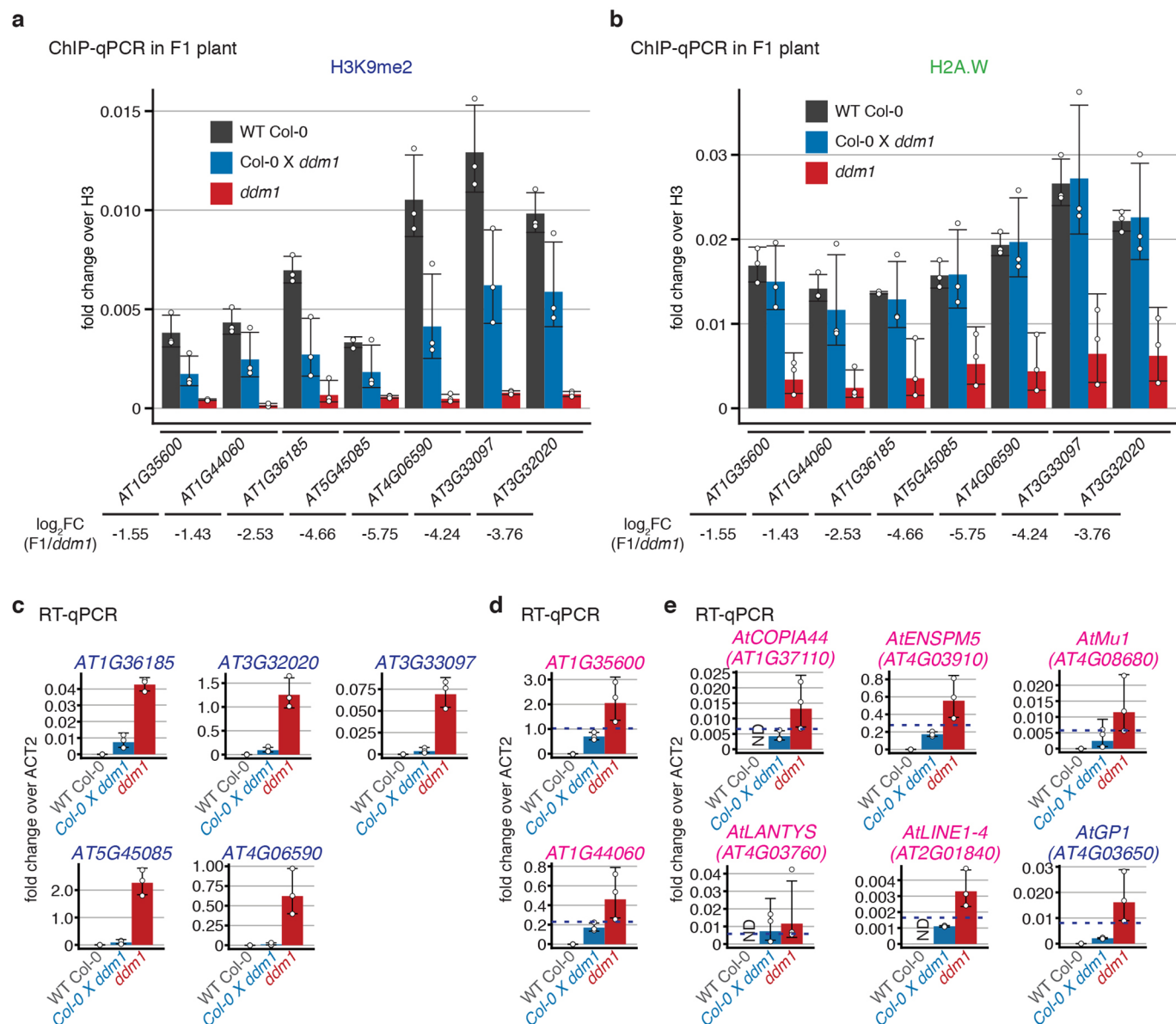
Extended Data Fig. 3 | Genomic distribution of H1, H3K9me2 and H2A.W in *ddm1*. Related to Figs. 2 and 3. **a, Aggregate profile plots (top) and heatmaps (bottom) of H2A.W, H2A.X, and H1 in Col-0 (grey) and *ddm1* (red) over heterochromatic regions, defined by H3K9me2 enrichment. Antibodies against H1 used in this study recognizes both H1.1 and H1.2. **b**, IGV genome browser snapshot of H1 (black), H2A.W (green), and H3K9me2 (blue) in wild type Col-0 and *ddm1*. Coloured and grey shading indicate an enriched or depleted signal, respectively. Annotated protein coding genes, TE fragments, and TEs are shown as boxes colored black, pink, or light blue, respectively. **c**, Aggregate profile plots (top) and heatmaps (bottom) showing distribution of H3K9me2 and H2A.W over TE fragments not including TEs in wild type and *ddm1*. **d**, Chromosomal distributions of TEs classified as group 1 (orange; $n=3,257$) and group 2 (grey; $n=646$) in Fig. 3a for each chromosome. Green boxes represent pericentric heterochromatin. **e**, **f** IGV genome browser snapshot of H2A.W (green), H3K9me2 (blue), H3K36me3 (orange), and H3K27me3 (purple) over TEs in wild type Col-0 and *ddm1*. Coloured and grey shading indicate an enriched or depleted signal, respectively.**



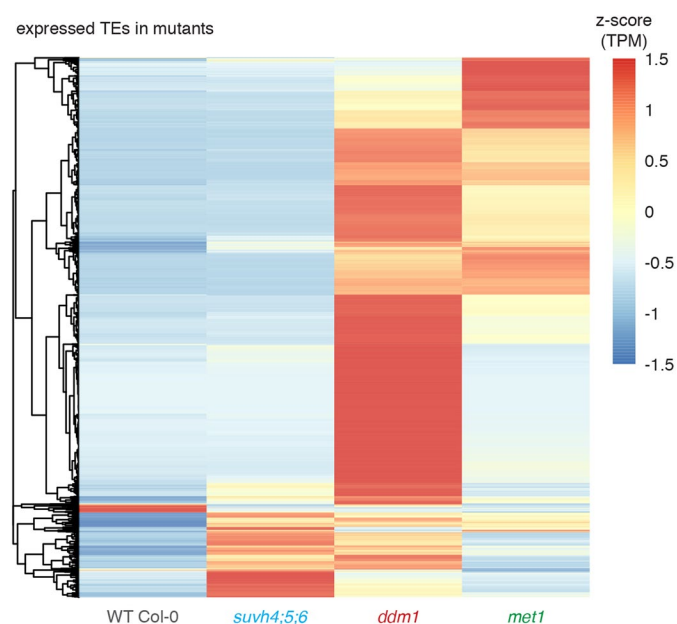
Extended Data Fig. 4 | RT-qPCR analyses of TEs in complementation lines expressing DDM1 mutant forms. Related to Fig. 4. a, RT-PCR analyses of transgenic lines. cDNA (+RT) and DNaseI-treated RNA (-RT) were prepared from each plant and used as template for RT-PCR analyses. ACT2 was used as a control for RT-PCR. PCR products were separated by agarose gel and visualized by SYBR Green I staining. **b**, Bar plots showing the results of RT-qPCR as fold change of enrichment over ACT2 at TEs analyzed in Extended Data Fig. 6c,d from wild type Col-0 (grey), *ddm1* (red), *ddm1* expressing DDM1-3xFLAG (purple), *ddm1* expressing DDM1 ΔN -3xFLAG (green), *ddm1* expressing DDM1 ΔC -3xFLAG (orange), or *ddm1* expressing DDM1 ΔNC -3xFLAG (brown), respectively. Presented data are means of $n=3$ independent biological replicates (each biological replicate has 2 technical repeats); error bars represent SD within biological replicates. Source data are provided in Source Data Extended Data Fig. 4.



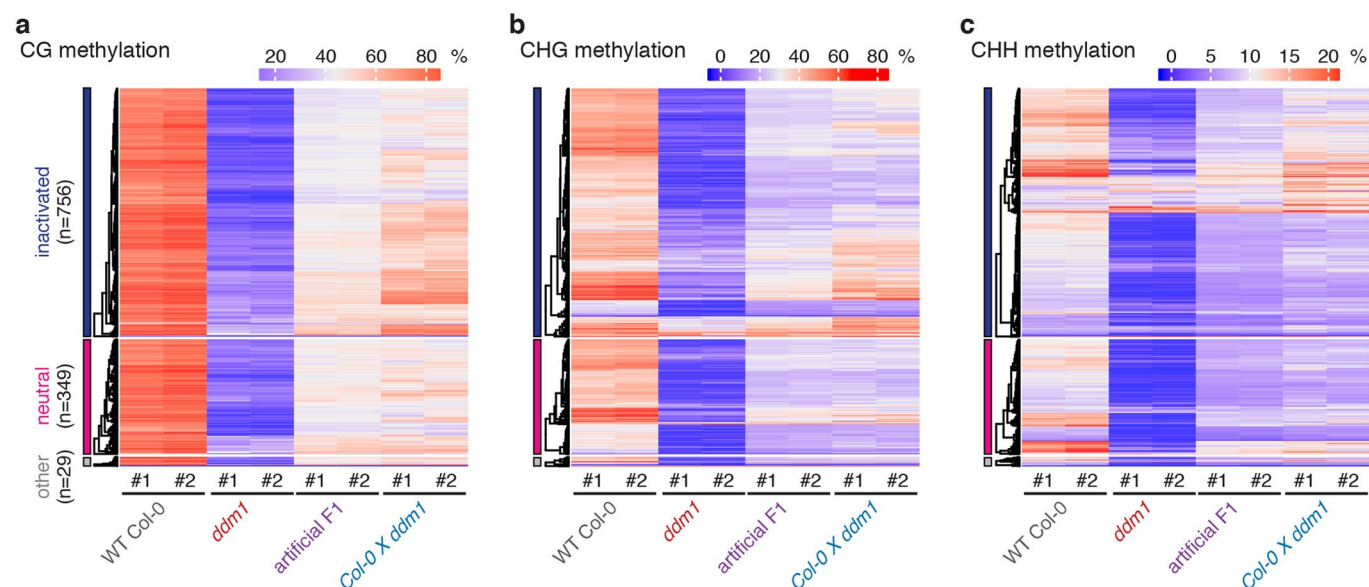
Extended Data Fig. 5 | Genomic distribution of H2A.W in complementation lines. Related to Fig. 4. **a, b, d, e** IGV genome browser snapshot of the distribution for H2A.W and H3K9me2 over TEs (panels a and b), protein coding gene (panel d) and TE fragment (panel e) in wild type, *ddm1*, *ddm1* expressing DDM1 FL, DDM1 Δ N, DDM1 Δ C, and DDM1 Δ NC. **c**, Aggregate profile plots (top) and heatmaps (bottom) showing distribution of H2A.W over TEs in wild type, *ddm1*, *ddm1* expressing DDM1 FL, DDM1 Δ N, DDM1 Δ C, and DDM1 Δ NC.



Extended Data Fig. 6 | ChIP-qPCR analyses of H3K9me2 or H2A.W over TEs and Expression of TEs in F1 *DDM1/ddm1*. Related to Fig. 5. a, b, Bar plots showing the results of ChIP-qPCR as fold change of H3K9me2 (panel a) or H2A.W (panel b) enrichment over H3 at TEs analyzed in Extended Data Fig. 6c, d from wild type Col-0 (grey), F1 *DDM1/ddm1* (blue), or *ddm1* (red), respectively. Presented data are means of $n=3$ independent biological replicates (each biological replicate has 3 technical repeats); error bars represent SD within biological replicates. The numbers below each TE show the expression level of TEs ($\log_2\text{FoldChange}(F1/ddm1)$) compared to *ddm1* mutant plants. **c–e**, Bar plots showing relative expression levels of TEs classified as inactivated (blue) or neutral (pink) measured by RT-qPCR from wild type Col-0 (grey), F1 *DDM1/ddm1* (blue), or *ddm1* (red), respectively. The expression levels of each TE are presented as a fold change over *ACT2*. Presented data are means of $n=3$ independent biological replicates (each biological replicate has 2 technical repeats); error bars represent SD within biological replicates. The dashed lines represent 50% of the *ddm1* fold change over *ACT2*. Source Data are provided in Source Data Extended Data Fig. 6.



Extended Data Fig. 7 | Transcriptome of TEs in *ddm1* and other mutants. Related to Fig. 6. Heatmap of the z-transformed expression levels of differentially expressed TEs ($n=1,878$) in mutants from wild type Col-0, *suvh4;5;6*, *met1*, and *ddm1*.



Extended Data Fig. 8 | DNA methylation of TEs in *ddm1* and F1 cross Col-0 x *ddm1*. Related to Fig. 7. Heatmaps showing weighted DNA methylation levels for TEs sorted in categories 'Inactivated', 'neutral' and 'other' in wild type Col-0, *ddm1*, artificial F1 cross [(Col-0 + *ddm1*)/2] and F1 cross Col-0 X *ddm1* for **a** CG, **b** CHG, and **c** CHH contexts.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | | |
|-------------------------------------|--|
| n/a | Confirmed |
| ☐ | ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| ☒ | ☐ A description of all covariates tested |
| ☐ | ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| ☒ | ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☐ | ☒ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Bowtie2 version 2.1.0.
fastqc v.0.11.8
fastx-toolkit v.0.0.14
Kallisto v0.46
tximport v1.12.3

Data analysis

deeptools v3.1.2
bedtools v2.27.1
DANPOS2 v2.2.2
ComplexHeatmap v2.3.2
DESeq2 v1.22.2
ggplot2 v3.1.0
R v3.5.1
R v3.6.1
methylypy (<https://github.com/yupenghe/methylypy>)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

GEO data:

Go to <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE150436>

This Super series is a part of 3 Subseries

* You may also cite the SubSeries that are linked to GSE150436:

RNA-seq (Dataset 1) :

<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE150433>

ChIP-seq :

<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE150434>

RNA-seq (Dataset 2) :

<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE150435>

SRA Data:

Data for EMseq and ChIP-seq

To view data on SRA please access

BioProject ID: PRJNA689609

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The sample size was based on similar studies as per usage for each specific procedure and the sample size meets the criteria for sound statistical results.
Data exclusions	a) RNA-seq: As detailed in the methods, for differential gene expression analysis of mutant transcriptomes, only transcripts that had 10 counts or more in at least one sample were included such that only genes with detectable expression were analysed. Such quality control steps are customary in the field and were pre-established. b) ChIP-seq: As detailed in the methods, reads were filtered with FASTX toolkit. Reads with phred score less than 30 were removed. Such quality control steps are customary in the field and were pre-established. c) Enzymatic methylation seq (EM-seq) : No data was excluded as detailed in the methods, The sequenced EM-seq treated reads were mapped to TAIR 10 assembly (https://www.arabidopsis.org/index.jsp) along with chloroplast genome using the methylpy (https://github.com/yupenghe/methylpy) pipeline Conversion rates were calculated based on methylation levels present on reads mapped to the chloroplast genome.
Replication	a) RNA-seq: Three biological replicates were used for RNA-seq analysis and were highly correlated. b) ChIP-seq: Two biological replicates were used for ChIP-seq analysis of each histone variant and histone mark and were highly correlated. c) Enzymatic methylation seq (EM-seq) : Two biological replicates were used for EM-seq analysis for each genotype and were highly correlated. For all other experiments at least three highly correlated replicates were obtained.
Randomization	Randomization was not relevant to any of the experiments because the genetic background of Arabidopsis inbred lines is uniform
Blinding	Blinding strategies are not relevant in the context of the experiments presented.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

Antibody, Clonality, Manufacturer, Catalogue number, Dilutions used
 anti-H3K36me3, mouse polyclonal, Abcam, ab9050, used on for ChIP
 anti-H3K27me3, rabbit polyclonal, Millipore 07-449, only used for ChIP
 anti-H3K9me2, mouse polyclonal, Abcam ab1220, 1:1,000 for western blot
 anti-H3, rabbit polyclonal, Abcam, ab1791, 1:1,000 for western blot
 anti-H4, rabbit polyclonal, Abcam, ab10158, 1:2,000 for western blot
 anti-H2A.X, rabbit polyclonal, homemade, 1:1,000 for western blot
 anti-H2A.W, rabbit polyclonal, homemade, 1:1,000 for western blot
 anti-H1, rabbit polyclonal, homemade, 1:1,000 for western blot
 anti-rabbit IgG-HRP conjugate, goat polyclonal, Bio-Rad, 170-6515, 1:10,000 for western blot

Validation

anti-H3K36me3 (Abcam, ab9050) is validated for use in Arabidopsis thaliana and for western blot and ChIP-seq on the manufacturer's website.
 anti-H3k27me3, (Millipore 07-449) is validated for use in Arabidopsis thaliana and for western blot and ChIP-seq on the manufacturer's website.
 anti-H3K9me2, (Abcam, ab1220) is validated for use in Arabidopsis thaliana and for western blot and ChIP-seq on the manufacturer's website.
 anti-H3 (Abcam, ab1791) is validated for use in Arabidopsis thaliana and for western blot and ChIP-seq on the manufacturer's website.
 anti-H4 (Abcam, ab10158) is validated for use in Arabidopsis thaliana and for western blot on the manufacturer's website.
 anti-H2A.X and H2A.W (homemade) are validated for use in Arabidopsis thaliana and for western blot on previous study (Yalegandula et al., 2014, Cell).
 anti-H1 (homemade) is validated for use in Arabidopsis thaliana as shown in Supplementary Information.

ChIP-seq

Data deposition

- ☒ Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).
☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links

May remain private before publication.

To view GEO accession GSE150434:
 Go to <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE150434>

To view data on SRA please access
 BioProject ID: PRJNA689609
<https://dataview.ncbi.nlm.nih.gov/object/PRJNA689609?reviewer=jpc9eic17kq8f5u2g16b9ppgul>

Files in database submission

GSM4550558 Col_Input_Rep1
 GSM4550559 Col_H3_Rep1
 GSM4550560 Col_H2AX_Rep1
 GSM4550561 Col_H2AW6_Rep1
 GSM4550562 Col_H3K9me2_Rep1
 GSM4550563 ddm1_Input_Rep1
 GSM4550564 ddm1_H3_Rep1
 GSM4550565 ddm1_H2AX_Rep1
 GSM4550566 ddm1_H2AW6_Rep1
 GSM4550567 ddm1_H3K9me2_Rep1
 GSM4550568 Col_Input_Rep2
 GSM4550569 Col_H3_Rep2
 GSM4550570 Col_H2AX_Rep2
 GSM4550571 Col_H2AW6_Rep2
 GSM4550572 Col_H3K9me2_Rep2
 GSM4550573 ddm1_Input_Rep2
 GSM4550574 ddm1_H3_Rep2
 GSM4550575 ddm1_H2AX_Rep2

GSM4550576 ddm1_H2AW6_Rep2
GSM4550577 ddm1_H3K9me2_Rep2
GSM4550578 Col_Input_set2_Rep1
GSM4550579 Col_H1_set2_Rep1
GSM4550580 ddm1_Input_set2_Rep1
GSM4550581 ddm1_H1_set2_Rep1
GSM4550582 Col_Input_set2_Rep2
GSM4550583 Col_H1_set2_Rep2
GSM4550584 ddm1_Input_set2_Rep2
GSM4550585 ddm1_H1_set2_Rep2
SAMN17213359 Col0 Input Rep1
SAMN17213360 Col0 H3 Rep1
SAMN17213361 Col0 H1 Rep1
SAMN17213362 Col0 H3K9me2 Rep1
SAMN17213363 Col0 H2AX Rep1
SAMN17213364 Col0 H2AW6 Rep1
SAMN17213365 ddm1_2ndgen Input Rep1
SAMN17213366 ddm1_2ndgen H3 Rep1
SAMN17213367 ddm1_2ndgen H1 Rep1
SAMN17213368 ddm1_2ndgen H3K9me2 Rep1
SAMN17213369 ddm1_2ndgen H2AX Rep1
SAMN17213370 ddm1_2ndgen H2AW6 Rep1
SAMN17213371 DDM1_FL Input Rep1
SAMN17213372 DDM1_FL H3 Rep1
SAMN17213373 DDM1_FL H1 Rep1
SAMN17213374 DDM1_FL H3K9me2 Rep1
SAMN17213375 DDM1_FL H2AX Rep1
SAMN17213376 DDM1_FL H2AW6 Rep1
SAMN17213377 DDM1_dN Input Rep1
SAMN17213378 DDM1_dN H3 Rep1
SAMN17213379 DDM1_dN H1 Rep1
SAMN17213380 DDM1_dN H3K9me2 Rep1
SAMN17213381 DDM1_dN H2AX Rep1
SAMN17213382 DDM1_dN H2AW6 Rep1
SAMN17213383 DDM1_dC Input Rep1
SAMN17213384 DDM1_dC H3 Rep1
SAMN17213385 DDM1_dC H1 Rep1
SAMN17213386 DDM1_dC H3K9me2 Rep1
SAMN17213387 DDM1_dC H2AX Rep1
SAMN17213388 DDM1_dC H2AW6 Rep1
SAMN17213389 DDM1_dNC Input Rep1
SAMN17213390 DDM1_dNC H3 Rep1
SAMN17213391 DDM1_dNC H1 Rep1
SAMN17213392 DDM1_dNC H3K9me2 Rep1
SAMN17213393 DDM1_dNC H2AX Rep1
SAMN17213394 DDM1_dNC H2AW6 Rep1
SAMN17213395 Col0 Input Rep2
SAMN17213396 Col0 H3 Rep2
SAMN17213397 Col0 H1 Rep2
SAMN17213398 Col0 H3K9me2 Rep2
SAMN17213399 Col0 H2AX Rep2
SAMN17213400 Col0 H2AW6 Rep2
SAMN17213401 ddm1_2ndgen Input Rep2
SAMN17213402 ddm1_2ndgen H3 Rep2
SAMN17213403 ddm1_2ndgen H1 Rep2
SAMN17213404 ddm1_2ndgen H3K9me2 Rep2
SAMN17213405 ddm1_2ndgen H2AX Rep2
SAMN17213406 ddm1_2ndgen H2AW6 Rep2
SAMN17213407 DDM1_FL Input Rep2
SAMN17213408 DDM1_FL H3 Rep2
SAMN17213409 DDM1_FL H1 Rep2
SAMN17213410 DDM1_FL H3K9me2 Rep2
SAMN17213411 DDM1_FL H2AX Rep2
SAMN17213412 DDM1_FL H2AW6 Rep2
SAMN17213413 DDM1_dN Input Rep2
SAMN17213414 DDM1_dN H3 Rep2
SAMN17213415 DDM1_dN H1 Rep2
SAMN17213416 DDM1_dN H3K9me2 Rep2
SAMN17213417 DDM1_dN H2AX Rep2
SAMN17213418 DDM1_dN H2AW6 Rep2
SAMN17213419 DDM1_dC Input Rep2
SAMN17213420 DDM1_dC H3 Rep2
SAMN17213421 DDM1_dC H1 Rep2
SAMN17213422 DDM1_dC H3K9me2 Rep2
SAMN17213423 DDM1_dC H2AX Rep2
SAMN17213424 DDM1_dC H2AW6 Rep2

SAMN17213425 DDM1_dNC Input Rep2
 SAMN17213426 DDM1_dNC H3 Rep2
 SAMN17213427 DDM1_dNC H1 Rep2
 SAMN17213428 DDM1_dNC H3K9me2 Rep2
 SAMN17213429 DDM1_dNC H2AX Rep2
 SAMN17213430 DDM1_dNC H2AW6 Rep2

Genome browser session
 (e.g. [UCSC](#))

Processed and analysed genomic data (BigWig files) has been deposited in GEO and is publicly accessible

Methodology

Replicates

All ChIP-seq datasets had 2 biological replicates.

Sequencing depth

All samples were sequenced single-end 50 with depth ranging between ~5-27.5 million reads. Alignment rate ranged between 12-90%, with the total number of reads resulting in >3x coverage of the Arabidopsis genome. Mapping statistics are detailed below:

Sample	Total	Aligned	Uniquely	Mapped	Run	Type
Col_H1_set2_Rep1	27503123	20021311	SR50			
Col_H1_set2_Rep2	26022125	19262294	SR50			
Col_H2AW6_Rep1	12555406	7005453	SR50			
Col_H2AW6_Rep2	15939799	8301641	SR50			
Col_H2AX_Rep1	8605520	6726250	SR50			
Col_H2AX_Rep2	10368773	8335591	SR50			
Col_H3K9me2_Rep1	8529845	4034985	SR50			
Col_H3K9me2_Rep2	14738477	6610932	SR50			
Col_H3_Rep1	7739364	5420255	SR50			
Col_H3_Rep2	5016721	3870904	SR50			
Col_Input_Rep1	15539227	11139340	SR50			
Col_Input_Rep2	17226342	11232192	SR50			
Col_Input_set2_Rep1	14236393	10941625	SR50			
Col_Input_set2_Rep2	9859939	7768824	SR50			
ddm1_H1_set2_Rep1	17481815	13134299	SR50			
ddm1_H1_set2_Rep2	9848786	7777209	SR50			
ddm1_H2AW6_Rep1	14520341	5639682	SR50			
ddm1_H2AW6_Rep2	18197054	7655186	SR50			
ddm1_H2AX_Rep1	10172891	5597309	SR50			
ddm1_H2AX_Rep2	10795056	8213285	SR50			
ddm1_H3K9me2_Rep1	13600456	4340452	SR50			
ddm1_H3K9me2_Rep2	18447520	2913684	SR50			
ddm1_H3_Rep1	10379095	6459543	SR50			
ddm1_H3_Rep2	12352154	8481640	SR50			
ddm1_Input_Rep1	24588427	17374643	SR50			
ddm1_Input_Rep2	16225107	10634611	SR50			
ddm1_Input_set2_Rep1	10362120	8139821	SR50			
ddm1_Input_set2_Rep2	14596227	11132812	SR50			
Col0 Input Rep1	13234731	8839068	SR100			
Col0 H3 Rep1	14007164	9788148	SR100			
Col0 H1 Rep1	12526930	8904491	SR100			
Col0 H3K9me2 Rep1	13420677	8857869	SR100			
Col0 H2AX Rep1	10908766	8341343	SR100			
Col0 H2AW6 Rep1	11786175	7537411	SR100			
ddm1_2ndgen Input Rep1	14725076	9710317	SR100			
ddm1_2ndgen H3 Rep1	13893199	9289187	SR100			
ddm1_2ndgen H1 Rep1	12075391	8281865	SR100			
ddm1_2ndgen H3K9me2 Rep1	13854542	9685987	SR100			
ddm1_2ndgen H2AX Rep1	12795107	9398991	SR100			
ddm1_2ndgen H2AW6 Rep1	12474538	7792899	SR100			
DDM1_FL Input Rep1	12289236	8564068	SR100			
DDM1_FL H3 Rep1	14125381	9830110	SR100			
DDM1_FL H1 Rep1	11936902	8226286	SR100			
DDM1_FL H3K9me2 Rep1	13116683	8721661	SR100			
DDM1_FL H2AX Rep1	11592884	8627257	SR100			
DDM1_FL H2AW6 Rep1	10340124	6063557	SR100			
DDM1_dN Input Rep1	13290418	8805640	SR100			
DDM1_dN H3 Rep1	12464139	8644979	SR100			
DDM1_dN H1 Rep1	11458492	7992485	SR100			
DDM1_dN H3K9me2 Rep1	14037882	9501095	SR100			
DDM1_dN H2AX Rep1	11188095	8364418	SR100			
DDM1_dN H2AW6 Rep1	11341178	6710918	SR100			
DDM1_dC Input Rep1	14799117	9730313	SR100			
DDM1_dC H3 Rep1	14321292	10106691	SR100			
DDM1_dC H1 Rep1	13503049	9439127	SR100			
DDM1_dC H3K9me2 Rep1	13845302	9431385	SR100			
DDM1_dC H2AX Rep1	13013288	9639510	SR100			
DDM1_dC H2AW6 Rep1	13947209	8196906	SR100			

DDM1_dNC Input Rep1 16190922 9810552 SR100
 DDM1_dNC H3 Rep1 13412061 9684707 SR100
 DDM1_dNC H1 Rep1 13657768 9497811 SR100
 DDM1_dNC H3K9me2 Rep1 16246032 10832124 SR100
 DDM1_dNC H2AX Rep1 12492323 9469689 SR100
 DDM1_dNC H2AW6 Rep1 14163128 8931833 SR100
 Col0 Input Rep2 14890230 10263125 SR100
 Col0 H3 Rep2 13392782 9886476 SR100
 Col0 H1 Rep2 12841326 9374715 SR100
 Col0 H3K9me2 Rep2 15696554 10229982 SR100
 Col0 H2AX Rep2 12654910 9985949 SR100
 Col0 H2AW6 Rep2 12952988 8415375 SR100
 ddm1_2ndgen Input Rep2 12020391 8016441 SR100
 ddm1_2ndgen H3 Rep2 15812571 10596721 SR100
 ddm1_2ndgen H1 Rep2 13074741 8949211 SR100
 ddm1_2ndgen H3K9me2 Rep2 14732154 10094521 SR100
 ddm1_2ndgen H2AX Rep2 13133145 9874805 SR100
 ddm1_2ndgen H2AW6 Rep2 13379028 8205736 SR100
 DDM1_FL Input Rep2 13468456 8876116 SR100
 DDM1_FL H3 Rep2 15299720 11102992 SR100
 DDM1_FL H1 Rep2 13650178 9471337 SR100
 DDM1_FL H3K9me2 Rep2 18459330 12922073 SR100
 DDM1_FL H2AX Rep2 11547528 8830412 SR100
 DDM1_FL H2AW6 Rep2 15466936 9509637 SR100
 DDM1_dN Input Rep2 14345777 9789594 SR100
 DDM1_dN H3 Rep2 14643027 10077584 SR100
 DDM1_dN H1 Rep2 14697149 10150921 SR100
 DDM1_dN H3K9me2 Rep2 15596759 10815282 SR100
 DDM1_dN H2AX Rep2 12325483 9166424 SR100
 DDM1_dN H2AW6 Rep2 12613240 7076990 SR100
 DDM1_dC Input Rep2 11974640 7791434 SR100
 DDM1_dC H3 Rep2 11894346 8344541 SR100
 DDM1_dC H1 Rep2 12187630 9230489 SR100
 DDM1_dC H3K9me2 Rep2 18463116 12817151 SR100
 DDM1_dC H2AX Rep2 12292252 9244917 SR100
 DDM1_dC H2AW6 Rep2 13547771 7957218 SR100
 DDM1_dNC Input Rep2 19742022 12853759 SR100
 DDM1_dNC H3 Rep2 11797477 8320082 SR100
 DDM1_dNC H1 Rep2 14211500 9551617 SR100
 DDM1_dNC H3K9me2 Rep2 16655074 11522353 SR100
 DDM1_dNC H2AX Rep2 16402246 12059492 SR100
 DDM1_dNC H2AW6 Rep2 13984805 8337167 SR100

Antibodies

Antibody, Clonality, Manufacturer, Catalogue number, Amount used

anti-H3, rabbit polyclonal, Abcam, ab1791, 5 µg for ChIP-seq
 anti-H3K9me2, mouse monoclonal, Abcam, ab1220, 5 µg for ChIP-seq
 anti-H3K36me3, rabbit polyclonal, Abcam, ab9050, 5 µg for ChIP-seq
 anti-H3K27me3, rabbit polyclonal, Millipore, 07-449, 5 µg for ChIP-seq
 anti-H2A.X, rabbit polyclonal, homemade, 5 µg for ChIP-seq
 anti-H2A.W, rabbit polyclonal, homemade, 5 µg for ChIP-seq
 anti-H1, rabbit polyclonal, homemade, 5 µg for ChIP-seq

Peak calling parameters

No peaks were called on this data. Normalized Bigwig files were generated as follows.

As detailed in the methods;
 Reads were filtered and then aligned to the TAIR10 Arabidopsis genome using Bowtie2 version 2.1.0. with default settings.
 The bamCompare function from Deeptools was used to normalize ChIP samples to Input or H3. log2ratio (ChIP/ (Input or H3)) bigwig files were then used to generate metaplots and heatmaps across the different features in the genome.
).
 Bedtools map was used to calculate average Log2(ChIP/H3) enrichment over TEs in all genotypes. Log2Fold Change (FC) (ddm1/Col-0) and (DDM1 FL/ddm1) for each ChIP was then calculated in R v3.5.1. This data was used to generate scatterplots and calculate correlations between TE expression and log2FC (ddm1/Col-0) for ChIP (H2A.W).

Data quality

Reads were filtered using FASTX (fastx-toolkit v.0.0.14)
 Command used : fastq_quality_filter -q 30 -p 90 -Q33 -i \${input} -v -o \${id}_filtered.fastq

Software

fastx-toolkit v.0.0.14
 deeptools v3.1.2
 samtools v1.9
 Bowtie2 version 2.1.0.
 bedtools v2.27.1
 fastqc v.0.11.8

DANPOS2 v2.2.2
R v3.5.1

4 Discussion

4.1 DDM1 deposits H2A.W to silence transposons

The collaborative work with my colleagues show that the chromatin remodeler DDM1 binds to the heterochromatic variant H2A.W and mediates its deposition to chromatin on certain transposons and influence their transcriptional status. Two distinct regions in DDM1 are important for binding H2A.W-H2B heterodimers. Both these regions are highly conserved in LSH and HELLS, the mammalian homologs of DDM1. In LSH deposits macroH2A to heterochromatin indicating the conserved mechanism with DDM1. In previous studies, *ddm1-2* mutant was associated with global loss of CG methylation and H3K9me2. As defined in the first part of my thesis, H2A.W also acts as an additional layer of heterochromatin and is tightly linked with transposons. We could show that H2A.W is specifically lost on certain transposons in the *ddm1-2* background. And the total loss of H2A.W could not be explained by transcription alone. Furthermore, our data confirmed that this loss of H2A.W could be rescued by overexpressing full-length DDM1 in the mutant background. And this rescue was independent of H3K9me2. Additionally, our data showed that the two conserved binding regions in DDM1 are necessary for H2A.W deposition and transposon repression. We also showed that DDM1 acts in trans to deposit H2A.W and repress TEs in an F1 plants from the backcross of *ddm1* x WT. Expression analysis showed that 70% of TEs de-repressed in *ddm1-2* recovered H2A.W enrichment and were re-silenced in F1 plants. What drives DDM1 to target expressed TEs remains unknown but at least we established that this targeting does not depend on H3K9me1/2. We conclude from this study that H2A.W is an important part of

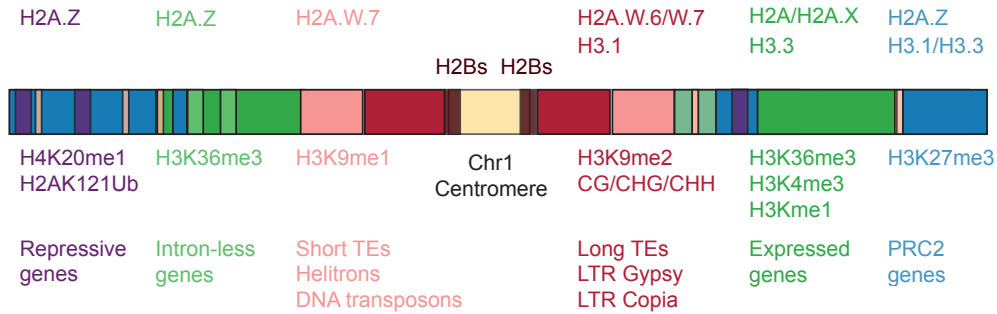
heterochromatin and recent evidence also shows that H2A.W and H3K9me act in synergy to maintain transposons silencing [Bourguet et al., 2022, Bourguet et al., 2021].

4.2 DDM1-H2A.W mediated silencing and other players

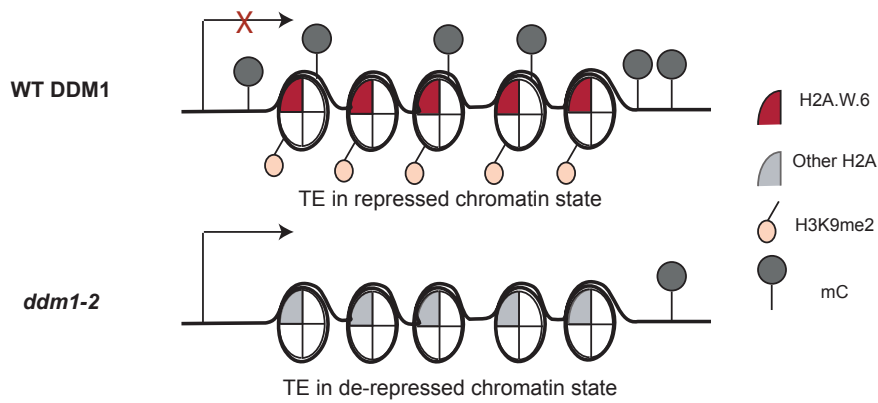
In our study we also investigated the role of the redundancy between MET1 (Methyltransferase 1), DDM1, SUVH4/5/6 (Histone methyltransferase depositing H3K9me2) controlled transposons repression. *ddm1-2* exhibits global loss of CG methylation, this could be partly attributed to the loss of H3K9me2 and its feedback loop with DNA methylation. But mechanisms linking DNA methylation and action of DDM1 still need to be uncovered. The mutants *met1* and *ddm1-2* show the strongest derepression of transposons and in *ddm1*, CG and non CG methylation is much reduced [Kakutani, 1997, Vongs et al., 1993, Wendte and Schmitz, 2018] and as expected a large set of TEs are deregulated in both *ddm1-2* and *met1*. This observation led to the idea that CG methylation is involved in TE silencing in the DDM1 pathway. Yet, several observations also suggest that DDM1 silences TEs in a CG methylation independent manner. A subset of TEs is derepressed in *ddm1-2* but not in *met1*. Re-silencing of TEs in F1 plants was not entirely explained by re-establishing DNA methylation over TEs. Importantly, mutants deprived of DDM1 and H2A.W have the same impact on TE expression than DDM1 showing that DDM1 acts upstream of H2A.W to silence TEs (Bourguet et al.). Yet, the link between DDM1 and MET1 remains unclear. It is possible as proposed that DDM1 controls access of MET1 and other DNA methyltransferases to TEs. This could be mediated by the nucleosome remodelling activity of DDM1. Because DNA methylation is required to recruit H3K9 methyltransferases in Arabidopsis, the lack of DNA methylation in *ddm1* would cause a loss of H3K9me1/2. To investigate further the role of DNA methylation in DDM1-H2A.W mediated silencing we need to look at the role of MET1 in the recruitment of H2A.W and to find the mechanisms that recruits DDM1 to deposit a specific heterochromatin state at its targets.

4.2 DDM1-H2A.W mediated silencing and other players

A Chromatin Landscape in WT Col-0



B DDM1 maintains specific TEs in repressed state by depositing H2A.W



C Chromatin Landscape in *ddm1-2*

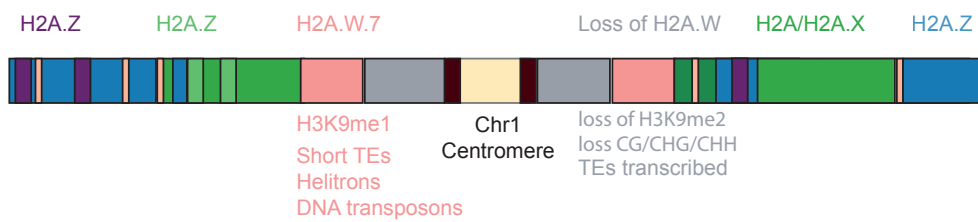


Figure 4.1: A) Schematic representation summarizing the chromatin landscape defined using chromosome 1 of Arabidopsis genome. B) Schematic representation of transposon landscape silenced by DDM1-H2A.W C) Hypothesized representation of chromatin landscape in *ddm1-2*

4.3 H2A variants are the major drivers in defining chromatin landscapes

Beyond the particular dissection of the role of DDM1 in the deposition of H2A.W, our analysis of combinations of all histone variants with the most studied histone PTMs shows that histone variants contribute strongly in establishing the various domains of chromatin in *Arabidopsis*. We have confirmed the strong association between H2A variants and PTMs. H2A.W.7 and H3K9me1 marked all heterochromatic states. Facultative heterochromatic states are defined by occupancy of H2A.Z associated with K121Ub and H3K27me3 supporting further previous results [Dai et al., 2017, Sura et al., 2017, Zhou et al., 2017]. To the exception of the TSS +1 nucleosomes, euchromatin is primarily occupied by H2A and H2A.X in association with various acetylated forms of H3 and H4 and H3K36me3. By contrast H3 variants do not establish strong associations with specific PTMs, although H3.1 is the preferred substrate of the deposition of H3K27me1 by ATXR5/6 [Jacob et al., 2009] and our data also outline a potential link between H3.3 and the deposition of H3K36me2.

The origins of the specific association between H2A variants with specific PTMs may be explained by the impact of H2A variants on the biophysical properties of the nucleosome. For example arrays of nucleosomes containing H2A.Z are the preferred substrates for PRC2. Whether such a concept also applies for the deposition of H3K9me or H3K36me remains to be challenged by *in vitro* experiments. Alternatively it is possible that a reader of the PTM influences the preferential deposition of a certain type of H2A variant. *In silico* analyses based on ChromHMM also highlights the importance of histone variants over histone PTMs in the definition of chromatin states. Taking into account primarily PTMs does outline the broad division of the chromatin into euchromatin, facultative and constitutive heterochromatin, but does not enable to define the difference between chromatin states within these domains. From our experience both PTMs and histone variants should be considered to have a comprehensive view of the chromatin organisation in any organism and the definition of chromatin states should thus be revised.

4.4 The new concept of chromatin landscape as assembly of chromatin states

Unsupervised clustering of chromatin states led to realise that they do assemble with each other in specific combinations which define chromatin landscape over genes and transposons. Each chromatin landscape is associated with a specific set of biological traits which support the uniqueness and biological significance of each chromatin landscape. Seven CLs define different groups of transposons distinguished by their location, family of origin and associated mode of transcriptional repression, providing supporting evidence for the selective adaptation of transposons to distinct silencing machineries (Bourguet et al. n.d. See reviews from the Feschotte lab). Shorter TEs in intergenic regions are occupied by landscapes distinct from the five landscapes that occupy constitutive heterochromatin. These TEs are associated with protein-coding genes and we hypothesize that they could provide cis-elements that regulate gene expression. This hypothesis is very likely in cases of TEs associated with Polycomb domains containing one or several genes because Polycomb PTMs H3K27me3 and H2AK121ub tend to spread across large domains of coregulated genes.

Repressed genes of facultative heterochromatin are associated with variations of chromatin landscapes where the gene body is covered by one chromatin state while a distinct state occupies the 3' and 5' regions. This symmetrical arrangement of chromatin states is strongly associated with transcriptional repression. By contrast asymmetric associations of chromatin states form the four distinct CL (CL7-10) that exist across actively transcribed protein-coding genes. (described in detail in Table 2). The origin of this asymmetrical organisation and the biological significance of the boundaries from one chromatin state to the next for each landscape is still unclear. We suggest that the limited number of CLs are produced by dedicated mechanisms of transcription which confers the directionality from 5' to 3' of the gene. The directionality of such CL on active genes could be set up by complexes like SWR1 during transcription. SWR1 in an ATP-dependent manner replaces nucleosomal H2A to H2A.Z [Luk et al., 2010].

4 Discussion

Besides CL associated with transcription, the strong association between H2A.Z and H2A.W with CLs of transcriptional repression is not explained. These associations may be the result from the yet unknown deposition pathways for H2A variants. Unlike chaperones specific for H3 deposition, chaperones for each type of H2A variant have not been identified [Nie et al., 2014]. But at least dynamics of H2A.Z are mediated by SWR1 [Aslam et al., 2019], and INO80 [Aslam et al., 2019] and we have identified one remodeler responsible for the deposition of H2A.W [Publication II - [Osakabe et al., 2021]]. We propose that additional remodelers participate in determining the architecture of the different CLs together with the chaperones that control the specific deposition of H3.1 or H3.3.

While there have been several studies identifying the epigenetic signature of the Transcriptional start sites (TSS), very few studies in plants have focused on identifying the epigenetic signature at the Transcription end site (TES). There have been studies linking H3.3 enrichment at 3' of gene bodies and we now also show H2Bs describe a specific state at TES across all CLs. Hence how H2Bs could interact with Pol-II during transcription termination needs further investigation.

4.5 The impact of chromatin landscapes on transcriptional activity

In our study, we describe four types of CLs that are associated with actively transcribed genes. Genes associated with CL7-CL10 genes are distinguished not only by the profiles of the PTMs and histone variants studied but also by specific features such as gene length and the presence or absence of introns, suggesting that the constraints of the gene architecture itself could limit the nature of the associated CL. Very short genes might have hardly enough space to accommodate the chromatin states required for the promoter and the terminator, such that the gene body remains limited to an extension of the promoter chromatin state, which is observed in CL9. It is possible that the separation between introns and exons requires specific marks and that these participate in the pausing of RNA

4.5 *The impact of chromatin landscapes on transcriptional activity*

Pol II. This is pure speculation and a more extensive study of the intron-exon boundary in *Arabidopsis* but also in other species would be required to test this hypothesis. Overall each CL is associated with the distinct range of transcript levels from high in CL7 to low in CL10. Overall the trends observed correlated with well established trends of enrichment of individual histone PTMs with transcription. However, within each group of genes associated with a particular CL, we no longer observed such correlation. The circa 10,000 genes with CL7 display a considerable variation between their transcript levels. Yet we observed no correlation with the level of transcripts and the profiles of most histone PTMs, histone variants and gene body mCG methylation while there was a clear correlation with chromatin accessibility and levels of enrichment of RNA Pol II. This surprising result indicates that the chromatin landscapes represent a source of information that constrains the levels of activity of RNA Pol II within a specific range and that this activity does not cause a feed-back on the composition of the chromatin landscape.

As an exception to the rule, the levels of enrichment of H2Bub changed with the transcript levels. Enrichment in H2Bub was the strongest indicator of the transcript levels irrespective of the CL. This could be explained by the fact that H2Bub deposited by HUB1 is important for recruiting factors involved in transcriptional elongation by its interaction with H3K4me3 [Buratowski, 2009, To and Kim, 2014]. The association of HUB1 with transcription elongation factors highlights the importance of this modification for productive transcriptional elongation. The ubiquitination of H2B is strongly inhibited by the C-terminal tail and the specific acidic patch of H2A.Z in mammals [Wojcik et al., 2018]. The strong opposition of H2A.Z and H2Bub in *Arabidopsis* supports the likely conservation of the inhibition of ubiquitination of H2B by the H2A.Z tail and the specific acidic patch is shown in mammals. Strikingly, genes in CL9 are expressed to comparable levels such as CL7 but lack H2Bub and are covered with H2A.Z suggesting shorter intron-less genes might not require H2Bub during their elongation. Again, pointing to the inherent structural constraints that could associate them with certain CL.

Overall my investigations of chromatin states led to identify a higher order form of chromatin assembly that is now termed chromatin landscapes. The landscape arranged in a

4 Discussion

symmetrical manner all associate with repressed transcriptional states while asymmetrical arrangements of chromatin states mark transcriptional activity. Each CL associates with a strong degree of specificity with a distinct mode of transcriptional control and more surprisingly with a distinct type of TE or gene relative to its family of origin or function of the encoded protein, respectively. Such specific associations support that CL are to a large degree instructive.

The origins of the formation of the CL remains unclear. One possibility is that cis-coded information to be identified signals and directs the activity of chromatin modifiers and remodelers. A second non-exclusive possibility is that elements sharing the same CL associate in space into separated compartments of the nucleus that host a specific concentrations of chromatin regulators. Confrontation of our data with HiC information will provide clues to test this latter hypothesis, which machine-learning strategies may enable to identify new cis-elements that direct the activity of chromatin regulators. Are the CLs described conserved across species and if so to which extent? Chromatin modifiers and remodelers are conserved across a wide range of eukaryotes and several histone PTMs share comparable association with euchromatin or heterochromatin. It still remains to test to which degree histone variants influence the organisation of CL in other plant species and animals. In addition, How these CL change during the different cell cycle stages as well as during distinct developmental stages needs to be investigated in the future studies. This will require building new genomic resources and the application of the *in silico* strategy used in this study.

5 Epilogue

In summary, my thesis work and the publications described here (Publication I and II) indicate that histone variants play a crucial role in defining major chromatin domains now defined as chromatin landscapes (CL) in Arabidopsis. In (Publication I) we described the novel chromatin landscape using histone variants and PTMs. Our study also suggests PTMs and histone variants together could be instructive for transcription or repression. This part of my work will be an essential resource for plant epigenetics and future studies. In the second part of my thesis (Publication II), we discovered the mechanism of action of a chromatin remodeler DDM1. We showed that DDM1 binds to Histone variant H2A.W and deposits H2A.W at specific transposon loci to maintain transposons in a silenced chromatin landscape. Our study was the first evidence linking DDM1 to its direct function in heterochromatin. Second, we show that chromatin remodelers like DDM1 can be crucial in establishing and maintaining such chromatin landscapes. Hence, DDM1-H2A.W together could contribute an additional layer of transposon repression. Finally, my work showed that histone variants do play a significant role in defining the chromatin state landscape. Furthermore, we could show an example of how chromatin remodelers like DDM1 could establish and maintain these chromatin landscapes.

Bibliography

- [Ahmad and Henikoff, 2002] Ahmad, K. and Henikoff, S. (2002). The histone variant h3.3 marks active chromatin by replication-independent nucleosome assembly. *Molecular cell*, 9(6):1191–1200.
- [Allfrey et al., 1964] Allfrey, V. G., Faulkner, R., and Mirsky, A. (1964). Acetylation and methylation of histones and their possible role in the regulation of rna synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, 51(5):786.
- [Aslam et al., 2019] Aslam, M., Fakher, B., Jakada, B. H., Cao, S., and Qin, Y. (2019). SWR1 Chromatin Remodeling Complex: A Key Transcriptional Regulator in Plants. *Cells*, 8(12).
- [Ausió and Abbott, 2002] Ausió, J. and 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.
- [Baldi et al., 2018] Baldi, S., Jain, D. S., Harpprecht, L., Zabel, A., Scheibe, M., Butter, F., Straub, T., and Becker, P. B. (2018). Genome-wide rules of nucleosome phasing in drosophila. *Molecular cell*, 72(4):661–672.
- [Bannister et al., 2005] Bannister, A. J., Schneider, R., Myers, F. A., Thorne, A. W., Crane-Robinson, C., and Kouzarides, T. (2005). Spatial distribution of di-and tri-methyl lysine 36 of histone h3 at active genes. *Journal of Biological Chemistry*, 280(18):17732–17736.

Bibliography

- [Barski et al., 2007] Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell*, 129(4):823–837.
- [Benoit et al., 2019] Benoit, M., Simon, L., Desset, S., Duc, C., Cotterell, S., Poulet, A., Le Goff, S., Tatout, C., and Probst, A. V. (2019). Replication-coupled histone h3. 1 deposition determines nucleosome composition and heterochromatin dynamics during arabidopsis seedling development. *New Phytologist*, 221(1):385–398.
- [Bergmüller et al., 2007] Bergmüller, E., Gehrig, P. M., and Gruissem, W. (2007). Characterization of post-translational modifications of histone h2b-variants isolated from arabidopsis thaliana. *Journal of proteome research*, 6(9):3655–3668.
- [Bernstein et al., 2004] Bernstein, B. E., Liu, C. L., Humphrey, E. L., Perlstein, E. O., and Schreiber, S. L. (2004). Global nucleosome occupancy in yeast. *Genome biology*, 5(9):1–11.
- [Borg et al., 2020] Borg, M., Jacob, Y., Susaki, D., LeBlanc, C., Buendía, D., Axelsson, E., Kawashima, T., Voigt, P., Boavida, L., Becker, J., Higashiyama, T., Martienssen, R., and Berger, F. (2020). Targeted reprogramming of H3K27me3 resets epigenetic memory in plant paternal chromatin. *Nature cell biology*, 22(6):621–629.
- [Bourguet et al., 2021] 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., and Mathieu, O. (2021). The histone variant H2A.W and linker histone H1 co-regulate heterochromatin accessibility and DNA methylation. *Nature Communications*, 12(1).
- [Bourguet et al., 2022] Bourguet, P., Yelagandula, R., To, T. K., Osakabe, A., Alishe, A., Lu, R. J.-H., Kakutani, T., Chen, P.-Y., and Berger, F. (2022). The histone variant h2a. w cooperates with chromatin modifications and linker histone h1 to maintain transcriptional silencing of transposons in arabidopsis. *bioRxiv*.

- [Bowman and Poirier, 2015] Bowman, G. D. and Poirier, M. G. (2015). Post-translational modifications of histones that influence nucleosome dynamics. *Chemical Reviews*, 115:2274 – 2295.
- [Buratowski, 2009] Buratowski, S. (2009). Progression through the rna polymerase ii ctd cycle. *Molecular cell*, 36(4):541–546.
- [Chereji and Morozov, 2014] Chereji, R. V. and Morozov, A. V. (2014). Ubiquitous nucleosome crowding in the yeast genome. *Proceedings of the National Academy of Sciences*, 111(14):5236–5241.
- [Creyghton et al., 2010] Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., Hanna, J., Lodato, M. A., Frampton, G. M., Sharp, P. A., et al. (2010). Histone h3k27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences*, 107(50):21931–21936.
- [Dai et al., 2017] Dai, X., Bai, Y., Zhao, L., Dou, X., Liu, Y., Wang, L., Li, Y., Li, W., Hui, Y., Huang, X., et al. (2017). H2a. z represses gene expression by modulating promoter nucleosome structure and enhancer histone modifications in arabidopsis. *Molecular plant*, 10(10):1274–1292.
- [Dawe and Henikoff, 2006] Dawe, R. K. and Henikoff, S. (2006). Centromeres put epigenetics in the driver’s seat. *Trends in biochemical sciences*, 31(12):662–669.
- [De Wulf and Earnshaw, 2009] De Wulf, P. and Earnshaw, W. C. (2009). *The Kinetochores*. Springer.
- [Di Croce and Helin, 2013] Di Croce, L. and Helin, K. (2013). Transcriptional regulation by polycomb group proteins. *Nature structural & molecular biology*, 20(10):1147–1155.
- [Dou et al., 2006] Dou, Y., Milne, T. A., Ruthenburg, A. J., Lee, S., Lee, J. W., Verdine, G. L., Allis, C. D., and Roeder, R. G. (2006). Regulation of mll1 h3k4 methyltransferase activity by its core components. *Nature structural & molecular biology*, 13(8):713–719.

Bibliography

- [Downs et al., 2007] Downs, J. A., Nussenzweig, M. C., and Nussenzweig, A. (2007). Chromatin dynamics and the preservation of genetic information. *Nature*, 447(7147):951–958.
- [Du et al., 2012] Du, J., Zhong, X., Bernatavichute, Y. V., Stroud, H., Feng, S., Caro, E., Vashisht, A. A., Terragni, J., Chin, H. G., Tu, A., et al. (2012). Dual binding of chromomethylase domains to h3k9me2-containing nucleosomes directs dna methylation in plants. *Cell*, 151(1):167–180.
- [D’Arcy et al., 2013] D’Arcy, S., Martin, K. W., Panchenko, T., Chen, X., Bergeron, S., Stargell, L. A., Black, B. E., and Luger, K. (2013). Chaperone nap1 shields histone surfaces used in a nucleosome and can put h2a-h2b in an unconventional tetrameric form. *Molecular cell*, 51(5):662–677.
- [Ehara et al., 2019] Ehara, H., Kujirai, T., Fujino, Y., Shirouzu, M., Kurumizaka, H., and Sekine, S.-i. (2019). Structural insight into nucleosome transcription by rna polymerase ii with elongation factors. *Science*, 363(6428):744–747.
- [Elsässer et al., 2012] Elsässer, S. J., Huang, H., Lewis, P. W., Chin, J. W., Allis, C. D., and Patel, D. J. (2012). Daxx envelops a histone h3. 3-h4 dimer for h3. 3-specific recognition. *Nature*, 491(7425):560–565.
- [Elsässer et al., 2015] Elsässer, S. J., Noh, K.-M., Diaz, N., Allis, C. D., and Banaszynski, L. A. (2015). Histone h3. 3 is required for endogenous retroviral element silencing in embryonic stem cells. *Nature*, 522(7555):240–244.
- [ENCODE Project Consortium, 2011] ENCODE Project Consortium (2011). A user’s guide to the encyclopedia of dna elements (encode). *PLoS biology*, 9(4):e1001046.
- [English et al., 2006] English, C. M., Adkins, M. W., Carson, J. J., Churchill, M. E., and Tyler, J. K. (2006). Structural basis for the histone chaperone activity of asf1. *Cell*, 127(3):495–508.
- [Ernst and Kellis, 2010] Ernst, J. and Kellis, M. (2010). Discovery and characterization of

- chromatin states for systematic annotation of the human genome. *Nature biotechnology*, 28(8):817–825.
- [Ernst and Kellis, 2012] Ernst, J. and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. *Nature Methods*.
- [Ernst and Kellis, 2017] Ernst, J. and Kellis, M. (2017). Chromatin-state discovery and genome annotation with ChromHMM. *Nature protocols*, 12(12):2478–2492.
- [Faast et al., 2001] Faast, R., Thonglairoam, V., Schulz, T. C., Beall, J., Wells, J. R., Taylor, H., Matthaei, K., Rathjen, P. D., Tremethick, D. J., and Lyons, I. (2001). Histone variant h2a. z is required for early mammalian development. *Current Biology*, 11(15):1183–1187.
- [Feingold and Pachter, 2004] Feingold, E. and Pachter, L. (2004). The encode (encyclopedia of dna elements) project. *Science*, 306(5696):636–640.
- [Flemming, 1882] Flemming, W. (1882). *Zellsubstanz, kern und zelltheilung*. Vogel.
- [Forties et al., 2011] Forties, R. A., North, J. A., Javaid, S., Tabbaa, O. P., Fishel, R., Poirier, M. G., and Bundschuh, R. (2011). A quantitative model of nucleosome dynamics. *Nucleic acids research*, 39(19):8306–8313.
- [Franklin and Zweidler, 1977] Franklin, S. G. and Zweidler, A. (1977). Non-allelic variants of histones 2a, 2b and 3 in mammals. *Nature*, 266(5599):273–275.
- [Gerstein et al., 2010] Gerstein, M. B., Lu, Z. J., Van Nostrand, E. L., Cheng, C., Arshinoff, B. I., Liu, T., Yip, K. Y., Robilotto, R., Rechtsteiner, A., Ikegami, K., et al. (2010). Integrative analysis of the caenorhabditis elegans genome by the modencode project. *Science*, 330(6012):1775–1787.
- [Goldberg et al., 2010] 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., et al. (2010). Distinct factors control histone variant h3. 3 localization at specific genomic regions. *Cell*, 140(5):678–691.

Bibliography

- [Hake and Allis, 2006] Hake, S. B. and Allis, C. D. (2006). Histone H3 variants and their potential role in indexing mammalian genomes: the “H3 barcode hypothesis”. *Proc. Natl Acad. Sci. USA*, 103(17):6428–6435.
- [Han et al., 2011] Han, W., Li, X., and Fu, X. (2011). The macro domain protein family: structure, functions, and their potential therapeutic implications. *Mutation Research/Reviews in Mutation Research*, 727(3):86–103.
- [Hartzog et al., 2002] Hartzog, G. A., Speer, J. L., and Lindstrom, D. L. (2002). Transcript elongation on a nucleoprotein template. *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression*, 1577(2):276–286.
- [Heitz, 1929] Heitz, E. (1929). *Heterochromatin, chromocentren, chromomeren*. Komm. Fischer.
- [Henikoff et al., 2001] Henikoff, S., Ahmad, K., and Malik, H. S. (2001). The centromere paradox: stable inheritance with rapidly evolving dna. *Science*, 293(5532):1098–1102.
- [Herz et al., 2012] Herz, H.-M., Mohan, M., Garruss, A. S., Liang, K., Takahashi, Y.-h., Mickey, K., Voets, O., Verrijzer, C. P., and Shilatifard, A. (2012). Enhancer-associated h3k4 monomethylation by trithorax-related, the drosophila homolog of mammalian mll3/ml4. *Genes & development*, 26(23):2604–2620.
- [Hume Stroud et al., 2014] Hume Stroud, T. D., Du, J., Zhong, X., Feng, S., Johnson, L., Patel, D. J., and Jacobsen, S. E. (2014). The roles of non-cg methylation in arabidopsis. *Nature structural & molecular biology*, 21(1):64.
- [Jackson and Gorovsky, 2000] Jackson, J. and Gorovsky, M. (2000). Histone h2a. z has a conserved function that is distinct from that of the major h2a sequence variants. *Nucleic acids research*, 28(19):3811–3816.
- [Jacob et al., 2009] Jacob, Y., Feng, S., LeBlanc, C. A., Bernatavichute, Y. V., Stroud, H., Cokus, S., Johnson, L. M., Pellegrini, M., Jacobsen, S. E., and 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.
- [Jarillo and Piñeiro, 2015] Jarillo, J. A. and Piñeiro, M. (2015). H2a. z mediates different aspects of chromatin function and modulates flowering responses in arabidopsis. *The plant journal*, 83(1):96–109.
- [Jiang et al., 2020] Jiang, D., Borg, M., Lorković, Z. J., Montgomery, S. A., Osakabe, A., Yelagandula, R., Axelsson, E., and Berger, F. (2020). The evolution and functional divergence of the histone h2b family in plants. *PLoS genetics*, 16(7):e1008964.
- [Johnson et al., 2007] Johnson, D. S., Mortazavi, A., Myers, R. M., and Wold, B. (2007). Genome-wide mapping of in vivo protein-dna interactions. *Science*, 316(5830):1497–1502.
- [Kakutani, 1997] Kakutani, T. (1997). Genetic characterization of late-flowering traits induced by dna hypomethylation mutation in arabidopsis thaliana. *The plant journal*, 12(6):1447–1451.
- [Kato et al., 2003] Kato, M., Miura, A., Bender, J., Jacobsen, S. E., and Kakutani, T. (2003). Role of cg and non-cg methylation in immobilization of transposons in arabidopsis. *Current Biology*, 13(5):421–426.
- [Kawashima et al., 2015] Kawashima, T., Lorković, Z. J., Nishihama, R., Ishizaki, K., Axelsson, E., Yelagandula, R., Kohchi, T., and Berger, F. (2015). Diversification of histone h2a variants during plant evolution. *Trends in plant science*, 20(7):419–425.
- [Kim et al., 2011] Kim, S., Kim, H., Fong, N., Erickson, B., and Bentley, D. L. (2011). Pre-mrna splicing is a determinant of histone h3k36 methylation. *Proceedings of the National Academy of Sciences*, 108(33):13564–13569.
- [Kizer et al., 2005] Kizer, K. O., Phatnani, H. P., Shibata, Y., Hall, H., Greenleaf, A. L., and Strahl, B. D. (2005). A novel domain in set2 mediates rna polymerase ii interaction

Bibliography

- and couples histone h3 k36 methylation with transcript elongation. *Molecular and cellular biology*, 25(8):3305–3316.
- [Kornberg and Klug, 1981] Kornberg, R. D. and Klug, A. (1981). The nucleosome. *Scientific American*, 244(2):52–65.
- [Kornberg and Lorch, 2020] Kornberg, R. D. and Lorch, Y. (2020). Primary role of the nucleosome. *Molecular Cell*, 79(3):371–375.
- [Lai and Pugh, 2017] Lai, W. K. and Pugh, B. F. (2017). Understanding nucleosome dynamics and their links to gene expression and dna replication. *Nature reviews Molecular cell biology*, 18(9):548–562.
- [Lauberth et al., 2013] Lauberth, S. M., Nakayama, T., Wu, X., Ferris, A. L., Tang, Z., Hughes, S. H., and Roeder, R. G. (2013). H3k4me3 interactions with taf3 regulate preinitiation complex assembly and selective gene activation. *Cell*, 152(5):1021–1036.
- [Lewis et al., 2010] Lewis, P. W., Elsaesser, S. J., Noh, K.-M., Stadler, S. C., and Allis, C. D. (2010). Daxx is an h3. 3-specific histone chaperone and cooperates with atrx in replication-independent chromatin assembly at telomeres. *Proceedings of the National Academy of Sciences*, 107(32):14075–14080.
- [Li et al., 2010] Li, A., Yu, Y., Lee, S.-C., Ishibashi, T., Lees-Miller, S. P., and Ausió, J. (2010). Phosphorylation of histone h2a. x by dna-dependent protein kinase is not affected by core histone acetylation, but it alters nucleosome stability and histone h1 binding. *Journal of biological chemistry*, 285(23):17778–17788.
- [Li et al., 2007] Li, B., Carey, M., and Workman, J. L. (2007). The role of chromatin during transcription. *Cell*, 128(4):707–719.
- [Li et al., 2009] Li, B., Jackson, J., Simon, M. D., Fleharty, B., Gogol, M., Seidel, C., Workman, J. L., and Shilatifard, A. (2009). Histone h3 lysine 36 dimethylation (h3k36me2) is sufficient to recruit the rpd3s histone deacetylase complex and to repress spurious transcription. *Journal of Biological Chemistry*, 284(12):7970–7976.

- [Lisch et al., 2009] Lisch, D. et al. (2009). Epigenetic regulation of transposable elements in plants. *Annual review of plant biology*, 60(1):43–66.
- [Long et al., 2019] Long, M., Sun, X., Shi, W., Yanru, A., Leung, S. T., Ding, D., Cheema, M. S., MacPherson, N., Nelson, C. J., Ausio, J., et al. (2019). A novel histone h4 variant h4g regulates rdna transcription in breast cancer. *Nucleic Acids Research*, 47(16):8399–8409.
- [Lorković and Berger, 2017] Lorković, Z. J. and Berger, F. (2017). Heterochromatin and DNA damage repair: Use different histone variants and relax. *Nucleus*, 8(6):583–588.
- [Luger et al., 1997] Luger, K., Mä Der, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. Technical report.
- [Luk et al., 2010] Luk, E., Ranjan, A., FitzGerald, P. C., Mizuguchi, G., Huang, Y., Wei, D., and Wu, C. (2010). Stepwise histone replacement by swr1 requires dual activation with histone h2a. z and canonical nucleosome. *Cell*, 143(5):725–736.
- [Lyons and Zilberman, 2017] Lyons, D. B. and Zilberman, D. (2017). DDM1 and lsh remodelers allow methylation of DNA wrapped in nucleosomes. *eLife*.
- [Lyons et al., 2016] Lyons, S. M., Cunningham, C. H., Welch, J. D., Groh, B., Guo, A. Y., Wei, B., Whitfield, M. L., Xiong, Y., and Marzluff, W. F. (2016). A subset of replication-dependent histone mrnas are expressed as polyadenylated rnas in terminally differentiated tissues. *Nucleic acids research*, 44(19):9190–9205.
- [Malik and Henikoff, 2003] Malik, H. S. and Henikoff, S. (2003). Phylogenomics of the nucleosome. *Nature structural & molecular biology*, 10(11):882–891.
- [Mannironi et al., 1989] Mannironi, C., Bonner, W. M., and 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.

Bibliography

- [Marzluff, 2005] Marzluff, W. F. (2005). Metazoan replication-dependent histone mrnas: a distinct set of rna polymerase ii transcripts. *Current opinion in cell biology*, 17(3):274–280.
- [Marzluff et al., 2002] Marzluff, W. F., Gongidi, P., Woods, K. R., Jin, J., and Maltais, L. J. (2002). The human and mouse replication-dependent histone genes. *Genomics*, 80(5):487–498.
- [Mito et al., 2005] Mito, Y., Henikoff, J. G., and Henikoff, S. (2005). Genome-scale profiling of histone h3. 3 replacement patterns. *Nature genetics*, 37(10):1090–1097.
- [Mito et al., 2007] Mito, Y., Henikoff, J. G., and Henikoff, S. (2007). Histone replacement marks the boundaries of cis-regulatory domains. *Science*, 315(5817):1408–1411.
- [Muthurajan et al., 2011] Muthurajan, U. M., McBryant, S. J., Lu, X., Hansen, J. C., and Luger, K. (2011). The linker region of macroh2a promotes self-association of nucleosomal arrays. *Journal of Biological Chemistry*, 286(27):23852–23864.
- [Natsume et al., 2007] Natsume, R., Eitoku, M., Akai, Y., Sano, N., Horikoshi, M., and Senda, T. (2007). Structure and function of the histone chaperone cia/asf1 complexed with histones h3 and h4. *Nature*, 446(7133):338–341.
- [Nie et al., 2014] Nie, X., Wang, H., Li, J., Holec, S., and 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.
- [Osakabe et al., 2021] Osakabe, A., Jamge, B., Axelsson, E., Montgomery, S. A., Akimcheva, S., Kuehn, A. L., Pisupati, R., Lorković, Z. J., Yelagandula, R., Kakutani, T., and Berger, F. (2021). The chromatin remodeler DDM1 prevents transposon mobility through deposition of histone variant H2A.W. *Nature cell biology*, 23(4):391–400.
- [Osakabe et al., 2018] Osakabe, A., Lorković, Z. J., Kobayashi, W., Tachiwana, H., Yelagandula, R., Kurumizaka, H., and Berger, F. (2018). Histone H2A variants confer

- specific properties to nucleosomes and impact on chromatin accessibility. *Nucleic acids research*, 46(15):7675–7685.
- [Park, 2009] Park, P. J. (2009). Chip-seq: advantages and challenges of a maturing technology. *Nature reviews genetics*, 10(10):669–680.
- [Pavri et al., 2006] Pavri, R., Zhu, B., Li, G., Trojer, P., Mandal, S., Shilatifard, A., and Reinberg, D. (2006). Histone h2b monoubiquitination functions cooperatively with fact to regulate elongation by rna polymerase ii. *Cell*, 125(4):703–717.
- [Pengelly et al., 2015] Pengelly, A. R., Kalb, R., Finkl, K., and Müller, J. (2015). Transcriptional repression by prc1 in the absence of h2a monoubiquitylation. *Genes & development*, 29(14):1487–1492.
- [Petruk et al., 2001] Petruk, S., Sedkov, Y., Smith, S., Tillib, S., Kraevski, V., Nakamura, T., Canaani, E., Croce, C. M., and Mazo, A. (2001). Trithorax and dcbp acting in a complex to maintain expression of a homeotic gene. *Science*, 294(5545):1331–1334.
- [Pikaard and Scheid, 2014] Pikaard, C. S. and Scheid, O. M. (2014). Epigenetic regulation in plants. *Cold Spring Harbor perspectives in biology*, 6(12):a019315.
- [Pinto and Flaus, 2010] Pinto, D. M. S. and Flaus, A. (2010). Structure and function of histone h2ax. *Genome stability and human diseases*, pages 55–78.
- [Plys et al., 2019] Plys, A. J., Davis, C. P., Kim, J., Rizki, G., Keenen, M. M., Marr, S. K., and Kingston, R. E. (2019). Phase separation of polycomb-repressive complex 1 is governed by a charged disordered region of cbx2. *Genes & development*, 33(13-14):799–813.
- [Pokholok et al., 2005] Pokholok, D. K., Harbison, C. T., Levine, S., Cole, M., Hannett, N. M., Lee, T. I., Bell, G. W., Walker, K., Rolfe, P. A., Herbolzheimer, E., et al. (2005). Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell*, 122(4):517–527.

Bibliography

- [Robertson et al., 2007] Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., Zeng, T., Euskirchen, G., Bernier, B., Varhol, R., Delaney, A., et al. (2007). Genome-wide profiles of stat1 dna association using chromatin immunoprecipitation and massively parallel sequencing. *Nature methods*, 4(8):651–657.
- [Rothbart and Strahl, 2014] Rothbart, S. B. and Strahl, B. D. (2014). Interpreting the language of histone and dna modifications. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1839(8):627–643.
- [Roudier et al., 2011] 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., Dérozier, S., Brunaud, V., Aubourg, S., Schnittger, A., Bowler, C., Martin-Magniette, M. L., Robin, S., Caboche, M., and Colot, V. (2011). Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. *The EMBO journal*, 30(10):1928–1938.
- [Schmitges et al., 2011] Schmitges, F. W., Prusty, A. B., Faty, M., Stützer, A., Lingaraju, G. M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S., et al. (2011). Histone methylation by prc2 is inhibited by active chromatin marks. *Molecular cell*, 42(3):330–341.
- [Schuettengruber and Cavalli, 2009] Schuettengruber, B. and Cavalli, G. (2009). Recruitment of polycomb group complexes and their role in the dynamic regulation of cell fate choice. *Development*, 136(21):3531–3542.
- [Sequeira-Mendes et al., 2014] Sequeira-Mendes, J., Aragüez, I., Peiró, R., Mendez-Giraldez, R., Zhang, X., Jacobsen, S. E., Bastolla, U., and 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.
- [Shilatifard, 2012] Shilatifard, A. (2012). The compass family of histone h3k4 methylases: mechanisms of regulation in development and disease pathogenesis. *Annual review of biochemistry*, 81:65.

- [Simon and Kingston, 2013] Simon, J. A. and Kingston, R. E. (2013). Occupying chromatin: Polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put. *Molecular cell*, 49(5):808–824.
- [Stack et al., 1977] Stack, S. M., Brown, D. B., and Dewey, W. (1977). Visualization of interphase chromosomes. *Journal of cell science*, 26(1):281–299.
- [Strahl and Allis, 2000] Strahl, B. D. and Allis, C. D. (2000). The language of covalent histone modifications. *Nature*, 403(6765):41–45.
- [Struhl and Segal, 2013] Struhl, K. and Segal, E. (2013). Determinants of nucleosome positioning. *Nature structural & molecular biology*, 20(3):267–273.
- [Sura et al., 2017] Sura, W., Kabza, M., Karlowski, W. M., Bieluszewski, T., Kus-Slowinska, M., Pawełszek, Ł., Sadowski, J., and Ziolkowski, P. A. (2017). Dual role of the histone variant h2a. z in transcriptional regulation of stress-response genes. *The Plant Cell*, 29(4):791–807.
- [Talbert and Henikoff, 2010] Talbert, P. B. and Henikoff, S. (2010). Histone variants — ancient wrap artists of the epigenome. *Nature Reviews Molecular Cell Biology* 2010 11:4, 11(4):264–275.
- [Talbert and Henikoff, 2014] Talbert, P. B. and Henikoff, S. (2014). Environmental responses mediated by histone variants. *Trends in cell biology*, 24(11):642–650.
- [Talbert and Henikoff, 2017] Talbert, P. B. and Henikoff, S. (2017). Histone variants on the move: substrates for chromatin dynamics. *Nature reviews Molecular cell biology*, 18(2):115–126.
- [Talbert and Henikoff, 2021] Talbert, P. B. and Henikoff, S. (2021). Remodeller–variant tag team slams transposons. *Nature Cell Biology*, 23(4):297–298.
- [Talbert et al., 2002] Talbert, P. B., Masuelli, R., Tyagi, A. P., Comai, L., and Henikoff, S. (2002). Centromeric localization and adaptive evolution of an arabidopsis histone h3 variant. *The Plant Cell*, 14(5):1053–1066.

Bibliography

- [Tang et al., 2006] Tang, Y., Poustovoitov, M. V., Zhao, K., Garfinkel, M., Canutescu, A., Dunbrack, R., Adams, P. D., and Marmorstein, R. (2006). Structure of a human asf1a–hira complex and insights into specificity of histone chaperone complex assembly. *Nature structural & molecular biology*, 13(10):921–929.
- [Tatavosian et al., 2019] Tatavosian, R., Kent, S., Brown, K., Yao, T., Duc, H. N., Huynh, T. N., Zhen, C. Y., Ma, B., Wang, H., and Ren, X. (2019). Nuclear condensates of the polycomb protein chromobox 2 (cbx2) assemble through phase separation. *Journal of Biological Chemistry*, 294(5):1451–1463.
- [To and Kim, 2014] To, T. K. and Kim, J.-M. (2014). Epigenetic regulation of gene responsiveness in arabidopsis. *Frontiers in Plant Science*, 4:548.
- [Turner, 2000] Turner, B. M. (2000). Histone acetylation and an epigenetic code. *Bioessays*, 22(9):836–845.
- [Urban and Zweidler, 1983] Urban, M. K. and Zweidler, A. (1983). Changes in nucleosomal core histone variants during chicken development and maturation. *Developmental biology*, 95(2):421–428.
- [Venkatesh and Workman, 2015] Venkatesh, S. and Workman, J. L. (2015). Histone exchange, chromatin structure and the regulation of transcription. *Nature reviews Molecular cell biology*, 16(3):178–189.
- [Vongs et al., 1993] Vongs, A., Kakutani, T., Martienssen, R. A., and Richards, E. J. (1993). Arabidopsis thaliana dna methylation mutants. *Science*, 260(5116):1926–1928.
- [Voon et al., 2015] Voon, H. P., Hughes, J. R., Rode, C., Inti, A., Jenuwein, T., Feil, R., Higgs, D. R., and Gibbons, R. J. (2015). Atrx plays a key role in maintaining silencing at interstitial heterochromatic loci and imprinted genes. *Cell reports*, 11(3):405–418.
- [Wagner and Carpenter, 2012] Wagner, E. J. and Carpenter, P. B. (2012). Understanding the language of lys36 methylation at histone h3. *Nature reviews Molecular cell biology*, 13(2):115–126.

- [Weber et al., 2016] Weber, B., Zicola, J., Oka, R., and Stam, M. (2016). Plant enhancers: a call for discovery. *Trends in plant science*, 21(11):974–987.
- [Wendte and Schmitz, 2018] Wendte, J. M. and Schmitz, R. J. (2018). Specifications of Targeting Heterochromatin Modifications in Plants. *Molecular plant*, 11(3):381–387.
- [Wirbelauer et al., 2005] Wirbelauer, C., Bell, O., and Schübeler, D. (2005). Variant histone h3. 3 is deposited at sites of nucleosomal displacement throughout transcribed genes while active histone modifications show a promoter-proximal bias. *Genes & development*, 19(15):1761–1766.
- [Wojcik et al., 2018] Wojcik, F., Dann, G. P., Beh, L. Y., Debelouchina, G. T., Hofmann, R., and Muir, T. W. (2018). Functional crosstalk between histone H2B ubiquitylation and H2A modifications and variants. *Nature Communications*, 9(1).
- [Wolffe and Hayes, 1999] Wolffe, A. P. and Hayes, J. J. (1999). Chromatin disruption and modification. *Nucleic acids research*, 27(3):711–720.
- [Wollmann et al., 2012] Wollmann, H., Holec, S., Alden, K., Clarke, N. D., Jacques, P. É., and Berger, F. (2012). Dynamic deposition of histone variant H3.3 accompanies developmental remodeling of the Arabidopsis transcriptome. *PLoS genetics*, 8(5).
- [Yelagandula et al., 2014] 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., and Berger, F. (2014). The histone variant H2A.W defines heterochromatin and promotes chromatin condensation in arabidopsis. *Cell*, 158(1):98–109.
- [Zhang et al., 2009] Zhang, X., Bernatavichute, Y. V., Cokus, S., Pellegrini, M., and Jacobsen, S. E. (2009). Genome-wide analysis of mono-, di-and trimethylation of histone h3 lysine 4 in arabidopsis thaliana. *Genome biology*, 10(6):1–14.
- [Zhou et al., 2011] Zhou, V. W., Goren, A., and Bernstein, B. E. (2011). Charting histone modifications and the functional organization of mammalian genomes. *Nature Reviews Genetics*, 12(1):7–18.

Bibliography

- [Zhou et al., 2015a] Zhou, W., Zhu, Y., Dong, A., and Shen, W. H. (2015a). Histone H2A/H2B chaperones: From molecules to chromatin-based functions in plant growth and development. *Plant Journal*, 83(1):78–95.
- [Zhou et al., 2015b] Zhou, W., Zhu, Y., Dong, A., and Shen, W.-H. (2015b). Histone h2a/h2b chaperones: from molecules to chromatin-based functions in plant growth and development. *The Plant Journal*, 83(1):78–95.
- [Zhou et al., 2017] Zhou, Y., Romero-Campero, F. J., Gómez-Zambrano, Á., Turck, F., and Calonje, M. (2017). H2a monoubiquitination in arabidopsis thaliana is generally independent of lhp1 and prc2 activity. *Genome biology*, 18(1):1–13.
- [Zilberman et al., 2008] Zilberman, D., Coleman-Derr, D., Ballinger, T., and Henikoff, S. (2008). Histone h2a. z and dna methylation are mutually antagonistic chromatin marks. *Nature*, 456(7218):125–129.