Chromatin reprogramming in *Arabidopsis thaliana* pollen

Toby Buttress

John Innes Centre

February 2021

A thesis submitted to the University of East Anglia for the degree of Doctor of Philosophy (PhD)

Contents

Abstract

Chromatin is extensively reprogrammed through male sexual lineage development in flowering plants. Such reprogramming results in highly dimorphic chromatin architectures in the nuclei of mature pollen. Sperm is highly compacted whilst the accompanying vegetative nucleus is globally decondensed. The pollen chromatin dimorphism phenomenon has long been characterised, but the underlying mechanisms that drive such differences are largely unknown. Here, the nuclear proteomes of sperm and vegetative nuclei were generated and mined for novel determinants of pollen chromatin structure.

In Chapter 2, H2B.8 is introduced as a novel sperm-expressed histone variant which is responsible for sperm chromatin compaction and is important for male fertility. Using super-resolution microscopy, H2B.8 is revealed to achieve compaction by aggregating euchromatin into small foci. Such foci form by phase separation and are dependent on an intrinsically disordered region in the histone tail. H2B.8 is deposited to inactive euchromatic regions and achieves compaction without affecting transcription. Evolutionarily, H2B.8 is specific to flowering plants and provides a mechanism by which sperm compacts in eukaryotes in the absence of protamines. Generally, these results demonstrate a novel mechanism of nuclear condensation via aggregation of euchromatic regions that are transcriptionally inactive.

Chapter 3 concerns the vegetative nucleus which undergoes active DNA demethylation by DEMETER (DME), preferentially targeting heterochromatic transposons. How DME gains access to such loci remains unknown. This work characterises Histone Acetyltransferase of the MYST family (HAM) proteins, which are present specifically in the vegetative cell nuclei and are required for normal DME-mediated cytosine demethylation, particularly in heterochromatin. Greater chromatin accessibility is provided at DME targets by HAM proteins, likely by putative H4 lysine 5 acetylation activity. HAM proteins facilitate activation of a subset of pollen expressed transposons but do not affect gene transcription. Collectively, this work demonstrates a developmental switch whereby an active chromatin modification mark, which usually targets genes in soma, targets transposons in the vegetative nucleus and relaxes heterochromatin for further epigenetic reprogramming at the DNA methylation level.

Work in collaboration with others

The project presented in Chapter 2 was undertaken in partnership with Dr Shengbo He (Feng Lab, John Innes Centre). Shengbo carried out parts of the confocal microscopy, including *in vivo* FRAP experiments, and assisted with molecular cloning. Shengbo also helped to optimise library preparation protocols, particularly native ChIP-seq in pollen. Mass spectrometry of purified nuclear proteins and histones was performed by Dr Gerhard Saalbach (Proteomics Facility, John Innes Centre). *In vitro* phase separation experiments were undertaken by Dr Liang Wang and Dr Lei Sun (Pilong Li Lab, Tsinghua University, China). The work presented in Chapter 3 was undertaken with assistance from students I supervised during my PhD, Aileen Magilin (Year in Industry student, Feng Lab, John Innes Centre) and Jiawen Chen (Rotation PhD student, John Innes Centre), and is being continued by Dr Yalin Liu (Feng Lab, John Innes Centre).

Acknowledgements

It has been an immense privilege to undertake my PhD at the John Innes Centre. I am incredibly grateful to all the support systems in place that allow we as students to flourish as scientists. Special mentions to Roy Dunford, Tim Hicks and the team in Laboratory Support for making lab life so smooth, Tim Wells and Horticultural Services for taking excellent care of my plants, Eva Wegel and Bioimaging for microscopy training and advice, Gerhard Saalbach for assistance with proteomics and all in computing for bioinformatic support. JIC is a wonderful environment to grow as a scientist. From the many wonderful seminars to fleeting conversations in coffee areas, the collaborative culture has helped to develop my ideas and broaden horizons.

I thank all my educators through school and university for their support and guidance. Particularly Colwyn Thomas at UEA for inspiring an interest in plant biology and encouraging me to pursue a PhD.

I would like to acknowledge the BBSRC for funding through the NRP DTP and the Graduate School Office for facilitating the programme. To my fellow PhD students, thanks for always being obliging in coffee breaks or visits to the pub, I'm excited to see where your careers go. The DTP offered the opportunity to gain experience in industry with RAGT Seeds, my thanks go to Chris Burt and the cereal genetics team for mentorship – I hope my wheat molecular markers are still in use!

During my PhD I have had the great honour of training students; I learnt a great deal from the experience and look forward to following your future scientific journeys, all the best to Jiawen Chen, Viktoria Hristova and Aileen Magilin. To my fellow Feng Lab PhD students Billy Aldridge, James Walker, Sam Deans and Judit Talas; thank you for the scientific discussions, experimental advice and most importantly, camaraderie.

I am incredibly appreciative of our groups post-docs who have always been on hand to offer their experience experimentally, analytically and in discussion. My thanks to Jingyi Zhang, Hongbo Gao, Martin Vickers, Jincheng Long, Shujuan Xu, Wenjing She, Yalin Liu, Shaoli Zhou and Chris Morgan. I reserve a special thanks to Shengbo He, my day-to-day supervisor from the start and an excellent scientist and friend. I have been very fortunate to receive advice and support from Daniel Zilberman and Sarah O'Connor as part of my supervisory committee.

To my supervisor Xiaoqi Feng, thank you giving me the opportunity to join the group and pursue such an exciting PhD project. You have given me the freedom and encouragement to explore challenging ideas, creating a wonderful group and collaborative environment to facilitate a thoroughly enjoyable experience. With your guidance, I have developed into an accomplished scientist with a keen interest across fundamental biology.

On a personal level, I thank my friends from school and university for the regular escapes from the academic bubble; you have all been amazing over the years! I am incredibly fortunate to have such a supportive family around me. To grandad, I am sorry that I have no need for your old carpentry tools, I am much more suited to a pipette. And nanna, I'd define myself as a molecular biologist not an agroecologist as you have been telling everyone! To mum and dad, you have always pushed me to be the best I can be and have given me every opportunity in life – I owe everything to you both. Finally, to Emily, thank you for being such a supportive partner over the years, from dragging me on holidays against my wishes to making sacrifices in your own life so I can pursue my PhD. It has been wonderful to grow in our respective careers and I cannot wait for the next chapter together.

Chapter 1 – Introduction

1.1 Introduction

Chromatin organisation is critical for determination of cellular function, mediating regulation of gene expression and stable silencing of transposable elements (TEs). A variety of epigenetic mechanisms contribute to chromatin organisation from higher-order nuclear compartmentalisation to modifications of individual nucleic acids. Such epigenetic factors are largely stable through somatic development, with minor alterations tuning gene expression within different cellular identities.

A drastic shift is observed in the switch from somatic to reproductive development, however. Slight alterations of chromatin modifications between somatic cell types are replaced with sweeping reprogramming events through the course of germline development (Kawashima and Berger, 2014). Such events offer exciting opportunities to study chromatin organisation mechanisms, with an added importance owing to the critical need to correctly pass on genetic regulatory information to the next generation.

The flowering plant *Arabidopsis thaliana* has emerged as a key model for the study of epigenetic regulation, owing to the ease of obtaining specific cell types and resilience to genetic mutations. In particular, the male sexual lineage is an intriguing system due to the vastly different chromatin states of sperm and vegetative nuclei within mature pollen (Kawashima and Berger, 2014; Borg and Berger, 2015). The male gamete, sperm, has highly compacted chromatin; a feature shared across eukaryotic lineages (Sassone-Corsi, 2002; Kimmins and Sassone-Corsi, 2005; Hao, Ni and Yang, 2019). The companion, the vegetative cell, has an extremely decondensed chromatin structure (Schoft *et al.*, 2009). Vast differences in chromatin organisation occur between these cell types despite being separated by a single mitotic division. The male sexual lineage of *Arabidopsis* is, therefore, an ideal system in which to examine fundamental principles of chromatin compaction in plants, with lessons perhaps applicable more widely across eukaryotes.

This chapter will introduce core elements of chromatin organisation in flowering plants before considering features specific to male sexual lineage development. Parallels and differences will be drawn to germline development in other plant lineages and eukaryotes more broadly. Female sex cell chromatin organisation is discussed along with post-fertilisation dynamics.

1.2 Features of somatic flowering plant chromatin

The basic subunit of chromatin is the nucleosome, an octameric complex of the four core histones, H2A, H2B, H3 and H4, wrapped with ~146 bp DNA (Talbert and Henikoff, 2010). Nucleosomes are assembled by the combination of a histone H3-H4 tetramer and two H2A-H2B dimers. The linker region of DNA between nucleosomes can be bound by histone H1 (Fyodorov *et al.*, 2018). Altogether, these components constitute the classic 'beads-on-a-string' description (Baldi, Korber and Becker, 2020).

Chromatin represents the canvas upon which all nuclear interactions take place, from transcription to replication (Talbert and Henikoff, 2016). Nucleosomes are a significant barrier to such processes, and so eukaryotes have evolved a vast array of proteins to modify basic chromatin conformation (Talbert and Henikoff, 2010). Collectively, the interactions of these processes define cellular function.

This section introduces determinants of somatic chromatin structure in the flowering plant *Arabidopsis thaliana* whilst drawing on literature from across eukaryotes. Consideration is given to cytologically visible macroscale domains through to highly localised modifications to the DNA itself.

1.2.1 Higher-order chromatin architecture

Chromatin can be broadly defined into two distinct regions, heterochromatin and euchromatin (Fig. 1.1A). Such terms were defined in 1928 by Emil Heitz, owing to the differential intensity of staining along chromosomes (Berger, 2019). Heitz went so far as to suggest dense heterochromatic regions are depleted of genes, a definition that holds true with modern findings (Berger, 2019). Indeed, heterochromatic domains are enriched for repetitive sequences, such as TEs, and largely situated adjacent to the centromere, termed pericentromere (Henderson and Jacobsen, 2007). Euchromatin is present along chromosome arms and is enriched with genes.

Cytological observations of nuclear architecture find heterochromatin domains, also called chromocenters, at the nuclear periphery or adjoining the nucleolus whilst euchromatin evenly occupies the nucleoplasm interior (Simon *et al.*, 2015) (Fig. 1.1A). Heterochromatin can be subdivided into constitutive and facultative forms, depending on its dynamics through development. Such chromatin arrangements are common across somatic eukaryotic nuclei (Simon *et al.*, 2015).

The development of Hi-C, a chromosome conformation capture technique, has enabled probing of three-dimensional chromatin architecture at the genome scale (Doğan and Liu, 2018). Numerous short-range contacts are identified within heterochromatic regions, consistent with the known condensed conformation (Feng *et al.*, 2014; Grob, Schmid and Grossniklaus, 2014). Euchromatic region contacts are far less frequent but can have long range intrachromosomal interactions, known as topologically associated domains (TADs). Prominent TADs are identified in plant species with large genomes, such as maize and rice, but are absent from *Arabidopsis* (Liu *et al.*, 2016, 2017). The occurrence of TADs often corresponds to contacts between enhancer regions (Doğan and Liu, 2018); the presence of bona fide enhancers remains controversial in *Arabidopsis* (Yan *et al.*, 2019; Pontvianne and Liu, 2020).

1.2.2 Phase separation in nuclear compartmentalisation

Phase separation has emerged as a core mechanism that governs chromatin organisation and nuclear function (Erdel and Rippe, 2018). The principles of phase separation are determined by the demixing of substances, resulting in distinct phases (Hyman, Weber and Jülicher, 2014). Phase transitions can take on a range of fluid states, the most common identified in biological systems is liquid-liquid phase separation (LLPS). The principles of phase separation have enabled the explanation of a range of biological phenomena, describing how membraneless organelles can form in cells (Hyman, Weber and Jülicher, 2014).

Driving forces underlying phase separation include the presence of proteins with intrinsically disordered regions (IDRs), intrinsically disordered proteins (IDPs) or the concentration of RNA molecules (Uversky, 2017). IDRs are defined by the inability to form fixed three-dimensional structures. Multivalent interactions between IDRs can drive the formation of aggregates with specific biophysical properties to create a barrier, negotiable by certain proteins but impenetrable to others (Erdel and Rippe, 2018) (Fig. 1.1B).

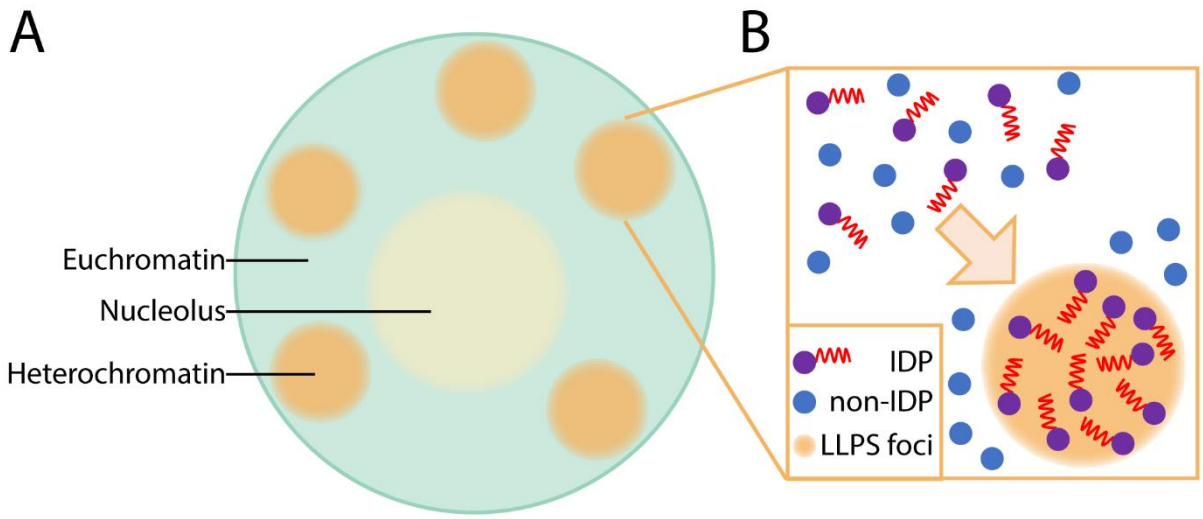


Figure 1.1: Phase separation as a mechanism of heterochromatin domain formation and maintenance. (**A**) Somatic eukaryotic nuclei are characterised by the presence of a nucleolus (yellow), euchromatic interior regions (green) and peripheral heterochromatic domains (orange). (**B**) Heterochromatic domains may form by liquid-liquid phase separation (LLPS; orange), driven by the presence of intrinsically disordered proteins (IDPs; purple/red). Phase separation can occlude certain proteins (such as non-IDP proteins; blue), explaining membraneless nuclear organisation.

Phase separation has been widely implicated in the formation of nuclear compartments. For example, under heat stress, the nucleolus serves as a membraneless reservoir of misfolded proteins which can be recovered once parity resumes (Feric *et al.*, 2016; Frottin *et al.*, 2019).

Regarding chromatin organisation, heterochromatin domains are widely understood to form by LLPS (Erdel and Rippe, 2018) (Fig. 1.1B), although the underlying components remain contentious in eukaryotes (Erdel *et al.*, 2020). HETEROCHROMATIN PROTEIN 1 (HP1) binds repressive histone modifications in animals, serving to scaffold heterochromatin into higher-order compacted structures (Wang *et al.*, 2019). In *Drosophila* and human, HP1 has been demonstrated to form condensates *in vitro* and govern liquid-like behaviour of heterochromatin domains *in vivo* (Larson *et al.*, 2017; Strom *et al.*, 2017). However, mouse HP1 fails to display hallmarks of LLPS despite close homology to metazoan counterparts (Erdel *et al.*, 2020). Comparable to HP1, Agenet Domain-Containing P1 (ADCP1, or ADGP1) in *Arabidopsis* bridges repressive histone marks and displays LLPS properties (Zhang *et al.*, 2018; Zhao *et al.*, 2019). Loss of ADCP1 partially impairs heterochromatin domain stability, indicating the protein is required but is not alone sufficient for domain formation or maintenance *in vivo* (Zhao *et al.*, 2019).

Linker histone H1 has also been implicated in heterochromatin LLPS. For example, animal H1 can form IDR-dependent liquid-like condensates *in vivo* and compact chromatin (Turner *et al.*, 2018; Gibson *et al.*, 2019; Shakya *et al.*, 2020). It is unknown whether *Arabidopsis* H1 has similar behaviour. Furthermore, chromatin itself is said to have liquid-like properties, nucleosome arrays can form droplets both *in vitro* and *in vivo* (Maeshima *et al.*, 2016; Gibson *et al.*, 2019). Although the relevance of such arrays to true biological conditions remains contentious (Erdel and Rippe, 2018), some lessons can be learned regarding chromatin behaviour. For example, acetylation of core histones has been implicated in nucleosome array droplet dissolution (Gibson *et al.*, 2019). Specific variants of core histones have not been demonstrated to promote or inhibit phase separation properties.

Chromatin LLPS is involved in heterochromatin domain assembly and maintenance but has not yet been implicated in large scale euchromatin dynamics. However, phase separation is noted as a mechanism to locally concentrate nuclear process components. For example, transcriptional machinery forms phase separated aggregates to increase RNA Pol II concentration for increased expression of underlying genes (Boehning *et al.*, 2018; Sabari *et al.*, 2018). Additionally, chromatin modifications can be mediated by phase separation, exampled by physical clustering of histone H2B ubiquitylation machinery (Gallego *et al.*, 2020). Collectively, phase separation offers a mechanistic explanation of nuclear compartmentalisation at both macro and micro scales.

1.2.3 Histone variants and modifications

The core histones embody the foundations of chromatin. Histones are highly basic proteins and are characterised by two domains, the N-terminal tail and the C-terminal body. Amino acid differences define histone variants and give rise to specific properties and deposition profiles (Talbert and Henikoff, 2010). Histones can also undergo post-translational modifications, primarily to the tail domain. Such modifications impact upon nuclear functions. Histone deposition and post-translational modifications are mediated by a vast array of enzymes and protein complexes; the mechanisms of which are extensively reviewed elsewhere (Clapier and Cairns, 2009; Gentry and Hennig, 2014; Han *et al.*, 2015; Xiao, Lee and Wagner, 2016). This section introduces *Arabidopsis* histone variants and the key modifications present in somatic chromatin, their genomic localisation, and their influence upon nuclear processes (Fig. 1.2).

H3 variants and associated modifications are the most extensively studied histones across eukaryotic lineages. Inactive euchromatic regions and heterochromatin domains are occupied by H3.1, whereas active euchromatin is enriched for the H3.3 variant (Stroud *et al.*, 2012) (Fig. 1.2A). H3.1 deposition is coupled to replication whilst H3.3 is incorporated throughout the cell cycle (Jiang and Berger, 2017). Centromeres are marked by a specific H3 variant, known as CenH3, the tail of which is required for kinetochore binding in cell division (Ravi *et al.*, 2010; Lermontova, Rutten and Schubert, 2011; Lermontova *et al.*, 2015) (Fig. 1.2A).

Modifications of H3 variants are well characterised. For example, H3K9me2 and H3K27me1 are strongly associated with constitutive heterochromatin (Jacob *et al.*, 2014). Such marks are enriched over pericentromeric TEs but are depleted from genic regions (Roudier *et al.*, 2011) (Fig. 1.2B). H3K27me3 domains cover inactive euchromatin, dynamically marking silenced genes through development (Mozgova, Köhler and Hennig, 2015; Xiao, Lee and Wagner, 2016) (Fig. 1.2B). Active genes are marked by H3K4me3, H3K36me3 and H3K27ac, serving to promote transcription (Roudier *et al.*, 2011; Xiao, Lee and Wagner, 2016) (Fig. 1.2B).

In contrast to H3, its tetrameric binding partner histone H4 has no variants and undergoes comparatively few well-defined modifications. However, acetylation of lysine residues of the H4 tail contribute to chromatin function. The modification is strongly associated with active transcription and permissive chromatin states (Roudier *et al.*, 2011) (Fig. 1.2B). *In vitro* studies have suggested tail acetylation can physically alter nucleosome compaction, causing a more open conformation (Shogren-Knaak *et al.*, 2006). In *Arabidopsis*, acetylation is detected on lysines 5, 8, 12 and 16 (Earley *et al.*, 2007).

A wider variety of histone variants is found in the H2A family (Fig. 1.2A). Besides the canonical H2A histones, variants exist with specific genomic localisations and roles (Kawashima *et al.*, 2015). The most well-established variant is H2A.Z, which deposits to promoters and the TSS of expressed genes (March-Díaz and Reyes, 2009) (Fig. 1.2B). H2A.Z deposition is important for mediating responses to external signals and reshaping the transcriptome through development (March-Díaz and Reyes, 2009). Constitutive heterochromatin is marked by H2A.W (Yelagandula *et al.*, 2014) (Fig. 1.2). Chromocenters are less defined in the absence of H2A.W, although this does not translate to gene or TE misregulation or developmental phenotypes (Bourguet *et al.*, 2020). H2A.X distribution is confined to sites of DNA damage and is in involved in the repair response pathway (Kawashima *et al.*, 2015). H2A variants undergo a variety of modifications but the functional significance of most remain undetermined (Zhang *et al.*, 2007).

The H2B family has a number of variants but they are largely uncharacterised (Jiang *et al.*, 2020; Khadka, Pesok and Grafi, 2020). However, much is known about ubiquitylation of H2Bs (Emre and Berger, 2004; Soares and Buratowski, 2013). Unlike the majority of previously introduced histone modifications which occur to the tail domain, H2B ubiquitylation (H2Bub) occurs to the body domain. This modification physically alters the binding affinity between DNA and histone, rendering nucleosomes less stable and the wider chromatin environment more accessible (Fierz *et al.*, 2011). H2Bub covers the bodies of highly expressed genes and is regarded to promote transcription (Roudier *et al.*, 2011) (Fig. 1.2B).

Linker histone H1 binds between nucleosomes and contributes to chromatin function. Three H1 variants exist in *Arabidopsis*. H1.1 and H1.2 (hereafter as H1) are considered canonical variants and are largely present through development whilst H1.3 is expressed under stress conditions (Hsieh *et al.*, 2016). H1 is present throughout chromatin but is specifically enriched in heterochromatic domains and is important for chromocenter structure (Rutowicz *et al.*, 2019) (Fig. 1.2). H1 regulates of a subset of genes and TEs, in addition to suppressing spurious intergenic transcripts (Choi *et al.*, 2020). Collectively, histone variants and associated modifications are critical to chromatin function in both active euchromatin and repressed heterochromatin.

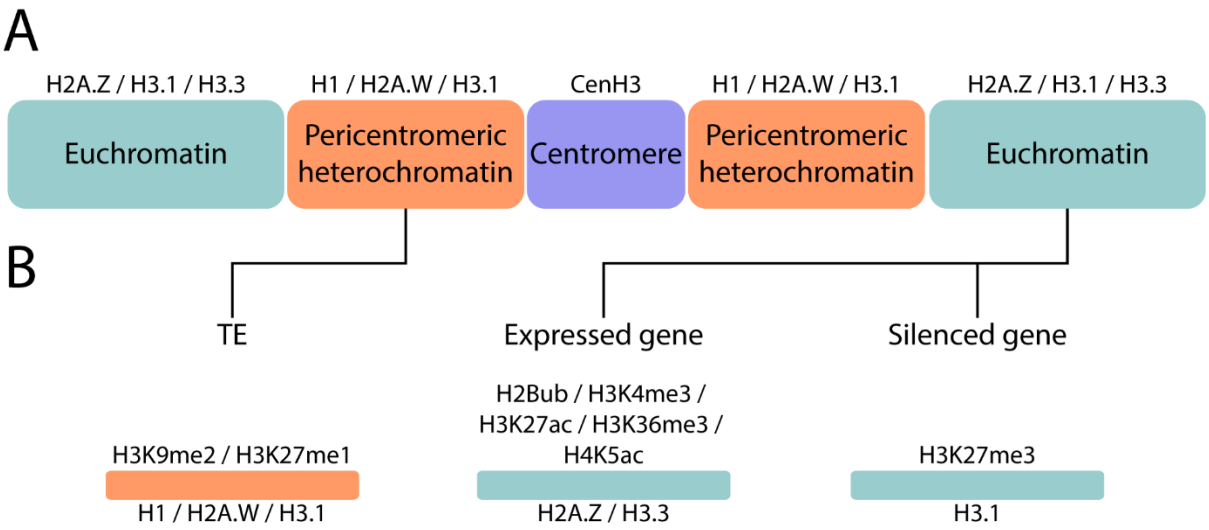


Figure 1.2: Histone variants and modifications in somatic chromatin. (**A**) Chromosomes consist of a centromere (purple), pericentromeric heterochromatin (orange) and euchromatin regions on the chromosome arms (green). Histone variants associated with each region are shown above. (**B**) The majority of TEs are situated in pericentromeric heterochromatin whilst genes are preferentially located in euchromatin. Histone variants (below) and modifications (above) associated with TEs, expressed genes and silenced genes are shown.

1.2.4 DNA methylation

DNA methylation refers to the addition of a methyl group ($-CH_3$) to carbon 5 of the aromatic ring of cytosine nucleic acids. In plants, DNA methylation occurs in three sequence contexts: CG, CHG and CHH (where H represents A, C or T) (Law and Jacobsen, 2010) (Fig. 1.3). DNA methylation is principally targeted to the bodies and edges of TEs, particularly those in heterochromatic regions with dense methylation in all three sequence contexts. Such methylation is required for the stable silencing of TEs (Chan, Henderson and Jacobsen, 2005). The bodies of certain genes are also targeted by DNA methylation in the CG context (Zilberman, 2017). Although the functional relevance of such methylation remains debatable, it is widely regarded to aid constitutive expression of housekeeping genes (Zilberman, 2017).

DNA methylation patterns are perpetuated via several mechanisms. Maintenance of CG methylation is coupled to replication; MET1 (DNA METHYLTRANSFERASE 1) and accessory proteins recognise hemimethylated CG pairs, where there is a methylated cytosine on the parent but not the daughter strand, and fill the methylation in accordingly (Law and Jacobsen, 2010) (Fig. 1.3). Non-CG methylation contexts, CHG and CHH, are maintained by CMT3 (CHROMOMETHYLASE3) and CMT2, respectively (Stroud *et al.*, 2014) (Fig. 1.3). Maintenance pathways are imperfect and so DNA methylation patterns are predicted to dilute through cell divisions (Hsieh *et al.*, 2016). However, *de novo* methylation activity is able to reimpose patterns in all three sequence contexts (Matzke and Mosher, 2014) (Fig. 1.3). This is achieved via RNA-directed DNA methylation (RdDM), whereby DRM (DOMAINS REARRANGED METHYLTRANSFERASE) enzymes are targeted to specific sequences by small interfering RNAs (siRNA). The RdDM pathway has been extensively reviewed previously (Matzke *et al.*, 2009; Law and Jacobsen, 2010; Matzke and Mosher, 2014).

Mechanistically, both non-CG maintenance and RdDM-mediated *de novo* DNA methylation pathways are intrinsically linked to other chromatin factors. For example, CMT enzymes require H3K9me2 modification to perpetuate CHG and CHH methylation patterns (Du *et al.*, 2012). In a feedback loop, histone methylation machinery is recruited by non-CG methylation (Stroud *et al.*, 2014), collectively reinforcing TE silencing in heterochromatic regions. Additionally, chromatin remodelling enzymes provide DNA methyltransferases access to compacted regions. DDM1 (DECREASED DNA METHYLATION 1) counters H1-mediated nucleosome compaction to facilitate DNA methylation (Zemach *et al.*, 2013). Moreover, DRD1 (DEFECTIVE IN RNA-DIRECTED DNA METHYLATION 1) is required for RdDM machinery access to heterochromatin domains (Kanno *et al.*, 2004; Huettel *et al.*, 2006). More recently, evidence suggests H2A.W is required for efficient DNA methylation in heterochromatin, similarly countering H1 functions (Bourguet *et al.*, 2020).

Plants have also evolved a suite of enzymes tasked with countering aberrant DNA methylation by undertaking active demethylation. In somatic tissues, the DNA glycosylase ROS1 (REPRESSOR OF SILENCING 1) functions at gene and TE boundaries, to prevent DNA methylation spread from the latter to the former and avoid unintentional repressive effects (Gong *et al.*, 2002; Zhu *et al.*, 2007). TEs targeted by ROS1 are primarily situated in repressed euchromatic regions, marked by H3K27me3 (Tang *et al.*, 2016). Collectively, DNA methylation dynamics and chromatin organisation are intrinsically linked, cooperating to enable gene expression whilst stably repressing TEs.

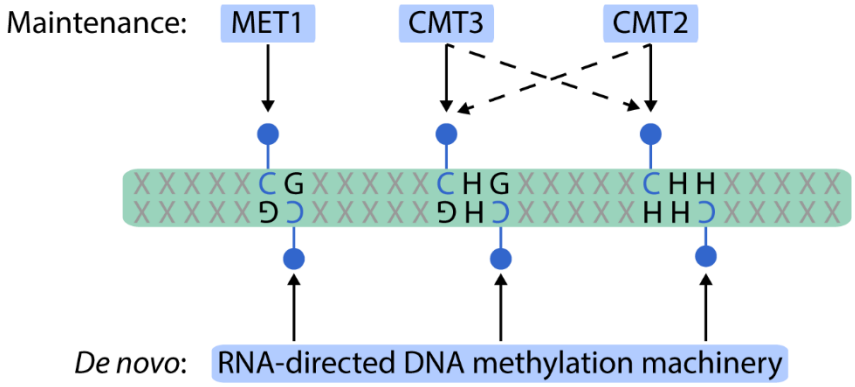


Figure 1.3: DNA methylation mechanisms in *Arabidopsis*. DNA methylation is maintained by MET1 in the CG context. Non-CG contexts, CHG and CHH (where H represents A, C or T), are maintained by CMT3 and CMT2. *De novo* DNA methylation in all three sequence contexts is undertaken by the RNA-directed DNA methylation pathway.

Altogether, this section has introduced core elements of chromatin organisation in *Arabidopsis* from macroscale heterochromatin domains and the principles of phase separation through to specific histone components of chromatin and modifications to the DNA itself. An appreciation of chromatin organisation in the soma is critical to the understanding of reprogramming in the male sexual lineage.

1.3 Male germline specification in flowering plants

The germline in angiosperms is initiated during floral organ morphogenesis. Male sexual lineage development has previously been extensively reviewed (Berger and Twell, 2011; Twell, 2011; Hackenberg and Twell, 2019). The male germline develops in the anthers, where somatic floral meristem cells differentiate into meiocytes in a position-dependent manner. Meiocytes undergo meiosis to produce four haploid daughter cells, in a tetrad arrangement (Fig. 1.4). Release from the tetrad enables differentiation to microspores. Microspores enter a highly asymmetric mitotic division, known as pollen mitosis I (PMI), resulting in a vegetative cell (VC) and generative cell as part of bicellular pollen (Fig. 1.4). The generative cell undergoes a second mitosis event (pollen mitosis II; PMII). This produces two sperm cells (SC), which, together with the VC, constitute mature pollen (Fig. 1.4). During fertilisation, the VC will produce the pollen tube to deliver the SCs to the female gametes. Sperm nuclei have highly compacted chromatin whilst the vegetative nuclear chromatin is globally decondensed (Kawashima and Berger, 2014; Borg and Berger, 2015).

The genetic pathways controlling male germline development are well understood. The R2R3 MYB transcription factor DUO1 (DUO POLLEN 1) serves as key regulator of gene activation in the male sexual lineage and PMII fails in its absence (Rotman *et al.*, 2005; Borg *et al.*, 2011; Higo *et al.*, 2018). DUO3 regulates a subset of DUO1 targets and is required for SC specification (Brownfield *et al.*, 2009). Furthermore, common cell proliferation pathways function in germline development, such as RBR (retinoblastoma-related) and E2F (Chen *et al.*, 2009; Yao *et al.*, 2018).

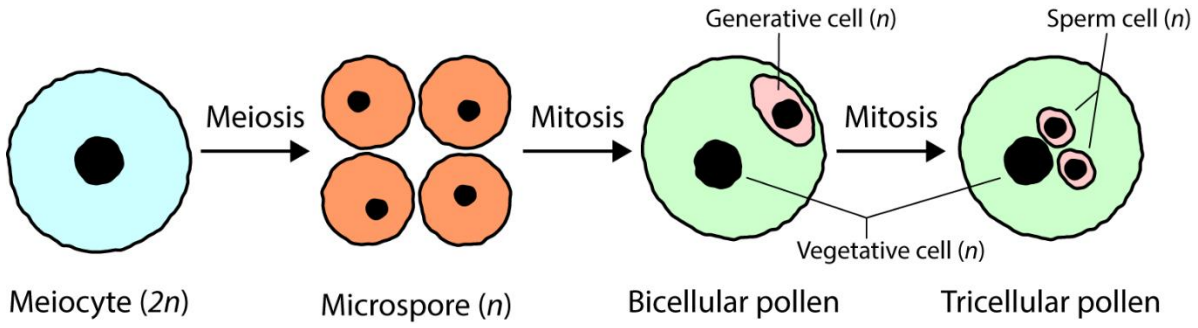


Figure 1.4: Male sexual lineage development in *Arabidopsis*. Meiocytes are differentiated in male reproductive tissue. Meiosis produces four haploid microspores. Each undergoes a highly asymmetric mitotic division to form the generative cell encased by the vegetative cell. The generative cell divides again by mitosis to produce two sperm.

Chromatin is extensively reprogrammed through male germline development to result in the highly dimorphic sperm and vegetative nuclei. Having introduced features of chromatin structure in somatic tissues and how sex cells are formed, the following sections present current literature regarding chromatin reprogramming in male germline development.

1.4 Chromatin compaction in flowering plant sperm

Sperm chromatin in flowering plants is highly condensed. This section will introduce the chromatin structure of sperm and the epigenetic reprogramming events in male sexual lineage development that contribute to it.

Chromocenters are readily observed in sperm, indicating that heterochromatin domains persist from soma rather than undergo reprogramming (Schoft *et al.*, 2009). Sperm nuclear size is much reduced when compared to somatic counterparts (Borg and Berger, 2015). Mutants displaying decondensed sperm chromatin have not been identified. As such, the overall mechanism of extreme sperm chromatin compaction remains unknown. However, other reprogramming events have been characterised.

The repressive euchromatic mark H3K27me3 is absent from sperm chromatin (Borg *et al.*, 2020). The removal is primarily due to activity of demethylase enzymes and the absence of histone methyltransferases (Borg *et al.*, 2020). Furthermore, a sperm-specific histone variant, H3.10 (also called MGH3; MALE GAMETE SPECIFIC HISTONE 3) (Okada *et al.*, 2005), is incorporated to chromatin and immunises loci against trimethylation via amino acid substitutions neighbouring lysine 27 (Borg *et al.*, 2020). Expression of H3.10 is under the direct control of DUO1 (Brownfield *et al.*, 2009). The variant was initially proposed to determine chromatin compaction in sperm (Borg and Berger, 2015), although no decompaction phenotype or fertility defects were reported in the mutant (Borg *et al.*, 2020). Incorporation of H3.10 has minimal impact on gene expression although overall H3K27me3 reprogramming is important for defining the transcriptional repertoire of sperm (Borg *et al.*, 2020). H3.10 is somewhat conserved in flowering plants (Anderson *et al.*, 2013), suggesting that H3K27me3 reprogramming is a feature common across the lineage.

Other sperm specific histone variants have been identified using mass spectrometry of lily species, including *Lilium longiflorum* and *Lilium davidii* (Ueda and Tanaka, 1995; Ueda *et al.*, 2000; Yang, Yang and Wang, 2016). Variants belonging to H2A, H2B and H3 families were reported in sperm but absent from somatic tissues. However, owing to the lack of genetic resources in lily, such variants have not been further explored.

Curiously, linker histone H1 is present in sperm but at lower levels than somatic tissues (Hsieh *et al.*, 2016). Despite the function of H1 to compact chromatin, sperm utilises the protein to a lesser extent. The significance of this downregulation remains to be determined.

DNA methylation exhibits interesting dynamics in male sexual lineage development. Compared to somatic cell types, CG methylation is more robustly maintained in sex cells (Hsieh *et al.*, 2016). Such reinforcement is thought to enable efficient transmission of DNA methylation patterns to the next generation. Furthermore, examination of DNA methylation profiles in meiocytes revealed the presence of specific and hypermethylated loci (Walker *et al.*, 2018). Such methylation was shown to regulate gene expression and is exampled by regulation of splicing of MPS1, essential for normal meiotic progression. Sexual lineage specific methylation persists from meiocytes to sperm (Walker *et al.*, 2018).

Collectively, mechanisms of epigenetic reprogramming have been identified in flowering plant sperm. However, how global compaction occurs remains unknown.

1.5 Decondensation of flowering plant vegetative cell chromatin

In contrast to the highly compacted chromatin structure of sperm, its companion within the pollen, the vegetative cell, has decidedly decondensed global chromatin organisation. This section will introduce established mechanisms of chromatin decompaction in the vegetative cell of pollen.

Centromeres are disassembled in the vegetative cell (VC) owing to the active removal of CenH3 by an AAA-ATPase chaperone, CDC48A (Mérai *et al.*, 2014). The VC is terminally differentiated and so the centromere is no longer required. CenH3 removal releases rRNA genes from silencing, subsequently enabling ribosome biogenesis. The increased protein production capability is critical for formation of the pollen tube and SC delivery in fertilisation (Mérai *et al.*, 2014).

Consistent with centromere disassembly, distinct H3K9me2-marked chromocenters are lost in the VC, although a single H3K27me1-marked domain can be observed (Schoft *et al.*, 2009). Moreover, H1 is absent from VC chromatin (Hsieh *et al.*, 2016; He *et al.*, 2019). The mechanisms by which such heterochromatic features are depleted from VC chromatin are unknown. However, as the VC is only separated by a single cell division, the process is likely active rather than passive.

Heterochromatin depletion is concomitant with an increase of active H3K4me3 modification (Pinon *et al.*, 2017). SDG2 (SET-DOMAIN GROUP 2) is principally responsible for trimethylation of H3 lysine 4 (Xiao, Lee and Wagner, 2016). Loss of SDG2 causes increased VC chromatin condensation, but not the rescue of chromocenters (Pinon *et al.*, 2017).

Loss of heterochromatin domains is accompanied by release of TE silencing (Slotkin *et al.*, 2009; Calarco *et al.*, 2012; Ibarra *et al.*, 2012). Activated TEs are situated in pericentromeric regions and normally marked by H3K9me2 (He *et al.*, 2019). Release of silencing is mediated by active DNA demethylation by DEMETER (DME), a DNA glycosylase specifically expressed in companion cells of gametes (Schoft *et al.*, 2011; Ibarra *et al.*, 2012). DME preferentially targets the edges of TEs and demethylation activity across all three sequence contexts is sufficient to release silencing (Ibarra *et al.*, 2012). The access of DME to heterochromatic sequences is partially due to loss of H1. Upon ectopic H1 expression in VC, a subset of TEs fail to undergo demethylation and subsequently do not activate (He *et al.*, 2019). However, H1 alone cannot fully explain TE expression in VC; therefore, other unknown mechanisms also contribute to the phenomenon.

TE activation in the VC provides templates for siRNA biogenesis (Slotkin *et al.*, 2009). Such siRNAs are understood to direct DNA methylation in sperm via the RdDM pathway (Calarco *et al.*, 2012; Ibarra *et al.*, 2012; Martínez *et al.*, 2016). DNA methylation at TEs is reinforced in sperm to ensure stable silencing across generations (Feng, Zilberman and Dickinson, 2013).

The core histone variant H3.14 is specifically expressed in gamete companion cells, such as the VC (Ingouff *et al.*, 2007, 2010). H3.14 differs from H3.3 by five amino acids, but none provide clear indications of involvement in heterochromatin disassembly (Borg and Berger, 2015). The role of H3.14 in companion cell chromatin is unknown.

Overall, VC chromatin decondensation is known to be caused by breakdown of chromocenters. The mechanism of centromere loss is established. However, the mechanisms that lead to removal of heterochromatic features remains unknown. Furthermore, DME access to TEs can only be partially explained by depletion of H1 and so other mechanisms exist to enable active demethylation of VC.

1.6 Metazoan sperm function and chromatin structure

Sperm chromatin compaction is ubiquitous across eukaryotic lineages (Hao, Ni and Yang, 2019). Unlike flowering plants, the mechanism in animals, such as mouse (*Mus musculus*) and human (*Homo sapiens*), is well established (Braun, 2001; Kimmins and Sassone-Corsi, 2005; Schagdarsurengin, Paradowska and Steger, 2012). This section will introduce sperm chromatin compaction in metazoans and draw comparisons to the process in *Arabidopsis*.

Mammals undergo near-global histone replacement by small, arginine-rich proteins called protamines during spermatogenesis (Balhorn, 2007). Before this, transitional histone variants are incorporated to replace somatic histones across the genome (Braun, 2001; Kimmins and Sassone-Corsi, 2005; Rathke *et al.*, 2014). For example, deposition of testis-specific H2B and H3 variants, known as TSH2B and H3T, respectively, promote nucleosome instability at early stages of sperm development (Tachiwana *et al.*, 2010; Montellier *et al.*, 2013). Furthermore, replacement of canonical linker histone H1 with the testis-specific variant H1T reduces chromatin compaction via lower binding affinity (Pérez-Montero, Carbonell and Azorín, 2016). Additionally, histone H4 hyperacetylation promotes histone eviction (Meistrich *et al.*, 1992). The combination of transitional histone variants and modifications leads to a loosened chromatin conformation.

Loosening of chromatin structure during early mammalian spermatogenesis enables the formation of DNA double strand breaks (DSBs) (Schagdarsurengin, Paradowska and Steger, 2012). The subsequent DSB repair process is coupled to incorporation of protamines, enabling a tight association with DNA and the formation of toroid-shaped looped structures. Protamines are intrinsically disordered and form gel-like phase separated condensates in sperm (Gou *et al.*, 2020).

Protamines are characterised by large contiguous blocks of arginine residues, which collectively constitute 50-80% of amino acid composition in mammals and 35-75% in other eukaryote lineages (Balhorn, 2007). The highly positive charge of arginine residues is conferred by the $-NH_2^+$ side chain, rendering protamines as basic proteins. The slight negative charge of DNA means that associations with protamines are very strong, with an overall neutral charge achieved (Balhorn, 2007). Protamines bind to the minor groove of DNA and induce the coiling of DNA into large toroidal structures. Furthermore, protamines are much smaller than individual histones (50-100 amino acids versus ~140-150 amino acids) and also the nucleosome octamer. As such, greater binding of DNA is achieved and a more highly compacted chromatin state is achieved (Balhorn, 2007).

Histones are evicted from ~85% of the genome and replaced with protamines (Balhorn, 2007); this near-global protamine incorporation is required for sperm viability (Cho *et al.*, 2001). Of the ~15% remaining histones, enrichment is identified at the promoters of genes important for early development (Hammoud *et al.*, 2009). These histones carry active and repressive modifications that correlate with expression post-fertilisation (Hammoud *et al.*, 2009; Teperek *et al.*, 2016). However, it remains to be determined whether this link is causal.

Mature sperm chromatin in mammals is highly compacted. Such compaction prevents all *de novo* transcription, with remaining RNA inherited from earlier developmental stages (Grunewald *et al.*, 2005). As such, mature sperm is limited in ability to respond to stimuli and is tasked solely with reaching female gametes (Kimmins and Sassone-Corsi, 2005). This is in contrast to flowering plants, which maintain the ability to transcribe albeit to a lesser extent than somatic tissues (Borges *et al.*, 2008).

The functional significance of sperm chromatin compaction is attributed to the reduction of the sperm nuclear size and corresponding head shape (Malo *et al.*, 2006; Maximiliano, Montserrat and Roldan, 2011). Mammalian sperm requires flagella-driven swimming for fertilisation. A small sperm head has been demonstrated to be important for swimming ability and is assumed to confer a competitive advantage in the race to fertilise (Malo *et al.*, 2006; Maximiliano, Montserrat and Roldan, 2011). Resultantly, sperm chromatin compaction is under intense evolutionary pressure. Protamines have enabled the ability to dramatically compact sperm chromatin but require the formation of potentially hazardous DSBs and loss of transcriptional capability (Balhorn, 2007).

Protamine-mediated sperm chromatin compaction is common across metazoans, but exceptions do exist. For example, *Caenorhabditis elegans* retains histones through spermatogenesis (Tabuchi *et al.*, 2018). Inheritance of H3K27me3 via paternal chromatin is required for a normal developmental programme in offspring (Kaneshiro, Rechtsteiner and Strome, 2019). Furthermore, zebrafish (*Danio rerio*) sperm chromatin has been demonstrated to acquire placeholder nucleosomes consisting of H2A.Z and the active histone modification H3K4me3 (Wu, Zhang and Cairns, 2011). Such placeholders do not affect transcription in sperm but are required for gene activation in early embryos (Murphy *et al.*, 2018). Despite the independent loss of protamines through evolution in certain lineages, sperm chromatin compaction remains consistent (Hao, Ni and Yang, 2019).

1.7 Non-flowering plant male reproduction and chromatin organisation

Similar to the majority of metazoans, non-seed plant sperm, such as that of pteridophytes and bryophytes, carries protamine-based chromatin and requires a flagellated morphology to enable swimming for fertilisation (Reynolds and Wolfe, 1984; Southworth and Cresti, 1997; Renzaglia and Garbary, 2001).

Protamines in non-seed plants, such as *Marchantia polymorpha*, share similar characteristics to metazoan homologs. For example, they are small in size and have contiguous blocks of arginine residues (Balhorn, 2007). Interestingly, evidence suggests that *Marchantia* protamines are formed by cleavage of precursor linker histone H1 (D'Ippolito *et al.*, 2019). This suggests an evolutionary relationship between protamine and H1, whether this extends throughout eukaryotic lineages remains to be determined.

Sperm chromatin of non-flowering seed plants, known as gymnosperms, is an outlier compared to the majority of eukaryotes; it does not undergo compaction to any extent during spermatogenesis (Southworth and Cresti, 1997). Accordingly, protamines have not been detected in non-flowering seed plant species (Southworth and Cresti, 1997). Gymnosperm lineages utilise VC-driven pollen tubes, similarly to angiosperms. However, certain species, such as *Ginkgo biloba*, also maintain flagellated sperm and are required to swim short distances during the fertilisation process (Norstog, Gifford and Stevenson, 2004). Gymnosperms represent an anomaly regarding sperm chromatin structure and also a conceptual intermediate in fertilisation mechanism from lineages with swimming sperm to those that use a pollen tube to deliver sperm.

Flowering plants have highly condensed sperm despite not requiring swimming for fertilisation. Protamines have not been identified in sperm chromatin of angiosperms (Southworth and Cresti, 1997), therefore compaction is achieved by another unknown mechanism.

Taxonomic group	Clade	Species	Compact sperm chromatin	Protamines	Pollen tube	Flagella
Seed plants	Angiosperms - eudicots	*Arabidopsis thaliana*	✓	×	✓	×
	Angiosperms - monocots	*Lilium davidii*	✓	×	✓	×
		Hordeum vulgare	✓	×	✓	×
	Angiosperms - basal	*Amborella trichopoda*	✓	×	✓	×
		Nymphaea colorata	✓	×	✓	×
	Gymnosperms	*Picea abies*	×	×	✓	×
		Ginkgo biloba	×	×	✓	✓
		Zamia sp.	×	×	✓	✓
Non-seed plants	Pteridophytes	*Marsilea* sp.	✓	✓	×	✓
	Bryophytes	*Marchantia polymorpha*	✓	✓	×	✓
		Physcomitrella patens	✓	✓	×	✓
	Lycophytes	*Selaginella moellendorffii*	✓	✓	×	✓

Figure 1.5: Sperm chromatin in compaction in plants and fertilisation mechanism. Comparison of plant lineage sperm chromatin compaction, the mechanism of compaction (protamines) and the method of sperm delivery to female gametes (pollen tube / flagella).

1.8 Chromatin organisation in female gametes and post-fertilisation

This section will introduce female sexual lineage development and draw parallels to epigenetic reprogramming in the male germline. Additionally, chromatin dynamics post-fertilisation will be discussed.

The female gametes develop from the products of meiosis, known as the megaspore (Berger and Twell, 2011). Then, three rounds of nuclear division take place to produce a multinucleate cell. The nuclei differentiate into the haploid egg cell (EC) and diploid central cell (CC), along with accessory cells known as synergids and antipodals.

Flowering plant reproduction utilises a process known as double fertilisation (Berger *et al.*, 2008; Hamamura *et al.*, 2011). The pollen tube, produced by the VC, delivers the two sperm to the female gamete. Individual sperm will fuse with the EC and CC separately. The former fusion results in the diploid zygote, which will develop into the embryo and ultimately the mature plant. The latter fusion produces the triploid endosperm, which provides nourishment for the developing embryo within the seed.

The chromatin structure within the female gametophyte has some similarity to the male development. The EC is substantially smaller in size than its companion, the CC, suggesting a somewhat condensed state (Pillot *et al.*, 2010). The EC maintains H3K9me2-marked heterochromatic domains despite undergoing histone H3 reprogramming (Ingouff *et al.*, 2010). On the other hand, the CC is depleted of distinct heterochromatic domains, similar to the VC (Pillot *et al.*, 2010), suggesting equivalent global chromatin decondensation. Indeed, the CC undergoes active DME-mediated DNA demethylation, which is thought to release TE silencing in a similar manner to the pollen (Ibarra *et al.*, 2012).

DME activity in the CC also establishes the DNA methylation profile of imprinted genes, which exhibit parent-of-origin biased expression in the endosperm (Hsieh *et al.*, 2011; Schoft *et al.*, 2011). Demethylation in the VC also overlaps imprinted genes. However, the majority of identified imprinted genes exhibit maternally biased expression rather than paternal (Hsieh *et al.*, 2011). Imprinted genes include transcription factors required for seed development and chromatin modifying enzymes (Hsieh *et al.*, 2011).

Recently, repression of FLC by H3K27me3 has been demonstrated to transmit through the female germline development and persist in the embryo following fertilisation (Luo *et al.*, 2020). This represents a mechanism by which memory of cold exposure can be inherited via maternal chromatin (Luo *et al.*, 2020). Borg *et al*. (2020) suggest that resetting of H3K27me3 in sperm forecasts the transcriptional programme in the next generation. However, direct evidence is lacking, and the effect could simply be owing to correlation; particularly given that histone H3 is entirely reset following fertilisation (Ingouff *et al.*, 2010).

Chromatin dynamics in female germline development and post-fertilisation remains somewhat unknown owing to the low numbers of cells and difficulty in obtaining material. As such, the male sexual lineage represents a core model for the study of chromatin reprogramming in germlines of flowering plants.

1.9 Thesis outline

This thesis addresses the chromatin dimorphism observed between sperm and vegetative nuclei in *Arabidopsis* pollen. To find candidate proteins that could partake in chromatin reprogramming in the male germline, a proteomics approach was used. The proteomes of sperm and vegetative nuclei were generated, and specific proteins were considered as candidates. This work focused on a potential contributor to sperm chromatin compaction (Chapter 2) and a candidate involved in vegetative chromatin decondensation (Chapter 3).

Chapter 2 introduces H2B.8 as a novel sperm-expressed histone variant that is required for normal male fertility. The variant deposits to silenced euchromatic regions and does not affect transcription. H2B.8 is demonstrated to undergo IDR-dependent phase separation *in vitro* and *in vivo*. This behaviour induces the formation of small euchromatic foci in nuclei, resulting in greater chromatin compaction and smaller nuclear size.

Chapter 3 implicates HAM (Histone Acetyltransferase of the MYST family) proteins as determinants of heterochromatic TE accessibility. HAM proteins are required for DME-mediated active demethylation at heterochromatic targets. In the absence of HAM proteins, a subset of pollen-expressed TEs fail to activate. Ultimately, HAM-mediated VC decondensation could be important for RdDM reinforcement of TE silencing in sperm and thus, across generations.

Chapters 2 and 3 consist of a brief introduction followed by several results sections and a short discussion to draw together the presented work. Each results chapter has a materials and methods section along with supplemental tables of the resources used. Finally, Chapter 4 draws general conclusions and discusses the presented results in the wider context of chromatin biology.

Chapter 2 – H2B.8 phase separates inactive euchromatin to drive sperm compaction in flowering plants

2.1 Introduction

Sperm chromatin compaction is a ubiquitous feature of eukaryotes. In the vast majority of metazoans and non-seed plants, sperm chromatin compaction is achieved by near-global replacement of histones by small, arginine-rich proteins known as protamines (Balhorn, 2007).

Incorporation of protamines through spermatogenesis is often preceded by the substitution of somatic histones with germline specific variants. Examples of such are the human testis-specific histones TSH2B and H3T, which destabilise the nucleosome and globally relax chromatin (Tachiwana *et al.*, 2010; Montellier *et al.*, 2013). This loosened state enables the formation of DNA double strand breaks (DSBs). The subsequent eviction of nucleosomes and DSB repair process incorporates protamines into tightly wound toroid loop structures (Braun, 2001; Kimmins and Sassone-Corsi, 2005). At this stage, chromatin is incredibly condensed, and the sperm head becomes very small.

Protamines occlude all *de novo* transcription, meaning mature sperm has no capacity to react to external stimuli should it need (Grunewald *et al.*, 2005). Additionally, the formation of DSBs is a gamble given the potential to introduce mutations in the germline (Schagdarsurengin, Paradowska and Steger, 2012). However, such risk is balanced evolutionarily, as compacting the sperm nucleus is thought to offer a speed advantage in the competitive motile fertilisation process (Malo *et al.*, 2006; Maximiliano, Montserrat and Roldan, 2011).

Unlike metazoans and non-seed plants, sperm of flowering plants are not motile (Southworth and Cresti, 1997). Instead, sperm are delivered by a pollen tube to the female gametes. Despite the immotile nature, sperm chromatin compaction is still observed across angiosperm lineages (Southworth and Cresti, 1997). Protamines have not been identified in sperm of flowering plants, so compaction is achieved by a different mechanism. This mechanism enables *de novo* transcription in sperm (Borges *et al.*, 2008), critical for normal functions.

The mechanism and functional significance of sperm chromatin compaction in flowering plants is unknown. This chapter seeks to address such questions by using proteomics to identify sperm and vegetative nuclei specific proteins in pollen that could contribute to the observed chromatin dimorphism. This chapter describes one such candidate, the previously uncharacterised histone variant H2B.8 (*HTB8*; AT1G08170). The role of H2B.8 in sperm chromatin compaction and fertility is determined, along with the mechanism by which H2B.8 functions.

2.2 H2B.8 is a novel histone variant in *Arabidopsis thaliana*

The chromatin states of sperm and vegetative nuclei in pollen are extremely dimorphic; the former has highly condensed chromatin, whilst the latter exhibits a very loose chromatin conformation. Here, FACS (Fluorescence-Activated Cell Sorting) isolation of sperm and vegetative nuclei followed by a novel protein extraction protocol enabled the profiling of nuclear proteomes. This analysis identified a number of potential candidates that could be involved in the observed chromatin dimorphism. This chapter describes one such candidate, the previously uncharacterised histone variant H2B.8 (*HTB8*; AT1G08170), and its contribution to sperm chromatin compaction.

H2B.8 specific peptides were identified in sperm but absent from vegetative nuclei across two replicates of mass spectrometry (Fig. 2.1A); therefore, H2B.8 was assumed to be sperm-specific in pollen. To attain a measure of the abundance of H2B.8 in sperm, canonical H2B peptide counts were compared to H2B.8-specific peptides. Accounting for the larger size of H2B.8 relative to canonical H2B (243 amino acids versus 151), H2B.8 constitutes 12.6% of total sperm histone H2B (Fig. 2.1A). Although, owing to the low input nature of the proteomics, this value is a relatively low-confidence estimate.

The recently published *Arabidopsis* proteome atlas (Mergner *et al.*, 2020) corroborates the presence of H2B.8 in pollen. Additionally, the atlas also detects H2B.8 peptides in mature seeds but does not detect the protein in any other examined tissue. To further explore the specificity of H2B.8, RNA-seq datasets were examined. In agreement with the mass spectrometry data, *HTB8* transcripts were not detected in any tissue besides sperm and mature seeds (Fig. 2.1B), such transcripts were found to be highly abundant in these samples. Therefore, it seemed likely that H2B.8 was specific to sperm and mature seeds.

To confirm the presence of H2B.8 and to establish its incorporation dynamics, a GFP fusion line was generated (*pHTB8::HTB8-eGFP*). Examination of the reporter line verified the mass spectrometry and RNA-seq data; GFP signals were detected specifically in sperm and mature seeds and are absent from other somatic tissues such as leaf and root (Fig. 2.1, C and D).

Excitingly, confocal imaging through male sexual lineage development reveals that H2B.8 is incorporated into chromatin following Pollen Mitosis II (PMII). It is at this point in pollen development that sperm chromatin undergoes extreme compaction and nuclei become very small in size (Fig. 2.1C). This emphasises the potential of H2B.8 as a candidate in sperm chromatin condensation.

Following fertilisation, H2B.8 is quickly depleted from chromatin and is cytologically undetectable at the 2-cell embryo stage (Fig. 2.1D). This removal dynamic mirrors previously published work on paternal histone turnover post-fertilisation (Ingouff *et al.*, 2010). Interestingly, H2B.8 returns to chromatin in mature seeds, with GFP signals detected in nuclei of all cell types (Fig. 2.1D). This expression of H2B.8 in mature seeds coincides with nuclear size reduction and chromatin compaction (Van Zanten *et al.*, 2011). This is similar to sperm, where the observation of H2B.8 is simultaneous with chromatin condensation and decreased nuclear size. To note, two recent papers reported H2B variant expression profiles and support the findings presented here (Jiang *et al.*, 2020; Khadka, Pesok and Grafi, 2020).

Given the unique expression patterns of H2B.8, and its correlation with developmental chromatin compaction events, it was taken forward as a candidate protein. The rest of this chapter explores whether H2B.8 is causal in chromatin condensation in sperm and if so, the mechanism by which it achieves this function. The role of H2B.8 in mature seeds is beyond the scope of this work, although it is discussed in section 2.13.

a

MS rep	H2B.8 peptides	cH2B peptides	H2B.8 peptides (%)	Normalised H2B.8 peptides (%)
1	14	51	21.5	14.6
2	12	61	16.4	10.9
Both	26	112	18.8	12.6

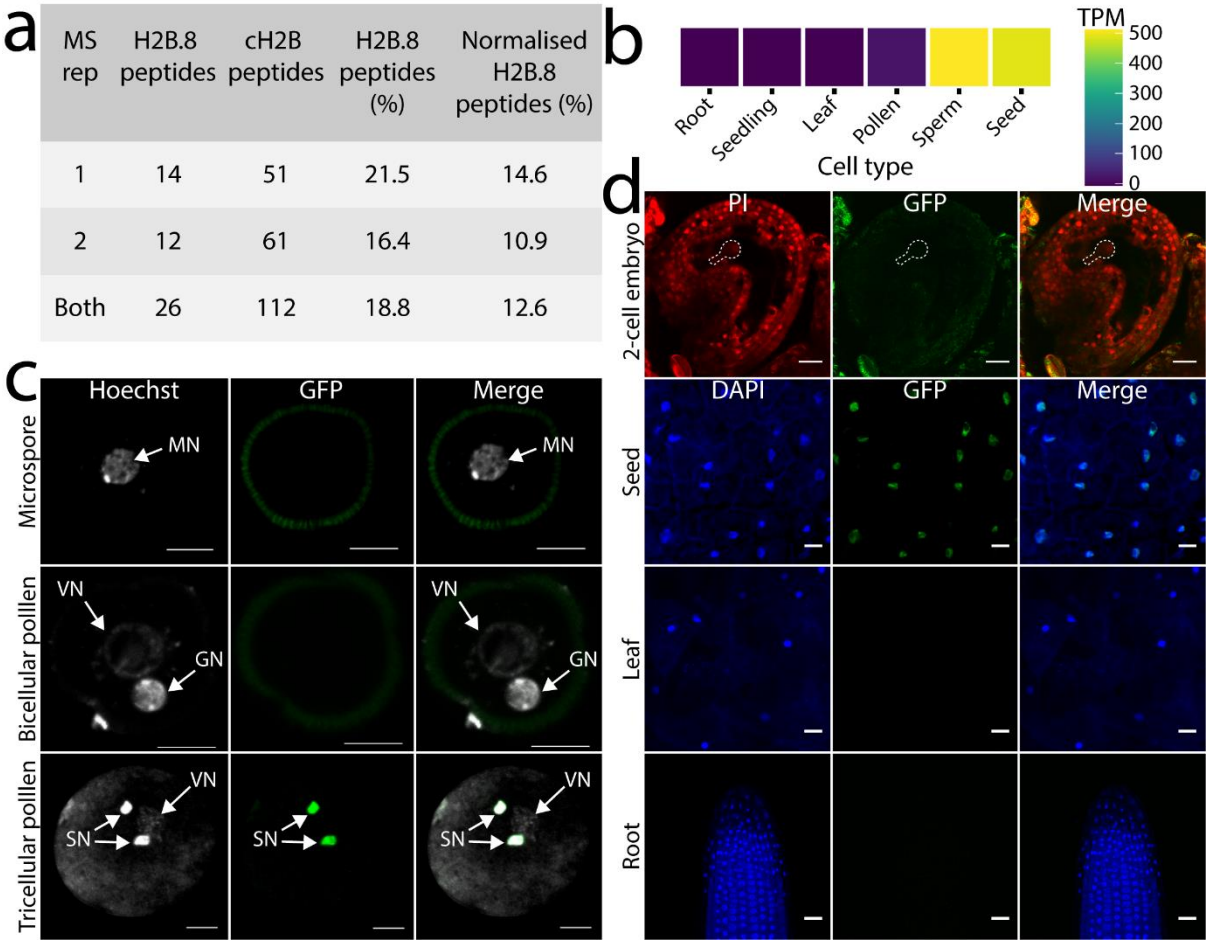


Figure 2.1: H2B.8 is a novel histone variant in *Arabidopsis thaliana* and is specific to sperm and mature seeds. (**A**) H2B.8 peptides are detected in sperm but not vegetative nuclei. H2B.8 protein comprises 12.6% of total histone H2B in sperm, adjusting for the larger size of H2B.8 compared to canonical H2Bs (243 amino acids versus 151). (**B**) *HTB8* transcripts are highly abundant in sperm and seed, but not detected in any other tissue. (**C**) Confocal images of H2B.8 (*pHTB8::HTB8-eGFP*) through male sexual lineage development. H2B.8 incorporates following Pollen Mitosis II and is therefore present in mature sperm in tricellular pollen. MN – microspore nucleus; GN – generative cell nucleus; VN – vegetative cell nucleus; SN – sperm cell nucleus (**D**) Confocal images of H2B.8 (*pHTB8::HTB8-eGFP*) in somatic tissues. H2B.8 is lost post-fertilisation and is therefore absent at the 2-cell embryo stage. Mature seeds have H2B.8 signals, but it is absent from other somatic tissues such as leaf and root. Scale bars are 5 µm in (C and D – mature seed) and 20 µm in (D – 2-cell embryo, leaf, root).

In conclusion, this section has shown how H2B.8 was identified using mass spectrometry of sperm and vegetative nuclei from *Arabidopsis* pollen (Fig. 2.1A). Analysis of RNA-seq data from various tissues revealed transcripts in sperm and mature seeds (Fig. 2.1B). GFP tagging of H2B.8 confirmed proteomic and transcriptomic data (Fig. 2.1, C and D). The following section compares *Arabidopsis* H2B variants and assesses features unique to H2B.8.

2.3 H2B.8 is unique among *Arabidopsis* H2B variants

H2B.8 is highly diverged from other H2B histones in *Arabidopsis*; three distinct properties distinguish this variant from canonical H2Bs. First, H2B.8 has a large N-terminal tail domain, the increased length is caused by a 90 amino acid intrinsically disordered region (IDR) (Fig. 2.2). Second, compared to canonical H2Bs, H2B.8 is enriched for histone body surface arginines (Fig. 2.2). Last, rather than a lysine at position 145 as part of a highly conserved C-terminal motif, H2B.8 has an asparagine residue (position 234) and adjacent amino acid substitutions (Fig. 2.2). The following sections will describe each of these features in greater detail.

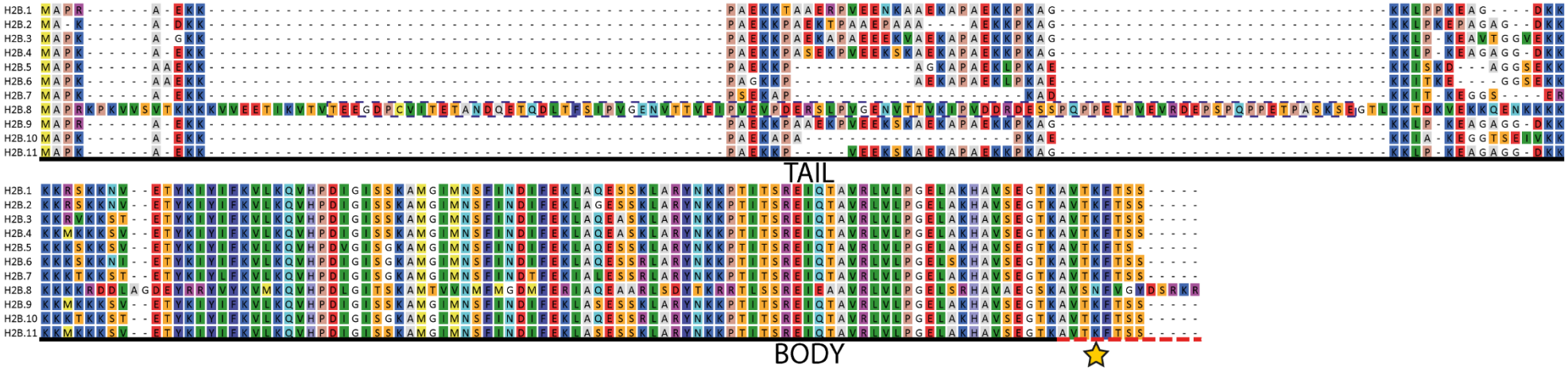

H2B.1
H2B.2
H2B.3
H2B.4
H2B.5
H2B.6
H2B.7
H2B.8
H2B.9
H2B.10
H2B.11
TAIL
BODY

Figure 2.2: Alignment of *Arabidopsis* H2B variants. *Arabidopsis thaliana* histone H2B variant alignment. The upper panel contains the N-terminal tail and the lower panel is the C-terminal body domain. H2B.8 has a large IDR in the N-terminal tail (blue dashed box), increased arginine residues (R - purple) in the histone body and a non-canonical C-terminal motif (red dashed section) with a K-to-N substitution (star) at position 145 (234 in H2B.8).

2.3.1 H2B.8 has a large intrinsically disordered region in the N-terminal tail

H2B.8 is much larger than canonical H2B variants due to its 90 amino acid IDR in the N-terminal tail domain (Fig. 2.3). IDRs are defined by their lack of defined three-dimensional structures and can be predicted according to protein sequence (Uversky, 2017). Recently, there has been a spike of interest in IDRs and IDPs (intrinsically disordered proteins) given their role in the formation of membraneless organelles by phase separation (Hyman, Weber and Jülicher, 2014). For example, IDRs and IDPs have been demonstrated to form structures such as P-bodies and stress granules in the cytosol (Hyman, Weber and Jülicher, 2014). Chromatin biology in particular has seen a great deal of interest in phase separation, with IDRs and IDPs explaining the formation of distinct nuclear compartments (Erdel and Rippe, 2018). A well characterised example is HETEROCHROMATIN PROTIEN 1 (HP1) in human and *Drosophila*, which binds methylated H3K9 histones and forms liquid-liquid phase separated (LLPS) domains of repressive chromatin (Larson *et al.*, 2017; Strom *et al.*, 2017).

The unique presence of an IDR in the N-terminal tail domain of H2B.8 (Fig. 2.3) offers the exciting possibility that the protein undergoes phase separation as a means of chromatin organisation in sperm. Further sections seek to explore the contribution of the IDR to protein function.

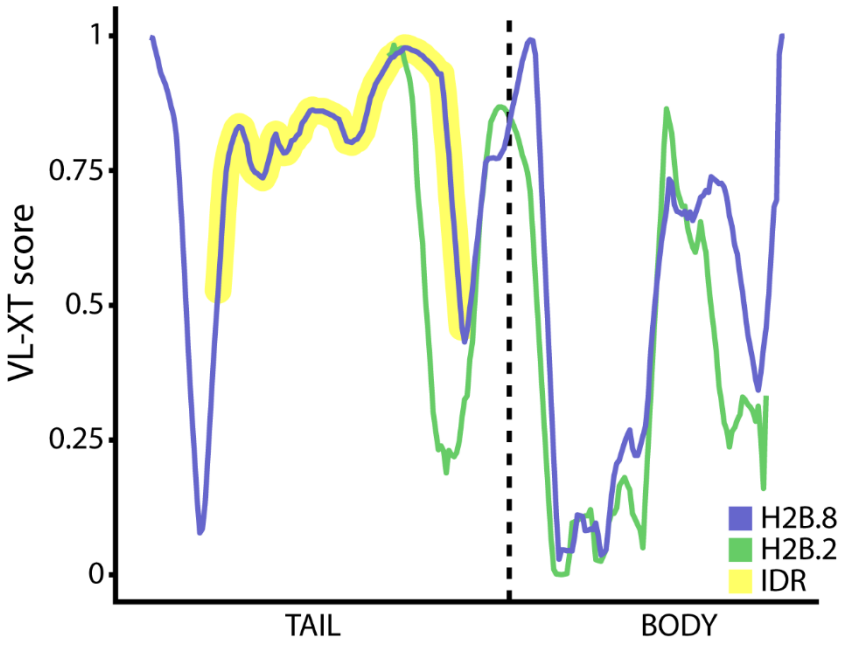


Figure 2.3: Intrinsic disorder predictions of H2B.8 and H2B.2. VL-XT score predictions of intrinsically disordered regions (yellow) for H2B.8 (blue) and a representative canonical H2B variant, H2B.2 (green). H2B.8 and H2B.2 are aligned at the transition between tail and body domains.

2.3.2 H2B.8 is enriched for histone body surface arginine residues

Another distinguishing feature of H2B.8 compared to canonical H2B variants is the increased presence of arginine residues (Fig. 2.4A). The total number of arginine residues in H2B.8 is approximately four times greater than the average among canonical H2Bs (Fig. 2.4A). Such arginine residues are primarily situated in the histone body domain, rather than the N-terminal tail (Fig. 2.4A).

To determine the location of these additional arginine residues, predicted structures of H2B.8 and a representative canonical H2B variant, H2B.2, were generated using Phyre2 (Kelley *et al.*, 2015). Resulting structures for the body domains were established with high confidence, owing to the evolutionary conservation of histone proteins across eukaryotes. Predicted structures were aligned to a published nucleosome core particle crystal structure (PDB = 1KX5) using PyMol (Schrödinger, 2015). Visualisation of arginine residues showed that they are preferentially found at DNA-histone contact points within the nucleosome for both H2B.8 and H2B.2 (Fig. 2.4B).

Arginine amino acids are highly positively charged owing to a $-NH_2^+$ side chain, whereas DNA is overall negatively charged. It is possible that the increased arginine in H2B.8 on the histone body surface may lead to a tighter association with DNA within the nucleosome. Such a property could give greater stability to H2B.8-containing nucleosomes and increase chromatin condensation. High arginine content is reminiscent of protamine structure. Contiguous blocks of arginines serve to anchor protamines to the phosphate backbone of DNA, inducing tight toroidal loops (Balhorn, 2007). It is tempting to speculate that the increased arginine content in the H2B.8 body domain at DNA-histone contacts could serve an analogous function and represent convergent evolution towards tight binding of DNA to structural proteins in sperm of eukaryotes. The contribution of H2B.8 arginine residues to histone protein function is explored in further sections.

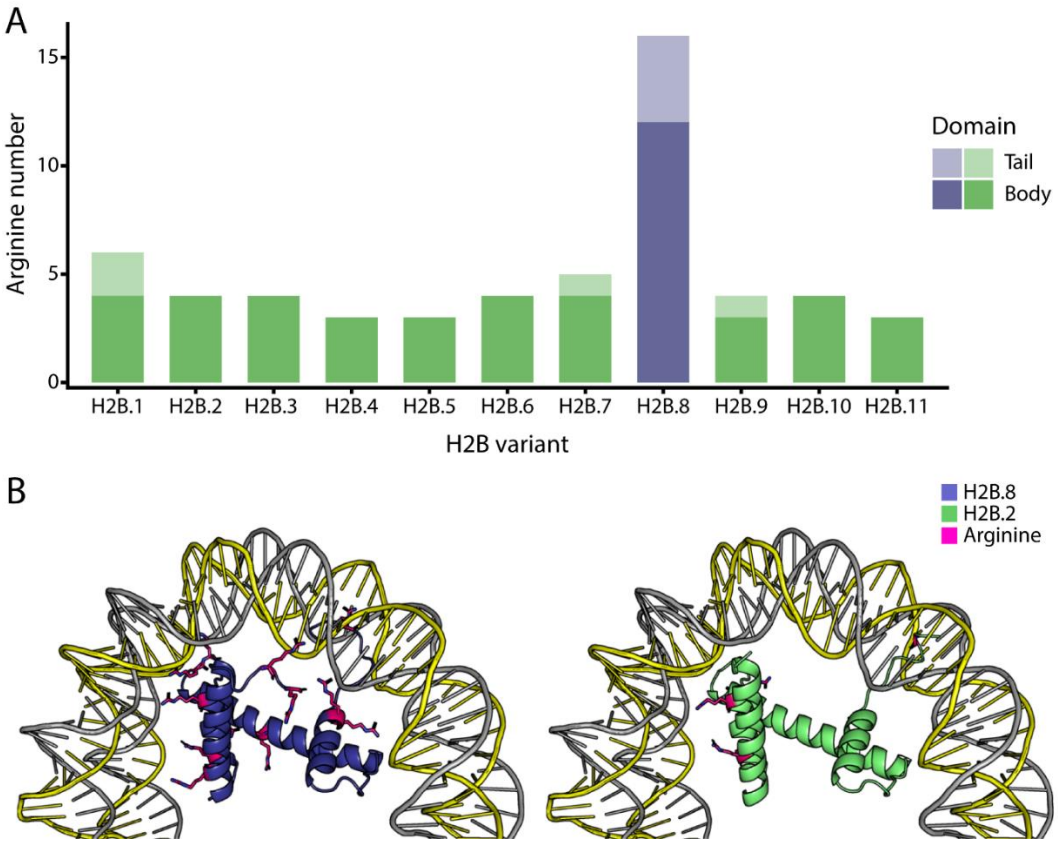


Figure 2.4: H2B.8 has increased histone body surface arginine residues. (**A**) H2B.8 (blue) has higher numbers of arginine residues compared to canonical H2B variants (green). Arginine residues are primarily located on the histone body (darker colour) rather than the tail domain (lighter colour). (**B**) Alignments of histone body predicted structures for H2B.8 (blue) and H2B.2 (green) to a published nucleosome core particle crystal structure (PDB = 1KX5). Arginine residues are shown as sticks (pink).

2.3.3 H2B.8 has a K-to-N substitution at position 145

The final distinguishing feature of H2B.8 is a K-to-N substitution at position 234 (145 in canonical H2B) as part of an abnormal C-terminal motif (Fig. 2.5). This motif (AVTKFTSS) is otherwise highly conversed across eukaryote lineages. The lysine residue at canonical position 145 (K123 in yeast, K120 in human) can be post-translationally modified with the addition of a ubiquitin moiety (Emre and Berger, 2004; Soares and Buratowski, 2013). This highly conserved histone modification has been shown to physically alter local chromatin structure, enabling a more open and accessible state (Fierz *et al.*, 2011). Accordingly, H2B ubiquitylation (H2Bub) often covers the bodies of highly expressed genes (Roudier *et al.*, 2011).

With the presence of an asparagine rather than lysine in H2B.8, it is highly likely that this histone variant cannot be ubiquitylated (Fig. 2.5). If this is the case, genomic loci with a high density of H2B.8 deposition would be immune to chromatin relaxation by H2Bub and exhibit lower gene expression.

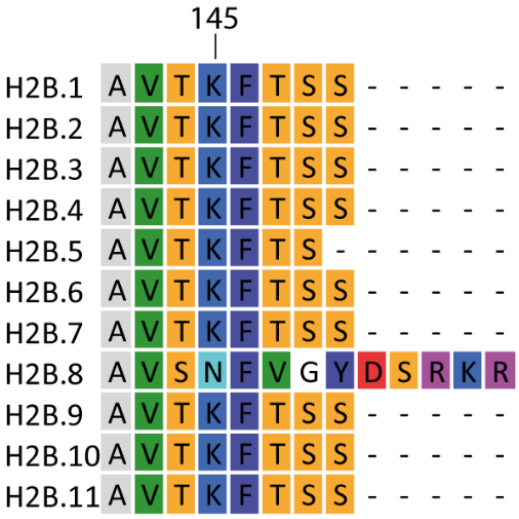


Figure 2.5: C-terminal motif alignment of *Arabidopsis* H2B variants. Alignment of *Arabidopsis* H2B variant C-terminal motifs. Canonical variants are highly conserved across this motif, whereas H2B.8 exhibits a high degree of divergence. Position 145 in canonical variants is a lysine residue, the equivalent residue (234) in H2B.8 is an asparagine.

This work seeks to establish whether there is a role for H2B.8 in sperm chromatin condensation, and the mechanism by which this occurs. This section presented three distinct features specific to H2B.8 compared to canonical H2Bs that could lead to chromatin compaction. These were a large IDR in the N-terminal tail (Section 2.3.1), an increased number of arginine residues at histone-DNA contacts (Section 2.3.2) and a K-to-N substitution at position 145 (Section 2.3.3). Considering the unique features of H2B.8 and its presence in sperm chromatin, the next section asks whether the histone variant is important for fertility.

2.4 Male fertility is impaired in *htb8* mutant

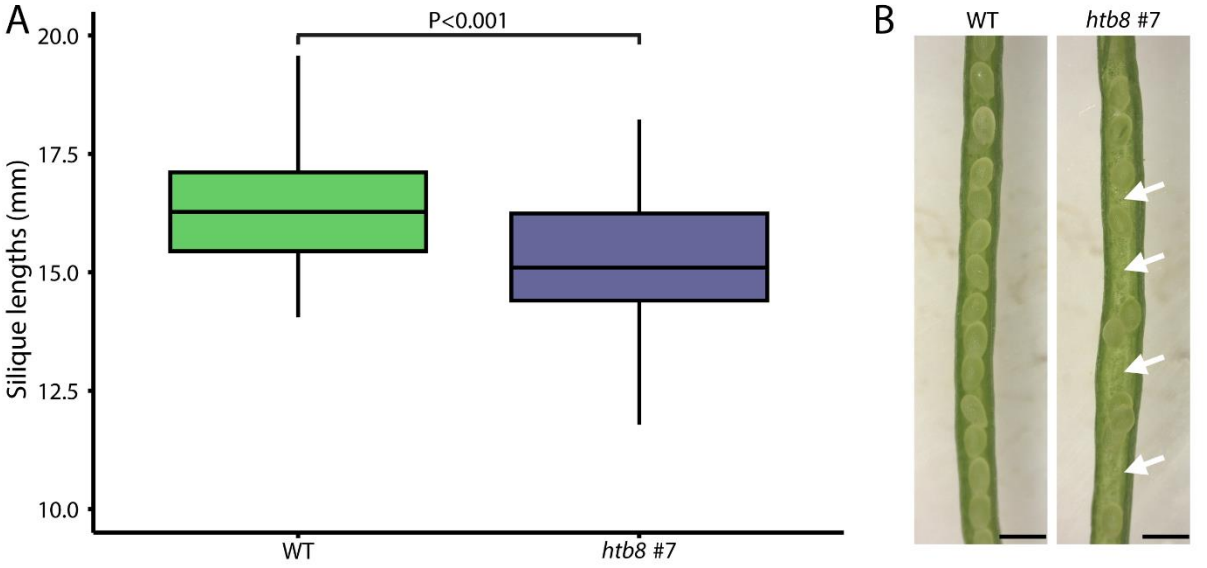
Given the exciting features of H2B.8 that could be involved in chromatin compaction in sperm, the effect of *HTB8* mutation on male fertility was examined.

T-DNA insertion lines were not found in the CDS of the *HTB8* locus, although two lines were identified with insertions < 1 kb upstream (SALKseq_057361 and Wiscseq_DsLox288D07), likely in the promoter. Assessment of these T-DNA lines by qRT-PCR revealed only a minor loss of *HTB8* expression relative to WT (~20% decrease). As such, T-DNA lines were deemed unsuitable for further analysis.

Null mutants were generated using the CRISPR-Cas9 system. Four guide RNAs targeted the 5' end of the *HTB8* locus and successful mutations were screened by Sanger sequencing. Two independent CRISPR mutants (*htb8* #7 and *htb8* #3) were identified with small deletions that incurred frame shifts leading to premature stop codons and ultimately truncated proteins (Supp. Fig. 2.1). The line, *htb8* #7 (hereafter as *htb8*, unless stated), was used for further experiments.

Initial assessment of the requirement of H2B.8 for male fertility considered silique lengths and seed abortion. Siliques are significantly shorter in the *htb8* mutant compared to WT, incurring a reduction of ~10.5% (Fig. 2.6A). Examination of seed set within siliques showed several developing embryos had been aborted in *htb8*, whilst this was rarely observed in WT (Fig. 2.6B). Collectively, loss of H2B.8 leads to a reduction in fertility although it is not essential for reproductive function.

To attain a more thorough evaluation of the contribution of H2B.8 to male fertility, a genetic test was employed. Heterozygous (*HTB8*/*htb8*) males were crossed to WT females and the segregation of alleles in the F1 progeny was determined (Fig. 2.6C). Offspring were 29.6% less likely to carry the *htb8* mutant allele than WT.



C

Maternal parent	Paternal parent	WT F1	Heterozygous F1	Ratio (WT : het)	*htb8* transmission (%)	P value
WT	*HTB8/htb8*	436	307	1 : 0.70	41.3	< 0.001

Figure 2.6: H2B.8 is required for normal male fertility in *Arabidopsis*. (**A**) Silique lengths in WT and *htb8* #7. Mutant siliques are significantly shorter than WT. (**B**) Aborted embryos are readily observed in siliques of *htb8* #7 but are rarely seen in WT. (**C**) Male transmission assay of *htb8* #7. F1 progeny are significantly less likely to inherit the mutant allele. Scale bars are 1 mm in (B). Boxplot (A) shows median (thick black bar) and first and third quartiles, with lower and upper whiskers extending to 1.5 times the interquartile range of the first and third quartiles, respectively. Statistical significance was determined in (A) by a Student's t-test; $P < 0.001$; WT N = 104; *htb8* #7 N = 123. Statistical significance was determined in (C) by a Chi-squared test; $P < 0.001$; N = 743.

Overall, phenotypic analysis has shown that H2B.8 is required for normal fertility in *Arabidopsis* (Fig. 2.6, A and B). Loss of H2B.8 incurs shorter siliques and an increase of aborted embryos. Using genetics, H2B.8 has been demonstrated to be required for male fertility (Fig. 2.6C). When the H2B.8 allele is absent, sperm are less likely to achieve fertilisation. The following section addresses whether H2B.8 impacts upon chromatin compaction in sperm.

2.5 Sperm chromatin compaction requires H2B.8

Having established the importance of H2B.8 to male fertility, it was next considered whether the histone variant has any effect on sperm chromatin compaction. Improper male fertility could be a result of incorrect chromatin packaging in sperm and the resultant detrimental effects on genome regulation or integrity.

2.5.1 Sperm nuclei are larger in *htb8* mutant

Comparing WT and *htb8* #7 sperm with confocal microscopy led to an interesting initial observation, namely that sperm without H2B.8 appear larger in size (Fig. 2.7A). Quantification of sperm nuclei from WT and *htb8* mutant pollen (lines #7 and #3) confirmed the preliminary observation. Mutant sperm are significantly larger than WT, equating to an increased area of ~38-44% (Fig. 2.7B). Independent mutant lines, *htb8* #7 and #3, are not different from one another (Fig. 2.7B). This suggests that the phenotype is caused directly by loss of H2B.8 rather than another unknown off-target mutation that would be unlikely to be shared between lines.

Further evidence that the effect is caused by H2B.8 is supplied by genetic complementation. The aforementioned GFP reporter (*pHTB8::HTB8-eGFP*) and a Myc tagged H2B.8 line (*pHTB8::HTB8-Myc*) were both able to fully complement sperm nuclear size to WT levels (Fig. 2.7B). Therefore, it is highly likely that loss of H2B.8 is directly responsible for increased sperm nuclear size.

Increased sperm nuclear size could be a direct consequence of impaired chromatin compaction in the *htb8* mutant background. The relationship between nuclear size and chromatin architecture has previously been established across eukaryotes (Macadangdang *et al.*, 2014).

As such, given the increased nuclear size in the *htb8* mutant background, it is tempting to propose that this effect is due to a failure to undergo correct chromatin condensation. The next section seeks to establish whether loss of H2B.8 affects sperm chromatin compaction.

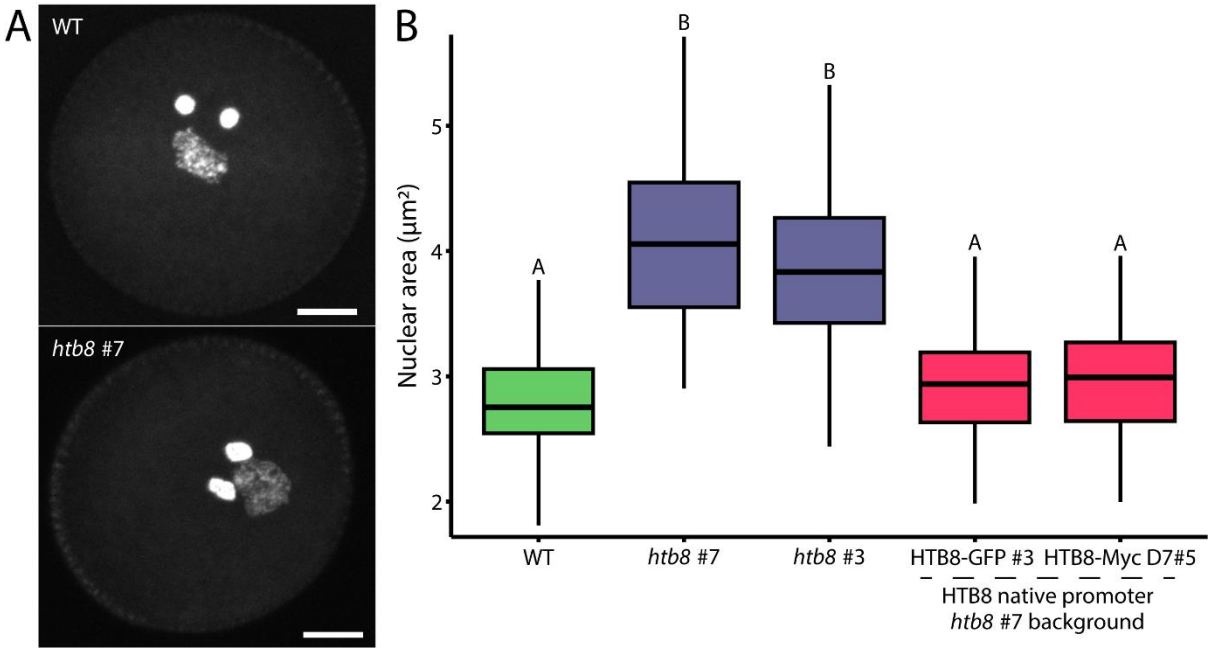


Figure 2.7: Sperm nuclei are larger in the *htb8* mutant. (**A**) Confocal images of WT and *htb8* #7 pollen stained with Hoechst. Two sperm nuclei and accompanying vegetative nucleus are observed in both. (**B**) Sperm nuclei are significantly larger in *htb8* mutant backgrounds compared to WT. The phenotype can be complemented by both eGFP and Myc tagged H2B.8 proteins, driven by the native promoter (*pHTB8*). Scale bars are 5 µm in (A). Boxplot (B) shows median (thick black bar) and first and third quartiles, with lower and upper whiskers extending to 1.5 times the interquartile range of the first and third quartiles, respectively. Statistical significance was determined in (B) by an ANOVA followed by individual Tukey tests; boxplots marked as A and B are not different within the group ($P > 0.001$) but are different between groups ($P < 0.001$); WT N = 80; *htb8* #7 N = 79; *htb8* #3 N = 77; HTB8-eGFP #3 N = 79; HTB8-Myc D7#5 N = 80.

2.5.2 Super-resolution microscopy reveals loss of distinct foci in *htb8* mutant

Chromatin differences were not readily observed using standard confocal microscopy, but this could be due to the highly dense nature of sperm chromatin and the small size of sperm nuclei rather than lack of biological difference. To get around these technical limits, super-resolution microscopy was used to observe chromatin in WT and *htb8* mutant sperm.

3D Structured Illumination Microscopy (3D-SIM) has been utilised extensively to observe nuclear dynamics in eukaryote model systems (Matsuda *et al.*, 2015; Cremer *et al.*, 2017; Ochs *et al.*, 2019). Such studies have achieved resolutions orders of magnitude beyond standard confocal techniques, thus giving greater insight into the functions of epigenetic proteins (Lakadamyali and Cosma, 2015). However, use of 3D-SIM in plants has been limited to studies of meiosis and microtubules (Vavrdová *et al.*, 2019; Morgan *et al.*, 2020). Here, 3D-SIM is employed to study chromatin structure in isolated sperm nuclei.

WT sperm nuclei are abundant with small foci throughout the nucleus, with larger foci at the nuclear periphery (Fig. 2.8A; Supp. Movie 2.1). The larger foci are assumed to be heterochromatic domains, due to their size and positioning (Simon *et al.*, 2015). However, the identity of the smaller foci within the nuclear interior is unknown. Such foci are very sharp in intensity with distinct troughs in between (Fig. 2.8A; Supp. Movie 2.1).

Examination of *htb8* sperm nuclei with 3D-SIM revealed that the small foci throughout the nucleoplasm were lost (Fig. 2.8B; Supp. Movie 2.1). Rather, chromatin is homogenous throughout the nucleus with no sharp peaks or troughs of intensity observed (Fig. 2.8B; Supp. Movie 2.1).

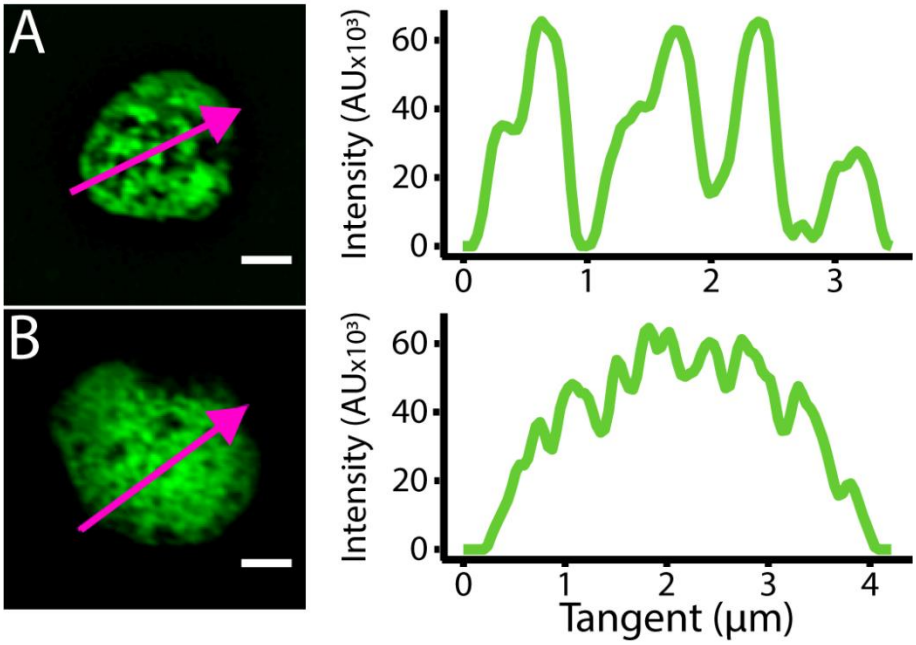


Figure 2.8: Distinct foci are lost in *htb8* mutant sperm. (**A**) 3D-SIM image of a WT sperm nucleus stained with SYBR Green DNA dye. WT sperm chromatin is characterised by small, distinct foci throughout the nucleoplasm. (**B**) *htb8* mutant sperm lose the foci characteristic of WT sperm, instead chromatin is homogenous across the nucleus. DNA is stained in the same way as (A). Scale bars are 1 µm in (A) and (B). Line plots associated with images (A) and (B) show intensity across the magenta tangent. See also Supp. Movie 2.1.

To obtain a quantitative measure of 3D sperm chromatin structural differences between WT and *htb8* mutant, the ImageJ plugin FociPicker3D was used (Du *et al.*, 2011). The FociPicker3D algorithm takes three-dimensional image data and defines foci according to absolute intensity versus background differences. Foci are defined as a minimum number of adjacent voxels with intensity exceeding a particular cut-off. Aggregate descriptive metrics for defined foci are provided in the output.

Analysis of WT and *htb8* 3D-SIM images reveals that foci in the mutant are significantly smaller in size (Fig. 2.9A). The smaller foci in *htb8* sperm are also much lower in intensity compared to those in WT (Fig. 2.9B). However, the *htb8* foci are more numerous than their WT counterparts (Fig. 2.9C). Collectively, this equates to total intensity being the same between WT and *htb8* (Fig. 2.9D). These descriptors show that chromatin structure is very different upon loss of H2B.8, whilst the overall chromatin volume is maintained.

These descriptors indicate that small, intense, and sharp foci observed in WT fail to form in the absence of H2B.8 (Fig. 2.9, A to D). Instead, analysis of the *htb8* mutant finds an increased number of smaller and less intense foci (Fig. 2.9, A to D). Altogether, these metrics agree with the previous qualitative observations (Fig. 2.8, A and B).

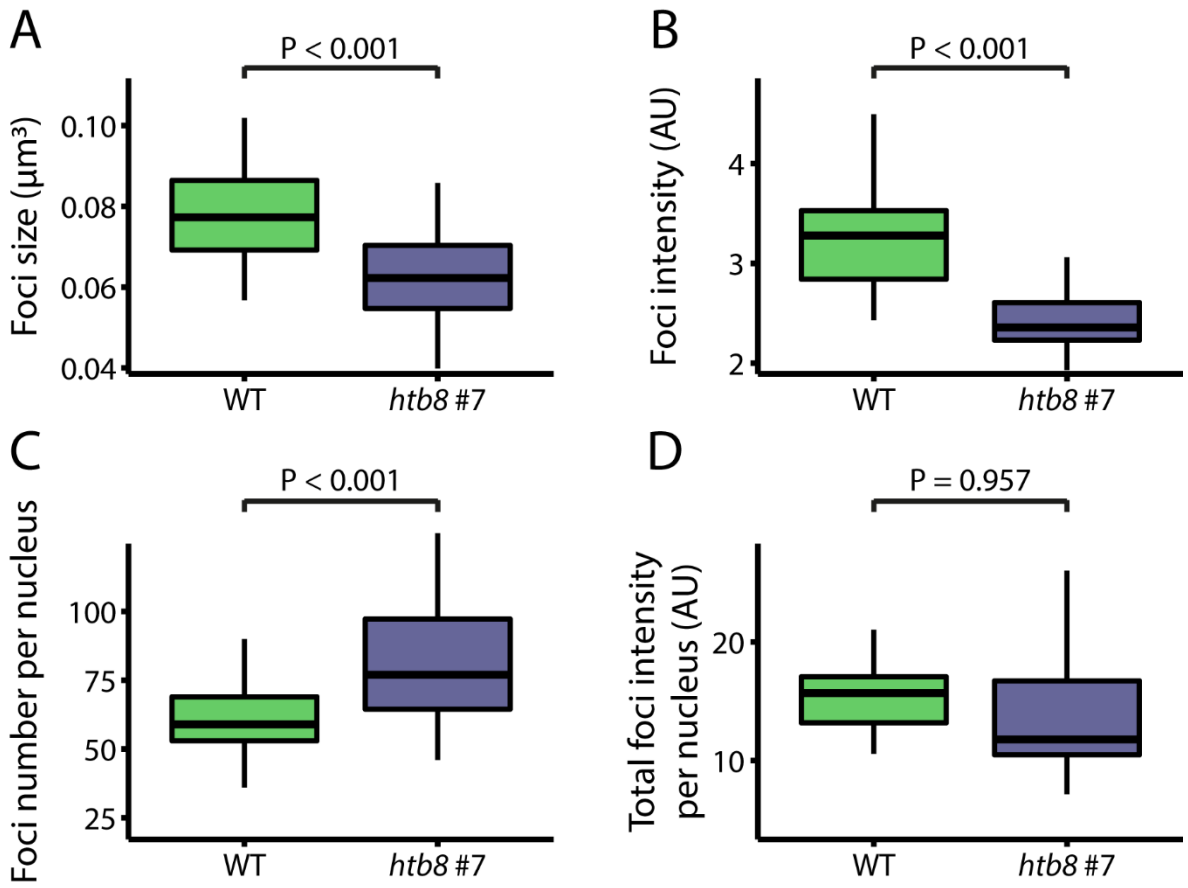


Figure 2.9: Foci are smaller, have lower intensity and are more numerous in *htb8* mutant sperm. (**A**) Foci defined in *htb8* mutant sperm are significantly smaller than those in WT. (**B**) The foci of *htb8* mutant sperm have significantly lower intensities than WT foci. (**C**) *htb8* mutant sperm nuclei have significantly more foci than WT sperm nuclei. (**D**) The total foci intensity per nucleus is not different between WT and *htb8* mutant sperm. Boxplots (A to D) show median (thick black bar) and first and third quartiles, with lower and upper whiskers extending to 1.5 times the interquartile range of the first and third quartiles, respectively. Statistical significance was determined in (A to D) by a Student's t-test; *P* values on figure; WT N = 49 nuclei; *htb8* #7 N = 48 nuclei.

In conclusion, this section has revealed a role for H2B.8 in defining sperm chromatin structure. H2B.8 is required for small sperm nuclear size; its loss incurs an increased size that is associated with decondensed chromatin. Super-resolution microscopy enabled the observation of small distinct foci in WT sperm. Such foci are lost in the *htb8* mutant, where chromatin becomes more homogenous throughout the nucleoplasm. It is likely that loss of these foci contributes to the increased sperm nuclear size in *htb8*. Further sections focus upon the small foci and consider the contribution of H2B.8 to their formation.

2.6 Euchromatic foci are specific to sperm and colocalise with H2B.8

Super-resolution microscopy of WT sperm nuclei revealed the presence of small, distinct foci throughout the nuclear interior. Such foci were absent in the *htb8* mutant which suggests that H2B.8 is critical for their formation. This section explores the uniqueness of the foci, their underlying composition and whether they are a sperm-specific feature of chromatin. Finally, it addresses whether H2B.8 could be involved directly in the formation of small, distinct foci in sperm.

2.6.1 Euchromatic foci are a unique feature to sperm chromatin

Given the unique nature of the small, distinct foci observed by 3D-SIM in the centre of sperm nuclei. It was next asked whether this was a general feature of *Arabidopsis* nuclei or unique to their sperm. To do so, nuclei from a range of *Arabidopsis* cell types were subjected to 3D-SIM. Furthermore, immunostaining against the characteristic heterochromatic mark H3K9me2 was used to assess the composition of chromatin underlying the foci.

The presence of small, distinct foci was confirmed in sperm nuclei (Fig. 2.10A; Supp. Movie 2.2). Larger foci exist at the nuclear periphery; as mentioned earlier, such features are indicative of heterochromatic domains. H3K9me2 signals colocalise with these large, peripheral foci (Fig. 2.10A; Supp. Movie 2.2); thus, confirming that they are heterochromatic in nature. This concurs with previous studies (Schoft *et al.*, 2009), suggesting that heterochromatin domains remain intact in sperm nuclei.

The absence of H3K9me2 from the small, distinct foci throughout the nucleoplasm implies that they are not heterochromatic (Fig. 2.10A; Supp. Movie 2.2). Euchromatic foci at this scale have never before been observed, suggesting an entirely novel chromatin architecture exists in *Arabidopsis* sperm nuclei.

Super-resolution imaging of vegetative nuclei did not identify euchromatic foci (Fig. 2.10B; Supp. Movie 2.2). Rather, chromatin is homogenous in appearance throughout the nucleus. In agreement with previous studies (Schoft *et al.*, 2009), heterochromatin domains were not found in the vegetative nucleus (Fig. 2.10B; Supp. Movie 2.2). Therefore, despite undergoing disassembly of heterochromatic foci and existing within pollen alongside sperm nuclei, vegetative nuclei do not have the distinctive euchromatic foci.

Somatic nuclei were also examined for the presence of euchromatic foci. Nuclei of leaf, roots and seedlings are characterised by large heterochromatic domains at the nuclear periphery along with the presence of the nucleolus in the nuclear interior (Fig. 2.10C; Supp. Movie 2.2; data shown for leaf). However, in all instances, euchromatic foci were not observed in somatic nuclei (Fig. 2.10C; Supp. Movie 2.2). Therefore, it can be inferred that euchromatic foci are specific to sperm in *Arabidopsis*.

Given the presence of H2B.8 in sperm, it is tempting to suggest that the unique histone variant could be implicit in the formation of euchromatic foci. Mature seeds also have H2B.8 expression, it would be interesting to explore whether nuclei at this developmental stage have similar euchromatic foci to sperm. However, examination of this cell type is beyond the scope of the current study.

This section has shown that heterochromatin domains persist in sperm rather than undergo reprogramming (Fig. 2.10A; Supp. Movie 2.2). Therefore, the global chromatin compaction and accompanying small nuclear size of sperm is unlikely to be caused by remodelling of heterochromatin regions. This suggests that the specific formation of small euchromatic foci could be implicit in global chromatin compaction.

Interestingly, analysis of WT and *htb8* mutant sperm nuclear size (Section 2.5.1) and chromatin structure (Section 2.5.2) evidenced a crucial role for H2B.8. Given that foci are lost from sperm in the *htb8* mutant and only occur in sperm itself, it is highly likely that H2B.8 is involved in their formation. However, it is unknown whether this is a direct or indirect effect.

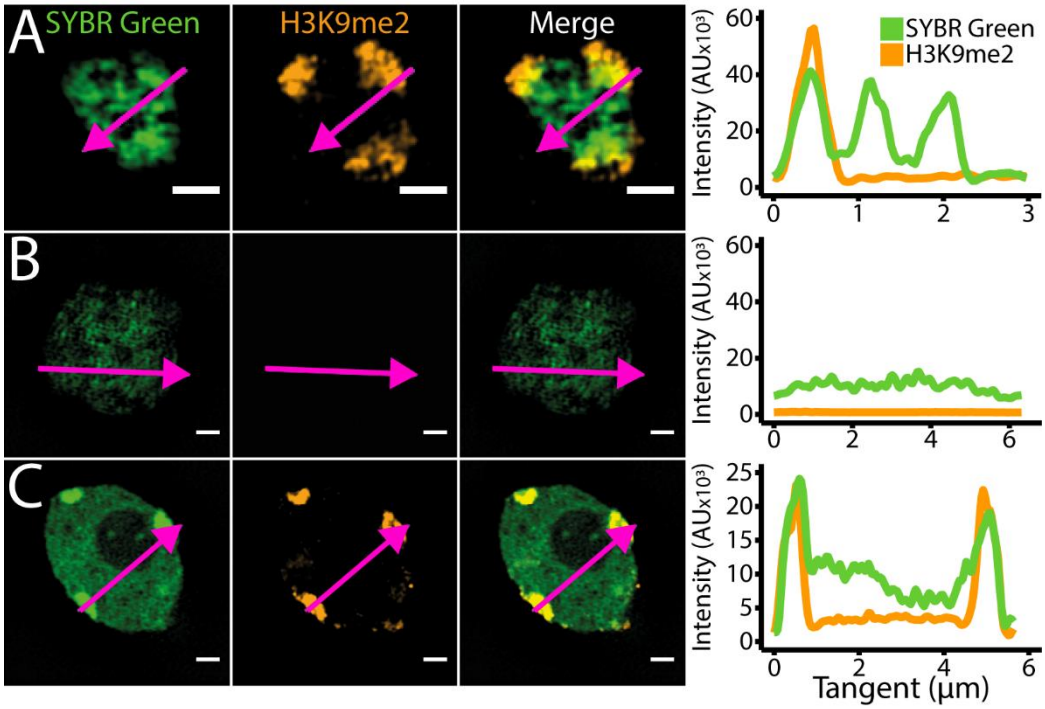


Figure 2.10: Small, euchromatic foci are a specific feature of sperm chromatin. (**A**) 3D-SIM image and associated line plot of a sperm nucleus with DNA stained by SYBR Green (green) and immunostained for the heterochromatic mark, H3K9me2 (orange). Sperm is characterised by small euchromatic foci throughout the nucleoplasm and large heterochromatic domains at the nuclear periphery. (**B**) Vegetative nuclei do not have heterochromatic domains or euchromatin foci; instead, chromatin is homogenous throughout the nucleoplasm. Staining is the same as (A). (**C**) Somatic nuclei, such as leaf, have defined heterochromatin domains at the nuclear periphery. Euchromatin is homogenous across the nucleus. Staining is the same as (A). Scale bars are 1 µm in (A to C). Line plots associated with images (A to C) show intensity for each fluorophore across the magenta tangent. See also Supp. Movie 2.2.

2.6.2 Euchromatic foci colocalise with H2B.8

Having established the unique presence of small euchromatin foci in sperm and their apparent reliance upon H2B.8, it was next considered whether the two directly colocalise.

The GFP reporter line (*pHTB8::HTB8-eGFP*) was imaged using super-resolution microscopy (Fig. 2.11; Supp. Movie 2.3). GFP signals denoting H2B.8 formed numerous small foci throughout the nucleoplasm, the majority of which colocalise with DAPI-stained foci in the nuclear interior (Fig. 2.11; Supp. Movie 2.3). A small number of H2B.8 foci colocalise with heterochromatin domains, marked by H3K9me2 (Fig. 2.11; Supp. Movie 2.3).

Collectively, it can be concluded that H2B.8 and small euchromatic foci colocalise with one another. Coupled with the evidence that euchromatic foci are lost in the *htb8* mutant and that they are uniquely present in sperm, it is clear that H2B.8 plays an essential role in the formation or maintenance of these puncta. The following sections examine the mechanism by which H2B.8 is involved in the establishment and maintenance of euchromatic foci.

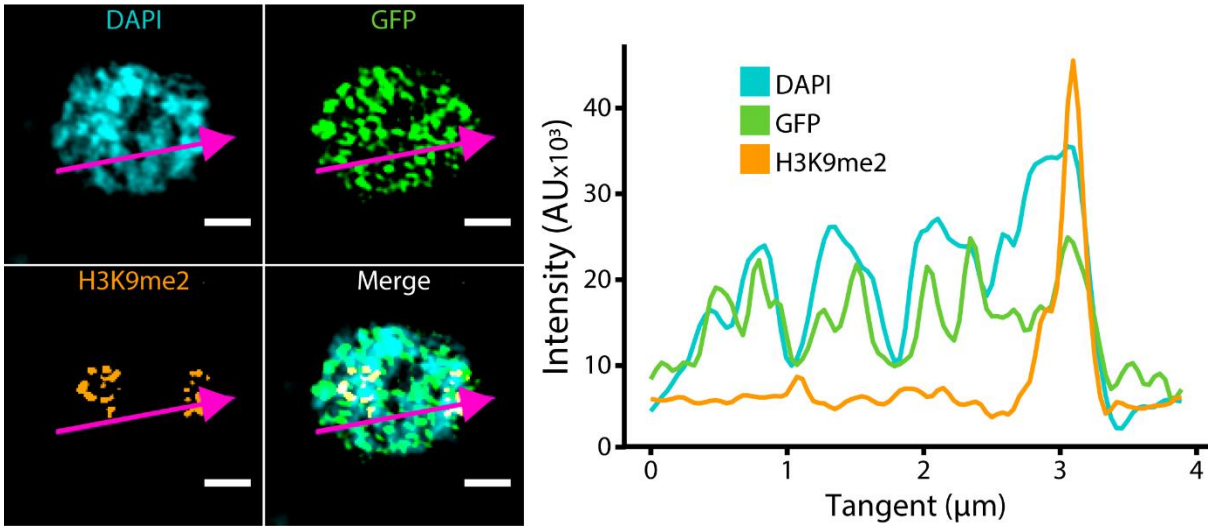


Figure 2.11: Euchromatic foci and H2B.8 foci colocalise in sperm. Super-resolution imaging of *pHTB8::HTB8-eGFP htb8* sperm reveals the presence of H2B.8 foci (green). Heterochromatic foci, demarcated by H3K9me2 (orange) and strong DAPI signals (blue), largely do not colocalise with H2B.8 foci. However, small euchromatic foci, indicated by weak DAPI peaks, and H2B.8 foci perfectly colocalise with one another. Scale bars are 1 µm. Line plots associated with image shows intensity for each fluorophore across the pink arrow. See also Supp. Movie 2.3.

2.7 H2B.8-mediated euchromatic foci are formed by phase separation

Biomolecular condensates formed by phase separation is an area of emerging interest in the field of chromatin biology (Erdel and Rippe, 2018). The presence of small euchromatic foci in sperm nuclei that colocalise with H2B.8 and are lost in the *htb8* mutant (Section 2.6), paired with the unique presence of an IDR in the N-terminal tail of H2B.8 (Section 2.3.1), suggests the exciting possibility that these foci may be formed by phase separation. Employing a range of *in vitro* and *in vivo* techniques, this section tests the hypothesis that H2B.8 mediates formation of euchromatic foci by phase separation.

2.7.1 H2B.8 forms IDR-dependent condensates *in vitro*

To test whether H2B.8 is able to undergo liquid-liquid phase separation under physiological conditions, the native protein was compared with a canonical H2B (H2B.2) and a recombinant H2B.8 with the IDR deleted (H2B.8ΔIDR) (Supp. Fig. 2.2). This section compares the three proteins abilities to phase separate *in vitro*.

H2B.8 forms condensates in the presence of DNA under physiological salt conditions (Fig. 2.12A). These puncta form at both low protein and DNA concentrations compared to previously characterised phase separating chromatin proteins, such as HP1 (Larson *et al.*, 2017; Strom *et al.*, 2017), indicating a high capability of H2B.8 to phase separate *in vitro* (Fig. 2.12A).

The ability of H2B.8 to phase separate is dependent on the N-terminal tail IDR. H2B.8ΔIDR fails to form condensates *in vitro*, irrespective of protein or DNA concentration (Fig. 2.12B). In fact, H2B.8ΔIDR poorly colocalises with DNA, demonstrated by the homogenous low levels of DAPI staining in solution that do not correspond to protein. This suggests a reduced binding affinity between the two. As such, it can be determined that the IDR is critical to H2B.8 formation.

As a control, H2B.2 was tested for the ability to phase separate *in vitro*. H2B.2 does not form puncta in the presence of DNA; rather, fibre-like precipitates are formed (Fig. 2.12C). Therefore, H2B.2 can associate with DNA but does not undergo phase separation *in vitro*.

To conclude, purified H2B.8 and H2B.2 protein can associate with DNA. However, only H2B.8, and not H2B.2, can undergo phase separation under physiological conditions at low protein and DNA concentrations *in vitro* (Fig. 2.12, A and C). The ability of H2B.8 to associate with DNA and ultimately form condensates is dependent on the N-terminal tail IDR (Fig. 2.12B). Therefore, H2B.8 forms IDR-dependent phase separated condensates *in vitro*.

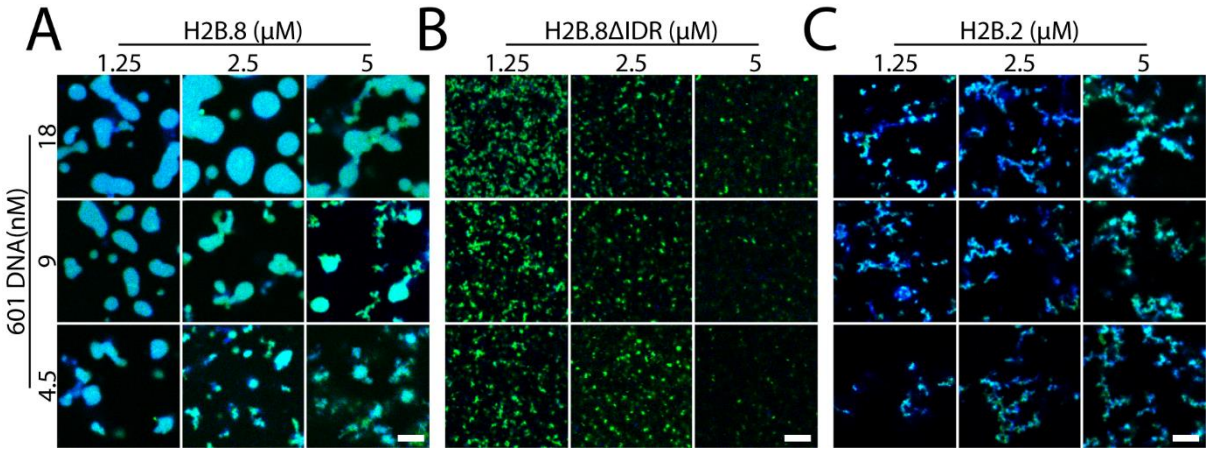


Figure 2.12: H2B.8 forms condensates in an IDR-dependent manner *in vitro*. (**A**) H2B.8 can form phase separated condensates with DNA under physiological salt conditions *in vitro*. Condensates are formed at low concentrations of protein and DNA, indicating a high propensity to phase separate. (**B**) H2B.8 without the N-terminal tail IDR (H2B.8ΔIDR) fails to form condensates under the same conditions *in vitro*, indicating a crucial role for the IDR in phase separation. (**C**) A representative canonical H2B variant, H2B.2, also fails to undergo phase separation with the same conditions *in vitro*. Images in (A to C) are merged of proteins labelled green and DNA labelled blue. Scale bars are 5 μm in (A to C).

2.7.2 Transient expression can form IDR-dependent foci *in vivo*

Next, the ability of H2B.8 to form condensates was tested *in vivo*, using *Agrobacterium*-mediated transient expression of tobacco (*Nicotiana benthamiana*) leaves. Unlike *Arabidopsis,* leaf nuclei of tobacco do not have distinct heterochromatic domains that stain brightly with DAPI. Instead, chromatin appears uniform throughout the nucleoplasm. As such, any foci formed and visible by DAPI staining are owing to the transient expression of the protein of interest.

H2B.8 (*p35S::HTB8-YFP*) expression is sufficient to form chromatin-associated condensates *in vivo* (Fig. 13A). Foci are small in size and abundant throughout the nucleus. This experiment shows that H2B.8 alone is able to compact chromatin and form condensates in plant nuclei; it is not reliant upon another unknown *Arabidopsis* sperm-specific factor to achieve function.

In agreement with *in vitro* data, the H2B.8 N-terminal tail IDR is critical for protein function. Transient expression of H2B.8ΔIDR (*p35S::HTB8ΔIDR-YFP*) fails to form foci *in vivo* (Fig. 2.13B). Instead, a large aggregate is observed, surrounding and within the nucleolus. Such a feature is indicative of protein misfolding and subsequent shuttling to the nucleolus for degradation. Therefore, H2B.8ΔIDR is likely to be unstable in tobacco cells.

Similar to its behaviour *in vitro*, transiently expressed H2B.2 (*p35S::HTB2-YFP*) associates with DNA but fails to form foci *in vivo* (Fig. 2.13C). H2B.2 protein is evenly distributed across chromatin, corresponding to DAPI signals. This experiment accounts for the potential effects of using the 35S promoter (*p35S*), which is very strong in plants. The foci formed by H2B.8 are not due to overexpression of protein, as this does not occur with H2B.2 (Fig. 2.13C).

Altogether, H2B.8 can condense chromatin *in vivo* via the formation of small foci, likely by phase separation (Fig. 2.13A). The IDR is required for H2B.8 function in tobacco, as H2B.8ΔIDR does not form discrete puncta but instead a single aggregate, likely to be a misfolding product (Fig. 2.13B). A representative canonical H2B, H2B.2, cannot form foci *in vivo* indicating the ability is specific to H2B.8 (Fig. 2.13C).

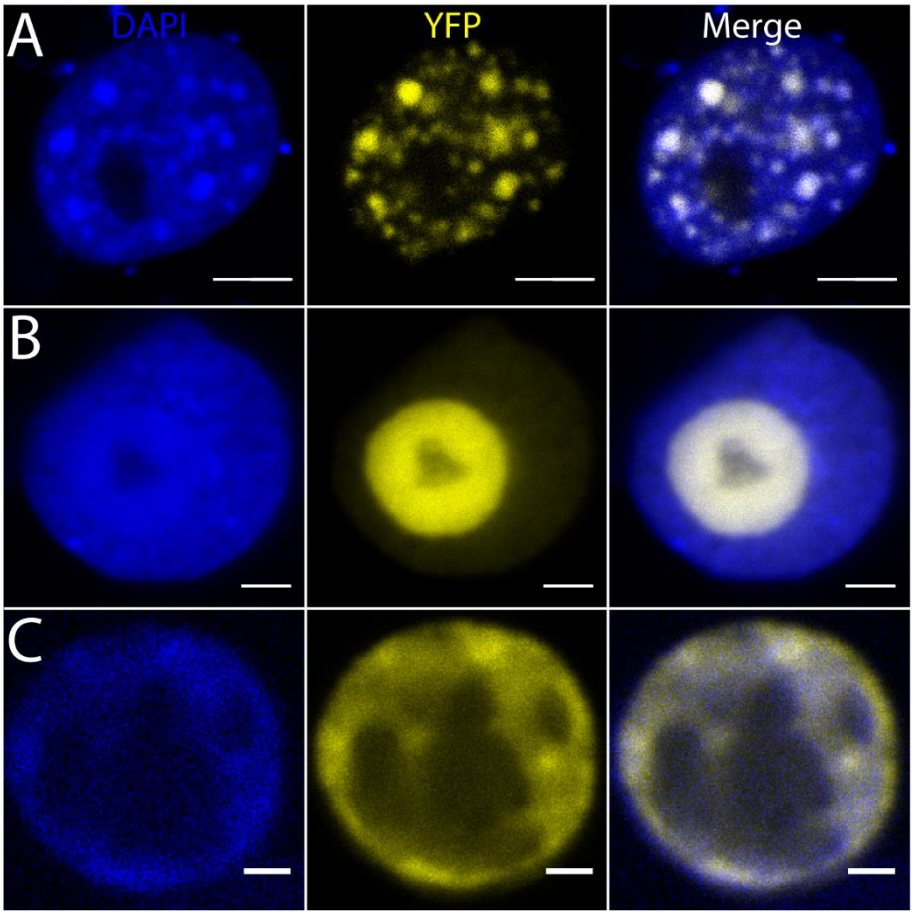


Figure 2.13: IDR-dependent foci are formed *in vivo*. (**A**) Transient expression of H2B.8 (*p35S::HTB8-YFP*; yellow) in *Nicotiana benthamiana* leaves stained with DAPI (blue). H2B.8 forms numerous small, chromatin-associated foci throughout the nucleus *in vivo*. (**B**) Formation of small foci is IDR-dependent, as H2B.8ΔIDR (*p35S::HTB8ΔIDR-YFP*) does not replicate H2B.8. Rather, a large aggregate is observed at the nucleolus, indicative of incorrect protein folding. (**C**) H2B.2 (*p35S::HTB2-YFP*), a representative canonical H2B, distributes evenly across chromatin and fails to form foci *in vivo*. Scale bars are 2 µm in (A to C).

2.7.3 H2B.8-mediated euchromatic foci can be reconstituted in soma

Having established that H2B.8 can form IDR-dependent foci *in vitro* under physiological salt conditions and *in vivo* under transient expression in tobacco, stable construction of such foci in *Arabidopsis* somatic tissues was attempted. Ectopic expression in soma would enable probing of the composition of H2B.8 foci and how they compare to the native foci in sperm.

Ectopic H2B.8 expression (*p35S::HTB8-eGFP*) in *Arabidopsis* forms foci (Fig. 2.14A; Supp. Movie 2.4), comparable to those observed in tobacco (Fig. 2.13A). Crucially, H2B.8-mediated foci in soma display similar characteristics to those observed in sperm (Fig. 2.11; Supp. Movie 2.3). Meaning, H2B.8 foci are small and distinct throughout the nucleoplasm and colocalise with small peaks of DAPI signal. H2B.8 foci in soma are also largely independent of heterochromatin domains, marked by H3K9me2 (Fig. 2.14A; Supp. Movie 2.4). Therefore H2B.8 forms small, euchromatic foci upon ectopic expression in *Arabidopsis* somatic tissues.

Despite vast developmental differences between sperm and soma, it is interesting to note that H2B.8 behaves in the same way. Again, as with transient expression in tobacco, this indicates that H2B.8 is sufficient to induce the formation of small euchromatic foci, rather than relying upon another unknown protein specific to sperm.

Interestingly, the localisation of H2B.8ΔIDR (*p35S::HTB8ΔIDR-YFP*) is vastly different between transient expression in tobacco (Fig. 2.13B) and stable ectopic expression in *Arabidopsis* (Fig. 2.14B; Supp. Movie 2.4). Unlike in tobacco, H2B.8ΔIDR is not localised at the nucleolus, indicating that correct protein folding has been achieved.

Remarkably, H2B.8ΔIDR has a near-inverse localisation to H2B.8 in somatic tissues (Fig. 2.14, A and B; Supp. Movie 2.4). H2B.8ΔIDR signals are largely deplete from euchromatic regions and instead colocalise with H3K9me2-enriched heterochromatin domains (Fig. 2.14B; Supp. Movie 2.4). Again, this validates a key role for the IDR in formation of biomolecular condensates. However, most strikingly, ectopic expression of H2B.8ΔIDR has demonstrated the requirement for the N-terminal tail IDR in correct deposition of H2B.8 to euchromatin.

How the global positioning entirely reverses according to the presence of the IDR in H2B.8 is difficult to explain. Logically, deletion of the IDR should cause the protein to act like a canonical H2B, with deposition to both euchromatin and heterochromatin, given that H2B.8ΔIDR and H2B.2 N-terminal tails are very similar (Supp. Fig. 2.2). Accordingly, this deposition profile is observed for H2B.2 (*p35S::HTB2-YFP*) when overexpressed in somatic tissues (Fig. 2.14C; Supp. Movie 2.4). H2B.2 signals perfectly correspond to DAPI intensity throughout the nucleus, with greater levels at heterochromatin regions (Fig. 2.14C; Supp. Movie 2.4).

Collectively, this indicates that H2B.8 localisation to euchromatin is dependent on the N-terminal tail IDR. Without the IDR, H2B.8 localises exclusively to heterochromatin despite the tail mimicking that of a canonical variant (Supp. Fig. 2.2). Therefore, the deposition of H2B.8ΔIDR to heterochromatin is likely due to features specific to the histone body domain; including the increased surface arginine residues (Section 3.2) and K-to-N substitution in the C-terminal motif (Section 3.3). Recently, arginine residues have been shown to provide greater viscosity to phase separated droplets compared to lysine (Fisher and Elbaum-Garfinkle, 2020). Such a biophysical property, specific to H2B.8 among H2B variants, could impact upon its nuclear distribution. Deposition of H2B.8 and contributions of specific features is addressed in greater detail in further sections.

Altogether, sections 2.7.1 to 2.7.3 have shown that H2B.8 is able to form condensates both *in vitro* and *in vivo*, either under transient or stable expression. Given the well-established roles of IDRs for phase separation (Uversky, 2017), it was considered whether the N-terminal tail IDR of H2B.8 was required for condensate formation. Indeed, H2B.8ΔIDR cannot form condensates *in vitro* and fails to form euchromatic foci *in vivo*. Therefore, the IDR is critical for the ability of H2B.8 to phase separate. Among *Arabidopsis* H2B variants, the high propensity to phase separate is likely specific to H2B.8, as a representative canonical variant, H2B.2, cannot form condensates *in vivo* or *in vitro*. Having established the unique ability of H2B.8 to phase separate, the biophysical properties of condensates were then probed.

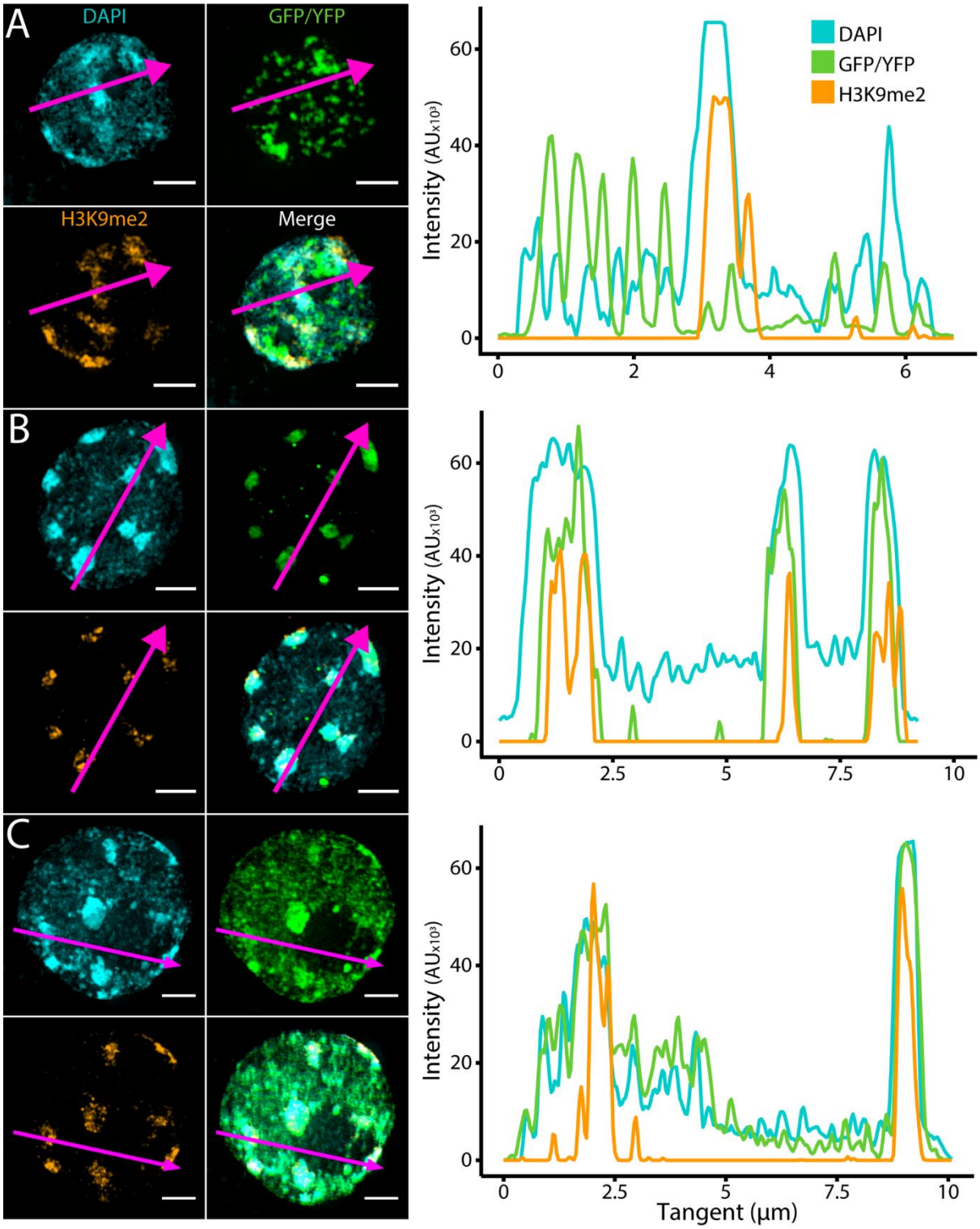

A
DAPI
GFP/YFP
H3K9me2
Merge
B
C
DAPI
GFP/YFP
H3K9me2
Intensity (AUx10³)
Tangent (µm)
0
20
40
60
0
2
4
6
2.5
5
7.5
10

Figure 2.14: H2B.8 forms stable euchromatic foci in somatic nuclei. (**A**) 3D-SIM of seedling nuclei ectopically expressing H2B.8 (*p35S::HTB8-eGFP*; green). H2B.8 forms foci throughout the nucleoplasm that largely colocalise with small DAPI-stained (blue) aggregates. Reconstituted H2B.8 foci are mostly independent of H3K9me2 (orange), indicating that they are euchromatic in composition. (**B**) Ectopic H2B.8ΔIDR (*p35S::HTB8ΔIDR-YFP*; green) colocalises with heterochromatic domains, deeply stained by DAPI, and is absent from euchromatic regions. (**C**) Expression of a representative canonical H2B variant, H2B.2 (*p35S::HTB2-YFP*; green), perfectly matches chromatin distribution, denoted by DAPI staining. This indicates no preference for heterochromatin or euchromatin. Scale bars are 2 µm in (A to C). Line plots associated with image shows intensity for each fluorophore across the magenta tangent. See also Supp. Movie 2.4.

2.7.4 H2B.8 condensates exhibit gel-like properties

Biophysical properties of H2B.8 puncta were characterised *in vitro*. Fluorescence recovery after photobleaching (FRAP) is often used as a measure of condensate mobility in phase separation studies (McSwiggen *et al.*, 2019). *In vitro*, condensates formed by H2B.8 in the presence of DNA under physiological salt conditions cannot recover to any extent following photobleaching (Fig. 2.15A; Supp. Movie 2.5). Such inability to recover indicates poor mobility dynamics, meaning condensates exhibit gel-like, rather than liquid-like properties.

As another measure of *in vitro* H2B.8 phase separation mobility, observations of puncta were made across a period of time. H2B.8 condensates fail to fully fuse upon contact, even over extended time frames (Fig. 2.15B; Supp. Movie 2.6). Were condensates liquid in nature, they would instantly fully fuse with one another (McSwiggen *et al.*, 2019). However, H2B.8 condensates maintain shape and only partially merge upon contact. This behaviour is another indication that H2B.8 condensates are gel-like *in vitro*.

Next, H2B.8 focus behaviour was assessed *in vivo*. Owing to the larger size of somatic nuclei compared to sperm, this analysis made use of the ectopic expression H2B.8 line (*p35S::HTB8-eGFP*). FRAP of H2B.8 foci in root nuclei showed similar behaviour to that observed *in vitro* (Fig. 2.15C; Supp. Movie 2.7). Foci failed to recover to any extent following photobleaching, indicative of gel-like condensates and agreeing with *in vitro* observations.

Collectively, biophysical analysis of H2B.8-mediated phase separation *in vivo* and *in vitro* has suggested that condensates are gel-like in nature. This is in contrast to the majority of phase transitions previously observed in nuclear dynamics, whereby condensates are liquid-like (Erdel and Rippe, 2018). However, protamines, causal of sperm chromatin condensation in most eukaryote lineages, have recently been shown to form gel-like condensates (Gou *et al.*, 2020). As such, an interesting parallel exists between protamines and H2B.8 function.

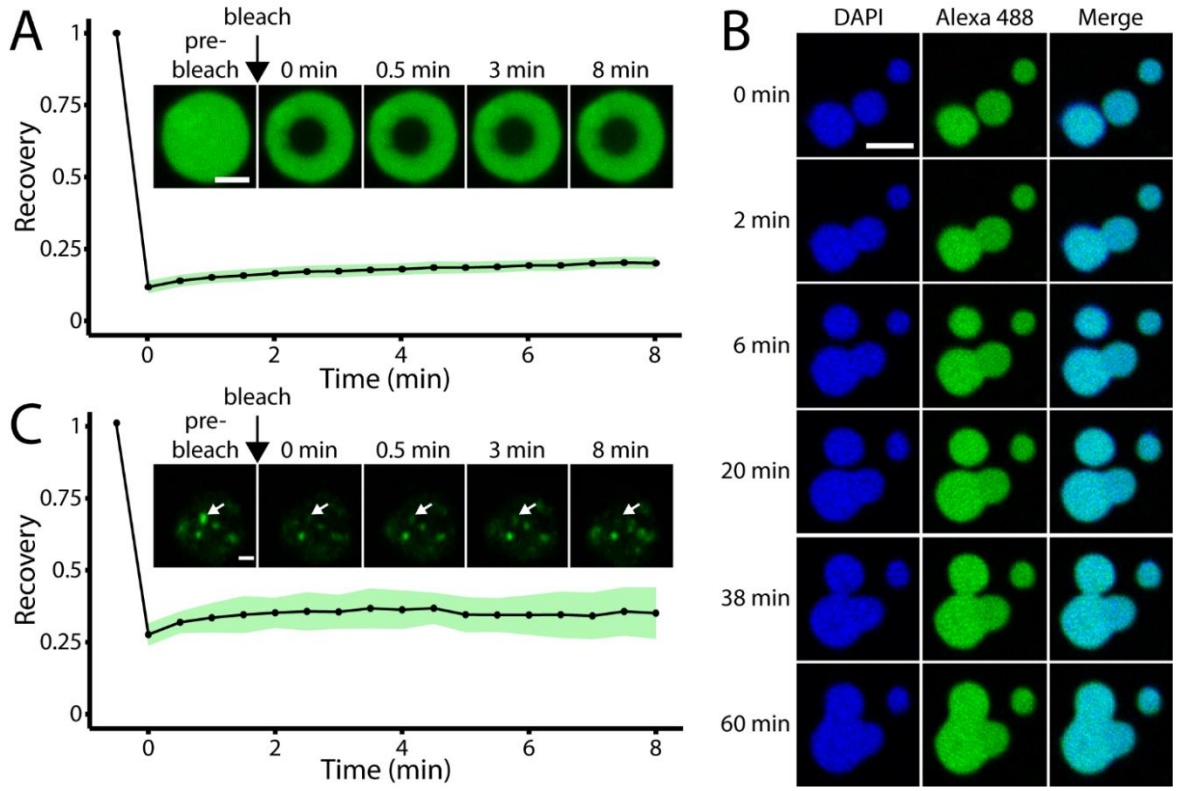


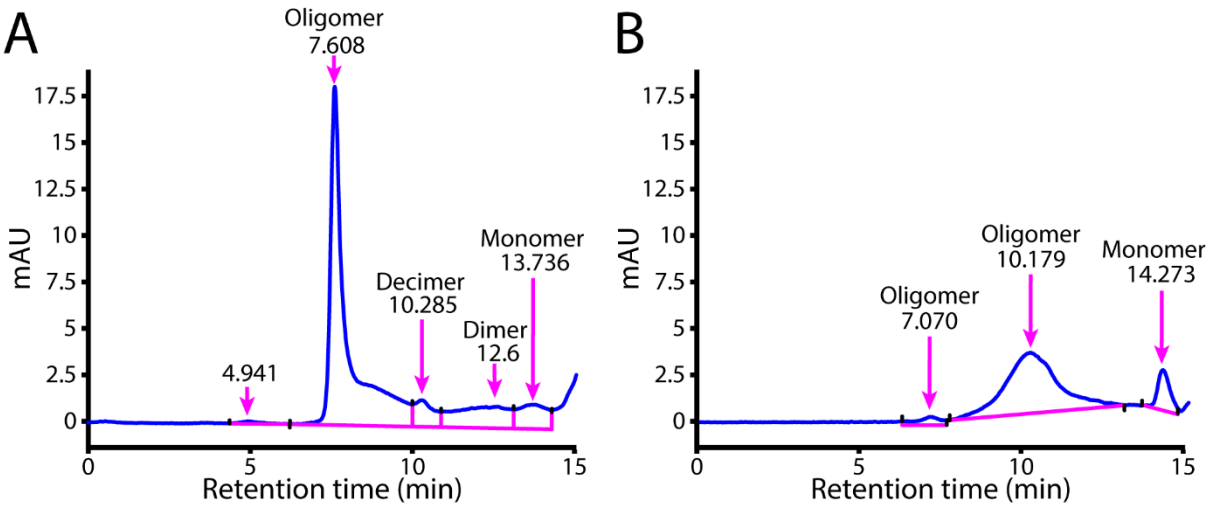
Figure 2.15: H2B.8-mediated foci exhibit gel-like behaviour. (**A**) Fluorescence recovery after photobleaching (FRAP) trace of H2B.8 condensates (Alexa 488; green) *in vitro*. H2B.8 condensates fail to recover following photobleaching, indicating gel-like properties *in vitro*. (**B**) Confocal images of H2B.8 (Alexa 488; green) plus DNA (DAPI; blue) phase separated puncta *in vitro* across 60 mins. Puncta do not fully fuse upon contact, suggesting gel-like behaviour. (**C**) FRAP trace of H2B.8 foci (GFP; green) in root nuclei ectopically expressing H2B.8 (*p35S::HTB8-eGFP*). Foci fail to recover following photobleaching, similar to *in vitro* data. Scale bars are 2 µm in (A and C) and 5 µm in (B). Line plots in (A) and (C) show mean proportional recovery relative to pre-bleach intensity (black dots). Green shaded area represents the standard deviation. (A) N = 8 foci. (C) N = 11 foci. See also Supp. Movies 2.5 – 7.

2.7.5 H2B.8 purifies as high-order aggregates

Having established that H2B.8-mediated foci are gel-like phase separated condensates; it was next considered how such aggregation could occur.

Purification of heterologously expressed H2B.8 and H2B.2 by High Performance Liquid Chromatography (HPLC) suggests that both proteins preferentially form oligomers (Fig. 2.16, A to C). However, H2B.8 oligomerises in complexes of more than 200 proteins (Fig. 2.16, A and C); whereas H2B.2 forms oligomers of ~20 proteins (Fig. 2.16, B and C). Therefore, H2B.8 has a propensity to oligomerise an order of magnitude greater than a representative canonical H2B variant.

How H2B.8 oligomerisation is achieved and how this relates to condensate formation remains unknown. However, HPLC data suggests that H2B.8 has a strong ability to bind to fellow proteins of the same type (Fig. 2.16, A and C). This ability to oligomerise is likely due to the IDR domain, if we consider its requirement for condensate formation (Sections 2.7.1 to 2.7.3).



C

Protein	Peak	MW (kDa)	Retention Time (min)	Mer
H2B.8	Calculated	28.3	NA	Monomer
	Peak1	24-32	13.7	Monomer
	Peak2	53-67	12.6	Dimer
	Peak3	270-300	10.3	Decamer
	Peak4 (main)	5500-7000	7.6	Oligomer
H2B.2	Calculated	16.8	NA	Monomer
	Peak1	14.4-19.7	14.273	Monomer
	Peak2 (main)	258-334	10.179	Oligomer
	Peak3	>7000	7.070	Oligomer

Figure 2.16: H2B.8 preferentially purifies as large oligomers. (**A**) High Performance Liquid Chromatography (HPLC) trace of H2B.8 purification. Peaks are annotated with -mer and retention time (min). H2B.8 preferentially purifies as >200-fold oligomers; displaying a propensity to form high-order aggregates. (**B**) HPLC trace, annotated as (A), of H2B.2 purification, a representative canonical H2B variant. H2B.2 principally forms ~20-fold oligomers; thus, forming aggregates an order of magnitude below H2B.8. (**C**) Oligomerisation table of data underlying HPLC traces of H2B.8 (A) and H2B.2 (B).

To conclude, this section has provided evidence for H2B.8-mediated euchromatic foci being formed by phase separation. *In vitro*, H2B.8 protein forms IDR-dependent condensates in the presence of DNA under physiological conditions (Section 2.7.1; Fig. 2.12). Similarly, *in vivo*, condensate formation by H2B.8 is dependent on the N-terminal tail IDR (Sections 2.7.2 and 2.7.3; Fig. 2.13 and 2.14). Ectopic expression in *Arabidopsis* somatic tissues also provides evidence that the IDR is critical for correct deposition to euchromatic regions (Section 2.7.3; Fig. 2.14). Biophysical analysis *in vivo* and *in vitro*, suggests that condensates are gel-like in nature and are therefore highly stable nuclear structures (Section 2.7.4; Fig. 2.15). Condensates are likely formed by the high propensity of H2B.8 to oligomerise into high-order aggregates (Section 2.7.5; Fig. 2.16). The oligomerisation property is thought to be due to the H2B.8 IDR. The contribution H2B.8-mediated IDR-dependent phase separated euchromatin foci to chromatin compaction is addressed in the following section.

2.8 The H2B.8 IDR is essential for chromatin compaction

Thus far, this work has demonstrated a role for H2B.8 in sperm chromatin compaction, likely through the formation of small, euchromatic foci by phase separation. The driver of such foci appears to be the N-terminal tail IDR. However, contributions of the increased histone body arginine residues and K-to-N substitution at position 145 remain unexplored. Furthermore, the ability of H2B.8 to affect global chromatin architecture more generally, rather than just in sperm, has not been addressed. This section seeks to assign roles to H2B.8 features in chromatin compaction and also assess whether H2B.8 can enact similarly on somatic nuclei chromatin structure as it does in sperm.

2.8.1 Ectopic H2B.8 reduces root nuclear size

A striking phenotype of *htb8* mutant sperm is their increased nuclear size. This change suggests that loss of H2B.8-mediated euchromatic foci can increase overall chromatin volume and ultimately increase the size of the nucleus. To further examine this effect, the ectopic expression line (*p35S::HTB8-eGFP*) was assessed for the effect of H2B.8 on somatic nuclear size.

As previously discussed, ectopic H2B.8 can induce the formation of euchromatic foci throughout the nucleoplasm (Section 2.7.3) in somatic nuclei. Therefore, H2B.8 is able to function in a comparable way at the local level, irrespective of cell type. Quantification of vascular cylinder cells in the elongation zone of the root tip provides evidence that H2B.8 expression is able to affect global chromatin organisation. Nuclei with H2B.8 are significantly smaller than WT (Fig. 2.17), equating to a decrease in area of 22.4%. Therefore, presence of H2B.8 is sufficient to cause smaller nuclear size, likely through its ability to form euchromatic foci.

Indeed, overexpression of H2B.8ΔIDR (*p35S::HTB8ΔIDR-YFP*) or H2B.2 (*p35S::HTB2-YFP*), both lacking the N-terminal tail IDR and therefore the ability to form euchromatic foci (Fig. 2.14), has no effect on root nuclear size (Fig. 2.17). Therefore, the ability of H2B.8 to reduce somatic nuclear size is not an artefact of protein overexpression, but a property specific to the histone variant in question.

Considering somatic nuclear morphology, the bulk of chromatin volume is provided by dispersed euchromatin rather than heterochromatic domains, which are highly condensed. As such, chromatin proteins acting upon heterochromatin are unlikely to achieve any further compaction. However, H2B.8 principally functions in euchromatin and is unique in its capability to condense such regions. Given the dispersed nature of somatic euchromatin, ectopic H2B.8 has a great opportunity to achieve chromatin condensation. Ultimately, global chromatin volume is reduced which manifests as smaller nuclear size (Fig. 2.17).

Interestingly, natural H2B.8 expression is not limited to sperm cells, H2B.8 is also found in mature seeds (Section 2.2). The occurrence of H2B.8 in these somatic cells coincides with chromatin compaction and reduction of nuclear size (Van Zanten *et al.*, 2011). A number of factors could be causal in the mature seed chromatin compaction event; although it does provide a further association of H2B.8 presence and subsequent alteration of nuclear size.

Collectively, this section has strengthened the link between H2B.8 and chromatin compaction. H2B.8 has been demonstrated to reduce somatic nuclear size, adding to its effect upon sperm. Function is dependent on the N-terminal tail IDR, suggesting H2B.8 reduces chromatin volume by forming phase separated euchromatic foci. Having established the importance of the IDR to H2B.8 function, the next section asks whether the increased histone body arginine residues and K-to-N substitution at position 145 also contribute to the chromatin condensation mechanism.

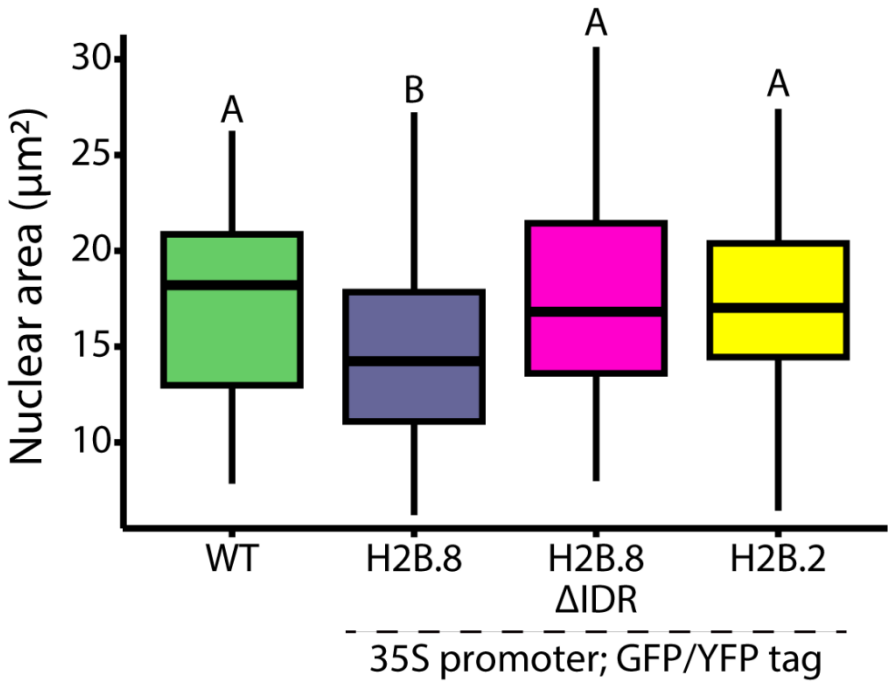


Figure 2.17: Ectopically expressed H2B.8 can reduce root nuclear size. Ectopic expression of H2B.8 (*p35S::HTB8-eGFP*) in somatic tissues can significantly reduce nuclear size of root vascular cylinder cells in the elongation zone. Similar expression of H2B.8ΔIDR (*p35S::HTB8ΔIDR-YFP*) and H2B.2 (*p35S::HTB2-YFP*) does not affect nuclear size and are no different from WT. The boxplot shows median (thick black bar) and first and third quartiles, with lower and upper whiskers extending to 1.5 times the interquartile range of the first and third quartiles, respectively. Statistical significance was determined by an ANOVA followed by individual Tukey tests; boxplots marked as A and B are not different within the group ($P > 0.001$) but are different between groups ($P < 0.001$); WT N = 81; H2B.8 N = 96; H2B.8ΔIDR N = 108; H2B.2 N = 129.

2.8.2 Chimeric proteins reveal roles of H2B.8 features in sperm nuclear size

To test the contributions of various H2B.8 features to protein function, several chimeric proteins were designed and transformed to the *htb8* mutant (Fig. 2.18A; Supp. Fig. 2.3). Relative complementation of nuclear size was used as a proxy for chromatin compaction, given the intrinsic link between the two phenotypes (Macadangdang *et al.*, 2014).

Domain swap chimeric proteins between H2B.8 and a canonical H2B variant, H2B.2, were generated (Fig. 2.18A; Supp. Fig. 2.3). These consisted of the H2B.8 tail and H2B.2 body (H2B.8T-2B) and vice versa (H2B.2T-8B). Such constructs enabled comparisons of the N-terminal tail domains with or without the IDR. Furthermore, the contributions of the H2B.8 increased histone body arginine and K-to-N substitution could be assessed independently of the IDR. Additional chimeric proteins sought to separate the roles of the histone body features (Fig. 2.18A; Supp. Fig. 2.3). H2B.8 was engineered with the C-terminal motif of H2B.2 (H2B.8-N234K), introducing a lysine residue that could be post-translationally ubiquitylated. In reverse, an asparagine residue was substituted at K145 of H2B.2 (H2B.2-K145N), rendering the protein unable to undergo ubiquitylation. Collectively, the H2B chimeric proteins enabled dissection of H2B.8 features and their contributions to sperm chromatin structure.

Sperm nuclear size is significantly larger than WT in two independent CRISPR mutants, *htb8* #7 and *htb8* #3 (Fig. 2.18B). This phenotype (*htb8* #7) can be complemented by both eGFP- and Myc-tagged H2B.8 under the native promoter (Fig. 2.18B). Quantification of sperm nuclear size in chimeric protein complementation lines revealed distinct patterns of importance for each H2B.8 feature, the general rules of which are discussed below. Pairwise *P* values can be found in the supplementary material (Supp. Fig. 2.4).

H2B.8-N234K, with IDR and histone body arginines, can fully rescue *htb8* #7 sperm nuclear size (Fig. 2.18B). This implies that the ability to be ubiquitylated does not affect H2B.8 function. Moreover, H2B.2-K145N, without IDR domain and histone body arginine residues, fails to reduce sperm nuclear size to any extent (Fig. 2.18B). Thus, loss of ubiquitylation potential of a histone variant does not contribute to sperm size. As such, the inability of H2B.8 to undergo ubiquitylation is likely to be dispensable for function in global chromatin condensation.

The H2B.2 tail with H2B.8 body domain, H2B.2T-8B, fails to rescue *htb8* #7 sperm nuclear size (Fig. 2.18B). This indicates that the IDR is required whilst the histone body arginine and K-to-N substitution alone are insufficient for function. As such, the N-terminal tail domain with IDR is most important to H2B.8 function. However, examination of H2B.8T-2B, with the IDR in an otherwise canonical H2B variant, cannot fully rescue the *htb8* #7 phenotype (Fig. 2.18B). Instead, sperm nuclear size is only partially restored to WT levels. Therefore, the IDR feature of H2B.8 alone is unable to achieve full protein function; this indicates a partial role for the histone body. Having previously established that the K-to-N substitution and accompanying incapability to undergo ubiquitylation is dispensable for H2B.8 function in sperm nuclear size, the role of the histone body is most likely due to surface arginine residues.

Collectively, analysis of H2B chimeric proteins abilities to rescue sperm nuclear size in *htb8* #7 has shown that the N-terminal IDR is most important, with a minor role for the histone body arginine residues (Fig. 2.18B). Furthermore, the IDR domain was demonstrated to be critical for H2B.8 function in chromatin compaction of somatic nuclei (Section 2.8.1; Fig. 2.17).

Previous results (Section 2.7) have shown that the IDR is critical for formation of phase separated euchromatic foci. However, the mechanism by which histone body arginine residues contribute to H2B.8 function is unknown. Perhaps the increased positive charge of the histone body, owing to the arginine residues, could help to stabilise H2B.8-containing nucleosomes through tighter association with negatively charged DNA. Such a stabilisation property could assist with formation of foci on a larger scale. Alternatively, the arginine residues themselves could contribute to phase separation properties of H2B.8. Recently, arginine amino acids have been implicated in increasing the viscosity of phase separated droplets *in vitro* (Fisher and Elbaum-Garfinkle, 2020). Therefore, arginine residues on the histone body surface of H2B.8 could contribute to the gel-like properties of IDR-mediated phase separated foci. Further work should seek to unravel the mechanism by which the H2B.8 body contributes to protein function.

Interestingly, no effect was observed upon sperm nuclear size for the K-to-N substitution in the C-terminal motif (Fig. 2.18B). However, this feature could be functioning on a more local scale and therefore not impact the nuclear size phenotype. Given the prediction that the asparagine residue immunises H2B.8 against ubiquitylation, the histone variant could have interesting deposition dynamics with H2Bub and subsequently, gene expression. The following section addresses exact H2B.8 localisation in chromatin.

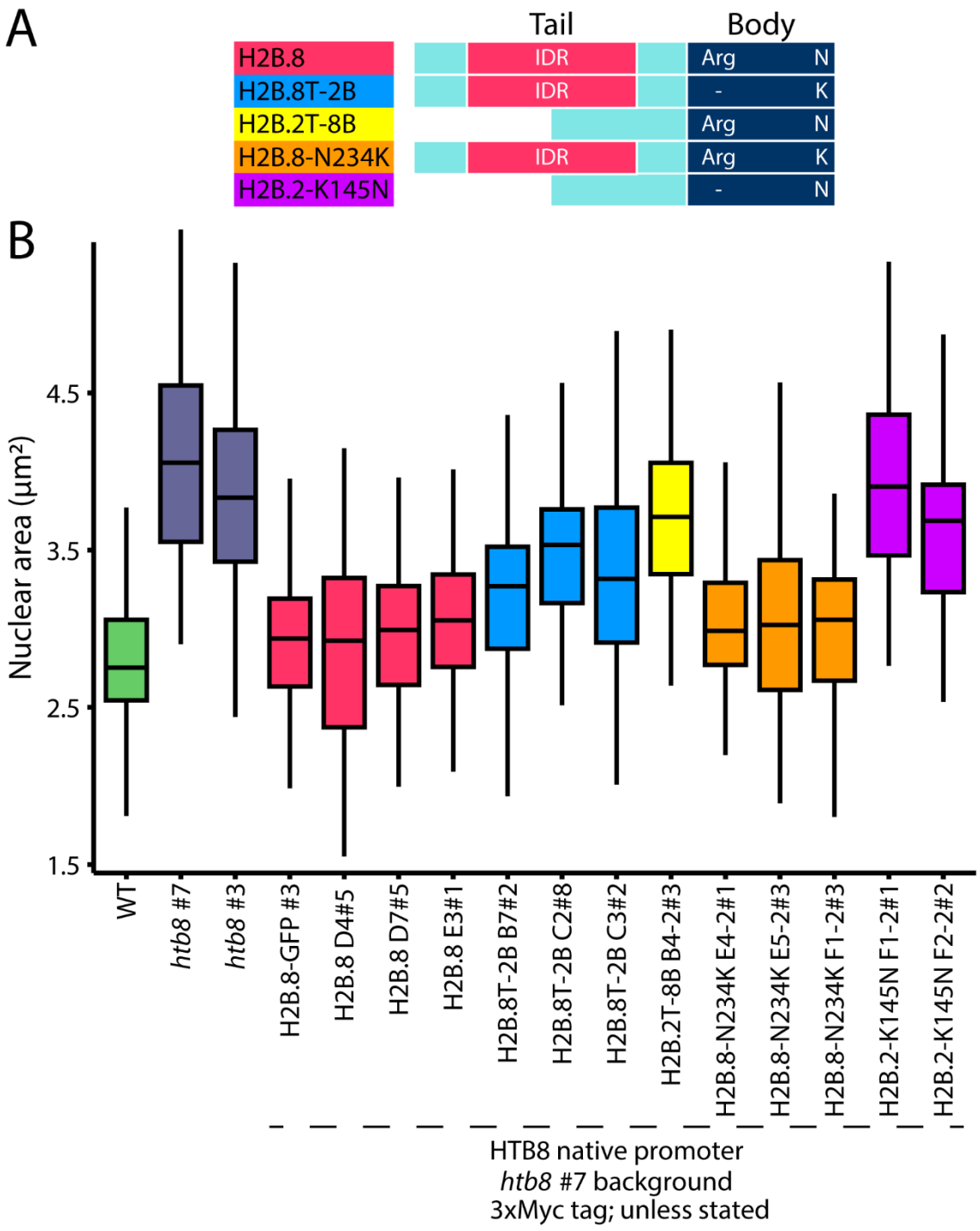


Figure 2.18: H2B.8 effect on sperm nuclear size is primarily determined by the IDR, with a minor role for body arginine residues. (**A**) Schematic of chimeric H2B proteins and their features. Chimeric constructs are denoted as H2B.8T-2B (H2B.8 tail; H2B.2 body), H2B.2T-8B (H2B.2 tail; H2B.8 body), H2B.8-N234K (H2B.8 with K at position to 234) and H2B.2-K145N (H2B.2 with N at position to 145). An alignment of chimeric H2Bs can be found in the supplementary (Supp. Fig. 2.3). (**B**) H2B.8 with eGFP or Myc tags can fully complement sperm nuclear size. H2B.8T-2B can partially restore *htb8* #7 to WT; whilst H2B.2T-8B cannot to any extent. H2B.8-N234K fully complements the phenotype whereas H2B.2-K145N fails to recover sperm nuclear size. Sperm nuclear size is primarily determined by the H2B.8 N-terminal IDR, with a lesser role for the C-terminal body containing elevated surface arginine whilst the K-to-N at position 234 is dispensable. Boxplot shows median (thick black bar) and first and third quartiles, with lower and upper whiskers extending to 1.5 times the interquartile range of the first and third quartiles, respectively. Statistical significance was determined by an ANOVA followed by individual Tukey tests; *P* values can be found in the supplementary (Supp. Fig. 2.4); N = 77-80 nuclei for each genotype.

2.9 H2B.8 deposits to inactive euchromatin

Using cytological techniques, H2B.8 has been demonstrated to preferentially localise to euchromatic regions (Section 2.6). However, the exact positioning of H2B.8 with regards to enrichment or depletion over genomic features remains unknown. This section addresses H2B.8 deposition using genomics.

2.9.1 H2B.8 is enriched at inactive euchromatin in sperm

To assess H2B.8 deposition in sperm, a native chromatin immunoprecipitation followed by sequencing (ChIP-seq) protocol was optimised for low input material. As H2B.8 is sperm-specific in the male germline, whole pollen could be used from the GFP-tagged line (*pHTB8::HTB8-eGFP*) for native ChIP-seq library construction. Two replicates were highly correlated with one another (Supp. Fig. 2.5A); replicate 1 was used for subsequent analysis.

Looking broadly at deposition along chromosomes, it is apparent that H2B.8 is depleted from constitutive heterochromatin, enriched at pericentromeric regions and present throughout euchromatic chromosome arms (Fig. 2.19A). This section will describe H2B.8 deposition patterns in detail and its distribution among genomic features.

Heterochromatin is characterised by the histone mark H3K9me2 in somatic tissues. Cytological evidence suggests that this remains the case in sperm (Fig. 2.10A). CUT&Tag (Kaya-Okur *et al.*, 2020) was optimised to profile histone modifications in FACS isolated sperm, rather than whole pollen, and used to examine H3K9me2 deposition. In agreement with cytological observations (Schoft *et al.*, 2009), H3K9me2 deposition is unchanged in sperm from somatic tissues with the two profiles strongly correlating (R = 0.853; Spearman's Rank Correlation Coefficient). This indicates that heterochromatin localisation does not undergo reprogramming over the course of male germline development.

Heterochromatin is enriched at the gene-poor and TE-rich pericentromeric region (Fig. 2.19A). Given the established relationship of H3K9me2 with TEs in soma, this mark was used to cluster TEs into three groups depending on enrichment (Supp. Fig. 2.6, C to E). Those with the lowest somatic H3K9me2 levels were classed as euchromatic (euTEs), whilst those with the highest were deemed heterochromatic (hetTEs). A third group of intermediate H3K9me2 marked TEs were also defined (intTEs). EuTEs are dispersed along chromosome arms, intTEs and hetTEs are enriched at pericentromeric regions neighbouring the centromere (Fig. 2.19A). Clustering using sperm H3K9me2 data gives almost identical groupings, further indicating that H3K9me2 is unchanged through the male germline.

Considering H2B.8 deposition in a broad context, the histone variant is enriched in pericentromeric regions with intTEs and is somewhat present along chromosome arms, rich with euTEs and genes. However, H2B.8 is largely absent from the pericentromeric hetTEs; accordingly, H2B.8 and H3K9me2 are weakly anticorrelated (R = -0.206; Spearman's Rank Correlation Coefficient).

Looking more closely at chromosome arms, an interesting deposition dynamic is observed. H2B.8 is highly enriched at euTEs and intTEs (Fig. 2.19B). Additionally, this enrichment is not restricted to TEs but rather spreads beyond and marks intergenic regions (regions without annotated genes or TEs). However, at genes, deposition is quite variable. As a rule, actively transcribed genes are highly depleted of H2B.8 (Fig. 2.19B), whereas genes without expression are marked by H2B.8 in some instances but not others. Having obtained an impression of H2B.8 deposition in sperm via the genome browser, downstream analysis sought to define overall trends.

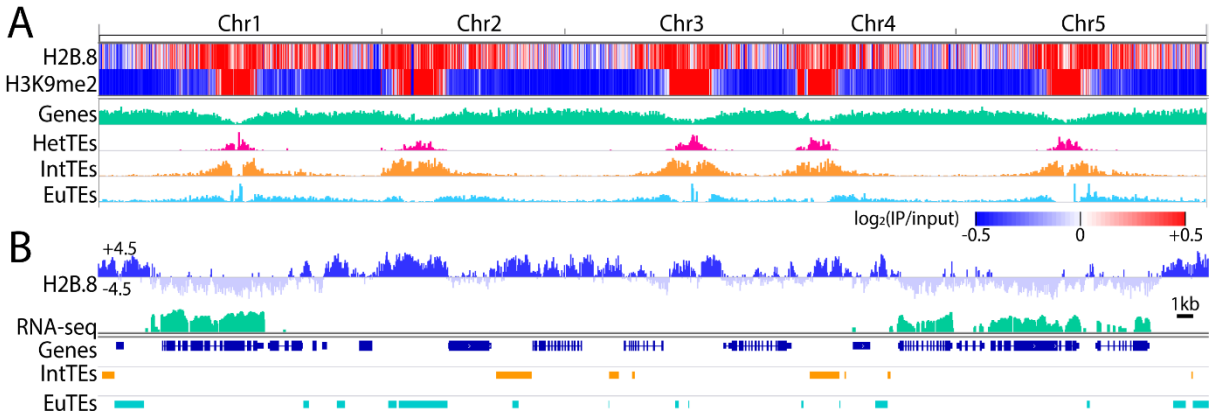


Figure 2.19: In sperm, H2B.8 is enriched at euchromatic/intermediate TEs and intergenic regions and is depleted from genes. (**A**) Whole genome browser view of sperm H2B.8 and H3K9me2 enrichment as a heatmap (log_2(IP/input); red – high; blue – low). Density tracks of genes (green), heterochromatic TEs (pink), intermediate TEs (orange) and euchromatic TEs (light blue). (**B**) Browser view of sperm H2B.8 enrichment (log_2(IP/input); blue) and sperm cell RNA-seq (log_2(FPKM); green) tracks over a genomic region on chromosome 1 (1:12,478,000-12,550,000) containing several genes (dark blue) and euchromatic/intermediate TEs (light blue and orange; respectively).

To further assess the relationship of H2B.8 with genes and transcription, profiles and heatmaps were generated (Fig. 2.20, A and D). Genes with the highest expression had the greatest depletion of H2B.8 from the gene body, almost all genes in this group were classified as depleted. However, genes with no/low expression largely had no preference for H2B.8, being neither enriched nor depleted (Fig. 2.20, A and D). This agrees with the observation made with the browser view. H2B.8 is readily able to deposit both upstream and downstream of genes, irrespective of expression. This indicates that the deposition dynamic at genes is specific rather than due to the wider chromatin environment. Overall, H2B.8 and gene transcription are moderately anticorrelated (R = -0.58; Spearman's Rank Correlation Coefficient).

Genome-wide, the bodies of both euTEs and intTEs are strongly marked by H2B.8 (Fig. 2.20, B and D). The vast majority of TEs within each category are enriched for H2B.8, equating to 74% of euTEs and 72% of intTEs. This supports the browser view observations. In contrast, hetTEs are largely neither enriched (3%) nor depleted (17%) of H2B.8 (Fig. 2.20, B and D), supporting the lack of preference of H2B.8 to highly heterochromatic regions. Collectively, H2B.8 is deposited to TEs in euchromatic or intermediate chromatin regions.

Finally, as described from browser view observations, H2B.8 is present in intergenic regions (Fig. 2.20, C and D). The heatmap of intergenic regions shows a clear boundary is formed in around half of cases, likely those adjacent to genes. The remainder of intergenic regions are likely to be adjoining TEs and therefore blur into the neighbouring features (Fig. 2.20C).

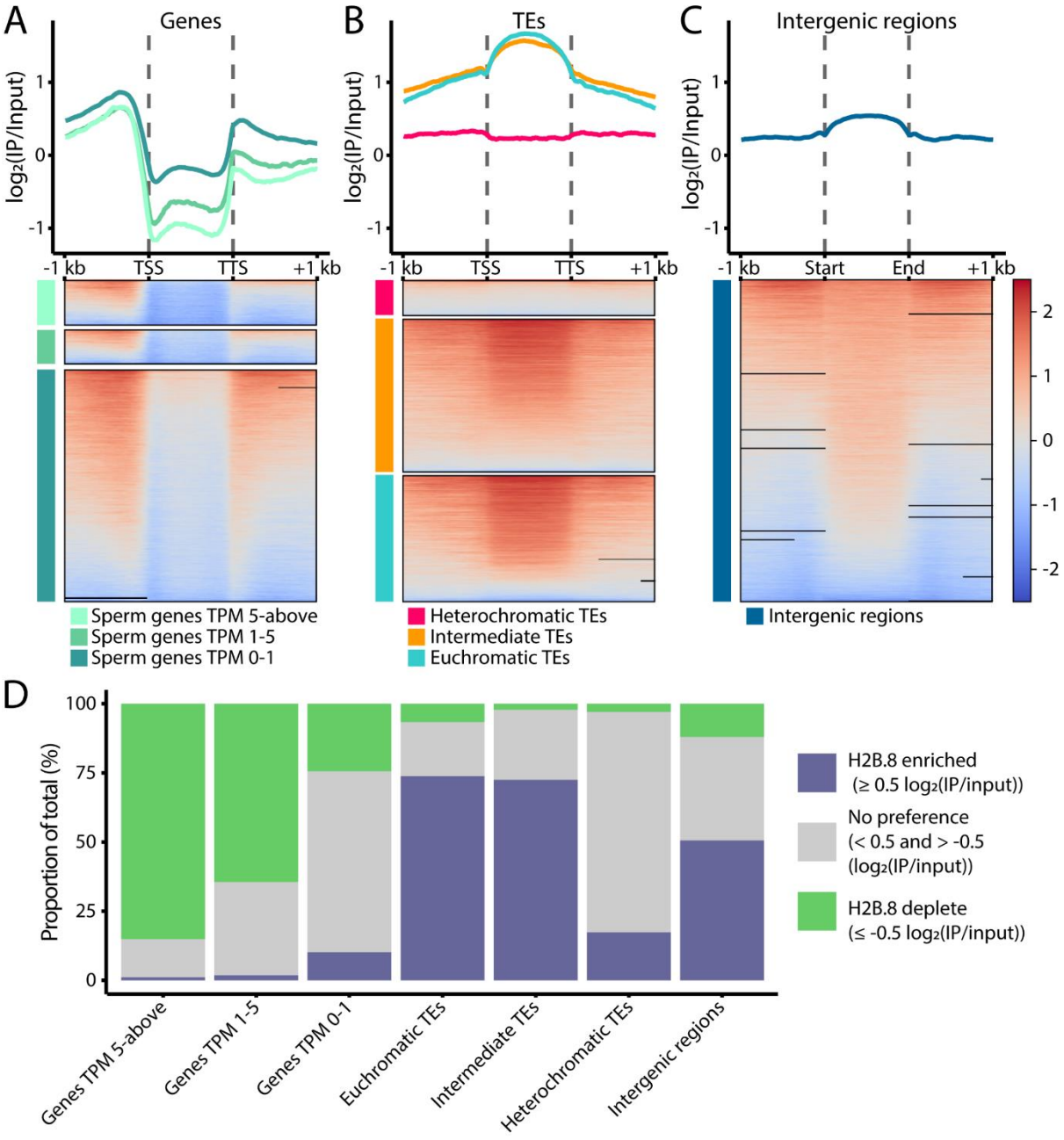

A
Genes
log2(IP/Input)
-1 kb
TSS
TTS
+1 kb
Sperm genes TPM 5-above
Sperm genes TPM 1-5
Sperm genes TPM 0-1
B
TEs
Heterochromatic TEs
Intermediate TEs
Euchromatic TEs
C
Intergenic regions
Start
End
Intergenic regions
D
Proportion of total (%)
H2B.8 enriched
(≥ 0.5 log2(IP/input))
No preference
(< 0.5 and > -0.5
(log2(IP/input))
H2B.8 deplete
(≤ -0.5 log2(IP/input))
Genes TPM 5-above
Genes TPM 1-5
Genes TPM 0-1
Euchromatic TEs
Intermediate TEs
Heterochromatic TEs
Intergenic regions

Figure 2.20: H2B.8 is preferentially enriched at euchromatic/intermediate TEs and intergenic regions in sperm. (**A** to **C**) Profile plots and associated heatmaps of average enrichment (log_2(IP/input); red – high; blue – low) of sperm H2B.8 (*pHTB8::HTB8-eGFP*) over genes (A) grouped by expression (light green – TPM 5-above; mid green – TPM 1-5; dark green – TPM 0-1), TEs (B) grouped by chromatin environment (light blue – euchromatic TEs; orange – intermediate TEs; pink – heterochromatic TEs) and intergenic regions (C; dark blue). Genes, TEs and intergenic regions are scaled to 1 kb lengths; 1 kb is shown upstream/downstream of the TSS/start and TTS/end, respectively. Heatmaps are sorted from high to low H2B.8 enrichment. (**D**) Proportion of genes (grouped by sperm cell expression), TEs (grouped by chromatin environment) and intergenic regions that are enriched (≥ 0.5; blue), depleted (≤ -0.5; green) or without preference (< 0.5 and > -0.5; grey) for sperm H2B.8 (log_2(IP/input); *pHTB8::HTB8-eGFP*).

Next, using the H2B.8 native ChIP-seq data, peak calling was attempted. Software including MACS2 (Zhang *et al.*, 2008), HOMER (Heinz *et al.*, 2010), epic2 (Stovner and Sætrom, 2019) and hiddenDomains (Starmer and Magnuson, 2016) were used to call peaks. However, each failed to capture the true nature of H2B.8 distribution; despite defining parameters to suit broad domains. For example, MACS2 could define sharp peaks between genes but then fail to capture the broad regions spanning TEs; whereas HOMER would scaffold between enriched regions and incorrectly include H2B.8 depleted genes. H2B.8 is highly abundant in chromatin, likely more so than the chromatin modifications that such programmes are designed to capture. Therefore, a different approach was taken to define H2B.8 peaks.

Using a combination of BEDtools (Quinlan and Hall, 2010) and deepTools (Ramírez *et al.*, 2016), H2B.8 peaks were defined. This analysis called 18416 peaks with a mean size of 2418.57 bp and median size of 1550 bp. Sperm H2B.8 peaks spanned 44,540,408 bp equating to 33.08% of the *Arabidopsis* genome.

H2B.8 peaks overlapped with 20% of non-expressed genes, but just 5% of those with moderate expression and 3% of highly expressed genes (Fig. 2.21). Approximately 80% of euchromatic and intermediate TEs were overlapped by H2B.8 peaks. Whereas heterochromatic TEs were overlapped in 18.5% of instances (Fig. 2.21).

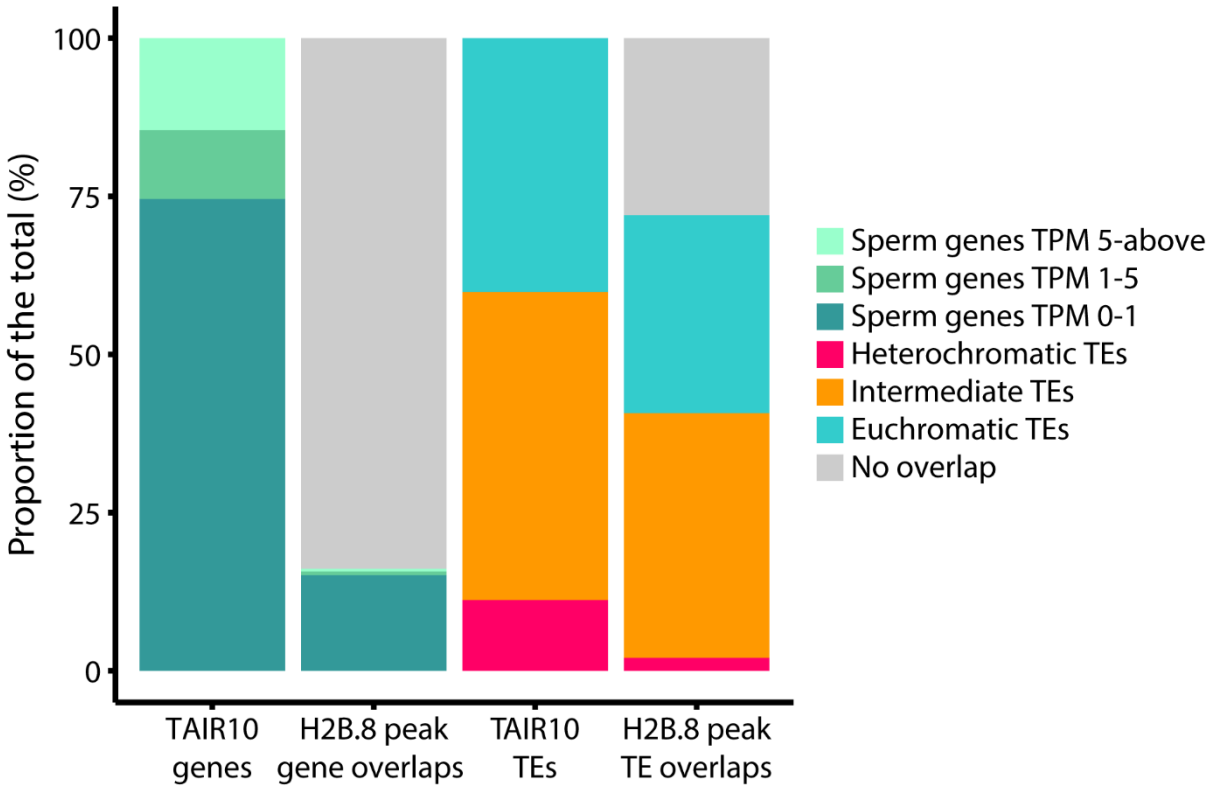


Figure 2.21: H2B.8 peaks overlap a subset of silent genes and most euchromatic/intermediate TEs sperm. The majority of genes in sperm cells are not expressed; a subset of which overlap with H2B.8 peaks. Such peaks scarcely overlap genes with any level of expression. H2B.8 peaks overlap the vast majority of euchromatic and intermediate TEs, whilst rarely marking heterochromatic TEs.

Taken together, this deposition pattern describes that H2B.8 is enriched at euTEs and intTEs but is not bound by their borders. Rather, H2B.8 can spread beyond such genomic features and therefore mark intergenic regions. However, upon meeting genes, a sharp boundary occurs whereby H2B.8 is depleted from genes, especially those with high expression. Having established the deposition profile of H2B.8 and its relationship with genomic features, it was next asked how the localisation compares to other histone variants/marks in sperm.

2.9.2 Sperm H2B.8 deposition is unique compared to other histone marks/variants

Many histone modifications have well-characterised somatic deposition profiles. However, whether such histone modifications are reprogrammed in sperm remains comparatively unknown. Recently, Borg *et al*. published several ChIP-seq data sets for active and repressive histone marks in sperm, along with the profile of H3.10, a sperm-specific H3 variant (Borg *et al.*, 2020). Reanalysis of the raw data (Supp. Table 2.3), along with the H3K9me2 data presented here can give an insight into chromatin reprogramming in sperm and how these profiles relate to H2B.8.

Heterochromatic marks such as H3K9me2 and H3K27me1 behave similarly in sperm as they do in soma; they are enriched at hetTEs and depleted from genes (Supp. Fig. 2.6; Supp. Fig. 2.7, B and C; Supp. Fig. 2.8, B and C). The two marks are correlated genome-wide between sperm and seedlings (R = 0.79 and 0.69 for H3K9me2 and H3K27me1, respectively; Spearman's Rank Correlation Coefficient). Coupled with cytological data, it can be concluded that genomic localisation of heterochromatin is largely unchanged from seedlings and through male germline development.

The histone mark H3K27me3 serves as a repressive chromatin modification over broad domains in gene-rich regions. In sperm, largely due to activity of demethylases and a moderate role of H3.10 deposition (Borg *et al.*, 2020), H3K27me3 is globally lost (Supp. Fig. 2.6; Supp. Fig. 2.7D; Supp. Fig. 2.8D). As such, H3K27me3 is thought to be reprogrammed during male sexual lineage development. Accordingly, H3K27me3 does not correlate with any other histone marks in sperm.

Histone modifications associated with active euchromatin seemingly have similar roles in sperm as is known in somatic tissues. H3K4me3 and H3K27ac marks correlate highly with one another (Fig. 2.22) and preferentially deposit to genes with the highest expression whilst being deplete from all TEs (Supp. Fig. 2.6; Supp. Fig. 2.7, E and F; Supp. Fig. 2.8, E and F). Active marks are likely to alter deposition depending on transcription in sperm cells; this behaviour is in accordance with established H3K4me3 and H3K27ac roles in somatic tissues and differences between cell types.

Collectively, analysis of previously characterised histone modifications has revealed that repressive heterochromatic and active euchromatic marks behave similarly in sperm as they do in somatic tissues. The repressive euchromatic mark, H3K27me3, undergoes reprogramming in male germline development, resulting in global erasure. Having established the patterns of available histone marks in sperm, it was next considered how these relate to H2B.8.

Broadly, H2B.8 clusters with repressive chromatin marks in sperm (Fig. 2.22). However, the associations are very weak in accordance with the established role of H2B.8 in marking euTEs and intTEs rather than hetTEs. Given the anticorrelation of H2B.8 with genes and principally, expression, it is unsurprising that active histone marks are also anticorrelated with H2B.8 (Fig. 2.22). Taken together, H2B.8 has a highly unique deposition profile in sperm, different from other known histone marks.

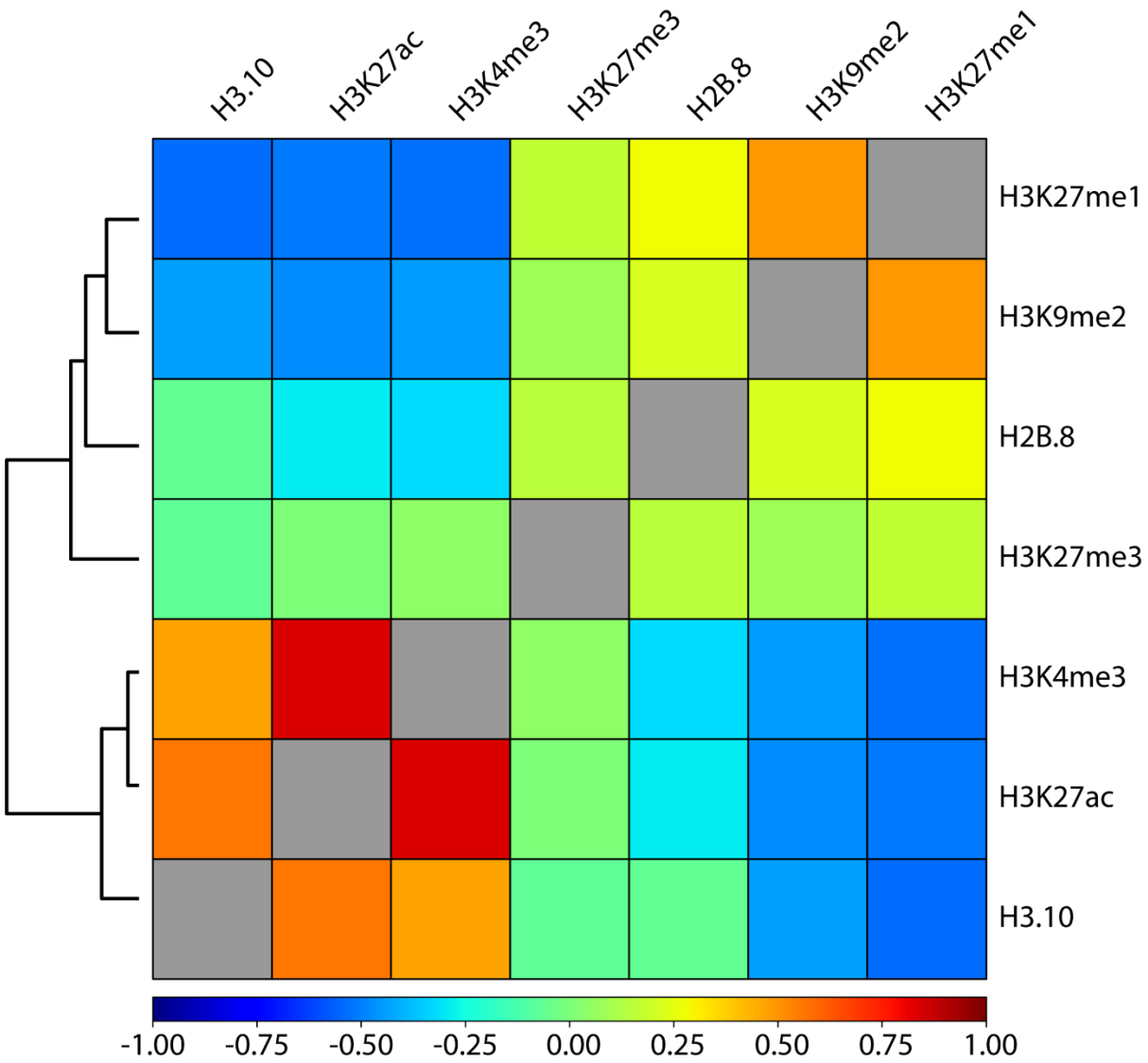


Figure 2.22: H2B.8 deposition clusters with other repressive chromatin marks in sperm. Heatmap of Spearman's rank correlation coefficients (*R*; red – strong positive correlation; blue – strong negative correlation; green – no correlation) between various sperm histone marks/variants enrichments (log_2(IP/input)) over 1 kb windows across the genome. Histone marks/variants are hierarchically clustered according to correlations.

2.9.3 Ectopic H2B.8 shows similar deposition pattern to sperm

The previous section established the profile of H2B.8 and its association with other chromatin marks in sperm. However, sperm is highly unique among *Arabidopsis* cell types, with substantially fewer genes expressed. To address how H2B.8 deposition could be affected by greater transcriptional activity, the ectopic expression H2B.8 line (*p35S::HTB8-eGFP*) was examined.

Native ChIP-seq was performed on ten-day-old seedlings expressing H2B.8. Since the two replicates were highly correlated with one another (Supp. Fig. 2.5B), a single replicate was taken forward for further analysis. Interestingly, global H2B.8 profiles in sperm and seedlings were highly correlated (Supp. Fig. 2.5C); this supports cytological observations that H2B.8 can form true euchromatic foci in seedlings, as well as in the sperm.

Similar to broad scale H2B.8 deposition in sperm, seedling H2B.8 is depleted from H3K9me2-rich pericentromeres and is present throughout the chromosome arms, abundant with genes and euTEs (Fig. 2.23A). Closer inspection of the genome browser shows that the majority of genes are expressed in seedlings (Fig. 2.23B). H2B.8 is depleted from genes with expression but can deposit to some non-transcribed genes. Comparable to observations in sperm, H2B.8 is enriched at euTEs and intTEs and can also evenly deposit to regions without annotated genomic features (Fig. 2.23B). Therefore, it seems that H2B.8 behaves similarly in seedlings as in sperm.

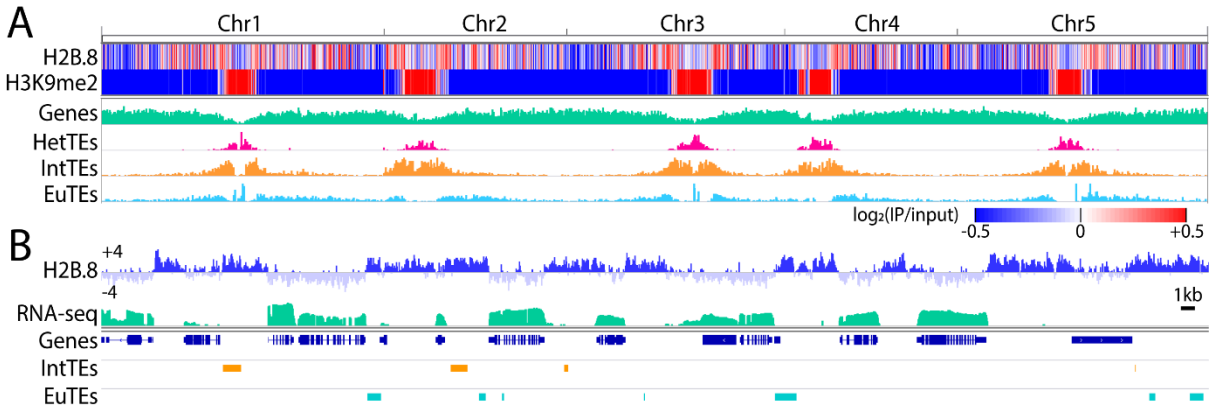


Figure 2.23: In seedlings, H2B.8 is similarly enriched at euchromatic/intermediate TEs and intergenic regions and is depleted from genes. (**A**) Whole genome browser view of seedling H2B.8 and H3K9me2 enrichment as a heatmap (log_2(IP/input); red – high; blue – low). Density tracks of genes (green), heterochromatic TEs (pink), intermediate TEs (orange) and euchromatic TEs (light blue). (**B**) Browser view of seedling H2B.8 enrichment (log_2(IP/input); blue) and seedling RNA-seq (log_2(FPKM); green) tracks over a genomic region on chromosome 1 (1:10,628,000-10,702,000) containing several genes (dark blue) and euchromatic/intermediate TEs (light blue and orange; respectively).

Analysis of seedling H2B.8 enrichment at genes, TEs and intergenic regions revealed similar patterns to those observed in sperm (Fig. 2.24, A to D). Genes with the highest expression exhibit the least enrichment of H2B.8, whilst those with no/low expression are marked in some cases but not others (Fig. 2.24, A and D). H2B.8 preferentially targets euchromatic and intermediate TEs, with no preference for heterochromatic TEs (Fig. 2.24, B and D). Regions without annotated genes or TEs, defined as intergenic, are also enriched for H2B.8 in seedlings.

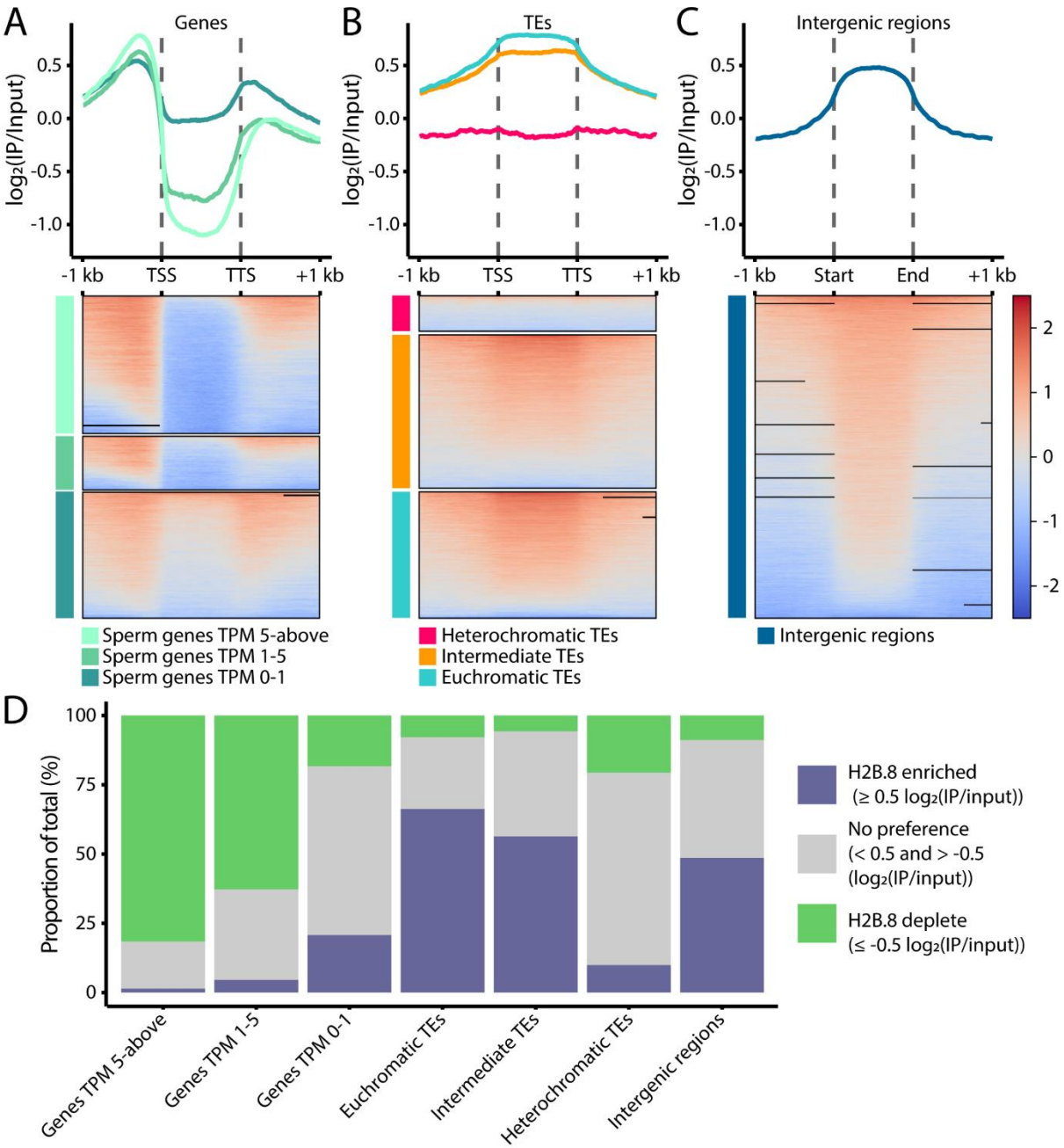


Figure 2.24: H2B.8 is preferentially enriched at euchromatic/intermediate TEs and intergenic regions in seedling. (**A** to **C**) Profile plots and associated heatmaps of average enrichment (log_2(IP/input); red – high; blue – low) of seedling H2B.8 (*p35S::HTB8-eGFP*) over genes (A) grouped by expression (light green – TPM 5-above; mid green – TPM 1-5; dark green – TPM 0-1), TEs (B) grouped by chromatin environment (light blue – euchromatic TEs; orange – intermediate TEs; pink – heterochromatic TEs) and intergenic regions (C; dark blue). Genes, TEs and intergenic regions are scaled to 1 kb lengths; 1 kb is shown upstream/downstream of the TSS/start and TTS/end, respectively. Heatmaps are sorted from high to low H2B.8 enrichment. (**D**) Proportion of genes (grouped by seedling expression), TEs (grouped by chromatin environment) and intergenic regions that are enriched (≥ 0.5; blue), depleted (≤ -0.5; green) or without preference (< 0.5 and > -0.5; grey) for seedling H2B.8 (log_2(IP/input); *p35S::HTB8-eGFP*).

Seedling H2B.8 peaks were called using the same parameters as for sperm. A total of 14935 peaks were identified, averaging a mean size of 2568.23 bp and a median size of 1800 bp. Peaks occupied 38,356,468 bp, corresponding to 28.49% of the *Arabidopsis* genome, a value slightly lower than for sperm, which could be due to more genes being expressed in seedlings and therefore reducing the genomic space into which H2B.8 can deposit. Interestingly, nuclear size upon presence of H2B.8 in seedlings results in a smaller area reduction (24%) compared to the difference between WT and *htb8* sperm (37%). This could be due to the smaller proportion of H2B.8 incorporation in seedlings compared to sperm. Indeed, extraction and mass-spectrometry quantification of histones from seedling nuclei revealed that H2B.8 proteins constitute a smaller proportion of the total histone H2B pool in seedling (9.1%) versus sperm (12.6%), despite being driven by a strong promoter (Supp. Fig. 2.9).

Similar to H2B.8 peaks in sperm, seedling peaks scarcely overlap with moderately (10%) or highly (4%) expressed genes but overlap with a subset of non-expressed (32%) genes (Fig. 2.25). Seedling H2B.8 peaks overlap fewer TEs than sperm peaks in all three clusters. However, euchromatic (68%) and intermediate (57%) TEs are still marked by H2B.8 in the majority of cases whilst heterochromatic (9%) TEs are rarely overlapped by H2B.8 peaks (Fig. 2.25).

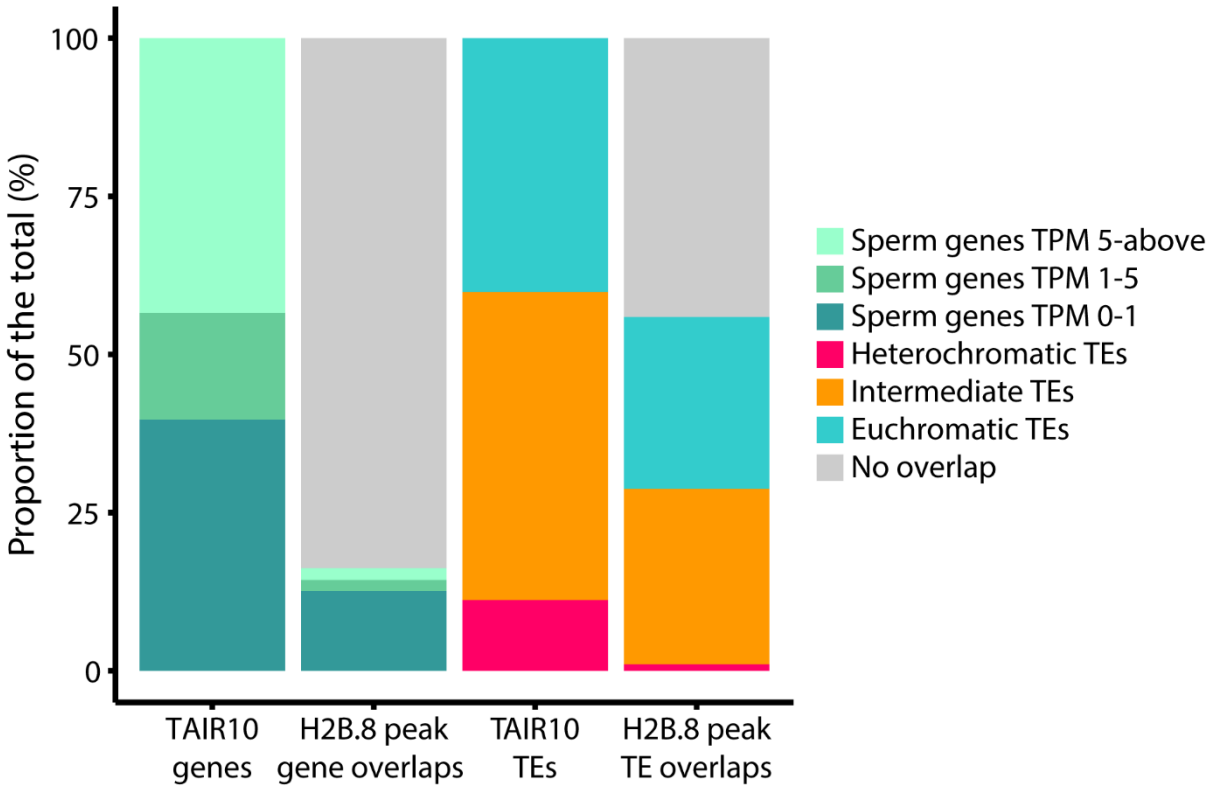


Figure 2.25: H2B.8 peaks overlap a subset of silent genes and most euchromatic/intermediate TEs in seedlings. The majority of genes are expressed in seedlings and are unmarked by H2B.8 peaks. A portion of non-expressed genes overlap with H2B.8 peaks. Euchromatic and intermediate TEs strongly overlap with H2B.8 peaks, whilst little overlap is observed for heterochromatic TEs.

Collectively, deposition of H2B.8 in seedlings follows the same patterns as in sperm, despite the greater transcriptional activity in the former. Therefore, localisation of H2B.8 is independent of any sperm-specific factor and follows common rules irrespective of cell type.

2.9.4 Seedling H2B.8 deposition clusters with repressive histone marks/variants

Distributions of histone modifications and variants are relatively unknown in sperm, however much more data exists for somatic tissues such as seedlings. Having established the profile of H2B.8 in seedlings, the associations with chromatin marks in seedlings were determined. Various available ChIP-seq datasets were downloaded and reanalysed (Supp. Table 2.3).

Heterochromatic marks, such as H3K9me2 and H3K27me1, are depleted from all genes and most transposons (Supp. Fig. 2.10; Supp. Fig. 2.11, B and C; Supp. Fig. 2.12, B and C). Enrichment is reserved for hetTEs, indicative of restriction of such marks to heterochromatic regions of the genome (Supp. Fig. 2.10C; Supp. Fig. 2.12, B and C). As mentioned previously, heterochromatic marks are largely unchanged in sperm cells compared to somatic tissues. Heterochromatin domains are highly robust through development, besides the noted loss in vegetative nuclei (Fig. 2.11B). Similar to sperm, seedling H2B.8 does not associate with heterochromatic marks and therefore occupies a different portion of the genome.

H3K27me3 serves as a repressive histone modification in euchromatic regions. In seedlings, genes with no/low expression are enriched for H3K27me3 across the gene body, whilst loci with transcription are depleted of the modification (Supp. Fig. 2.11D). Most genes in seedlings are expressed, so H3K27me3 is usually absent from genes (Supp. Fig. 2.10A). H3K27me3 enrichment is not observed at TEs of any chromatin class, in agreement with its widely established role in targeting genes (Supp. Fig. 2.10, B to E; Supp. Fig. 2.12D).

Given that H3K27me3 is absent from sperm, this analysis of seedling data provides an opportunity to establish whether H2B.8 can occupy similar genomic spaces. In seedlings, H2B.8 and H3K27me3 can both deposit to unexpressed genes. This suggests that H2B.8 could fulfil an analogous function in sperm, in the absence of H3K27me3. However, the profile of H2B.8 is far more expansive than H3K27me3, marking euTEs and intTEs. This means that the role of H2B.8 goes beyond replacing H3K27me3.

Active histone marks, such as H3K27ac, H2Bub and H3K4me3, strongly associate with genes with the highest expression and are absent from those without transcription (Supp. Fig. 2.11, E to G). TEs are heavily depleted of such histone modifications (Supp. Fig. 2.10, B to E; Supp. Fig. 2.12, E to G). As such, active histone marks behave similarly in both seedling and sperm and are strongly anticorrelated with H2B.8 deposition.

Profiles of several histone variants have also been established in somatic tissues. H3.1 and H2A.W are associated with repressed chromatin. Like H3K9me2 and H3K27me1, both variants principally occupy hetTEs and are depleted from all genes, irrespective of expression (Supp. Fig. 2.10; Supp. Fig. 2.11, H and K; Supp. Fig. 12, H and K). Accordingly, no association with seedling H2B.8 is found. H3.3 and H2A.Z are known to relate to active chromatin states. This analysis determined that both variants are highly depleted from all classes of TEs and deposit to genes, with incrementally greater enrichment according to gene expression (Supp. Fig. 2.10; Supp. Fig. 2.11, I and J; Supp. Fig. 2.12, I and J). As with H2B.8 and active histone modifications, H2B.8 is anticorrelated with active histone variants.

Genome-wide, seedling H2B.8 clusters with repressive histone modifications and variants (Fig. 2.26). This clustering also indicates that H2B.8 deposition is most similar to H3K27me3 rather than the heterochromatic marks, such as H3K9me2. However, as previously discussed, H2B.8 and H3K27me3 are only similar at non-expressed genes. H2B.8 deposition is far more wide-ranging, including enrichment at euTEs, intTEs and intergenic regions.

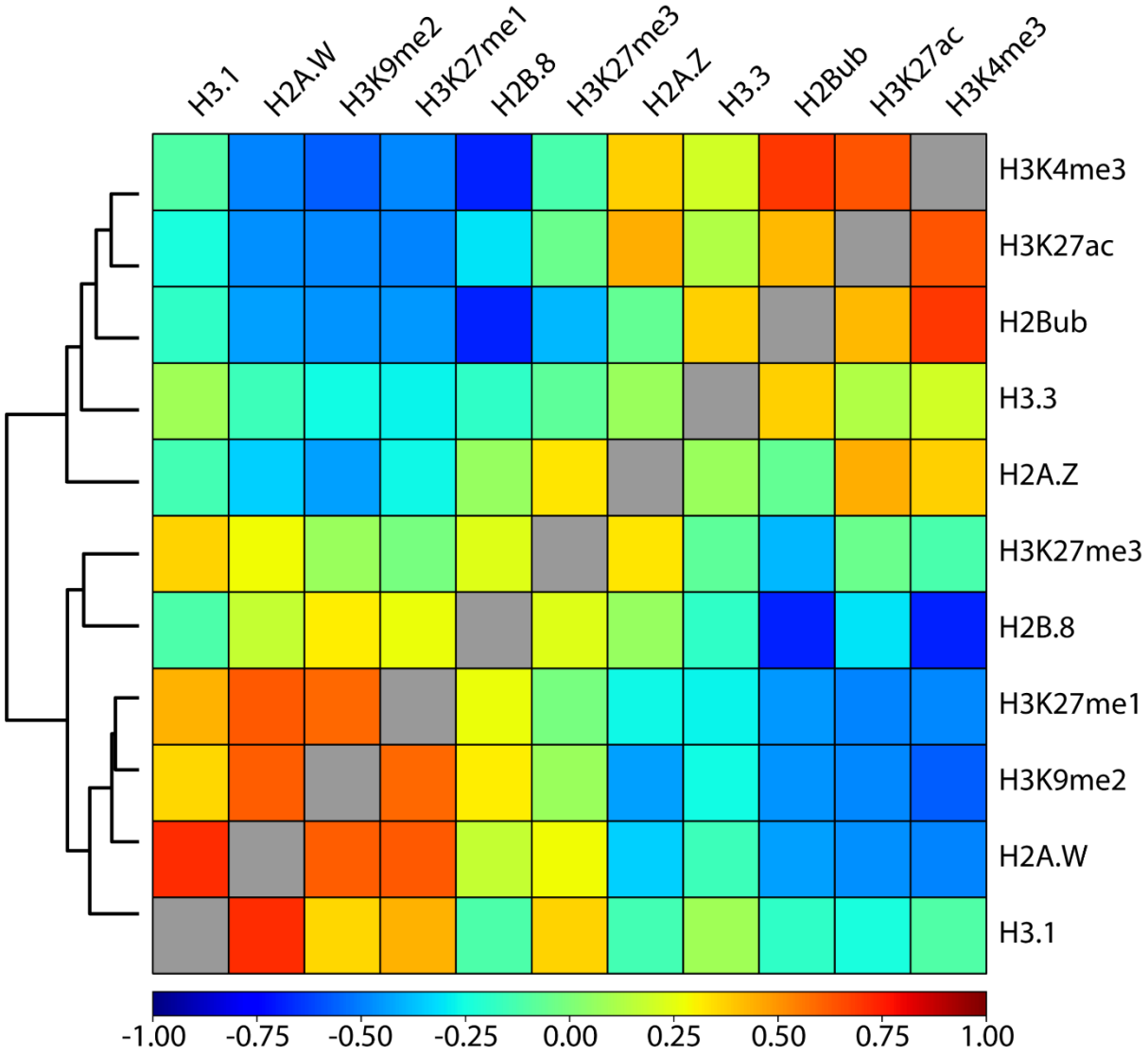


Figure 2.26: H2B.8 deposition clusters with other repressive chromatin marks in seedling. Heatmap of Spearman's rank correlation coefficients (*R*; red – strong positive correlation; blue – strong negative correlation; green – no correlation) between various seedling histone marks/variants enrichments (log_2(IP/input)) over 1 kb windows across the genome. Histone marks/variants are hierarchically clustered according to correlations.

To conclude, this section has revealed the genomic distribution of H2B.8 in sperm and seedlings; the profile in both cell types is very similar and follows the same rules. H2B.8 preferentially localises to euchromatic and intermediate TEs but not heterochromatic TEs. The histone variant can occupy genes with no/low expression but is excluded from genes with active transcription. H2B.8 also deposits to intergenic regions between genes and TEs in euchromatic chromosome arms. When compared to many different histone modifications and variants in sperm and seedlings, H2B.8 clusters with repressive chromatin marks but exhibits only a weak association. This indicates that H2B.8 distribution on chromatin is highly unique.

Having established the deposition profile of H2B.8, it was next asked whether local chromatin accessibility is altered depending on presence of the histone variant. This analysis used ATAC-seq (Assay for Transposase-Accessible Chromatin) to probe differences between WT and *htb8* mutant sperm.

2.10 H2B.8 deposition does not affect local chromatin accessibility

Earlier sections have described the role of H2B.8 in global chromatin organisation by the formation of small euchromatic foci by phase separation. However, whether H2B.8 affects local chromatin organisation is unknown. Having determined the deposition profile of H2B.8 in sperm, it was next asked if such sites become more accessible in *htb8* mutant sperm. Using ATAC-seq, genome-wide chromatin accessibility can be profiled.

First, chromatin accessibility at genes and TEs overlapped by H2B.8 peaks was assessed. In both cases, WT and *htb8* chromatin accessibility are highly correlated (Fig. 2.27, A and B; Supp. Fig. 2.13, H and I). This indicates that loss of H2B.8 does not affect local chromatin accessibility.

Additionally, analysis of genes grouped by expression, TEs grouped by chromatin environment and intergenic regions suggests that chromatin accessibility is unchanged from WT to *htb8* mutant sperm (Supp. Fig. 2.13, A to G). Therefore, it can be concluded that H2B.8 presence does not affect local chromatin accessibility in sperm.

This finding was to be expected given the features of H2B.8. The greater number of positively charged arginine residues on the histone body are proposed to increase nucleosome stability by strengthening the association with negatively charged DNA. However, such a proposed function would not necessarily be detected by ATAC-seq. This hypothesis suggests that linker DNA between nucleosomes would be unchanged; equal signals of accessibility would be called irrespective of H2B.8 presence.

Furthermore, H2B.8 deposition is proposed to protect nucleosomes from H2B ubiquitylation. *In vitro* studies of nucleosome arrays have suggested that H2Bub modifications can physically alter chromatin structure to a more open state (Fierz *et al.*, 2011). However, ChIP-seq analysis revealed that H2B.8 and H2Bub profiles are strongly anticorrelated; H2B.8 deposition is restricted to inactive genes whilst H2Bub marks highly expressed genes in both sperm and seedlings. Therefore, loss of H2B.8 could enable H2Bub deposition at previously marked genes, although this depends on whether such genes become expressed in the mutant.

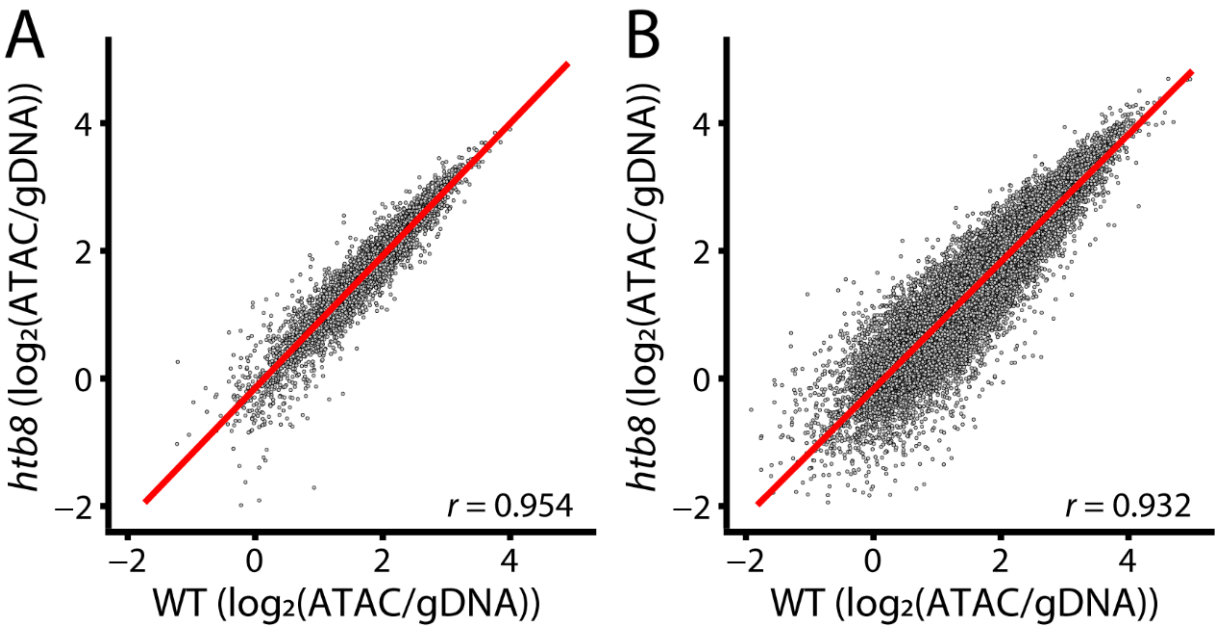


Figure 2.27: Local chromatin accessibility is unchanged at genes and TEs marked by H2B.8. (**A** and **B**) Local chromatin accessibility of WT and *htb8* mutant sperm nuclei was profiled by ATAC-seq (log_2(ATAC/gDNA)). Genes (A) and TEs (B) overlapped by H2B.8 peaks are plotted. Spearman's rank correlation coefficients (*R*) were used to account for abnormal distribution of the data; *R* values are labelled on the figure.

The question of gene and transposon expression differences between WT and *htb8* mutant sperm is addressed in the next section. H2B.8 does not affect local chromatin accessibility; however, it does mediate the formation of small euchromatic foci and change global chromatin architecture. It was considered whether disruption of three-dimensional chromatin structure and consequently loss of potential *trans* regulation could impact gene expression in the presence or absence of H2B.8.

2.11 Presence of H2B.8 does not affect transcription

H2B.8 deposits to euchromatic regions in sperm and seedlings and forms IDR-dependent phase separated foci to condense chromatin and reduce nuclear size. Native ChIP-seq revealed that H2B.8 is excluded from actively transcribed genes but can occupy some silenced genes. This section seeks to address whether loss of H2B.8 from sperm, or gain in seedlings, can affect transcription. Such analysis will indicate whether H2B.8 is able to silence genes itself, or if it is only able to deposit due to inactivity. Furthermore, given the appearance of euchromatic foci in nuclei with H2B.8, three-dimensional chromosome structure is thought to be much changed; therefore, consideration is given to disrupted *trans*-acting genome regulation.

2.11.1 Sperm cell gene and transposon expression is unchanged in *htb8* mutant

In *htb8* sperm nuclei euchromatic foci are lost, leading to a globally decondensed chromatin state and larger nuclear volume (Section 2.5). To test whether this drastic change in chromatin architecture affects gene and transposon transcription, RNA extraction followed by sequencing (RNA-seq) was performed on WT and *htb8* sperm cells.

Analysis of WT and *htb8* sperm cell RNA-seq data revealed that just one gene was downregulated in the mutant and no transposons were mis-regulated (Fig. 2.28, A and B). Gene and transposon expression in sperm cells is unchanged, irrespective of H2B.8 occupancy (Fig. 2.28, A and B). Therefore, the transcriptome of sperm cells is not dependent on H2B.8.

The fact that H2B.8-occupied genes do not become expressed in the *htb8* mutant background suggests that H2B.8 does not silence such genes. Instead, it appears that H2B.8 can deposit to genes in the absence of active transcription. Given the sharp boundaries of H2B.8 enrichment between genes and intergenic regions or TEs, it is possible that transcription itself is antagonistic to H2B.8 deposition. Therefore, in sperm, H2B.8 is dispensable for normal transcription.

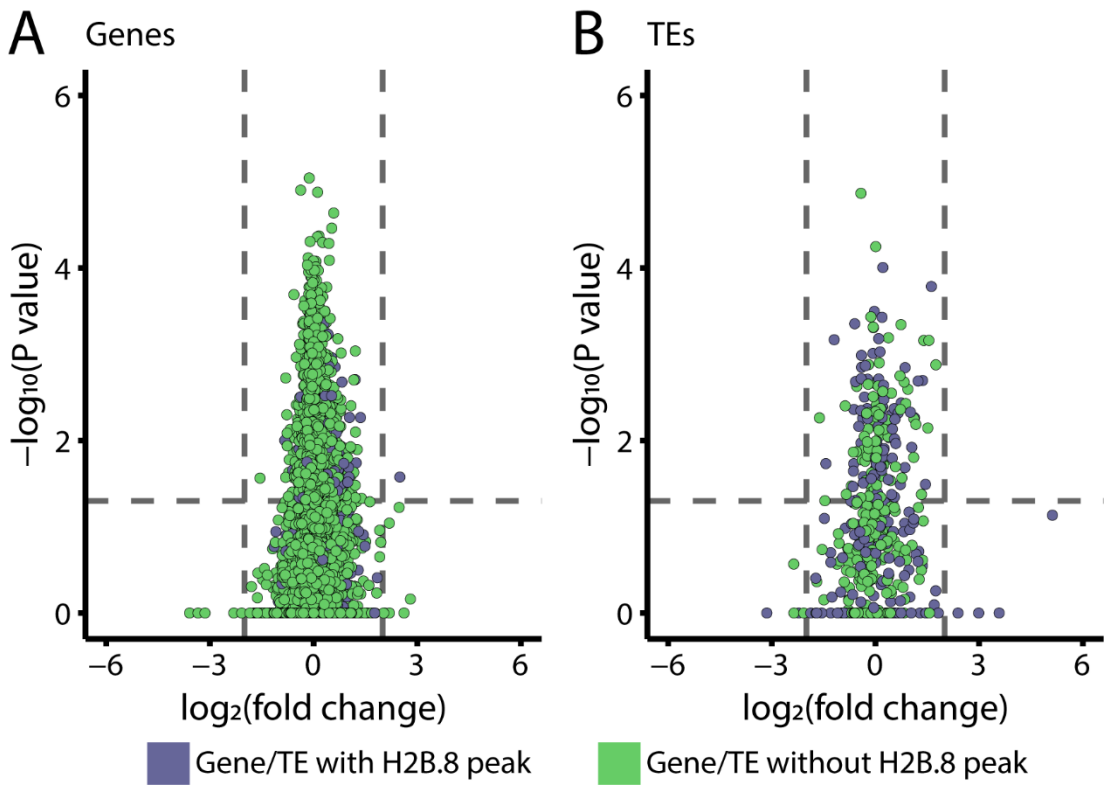


Figure 2.28: Loss of H2B.8 does not affect the sperm cell transcriptome. (**A** and **B**) Volcano plots of gene (A) and TE (B) expression between WT and *htb8* mutant sperm cells. Genes and TEs are defined as differentially expressed should they meet the following criteria: > 2 or < -2 $\log_2$(fold change) and $P < 0.05$ (likelihood ratio test; -$\log_{10}$ transformed in figure). Genes and TEs are coloured according to whether an H2B.8 peak overlaps (blue) or does not overlap (green) the genomic feature.

2.11.2 Ectopic H2B.8 does not alter gene and transposon expression in seedlings

In sperm, H2B.8 is able to condense chromatin structure without impacting upon normal gene expression. However, the number of genes expressed in sperm is substantially lower than other cell types (Fig. 2.29). In contrast to transcriptomes of other single cell types in the male sexual lineage, somatic single cell types and somatic tissues (Supp. Table 2.4), sperm cells express between a half and a third fewer genes (Fig. 2.29). This equates to expression of ~12% of genes in sperm cells, compared to 30-50% of genes in leaf or seedlings. Given the similarity of WT and *htb8* mutant sperm cell transcriptomes (Fig. 2.28, A and B), H2B.8 is clearly not causal of the comparative transcriptional quiescence observed.

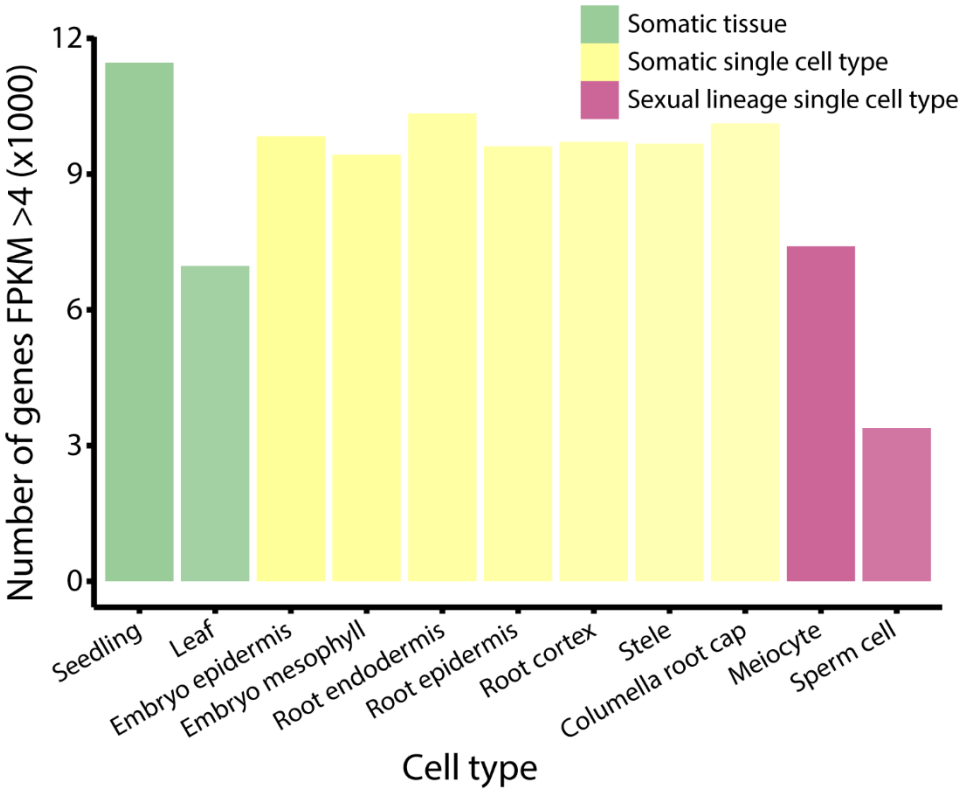


Figure 2.29: Sperm cells express far fewer genes than somatic cells. The number of expressed genes (defined as FPKM > 4) is far lower in sperm cells, compared to another sexual lineage single cell type (pink), somatic single cell types (yellow) and somatic tissues (green).

However, whether H2B.8 itself silences genes could be masked by the smaller scale of transcription in sperm; in other words, the occupied genes may not have the potential to be expressed despite the removal of H2B.8. To address this possibility, the impact of H2B.8 upon transcription in a far more active cell type was assessed. Ectopic H2B.8 in seedlings (*p35S::HTB8-eGFP*) could successfully aggregate chromatin to euchromatic foci and reduce nuclear size. RNA-seq was performed on WT and H2B.8-expressing seedlings.

A total of 23 genes were upregulated in seedlings ectopically expressing H2B.8 whilst 5 genes were downregulated (Fig. 2.30A). Regarding transposons, 8 were upregulated in the presence of H2B.8 whereas two were downregulated (Fig. 2.30B). Altogether, transcriptomes of WT and H2B.8-expressing seedlings are very similar.

Furthermore, of the few genes and transposons that were mis-expressed, there is no overall pattern for H2B.8 occupancy (Fig. 2.30, A and B). This indicates that H2B.8 deposition does not affect expression even in cell types with globally active transcription. As such, the possibility that H2B.8 silences genes can be excluded.

Additionally, despite H2B.8 inducing the formation of phase separated euchromatic foci and potentially disrupting genome regulation in-*trans*, seedling gene expression is unchanged in the presence of the histone variant. However, the *Arabidopsis* genome is very small compared to those of other flowering plants. Prominent topologically associated domains (TADs) are not observed in *Arabidopsis* when probing for three-dimensional genome contacts using Hi-C (Liu *et al.*, 2016). TADs are readily identified in flowering plant species with larger genomes, such as rice and maize, however (Liu *et al.*, 2017). Such species have many well-established enhancer elements, a feature of larger genomes that are largely uncharacterised in *Arabidopsis* (Yan *et al.*, 2019). Therefore, it is likely that *trans*-interactions and three-dimensional chromosome architecture are far more important to species with larger genomes. As such, it is important to consider that formation of H2B.8-mediated euchromatin foci may affect larger genomes differently.

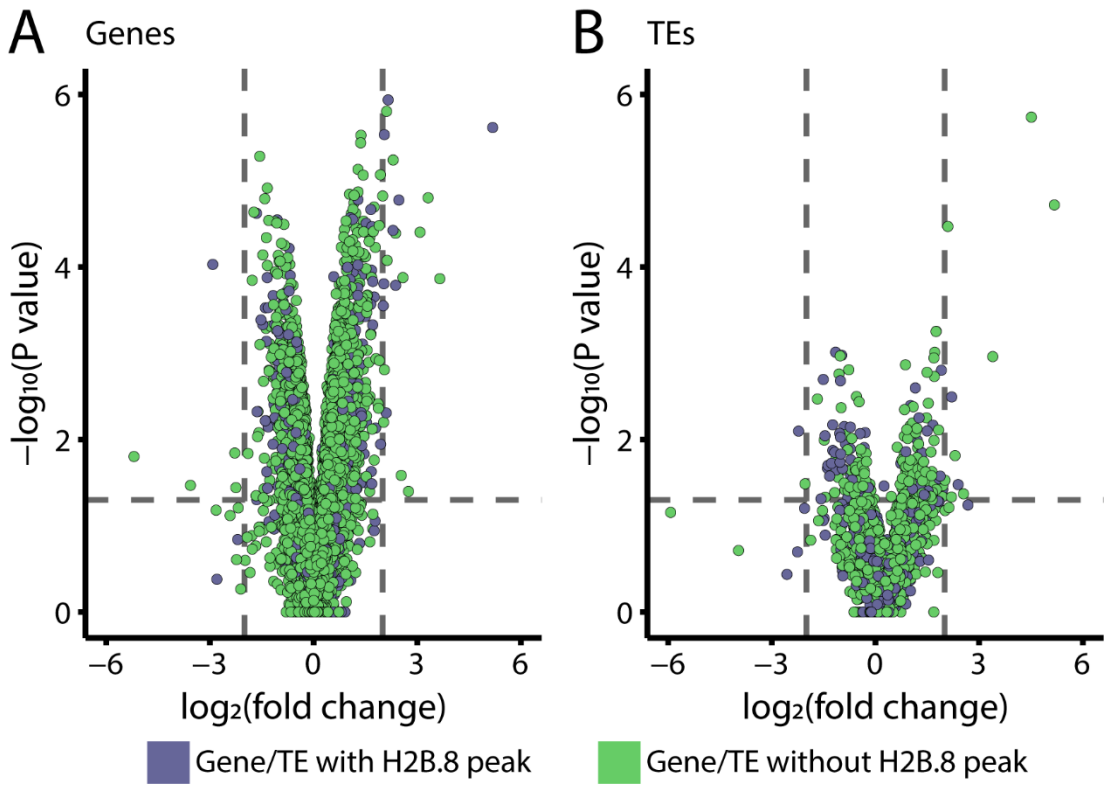


Figure 2.30: Ectopic expression of H2B.8 in seedlings does not alter the transcriptome. (**A** and **B**) Volcano plots of gene (A) and TE (B) expression between WT and H2B.8 expressing seedlings. Genes and TEs are defined as differentially expressed should they meet the following criteria: > 2 or < -2 $\log_2$(fold change) and $P < 0.05$ (likelihood ratio test; -$\log_{10}$ transformed in figure). Genes and TEs are coloured according to whether an H2B.8 peak overlaps (blue) or does not overlap (green) the genomic feature.

Collectively, in both sperm and seedlings with vastly different transcriptional activities, H2B.8 does not affect expression of genes or transposons. Therefore, H2B.8 is able to achieve chromatin condensation without affecting transcription.

Despite having reduced transcriptional activity, sperm cells still express a diverse set of genes that are critical for function (Borges *et al.*, 2008). This is in contrast to metazoans, where *de novo* transcription in mature sperm is entirely lost as a consequence of protamine-mediated chromatin condensation (Grunewald *et al.*, 2005). Therefore, it seems that H2B.8 has evolved as a novel mechanism to achieve sperm chromatin compaction without sacrificing the ability to transcribe genes. The following section addresses the evolution of H2B.8 using phylogenetics and compares chromatin compaction mechanisms among eukaryotes.

2.12 H2B.8 emerged in flowering plants

This work has established H2B.8 as a novel histone variant in *Arabidopsis*, present in sperm and mature seeds. H2B.8 functions to condense chromatin and reduce nuclear size by depositing to inactive euchromatic regions and forming IDR-dependent phase separated foci, without interrupting normal transcription. Overall, H2B.8-mediated chromatin compaction has been demonstrated to be important for male fertility. Considering the functional significance of H2B.8 to *Arabidopsis*, it was considered whether this could be a more general mechanism of sperm chromatin condensation among flowering plants.

To dissect the evolutionary history of H2B variants in plant lineages, histone protein sequences were downloaded from all available flowering plants (angiosperms; 54 species), along with representatives from other plant lineages, including non-flowering seed plants (gymnosperms; 4 species) and non-seed plants (such as bryophytes; 10 species). As out-groups, H2B variants from human and yeast (*Saccharomyces cerevisiae*) were included for phylogenetic tree construction.

The phylogeny revealed a distinct branch of H2B variants that included *Arabidopsis* H2B.8. This branch consisted of representatives from all flowering plants, except the most basal angiosperm, *Amborella trichopoda*. No non-angiosperm species had H2B variants that clustered on this phylogenetic branch. To check eukaryotes beyond this tree, BLAST was used to search with the H2B.8 protein sequence. The returned results did not identify any homologs outside of the angiosperm lineage.

Phylogenetic analysis revealed H2B.8 to be a flowering plant specific H2B variant. H2B.8 evolved in the angiosperm lineage following the divergence with the most basal member, *Amborella trichopoda* (Fig. 2.31). The fellow basal angiosperm lineage Nymphaeaceae (water lily) has an H2B.8 homolog, along with all eudicots and monocots examined. This indicates that H2B.8 evolved ~230 million years ago.

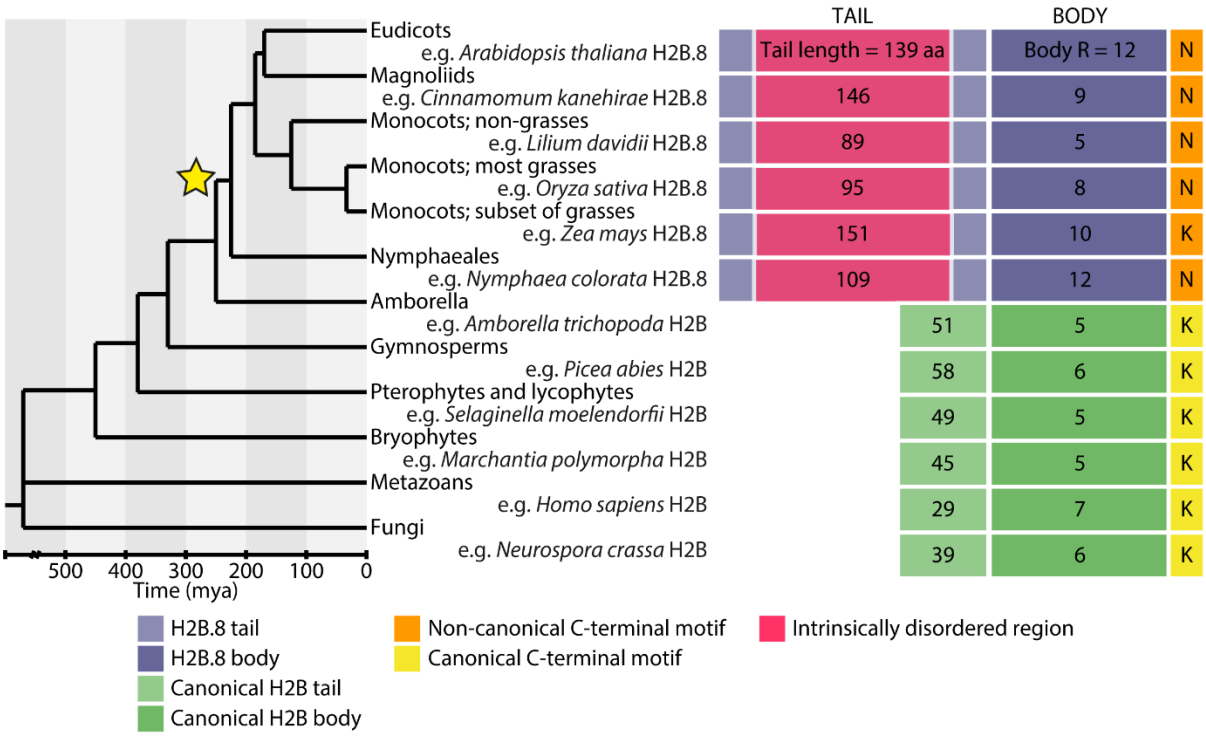


Figure 2.31: H2B.8 homologs are identified exclusively in angiosperm lineages. H2B.8 homologs (blue) are identified in all available angiosperm lineages beyond the split from *Amborella trichopoda* (star). H2B.8 homologs share the common features of a large IDR (pink) in the N-terminal tail, increased body arginine and a non-canonical C-terminal motif (orange); most with an asparagine at position 234 besides a small group of grasses, such as *Zea mays* which maintain a lysine residue. Non-angiosperm eukaryote lineages, from non-flowering seed plants and non-seed plants to metazoans and fungi, do not have H2B.8 homologs. For these lineages, a representative canonical H2B (green) is shown.

Identified H2B.8 homologs share similar features to the *Arabidopsis* homolog characterised in this work (Fig. 2.31; Supp. Table 2.5). Of the H2B.8 homologs identified; all share the larger size (Fig. 2.32B; Supp. Table 2.5). This larger size is due to the presence of an IDR insertion in the N-terminal tail domain (Fig. 2.32A). The IDRs do not share high homology between species, indicating that sequence similarity is dispensable for function. Lack of sequence homology is expected given previously reported IDRs (Uversky, 2017). Rather, the function is conserved between homologs by the inability to form defined secondary structures. The conservation of IDR presence among H2B.8 homologs is consistent with the data presented in Section 2.8, whereby the IDR of *Arabidopsis* H2B.8 is critical for the chromatin compaction mechanism.

The majority of identified H2B.8 homologs share the feature of increased arginine on the histone body at DNA-nucleosome contacts (51 out of 53 homologs; Fig. 2.31; Fig. 2.32C; Supp. Table 2.5). The highly conserved nature of this feature is consistent with the demonstrated importance of histone body arginine residues presented in Section 2.8.2.

The lysine-to-asparagine substitution at position 234 is mostly shared among H2B.8 homologs (45 out of 53 homologs; Fig. 2.31; Supp. Table 2.5). A small subset of grass species (Poaceae), including maize, maintain a lysine residue at this position. However, the lysine is positioned within a non-canonical C-terminal motif (Fig. 2.31). This lack of conservation could affect the ability of the lysine to undergo ubiquitylation as a post-translational modification, meaning it would behave similarly to H2B.8 homologs with an asparagine at 234. Experiments presented in Section 2.8.2 suggested the K-to-N substitution was least important for H2B.8 function, perhaps indicative of this feature being the least well conserved.

Interestingly, previous papers examining the histone variant composition of *Lilium* species identified H2B.8 homologs in the male gamete (Ueda and Tanaka, 1995; Ueda *et al.*, 2000; Yang, Yang and Wang, 2016). This is supportive of H2B.8 homologs being present in sperm across a diverse range of angiosperms, from eudicots to monocots.

Collectively, phylogenetic analysis of H2B variants across eukaryote lineages has revealed that H2B.8 is specific to flowering plants. Eukaryotes with swimming sperm, such as metazoans and non-seed plants, mostly utilise protamines to achieve extreme chromatin compaction at the expense of transcription. Flowering plants do not have motile sperm, but it seems compaction is still necessary for normal fertility. H2B.8 likely represents an independent invention of flowering plants to achieve a moderate level of condensation whilst enabling the required transcription.

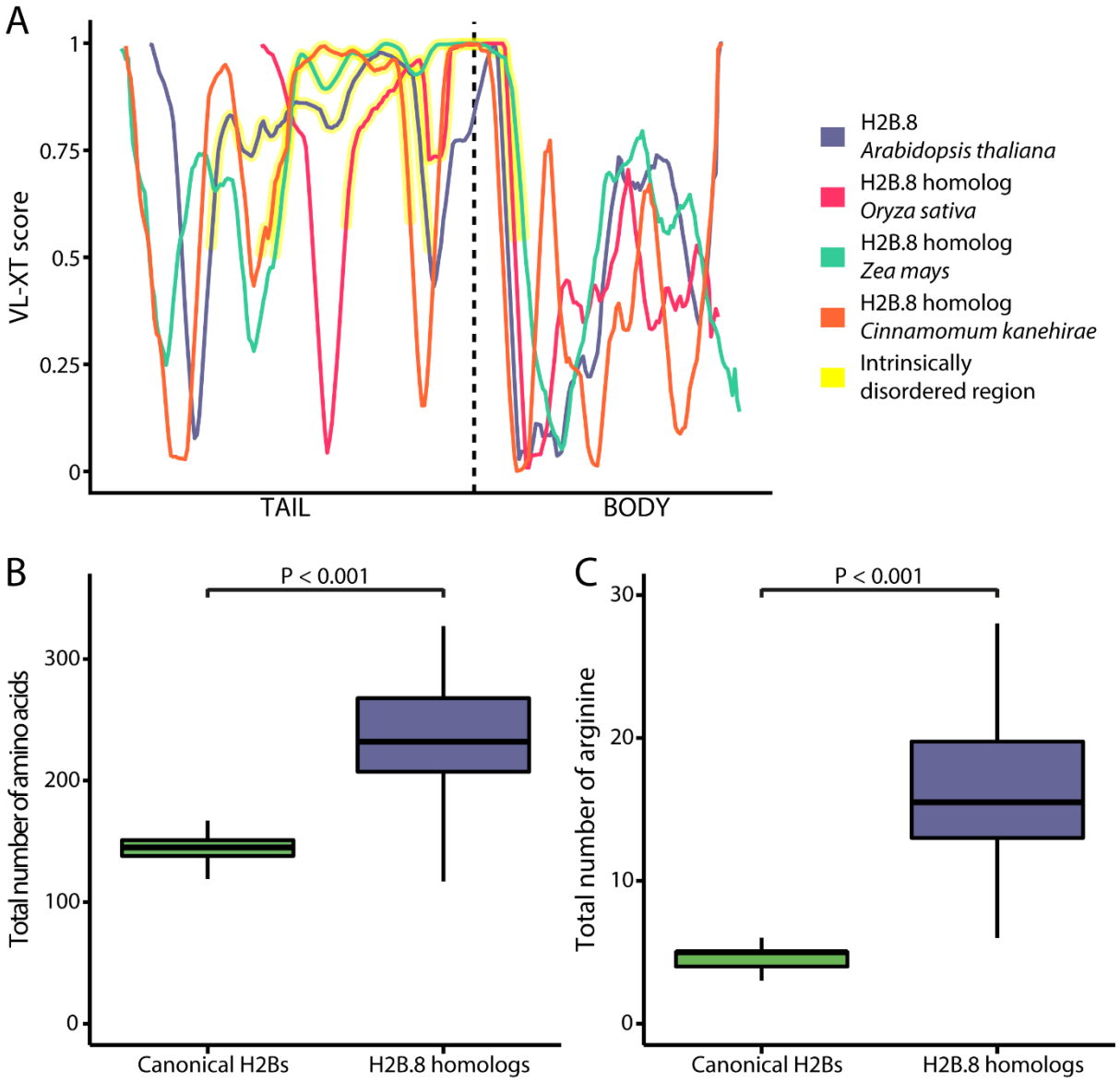


Figure 2.32: H2B.8 homologs share common features. (**A**) VL-XT score predictions of intrinsically disordered regions (yellow) for *Arabidopsis* H2B.8 (blue) and representative homologs. Histones are aligned at the transition between tail and body domains. (**B**) H2B.8 homologs (blue) are significantly larger than canonical variants (green), owing to the IDR insertion in the N-terminal tail. (**C**) H2B.8 homologs (blue) have significantly more arginine residues than canonical variants (green), largely owing to increased numbers in the histone body domain. Boxplots (B and C) show median (thick black bar) and first and third quartiles, with lower and upper whiskers extending to 1.5 times the interquartile range of the first and third quartiles, respectively. Statistical significance was determined in (B and C) by a Student's t-test; $P < 0.001$; H2B.8 homologs N = 58; canonical H2Bs N = 670.

2.13 Conclusion

This chapter has presented evidence for H2B.8 as a primary determinant of sperm chromatin compaction in the flowering plant, *Arabidopsis thaliana* (Fig. 2.33, A and B).

Using proteomics, H2B.8 was identified as a novel sperm specific histone variant. Screening of a reporter line confirmed its presence in sperm and also mature seeds, and absence from other somatic tissues. Analysis of *Arabidopsis* H2B variants revealed three distinct features unique to H2B.8; a large IDR in the N-terminal tail, increased arginine residues on the histone body and a K-to-N substitution at position 145/234 in the C-terminal motif.

H2B.8 was shown to be important for male fertility and crucial for chromatin structure in sperm, with increased nuclear size and loss of distinct foci in the mutant. Sperm foci were revealed to colocalise with H2B.8 and be euchromatic in nature. Functional analysis of the H2B.8 IDR demonstrated that this feature defined phase separation properties *in vitro* and *in vivo*. A minor role was established for the histone body arginine residues in determining sperm nuclear size.

ChIP-seq revealed H2B.8 to be localised in chromosome arms, preferentially depositing to euchromatic and intermediate TEs and spreading to neighbouring intergenic regions. H2B.8 is depleted from genes, only depositing to a subset of non-expressed genes. Accordingly, H2B.8 presence in sperm and ectopically in seedlings did not affect transcription.

Phylogenetic analysis revealed H2B.8 to be a flowering plant specific histone variant, evolving at the branch from the most basal angiosperm, *Amborella trichopoda*. Homologs are identified throughout angiosperm lineages, largely sharing the features of *Arabidopsis* H2B.8.

Collectively, H2B.8 represents a novel mechanism of chromatin compaction, whereby euchromatin is condensed into small foci by phase separation (Fig. 2.33, A and B). This mechanism accomplishes sperm chromatin compaction without sacrificing transcription, thus striking a balance that cannot be achieved by a protamine-mediated approach. Broadly, H2B.8 could represent a general mechanism in flowering plants with immotile sperm, where compaction is not so imperative as is the case for eukaryote lineages with motile sperm.

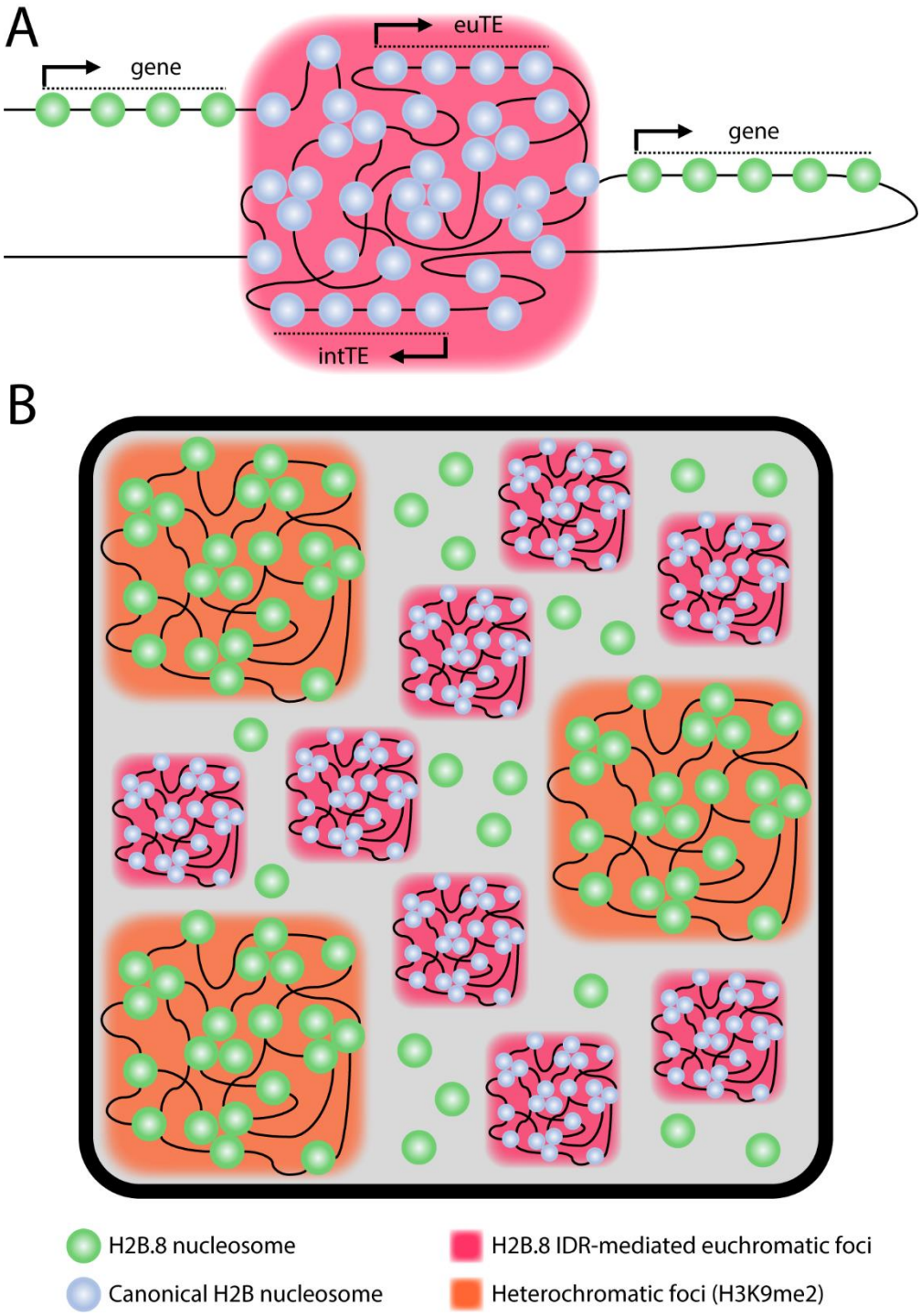


Figure 2.33: H2B.8 localises to inactive euchromatin and forms IDR-dependent phase separated foci. (**A**) H2B.8 (blue) deposits to euchromatic and intermediate TEs (euTEs; intTEs) and spreads to neighbouring intergenic regions. H2B.8 is largely excluded from genes, particularly those with active transcription, where canonical H2B variants (green) are maintained. H2B.8 forms IDR-dependent phase separated foci (pink). (**B**) Sperm chromatin compaction is achieved by H2B.8-mediated euchromatic foci formation throughout the nucleoplasm. Heterochromatic foci (orange) are maintained in sperm.

2.14 Discussion

Chromatin has previously been shown to have liquid-like properties, with nucleosome arrays undergoing phase separation (Gibson *et al.*, 2019). Several studies have identified and characterised proteins that facilitate heterochromatin domain formation by phase separation (Larson *et al.*, 2017; Strom *et al.*, 2017; Wang *et al.*, 2019). However, euchromatin aggregation has never before been demonstrated. Therefore, H2B.8 represents an entirely novel mechanism of chromatin organisation.

Furthermore, despite chromatin itself having liquid-like properties, specific histone variants have not previously been shown to promote phase separation compared to canonical counterparts. The ability of specific histone variants to undergo phase separation could be a more general mechanism of chromatin biology. Unlike phase separation proteins that act upon the chromatin fibre, histone variants are embedded within nucleosomes. As such, specific genomic regions can be more stably targeting owing to the deposition dynamics. Additionally, as a result of the stable rather than transient localisation, droplet dynamics could be less fluid; perhaps indicated by the gel-like properties of H2B.8-mediated foci.

On the topic of H2B.8 deposition and the IDR, evidence suggests that the IDR is critical for H2B.8 localisation to euchromatin. In fact, upon IDR deletion, H2B.8 solely inhabits heterochromatic regions rather than behaving like a canonical H2B and distributing evenly across chromatin (Fig. 2.14). This suggests that the biophysical properties of the IDR define euchromatic deposition. However, H2B.8ΔIDR essentially has a canonical H2B tail, so how it then deposits only to heterochromatin is likely due to features of the histone body domain. The increased body arginine residues and K-to-N substitution were only characterised to the extent that they determine sperm nuclear size, more work could be undertaken to establish greater roles for such features in H2B.8 function.

A further interesting discussion point is how does impaired chromatin compaction and increased nuclear size reduce male fertility. An imbalance of histones to protamines in mammalian sperm has been associated with male infertility (Cho *et al.*, 2001). This imbalance leads to incomplete chromatin compaction and subsequent aberrant transcription, often linked to sperm failure (Ostermeier *et al.*, 2002, 2004). However, *htb8* mutant sperm do not have different transcription compared to WT. As such, the suggestion of aberrant

transcription can be ruled out. An alternative hypothesis is that sperm must be condensed in order to correctly fuse with the female egg cell. Unfortunately, such an idea is difficult to test *in planta* and would require careful experimental design to explicitly test the hypothesis. Another possibility is that larger mutant sperm nuclei are more difficult to transport through the pollen tube and thus fail to reach the female gamete. Further experiments could utilise heterozygous H2B.8-eGFP in the *htb8* background, meaning pollen from the same plant has both mutant and complemented sperm. Live imaging would reveal the transport of nuclei through the pollen tube and determine whether *htb8* sperm exhibit transport difficulties.

Work in this chapter has focused upon the role of H2B.8 in sperm, however H2B.8 is also present in mature seeds. The study of ectopically expressed H2B.8 in seedlings has revealed that the histone variant behaves vary similarly in a vastly different cell type. Therefore, it is likely that H2B.8 in mature seeds would still deposit to inactive euchromatin and form small foci. Indeed, nuclei of mature seeds are known to become more condensed and smaller in size (Van Zanten *et al.*, 2011), correlating with the presence of H2B.8. The functional significance of such nuclear compaction in mature seeds is unknown. Future work could seek to establish whether H2B.8 is responsible for nuclear compaction and whether this impacts upon seed dormancy or other such processes.

Phylogenetic work in this chapter showed that H2B.8 is exclusively present in angiosperms. The function of H2B.8 in *Arabidopsis* has been addressed in this work, but genome sizes and chromatin architecture can be highly variable across lineages. For example, maize has a genome 20 times larger than *Arabidopsis* and is abundant with islands of heterochromatin throughout the chromosome arms, rather than being confined to the pericentromere (West *et al.*, 2014). How H2B.8 would deposit in such species and how such interspersed heterochromatin would affect euchromatic foci formation are interesting questions that need to be addressed.

2.15 Materials and methods

2.15.1 Plant growth conditions

Plants were grown under long day (16 hr light; 8 hr dark) conditions at 22 °C and 70% humidity. Seedlings were grown on germination medium (GM) plates without glucose under the same conditions.

2.15.2 Sperm and vegetative nuclei isolation by FACS

Sperm and vegetative nuclei were isolated from pollen by Fluorescence-Activated Cell Sorting (FACS) as previously described (Borges *et al.*, 2012). Briefly, open flowers were collected and vortexed with Galbraith buffer (45 mM $MgCl_2$, 30 mM sodium citrate, 20 mm MOPS, 0.1 % Triton X-100, pH 7.0). Released pollen was filtered through a 40 µm cell strainer to remove flower parts. Pollen were concentrated by centrifugation at 2600 x g for 5 mins at 4 °C and resuspended with 200 µl Galbraith buffer. An equal volume of glass beads (Merck) was added, and samples were vortexed to release nuclei. Glass beads were washed with Galbraith buffer and the samples were passed through 40 µm, 20 µm and then 10 µm cell strainers (Corning) sequentially. The volume was adjusted to 20 ml with Galbraith buffer and then centrifuged at 1500 x g for 5 mins at 4 °C to wash nuclei, the pellets were washed twice. The nuclei were resuspended in 1 ml Galbraith buffer and stained with 100x SYBR Green (10000x stock; Invitrogen). Nuclei were incubated at 4 °C for 10 mins in the dark before proceeding to FACS.

FACS was undertaken using a BD FACSMelody machine (BD Biosciences). Sperm and vegetative nuclei populations were formed owing to the smaller size and lesser SYBR Green staining of sperm nuclei (Fig. 2.34, A to C). Collected nuclei were centrifuged at 1500 x g for 5 mins at 4 °C, the supernatant was removed, and the pellet was stored at -70 °C.

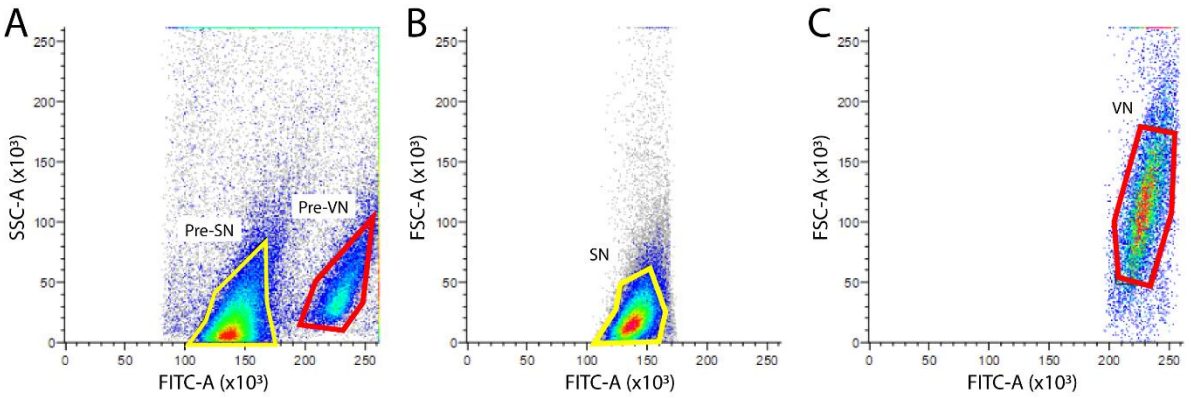


Figure 2.34: Example sperm and vegetative nuclei flow cytometry. (**A**) Sperm (yellow) and vegetative (red) nuclei were pre-sorted according to size (Side Scatter Area; SSC-A) and SYBR Green staining (FITC-A). (**B** and **C**) Populations were refined according to a second size metric (Forward Scatter Area; FSC-A).

2.15.3 Total protein extraction and mass spectrometry

Sperm and vegetative nuclei isolated by FACS were pooled and 0.45 volumes of 3.2 x lysis buffer (10% SDS, 100 mM TEAB, pH 7.55) was added. Nuclei were lysed at 95 °C for 5 min, then centrifuged at 13000 x g for 8 min at RT. Lysate was moved to a new tube. One tenth volume 12% phosphoric acid was added and mixed by pipetting. Then, six times volumes of S-Trap buffer (90% aqueous MeOH, 100 mM TEAB, pH 7.1) was added and mixed by pipetting. Protein was loaded to an S-Trap Micro column (Protifi) by centrifugation, 4000 x g for 30 s. The column was washed three times with 150 µl S-Trap buffer. Protein was digested on column with 4 µg trypsin in 50 mM TEAB at 47 °C for 1 hr. Peptides were eluted sequentially by centrifugation (4000 x g for 30 s) with 40 µl 50 mM TEAB, 40 µl 0.2% formic acid and 35 µl 50% ACN 0.2% formic acid.

Mass spectrometry was performed by Dr Gerhard Saalbach (Proteomics Facility, John Innes Centre) as follows. The eluted peptide solution was dried down, and the peptides dissolved in 0.1% TFA / 3% acetonitrile. Aliquots were analysed by nanoLC-MS/MS on an Orbitrap Fusion Tribrid mass spectrometer coupled to an UltiMate 3000 RSLCnano LC system (Thermo Fisher Scientific). The samples were loaded and trapped using a pre-column with 0.1% TFA at 20 µl/min for 3 min. The trap column was then switched in-line with the analytical column (nanoEase M/Z column, HSS C18 T3, 100 Å, 1.8 µm; Waters) for separation using the following long gradient of solvents A (water, 0.05% formic acid) and B (80% acetonitrile, 0.05% formic acid) at a flow rate of 0.3 µl/min : 0-3 min 3% B (trap only); 3-14 min linear increase B to 13%; 14-113 min increase B to 39%; 113-123 min increase B to 55%; followed by a ramp to 99% B and re-equilibration to 3% B.

Data were acquired with the following mass spectrometer settings in positive ion mode: MS1/OT: resolution 120K, profile mode, mass range *m/z* 300-1800, AGC $4e^5$, fill time 50 ms; MS2/IT: data dependent analysis was performed using HCD fragmentation with the following parameters: top30 in IT rapid, centroid mode, isolation window 1.6 Da, charge states 2-5, threshold $1.9e^4$, CE = 30, AGC target $1.9e^4$, max. inject time 35 ms, dynamic exclusion 1 count, 15 s exclusion, exclusion mass window ±5 ppm.

Recalibrated peaklists were generated with MaxQuant-1.6.1.0 (Tyanova, Temu and Cox, 2016) in LFQ mode using the TAIR10_pep_20101214 *Arabidopsis thaliana* protein sequence database (TAIR, 35386 entries) plus the MaxQuant contaminants database (245 entries). The quantitative LFQ results from MaxQuant with default parameters were used together with search results from an in-house Mascot Server-2.4.1 (Matrixscience) on the same databases. For this search a precursor tolerance of 6 ppm and a fragment tolerance of 0.6 Da was used. The enzyme was set to trypsin/P with a maximum of 2 allowed missed cleavages; oxidation (M), acetylation (protein N-term) were set as variable modifications; carbamido-methylation (C) as fixed modification. The Mascot search results were imported into Scaffold-4.11.0 (Proteome Software) using identification probabilities of 99% for proteins and 95% for peptides.

2.15.4 Generation and genotyping of CRISPR mutants

Null alleles of *HTB8* were generated by CRISPR-Cas9 (Castel *et al.*, 2019). CHOPCHOP (Labun *et al.*, 2019) was used to design four sgRNAs that target the 5' end of *HTB8* with low predictions of off-target effects. Using the Golden Gate cloning system (Castel *et al.*, 2019), sgRNAs were cloned to level one vectors (pICH47751, pICH47761, pICH47772, pICH47781) before assembly with FAST-Red (pICSL11015) and Cas9 driven by the *YAO* promoter (BCJJ345B) to a level two expression vector (pICSL4723). The construct was transformed via *Agrobacterium tumefaciens* strain GV3101 using floral dip (Clough and Bent, 1998) to WT Col-0 *Arabidopsis thaliana*. Successful T1 transformants were selected by FAST-Red expression.

Mutations were screened by Sangar sequencing. Line #7 had a single base deletion at 76 bp, leading to a premature stop codon after 33 amino acids (Supp. Fig. 2.1). Line #3 had a 12 bp deletion after 67 bp and another of 5 bp after 92 bp, translating as a truncated to 54 amino acids protein (Supp. Fig. 2.1). Lines #7 and #3 were taken to the next generation in order to obtain homozygous mutants. Absence of FAST-Red and validation by PCR confirmed the removal of the Cas-9.

htb8 #7 was genotyped using a Derived Cleaved Amplified Polymorphic Sequences (dCAPS) approach (Supp. Table 2.1). PCR products were digested by EcoNI, successful digestion with two products of 230 bp and 20 bp indicated the homozygous mutant allele. *htb8* #3 was genotyped by PCR, a single smaller band signalled a homozygous mutant (Supp. Table 2.1).

2.15.5 Molecular cloning and plant transformation

The H2B.8 reporter line constructs (*pHTB8::HTB8-eGFP* and *pHTB8::HTB8-Myc*) were generated using MultiSite Gateway Technology (Thermo Fisher Scientific). The *HTB8* promoter was taken as the immediate ~2 kb upstream of the gene and ligated to pDONR P4P1r. The *HTB8* locus was cloned to pDONR 207 without the stop codon. C-terminal tags of eGFP or 3xMyc were cloned into pDONR P2r-P3 from existing constructs as template. Entry clones were combined into the pK7m34GW vector. The expression clones were transformed to *Agrobacterium tumefaciens* strain GV3101 and then to the *htb8* #7 mutant by floral dip (Clough and Bent, 1998). Successful transformants were selected on GM plates with kanamycin. Expression of the constructs was assessed by confocal microscopy of the eGFP tag or visualisation of the Myc tag by immunostaining (Borg, Buendía and Berger, 2019). Selected lines were checked for transgene copy number to ensure for single insertion events.

The ectopic H2B.8 line (*p35S::HTB8-eGFP*) was similarly generated by MultiSite Gateway (Thermo Fisher Scientific). The 35S promoter was cloned to pDONR P4P1r from an existing expression clone and combined with the *HTB8* and eGFP clones to pK7m34GW. Transformation and selection were undertaken in the same way as for reporter lines.

Ectopic H2B.8ΔIDR (*p35S::HTB8ΔIDR-YFP*) was generated by overlapping PCR to remove the IDR sequence whilst H2B.8 (*p35S::HTB8-YFP*) and H2B.2 (*p35S::HTB2-YFP*) were simply cloned from gDNA. Products were ligated to pCAMBIA1300 vector backbone containing the 35S promoter and a C-terminal YFP using the In-Fusion system (Takara Bio). Constructs were used for transient expression in tobacco (Section 2.14.12) and also stably transformed to WT *Arabidopsis* as previously described.

Chimeric H2B sequences were generated and cloned to the pDONR 207 entry vector. Domain swaps were generated by overlapping PCR, whilst mutations were introduced by using primers with site differences. Entry vectors were ligated to the pK7m34GW expression clone with the *HTB8* promoter and 3xMyc sequences using MultiSite Gateway (Thermo Fisher Scientific). Transformation and selection were performed as per the Myc-tagged reporter line.

2.15.6 Male transmission assay

Male transmission of the *htb8* #7 allele was assessed genetically. Heterozygous *htb8* #7 pollen was used to fertilise WT females. The inheritance of mutant versus WT alleles was determined by dCAPS genotyping, as described in section 2.13.2. A total of 743 F1 progeny were genotyped and the ratio of WT to *htb8* #7 +/- was determined. Statistical significance was tested using the Chi-squared test in R-3.6.0.

2.15.7 Quantification of silique lengths and aborted seeds

Mature siliques from the main stem and above the final branch were removed from several individual plants for WT and *htb8* #7. Siliques were flattened and imaged with a ruler for scale. Quantification was undertaken using the line segment tool in ImageJ-1.51, with the scale set by the ruler. R was used for statistical analysis and ggplot2 of the Tidyverse suites was used for plotting of the data (Wickham *et al.*, 2019). To assess embryo abortion, mature siliques were dissected and imaged using a Leica S8AP0 stereo microscope.

2.15.8 Histone alignments and structural predictions

Alignments of histone protein sequences was performed using CLC Main Workbench software-8.1 (QIAGEN).

Predictions of intrinsic disorder were undertaken by PONDR (Romero *et al.*, 2001), using the VL-XT algorithm. Raw data was plotted using ggplot2 in R.

Structural predictions of the histone body domains of H2B.8 and H2B.2 were obtained using Phyre2 (Kelley *et al.*, 2015). Predictions were imported to Pymol-2.3.3 for further analysis (Schrödinger, 2015). Alignment with a crystal nucleosome structure (PDB = 1KX5) gave strong RMSD (Root-Mean-Square Deviation) values of 0.810 and 0.259 for the H2B.8 and H2B.2 body domains, respectively.

2.15.9 Confocal microscopy

For images of H2B.8 incorporation through male sexual lineage development (Fig. 2.1C), microspores and pollen were isolated as described previously (Borges *et al.*, 2012). In short, young buds were ground in Pollen Extraction Buffer (PEB; 10 mM $CaCl_2$, 2 mM MES, 1 mM KCl, 1% H_3BO_3, 10% sucrose, pH 7.5). The released material was filtered through a 35 µm cell strainer. Hoechst was added to the solution before centrifugation at 800 x g for 5 mins. The pellet was pipetted to a slide and covered with a coverslip. Images were obtained using a Leica SP8 confocal microscope.

Young embryos were obtained by performing crossing of *pHTB8::HTB8-eGFP* plants and then waiting ~8 hrs before imaging the fertilised products stained with propidium iodide (Pillot *et al.*, 2010; Hamamura *et al.*, 2011). The 2-cell embryo was the earliest developmental stage observed (Fig. 2.1D). Mature embryos were dissected from dry seeds using a stereo microscope. Mature embryos and seedlings (leaf and root; Fig. 2.1D) were stained in PBS with 0.1% Triton X-100 and 0.5 µg/ml DAPI for ~10 min prior to microscopic examination with a Zeiss 880 microscope, using Airyscan mode. Root nuclear size was quantified from images of root tips imaged in this way.

Whole pollen was imaged for quantification of sperm nuclear size. DAPI (2 µg/ml) was added to the pollen suspension before spinning down. The pellet was pipetted to a slide and covered with a coverslip. Images were obtained using a Leica SP5 confocal microscope.

2.15.10 Quantification of confocal images

Sperm nuclear size was quantified from whole pollen confocal images using a semi-automated pipeline in ImageJ as previously described (Kalyanikrishna, Mikulski and Schubert, 2020). Briefly, auto-threshold was used to obtain nuclei and then processed using Gaussian blur to smooth edges. The auto-threshold was repeated and then nuclei were selected using the wand tool. Measurements were then obtained for nuclear area (μm^2).

Somatic nuclei selected for analysis were vascular cylinder cells in the elongation zone of the root tip. Such nuclei were selected owing to the ability to accurately identify the cell type within the tissue. Using ImageJ, Z-stacks were divided into substacks of different cell layers within the root tip. Maximum intensity projections were then obtained to account for slight differences in the depth of nuclei. Images were then analysed in the same semi-automated way as per sperm nuclei.

Quantified nuclei were analysed in R. ANOVA was performed to assess for statistical significance within the data. The Shapiro-Wilk test was used to confirm normality in the data distribution. Pairwise comparisons were made using the Tukey test. Data was plotted using ggplot2 in R.

2.15.11 3D Structured Illumination Microscopy

Sperm and vegetative nuclei were isolated from pollen as described in section 2.14.2 and resuspended in 200 µl Galbraith buffer (45 mM $MgCl_2$, 30 mM sodium citrate, 20 mm MOPS, 0.1 % Triton X-100, pH 7.0). Nuclei were extracted from seedlings by finely chopping with a razor blade in lysis buffer (15 mM Tris pH 7.5, 2 mM EDTA, 0.5 mM spermine, 80 mM KCl, 20 mM NaCl, 0.1% Triton X-100). Suspension was filtered through a 35 µm filter to a 1.7 ml tube. Nuclei were pelleted at 500 x g for 3 mins and resuspended in 200 µl lysis buffer.

Nuclei were fixed in solution with 4% MeOH-free formaldehyde (Pierce) for 5 mins. HiQA No. 1.5H coverslips (CellPath) were washed with 10% HCl for 30 mins and then washed three times in H_2O for 5 mins to remove any impurities. Fixed nuclei were spun onto coverslips at 500 x g for 3 mins using a Shandon Cytospin 2. Fixation was repeated by blotting nuclei with 4% MeOH-free formaldehyde for 5 mins. Formaldehyde was removed and coverslips were washed three times in PBS, 3 mins per wash. Nuclei were blocked with 3% BSA in PBS with 0.1% Tween-20 (PBST) for 30 min in a humidified chamber. If performing immunostaining, antibodies were diluted 200-fold in 3% BSA in PBST and then blotted to nuclei on coverslips. Antibody incubation occurred overnight at 4 °C. Primary antibodies were removed by washing three times with PBST for 5 mins. Secondary antibodies were diluted similarly to primary and then added to nuclei. Incubation occurred for 1 hr at RT in a humidified chamber. If not performing immunostaining, the protocol resumes at this point. PBST washes were repeated as before. Nuclei were stained in the dark with either DAPI or SYBR Green at 2 mg/µl or 100 x dilution, respectively for 5 mins. DNA stain was removed by washing in H_2O for 5 mins. Coverslips were adhered to slides in 13 µl VECTASHIELD H-1000 mounting medium.

Nuclei were imaged using a 63x oil immersion lens on a Zeiss Elyra PS.1 super-resolution microscope. Three dimensional reconstructions for SIM were undertaken using Zeiss Zen Black software-3.2.

2.15.12 Quantification of 3D-SIM images

Super-resolution images are shown as maximum intensity projections throughout to account for the three-dimensional nature of 3D-SIM images. Intensity profiles associated with images are generated in ImageJ using the Plot Profile tool. Raw data was plotted using ggplot2 in R.

To obtain quantitative metrics between WT and *htb8* #7 sperm, FociPicker3D was used (Du *et al.*, 2011). Image scales were determined according to the raw data to correctly calculate voxel sizes. The tolerance setting was set to 2000 to account for the absolute intensity difference between the background and object. Raw outputs for 49 and 48 nuclei were analysed for WT and *htb8* #7, respectively. Statistical significance was determined by Student's t-test and data was plotted using ggplot2 in R.

2.15.13 Histone purification from *E. coli*

Histones H2B.8 and H2B.2 were cloned into the pET28a+ vector with a non-cleavable C-terminal 8×His-tag and then transformed to *Escherichia coli* strain BL21 (DE3; Tiangen). Cells were grown to OD 0.8 at 37 °C in LB media with 30 μg/ml Kanamycin. Histone expression was induced by addition of 0.1 mM isopropyl-β-d-thiogalactopyranoside (IPTG) and incubation overnight at 16 ℃. Cells were collected by centrifugation at 4000 rpm for 20 min and resuspended in lysis buffer (20 mM Tris-HCl, 500 mM NaCl, pH 8.0). Cells were lysed by ultrasonication and debris was pelleted at 20,000 × g. The supernatant was applied to a 5 ml HisTrap HP column (Cytiva) on AKTA pure (Cytiva). Target proteins were eluted at ~250 mM Imidazole concentration during gradient elution. The peaks eluted were applied to a Superdex 200 Increase 3.2/300 (Cytiva) gel filtration column, then dialyzed and concentrated using *in vit*ro phase separation assay buffer (20 mM HEPES, 150 mM NaCl, pH 8.0).

The H2B.8ΔIDR sequence was cloned to a modified pET11 expression vector (Novagen) as described (Fang *et al.*, 2019). The expression vector contains a solubility MBP tag followed by a TEV cleavage site and a GFP tag upstream of the insertion site and a non-cleavable C-terminal 8×His-tag at downstream of the insertion site. Proteins were expressed and lysed as before, besides using 50 μg/ml Ampicillin for selection. Purification was performed as previous, with target proteins were eluted by between 300mM and 500mM Imidazole concentration during gradient elution. For *in vitro* phase separation assays, the MBP tag was cleaved by incubating with ~0.02 mg/ml 6×Histag-TEV protease overnight at 4 °C, and the cleaved GFP-H2B.8ΔIDR was tested by western blot using His-tag antibody (Huaxingbio).

2.15.14 High performance liquid chromatography of purified histones

An Agilent 1200 HPLC system was used for analysing protein oligomer states. Chromatographic separation was achieved on a Bio CoreSEC-300 column (Thermo Scientific). The mobile phase was composed of 20 mM HEPES, 150 mM NaCl, pH 8.0. A flow rate of 0.26 ml/min was used.

2.15.15 *In vitro* phase separations assays

In vitro experiments were performed in phase separation assays buffer (20 mM HEPES, 100 mM NaCl, pH 7.4). *In vitro* experiments were recorded on 384 low-binding multi-well 0.17 mm microscopy plates (In Vitro Scientific) and sealed with optically clear adhesive film. Imaging was performed with a NIKON A1 microscope equipped with a 100x oil immersion objective. NIS-Elements AR Analysis was used to analyse images.

2.15.16 *In vitro* FRAP

In vitro FRAP experiments were carried out with a NIKON A1 microscope equipped with a 100x oil immersion objective. Droplets were bleached with a 488- or 561-nm laser pulse (3 repeats, 70% intensity, dwell time 1 s). Images were processed using the ImageJ StackReg plugin. Post-bleach intensity was normalised to pre-bleach levels to obtain a measure of recovery. Data was plotted using ggplot2 in R.

2.15.17 Transient expression in tobacco

The In-Fusion cloning system (Takara Bio) was used to generate *p35S::HTB8-YFP*, *p35S::HTB8ΔIDR-YFP*, and *p35S::HTB2-YFP* in the pCAMBIA1300 expression vector. Constructs were transformed to *Agrobacterium tumefaciens* strain GV3101 by heat shock.

Infiltration of *Nicotiana benthamiana* was undertaken as previously described (Sparkes *et al.*, 2006). Briefly, *Agrobacterium* was grown overnight at 28 °C in 200 ml Lysogeny broth (LB) with appropriate antibiotic. Bacteria were centrifuged at full speed for 10 mins. The pellet was resuspended with MMA (10 mM MES pH 5.6, 10 mM $MgCl_2$, 200 µM acetosyringone) to a final OD_{600} of ~0.6. Suspension were incubated for 3 hrs at RT before injection to *Nicotiana benthamiana*. Two days after infiltration, DAPI (1 µg/ml) was injected to transformed leaves immediately prior to imaging. Microscopy was undertaken using a Leica SP8 confocal microscope.

2.15.18 *In vivo* FRAP

Root nuclei expressing H2B.8 (*p35S::HTB8-eGFP*) were imaged using Airyscan mode with a Zeiss 880 confocal microscope. Individual H2B.8 foci were photobleached and imaged over time. Images were processed using the ImageJ StackReg plugin. Post-bleach intensity was normalised to pre-bleach levels to obtain a measure of recovery. Data was plotted using ggplot2 in R.

2.15.19 Pollen and seedling native ChIP-seq library preparation

Pollen was obtained by collecting open flowers and vertexing in Galbraith buffer. Pollen was concentrated and resuspended in Nuclei Isolation Buffer (NIB; 0.25 M sucrose, 15 mM PIPES pH 6.8, 5 mM MgCl2, 60 mM KCl, 15 mM NaCl, 1 mM CaCl2, 0.9% Triton X-100, 1 mM PMSF, 1x proteinase inhibitors Cocktail (Roche)). Nuclei were released from pollen by breakage with glass beads. Filtration through 40 µm and then 10 µm cell strainers (Corning) removed debris, leaving nuclei in suspension.

Nuclei were released from 10-day-old seedlings by grinding in liquid N_2 with a pestle and mortar. NIB was added to homogenise the powder to solution, for 15 mins. Debris was removed via filtration with two layers of miracloth.

Nuclei suspensions from pollen and seedlings were centrifuged at 4000 x g and 4 °C for 10 mins. The pellets were resuspended with ice-cold TM2 (50 mM Tris-HCl, 2 mM MgCl2, 0.25 M sucrose, 1 mM PMSF, 1 x proteinase inhibitors Cocktail). Nuclei were pelleted again by centrifugation at 4000 x g and 4 °C for 5 mins. The pellet was resuspended with MNase digestion buffer (50mM Tris-HCl pH 7.5, 5mM CaCl2, 0.25 M sucrose, 1 mM PMSF, 1 x proteinase inhibitors Cocktail) containing an appropriate amount of MNase enzyme (New England Biolabs), dependent on input material amount. Digestion occurred at 37 °C for 10 mins. The reaction was stopped by addition of EDTA, to a final concentration of 25 mM. One tenth volume of 1% Triton X-100 and 1% Sodium Deoxycholate was then added and the reaction was left on ice for 15 mins. Low salt buffer (50 mM Tris-HCL pH 7.5, 10 mM EDTA, 150 mM NaCl, 0.1% Triton X-100, 1 mM PMSF, 1 x proteinase inhibitors Cocktail) was added to dilute the mixture, it was rotated for 1 hr at 4 °C. Centrifugation removed nuclear debris, with released chromatin in the supernatant. Chromatin was immunoprecipitated by rotating with pre-washed GFP-Trap beads (Chromotek) overnight at 4 °C. Beads were washed four times, twice with low salt buffer and twice with high salt buffer (50 mM Tris-HCL pH 7.5, 10 mM EDTA, 300 mM NaCl, 0.1% Triton X-100, 1 mM PMSF). Bound chromatin was released from beads in elution buffer (0.1 M $NaHCO_3$, 1% SDS) by shaking at 65 °C for 15 mins. Protein and RNA was removed by digestion with Proteinase K and RNase A, respectively. Phenol-chloroform extraction was used to obtain DNA.

DNA was used as input for library preparation using the Nugen Ovation Ultralow System V2 (0344) kit. Library concentrations and fragment size distributions were determined by Bioanalyzer (Agilent). Libraries were sequenced paired-end, 2 x 38 bp reads. Sequencing was undertaken at the John Innes Centre, using the NextSeq 500 (Illumina).

2.15.20 Analysis of ChIP-seq data

Paired-end sequencing reads were mapped to TAIR10 using Bowtie2-2.3.4.1 (Langmead and Salzberg, 2012). Mononucleosomal fragments were retained for analysis (-p 16 --very-sensitive-local --no-mixed --no-discordant --no-unal --phred33 -I 130 -X 200). Bigwig files were generated by normalising mapped IP reads against respective input bam files. Two replicates for each experiment were generated and confirmed to be highly correlated, a single replicate was used for downstream analyses.

ChIP-seq profiles were visualised using IGV-2.6.2 (Thorvaldsdóttir, Robinson and Mesirov, 2012). deepTools-3.1.1 (Ramírez *et al.*, 2016) was used to generate data underlying metaplots and heatmaps. Data was plotted in R using a custom ggplot2 script.

To generate peaks, H2B.8 enrichment was calculated over 50 bp windows and those with > 0.5 $\log_2$(IP/input) were retained. Windows within 501 bp were merged using BEDtools-2.28.0 (Quinlan and Hall, 2010). Regions were filtered by size, with those < 201 bp removed from analysis. H2B.8 enrichment was then calculated over the new regions, those with < 0.5 $\log_2$(IP/input) were discarded. The remaining regions were defined as H2B.8 peaks. Overlaps with genomic features were determined using BEDtools; 50% of the feature was required to be overlapped by a peak to be defined as a true overlap.

Published ChIP-seq data (Supp. Table 2.3) was downloaded from the GEO (NCBI). Profiles were generated and analysed in the same way.

2.15.21 Seedling nuclei isolation by FACS

Nuclei were obtained from seedlings as per section 2.15.11. An equal volume of CyStain UV Precise T buffer (Sysmex) was added to resuspended nuclei and incubated on ice for 10 mins. FACS was then performed using a BD FACSMelody cell sorter (BD Biosciences). Nuclei were distinguished depending on DNA content, stained by DAPI (Fig. 2.35A). Nuclei of all ploidy levels were collected (Fig. 2.35B).

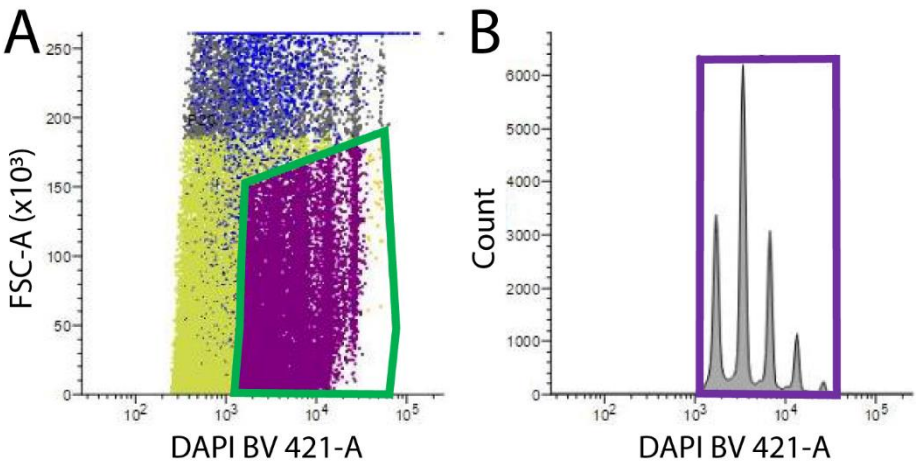


Figure 2.35: Example seedling nuclei flow cytometry. (**A**) Nuclei (green) are determined by DAPI-stained (DAPI BV 421-A) DNA content and size (Forward Scatter Area; FSC-A). (**B**) Nuclei of all ploidy levels are sorted (purple).

2.15.22 Histone protein extraction and mass spectrometry

Seedling nuclei in 1x PBS from FACS were pelleted at 6500 x g for 10 min at 4°C. The supernatant was removed, nuclei were resuspended in 0.2 N HCl and rotated overnight at 4°C. Samples were centrifuged at 6500 x g for 10 min at 4°C, the supernatant containing histone proteins was moved to a new tube. Acid was neutralised with 1/10 volumes 2M NaOH and 1/10 volumes 1M HEPES.

Mass spectrometry was performed by Dr Gerhard Saalbach (Proteomics Facility, John Innes Centre) as follows. The eluted peptide solution was dried down, and the peptides dissolved in 0.1% TFA / 3% acetonitrile. Aliquots were analysed by nanoLC-MS/MS on an Orbitrap Eclipse Tribrid mass spectrometer coupled to an UltiMate 3000 RSLCnano LC system (Thermo Fisher Scientific). The samples were loaded and trapped using a pre-column with 0.1% TFA at 15 µl/min for 4 min. The trap column was then switched in-line with the analytical column (nanoEase M/Z column, HSS C18 T3, 100 Å, 1.8 µm; Waters) for separation using the following gradient of solvents A (water, 0.1% formic acid) and B (80% acetonitrile, 0.1% formic acid) at a flow rate of 0.2 µl/min: 0-4 min 3% B (parallel to trapping); 4-10 min linear increase B to 8%; 10-60 min increase B to 25%; 60-80 min increase B to 38%; 80-90 min increase B to 60%; followed by a ramp to 99% B and re-equilibration to 3% B. Data were acquired with the following mass spectrometer settings in positive ion mode: MS1/OT: resolution 120K, profile mode, mass range *m/z* 300-1800, AGC $2e^5$, fill time 50 ms; MS2/IT: data dependent analysis was performed using parallel CID and HCD fragmentation with the following parameters: top20 in IT turbo mode, centroid mode, isolation window 1.6 Da, charge states 2-5, threshold $1.9e^4$, CE = 30, AGC target $1e^4$, max. inject time 35 ms, dynamic exclusion 1 count, 15 s exclusion, exclusion mass window ±5 ppm.

All MS/MS samples were analyzed in Proteome Discoverer-2.4.1.15 (Thermo Fisher Scientific) using Mascot-2.7.0 (Matrix Science) and Sequest version IseNode. Both Mascot and Sequest were set up to search the TAIR10_pep_20101214 *Arabidopsis thaliana* protein sequence database (TAIR, 35386 entries) and the MaxQuant contaminants database using the digestion enzyme trypsin with 2 missed cleavages, a fragment ion mass tolerance of 0.60 Da and a parent ion tolerance of 6.0 ppm. Carbamidomethyl of cysteine was specified in Mascot and Sequest as a fixed modification. Deamidation (N/Q), oxidation (M), and acetyl of the protein N-terminus were specified in Mascot and Sequest as variable modifications.

The output was imported into Scaffold-4.11.0 (Proteome Software) using identification probabilities of 99% for proteins and 95% for peptides.

2.15.23 Sperm nuclei ATAC-seq library preparation and sequencing

Sperm nuclei were isolated by FACS as per section 2.15.2. Nuclei were used immediately following centrifugation at 800 x g for 10 mins at 4 °C. A Tn5 transposase master mix was prepared (Illumina). The supernatant was removed, and pellets were resuspended with 5 µl Tn5 master mix. Reactions were incubated at 37 °C for 30 mins and then transferred to ice. For library amplification, 2x PCR master mix (Thermo Fisher Scientific) and primers P5 and P7 (Illumina) were added. Reactions then underwent 5 cycles of PCR before qPCR was used to assess the total cycle number required. Libraries then underwent total amplification and were finally purified with Agencourt AMPure beads (Beckman Coulter).

Library quality was determined by Bioanalyzer (Agilent). Libraries were sequenced paired-end using 2 x 38 bp reads. Sequencing was undertaken at the John Innes Centre, using the NextSeq 500 (Illumina).

2.15.24 Analysis of ATAC-seq

Bowtie2-2.3.4.1 (Langmead and Salzberg, 2012) was used to map paired-end reads to the TAIR10 reference genome. ATAC-seq libraries were normalised against a genomic DNA sequenced library to generate bigwig files. Four replicates for each genotype were confirmed to be highly correlated and then subsequently combined to increase genome coverage.

ATAC-seq profiles were visualised using IGV-2.6.2 (Thorvaldsdóttir, Robinson and Mesirov, 2012). Downstream analyses were undertaken using deepTools-3.1.1 (Ramírez *et al.*, 2016). Data was plotted in R using a custom ggplot2 script.

2.15.25 Sperm cell isolation by FACS

Sperm cells were isolated as described previously (Santos, Bispo and Becker, 2017). Briefly, pollen was collected from open flowers by vertexing in Sperm Cell Buffer (SCB; 1.3 mM H_3BO_3, 3.6 mM $CaCl_2$, 0.74 mM KH_2PO_4, 438 mM sucrose, 7mM MOPS, 0.83 mM $MgSO_4$, pH 7.5). Pollen was broken with glass beads (Merck) using gentle vertexing to release sperm cells into the buffer. Solution was spun through a 10 µm cell strainer (Corning) to a 50 ml tube. Unbroken pollen on the filter membrane was vortexed with glass beads again, this was repeated until all pollen was broken. The solution with sperm cells was made up to 20 ml and centrifuged at 190 x g for 20 mins at 4 °C, using low acceleration/deceleration. The supernatant was moved to a new tube as many sperm cells remain in suspension. The pellet was resuspended with 20 ml SCB and centrifugation was repeated, this washes the sperm cells and removes extraneous RNA molecules. The pellet was resuspended in 1 ml SCB and stained with 50x SYTOX Orange (5 mM stock, dilute 1000x for working concentration, Invitrogen) and 100x SYBR Green (10000x, Invitrogen). Sperm cells were incubated at 4 °C for 10 mins in the dark before proceeding to FACS.

Cell sorting was carried out using a BD FACSMelody machine (BD Biosciences). DNA containing particles, but not whole pollen, are first distinguished (Fig. 2.36A). Then populations are separated by staining with SYTOX Orange and SYBR Green; sperm cells are stained only by the latter, whereas sperm nuclei and vegetative nuclei are stained by both (Fig. 2.36B). Sperm cells were sorted into triazol buffer (ZYMO Research), using a ratio of 100k cells to 400 µl buffer. Sorted cells were immediately stored at -70 °C to preserve RNA for downstream library preparation. Purity was assessed by sorting 5K sperm cells and observing using a Leica DM6000 light microscope.

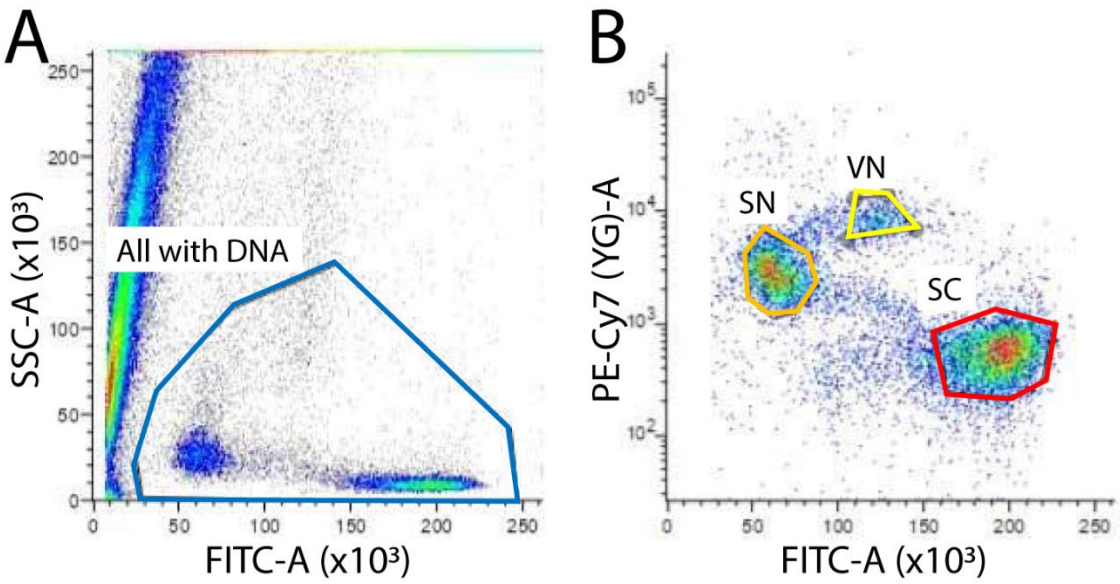


Figure 2.36: Example sperm cell flow cytometry. (**A**) Particles with DNA (blue) are pre-sorted by SYBR green staining (FITC-A) and size (Side Scatter Area; SSC-A), to exclude debris and whole pollen. (**B**) Populations of sperm cells (SC; red), sperm nuclei (SN; orange) and vegetative nuclei (VN; yellow) are separated depending on staining with SYBR green (FITC-A) and SYTOX Orange (PE-Cy7 (YG)-A). Sperm cells are stained by SYBR green but not by SYTOX Orange; enabling their separation from sperm nuclei and vegetative nuclei, which are both stained by both dyes.

2.15.26 Sperm cell RNA-seq library preparation and sequencing

Sperm cells in triazol buffer from FACS were sequentially loaded to columns from the Direct-zol RNA Microprep kit (ZYMO Research). DNA digestion was repeated twice, once on the column with the Direct-zol kit and again with the DNase digestion kit (QIAGEN). All extracted RNA was used for library preparation with the Nugen Universal RNA-seq kit (0364).

Concentrations and fragment sizes of libraries were assessed by Bioanalyzer (Agilent). The Nextseq 500 (Illumina), at the John Innes Centre, was used to sequence libraries single-end with 76 bp reads.

2.15.27 Seedling RNA-seq library preparation and sequencing

RNA was extracted from 100 mg 10-day-old seedlings using the RNeasy plant mini kit (QIAGEN). DNA digestion was undertaken using DNase digestion kit (QIAGEN). Libraries were prepared from 100 ng input RNA using the Nugen Universal RNA-seq kit (0364).

Library concentrations and fragment size distributions were determined by Bioanalyzer (Agilent). Libraries were sequenced single-end, 76 bp reads, at the John Innes Centre, using the NextSeq 500 (Illumina).

2.15.28 Analysis of RNA-seq data

For differential expression analysis between WT and *htb8* #7 sperm cells and WT and *p35S::HTB8-eGFP* seedlings, single-end sequencing reads were mapped to TAIR10 using TopHat-2.0.10 (Kim *et al.*, 2013). Two replicates were used for sperm cells, whilst three replicates were used for seedling RNA-seq analysis. Kallisto-0.43.0 (Bray *et al.*, 2016) and Sleuth-0.30.0 (Pimentel *et al.*, 2017) were used to obtain TPM and *P* values, respectively. Genes and TEs were required to have ≥ 2-fold change and $P < 0.05$ between samples to be categorised as differentially expressed. Volcano plots of fold change versus *P* values were generated using a custom script in R.

Analysis comparing gene expression between various cell types and tissues was undertaken using a pipeline allowing for comparisons of multiple datasets. Publicly available sequencing data (Supp. Table 2.4) was downloaded from the GEO (NCBI). Sequencing reads were mapped to TAIR10 using HISAT2-2.1.0 (Kim *et al.*, 2019). Cufflinks-2.2.1 (Trapnell *et al.*, 2012) and Ballgown-2.16.0 (Pertea *et al.*, 2016), from the tuxedo suite of bioinformatic tools, were used to generate FPKM values. Expressed genes were defined as those with ≥ 4 FPKM.

2.15.29 Phylogenetic tree construction

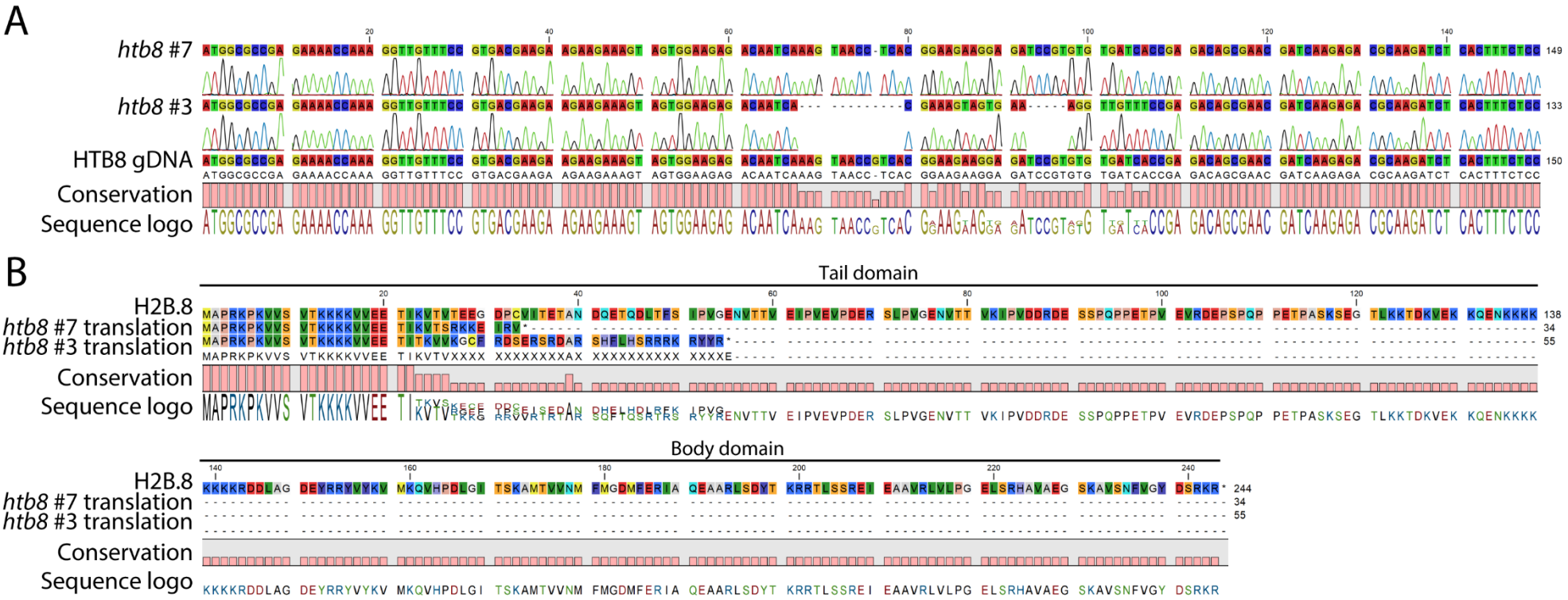
Plant H2B protein sequences were downloaded from Phytozome (Goodstein *et al.*, 2012), Congenie (Sundell *et al.*, 2015), Waterlily Pond (L. Zhang *et al.*, 2019), Magnoliid genomes (Chaw *et al.*, 2019; Chen *et al.*, 2019) and Uniprot (The UniProt Consortium, 2019). Human and yeast H2B sequences were obtained from Uniprot and used as out-groups for phylogenetics.

Sequences were imported to MEGA-X (Kumar *et al.*, 2018) and aligned using MUSCLE with default parameters. The phylogeny was generated using Neighbor-Joining testing, applying the Poisson model, and allowing for uniform substitution rates. H2B.8 homologs were identified owing to the distinct branch formed, separate from canonical H2B variants. Several representative H2B.8 homologs were searched using BLAST (Johnson *et al.*, 2008) to ask whether such homologs are specific to flowering plants.

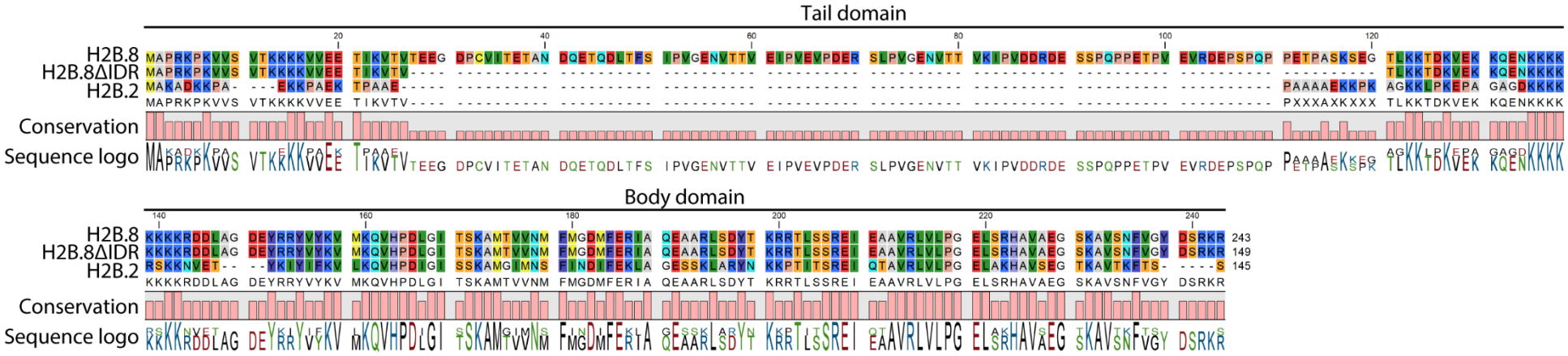
H2B.8 homolog sequences were extracted and analysed for shared features. The VL-XT algorithm of PONDR (Romero *et al.*, 2001) was used to identify IDRs. Amino acid compositions were calculated and analysed for differences in arginine residues.

2.16. Supplemental data

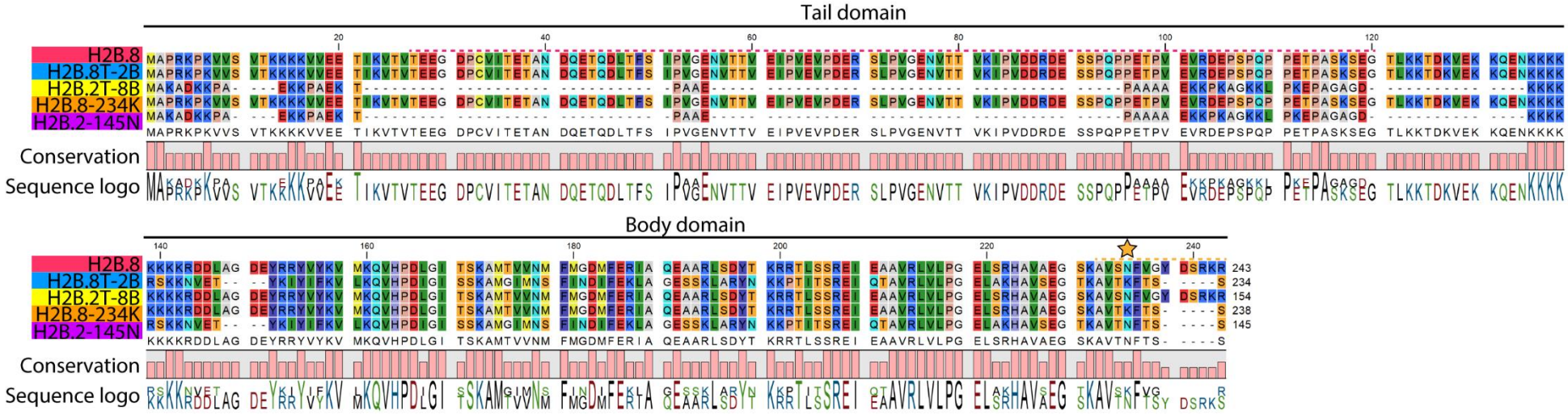
2.16.1 Supplemental figures


A
htb8 #7
htb8 #3
HTB8 gDNA
Conservation
Sequence logo
ATGGCGCCGA GAAAACCAAA GGTTGTTTCC GTGACGAAGA AGAAGAAAGT AGTGGAAGAG ACAATCAAAG TAACC-TCAC GGAAGAAGGA GATCCGTGTG TGATCACCGA GACAGCGAAC GATCAAGAGA CGCAAGATCT CACTTTCTCC
B
Tail domain
H2B.8
htb8 #7 translation
htb8 #3 translation
Conservation
Sequence logo
MAPRKPKVVS VTKKKKVVEE TIKVTVTEEG DPCVITETAN DQETQDLTFS IPVGENVTTV EIPVEVPDER SLPVGENVTT VKIPVDDRDE SSPQPPETPV EVRDEPSPQP PETPASKSEG TLKKTDKVEK KQENKKKK
Body domain
KKKKRDDLAG DEYRRYVYKV MKQVHPDLGI TSKAMTVVNM FMGDMFERIA QEAARLSDYT KRRTLSSREI EAAVRLVLPG ELSRHAVAEG SKAVSNFVGY DSRKR*

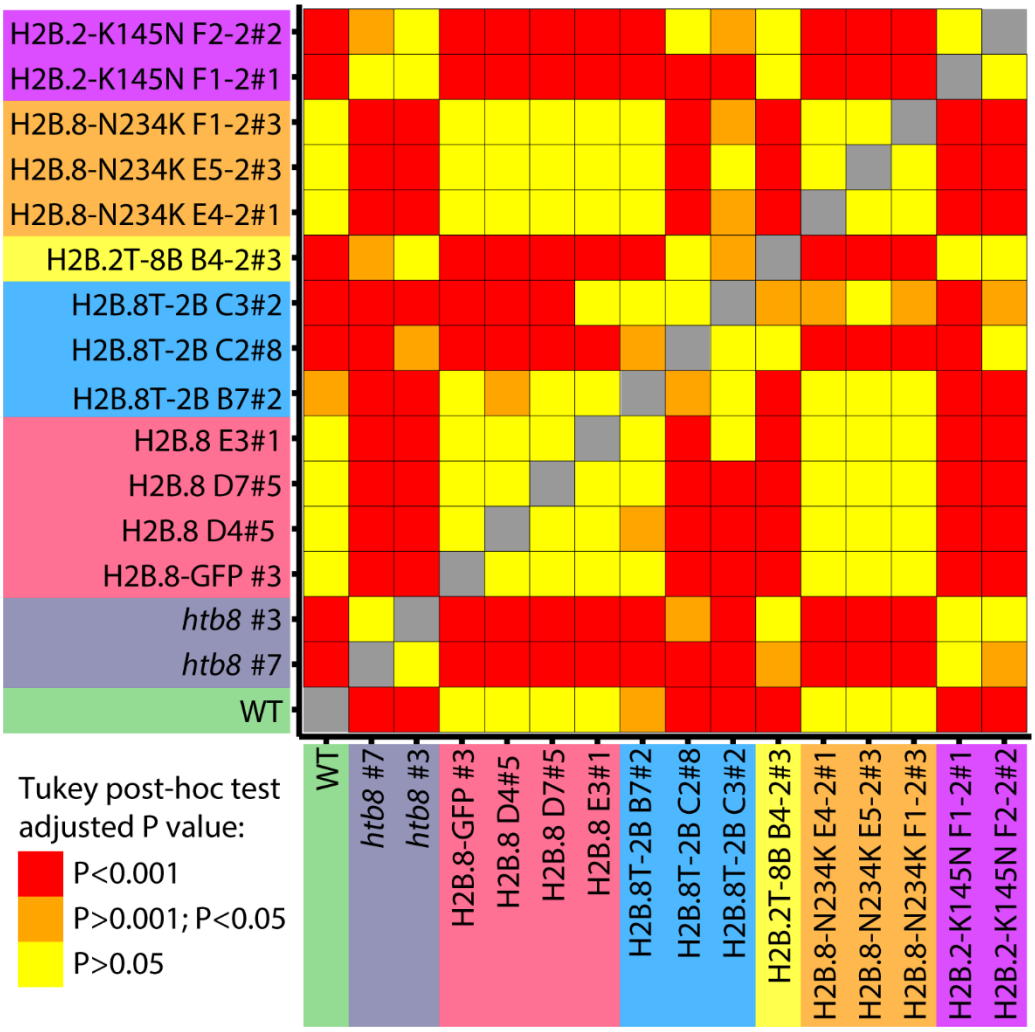
Supplemental Figure 2.1: Generation of *htb8* #7 and *htb8* #3 CRISPR mutants. (**A**) Alignment of Sanger sequencing results from two independent CRIPSR mutants to *HTB8*. *htb8* #7 has a single base deletion at 76 bp into the gene; *htb8* #3 has a 12 bp deletion after 67 bp and another of 5 bp after 92 bp. (**B**) Translations of *htb8* #7 and *htb8* #3 aligned to H2B.8 protein sequence. *htb8* #7 loses homology after 25 amino acids and stops prematurely after 33. *htb8* #3 is out of frame after 22 amino acids, the protein is truncated to 54 amino acids following a premature stop codon.


Tail domain
H2B.8
H2B.8ΔIDR
H2B.2
Conservation
Sequence logo
Body domain
H2B.8
H2B.8ΔIDR
H2B.2
Conservation
Sequence logo

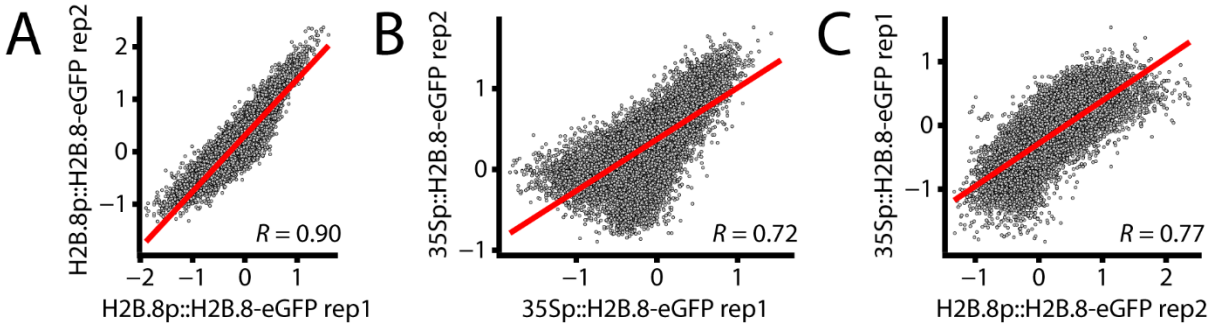
Supplemental Figure 2.2: H2B proteins tested in phase separation experiments. Alignment of protein sequences of H2B.8, H2B.8ΔIDR and H2B.2.


Tail domain
H2B.8
H2B.8T-2B
H2B.2T-8B
H2B.8-234K
H2B.2-145N
Conservation
Sequence logo
Body domain
H2B.8
H2B.8T-2B
H2B.2T-8B
H2B.8-234K
H2B.2-145N
Conservation
Sequence logo

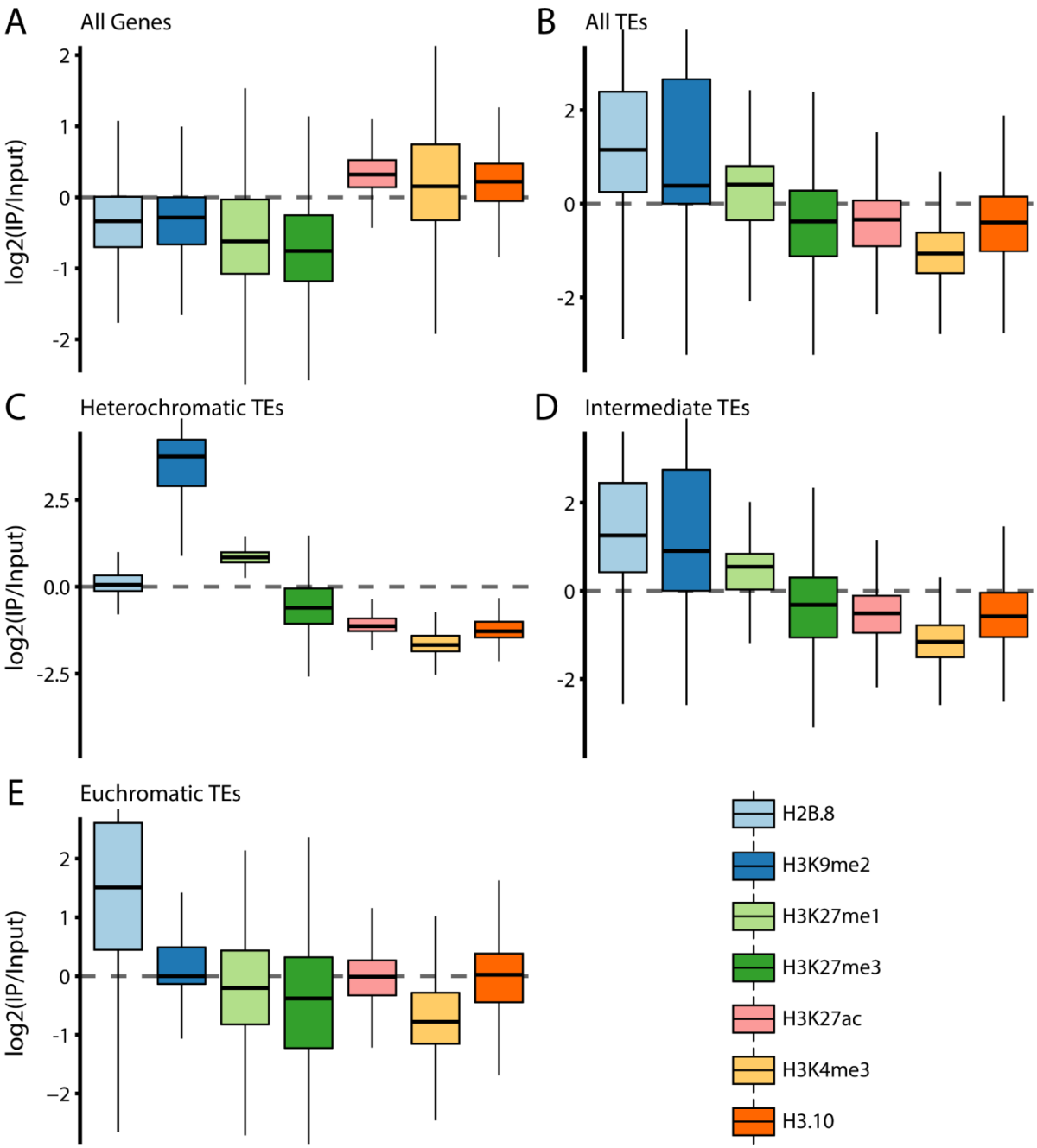
Supplemental Figure 2.3: Chimeric H2B proteins tested for complementation of *htb8* #7 sperm nuclear size. Alignment of protein sequences of H2B.8, H2B.8T-2B (H2B.8 tail; H2B.2 body), H2B.2T-8B (H2B.2 tail; H2B.8 body), H2B.8-N234K (H2B.8 with K at position to 234) and H2B.2-K145N (H2B.8 with N at position to 145).



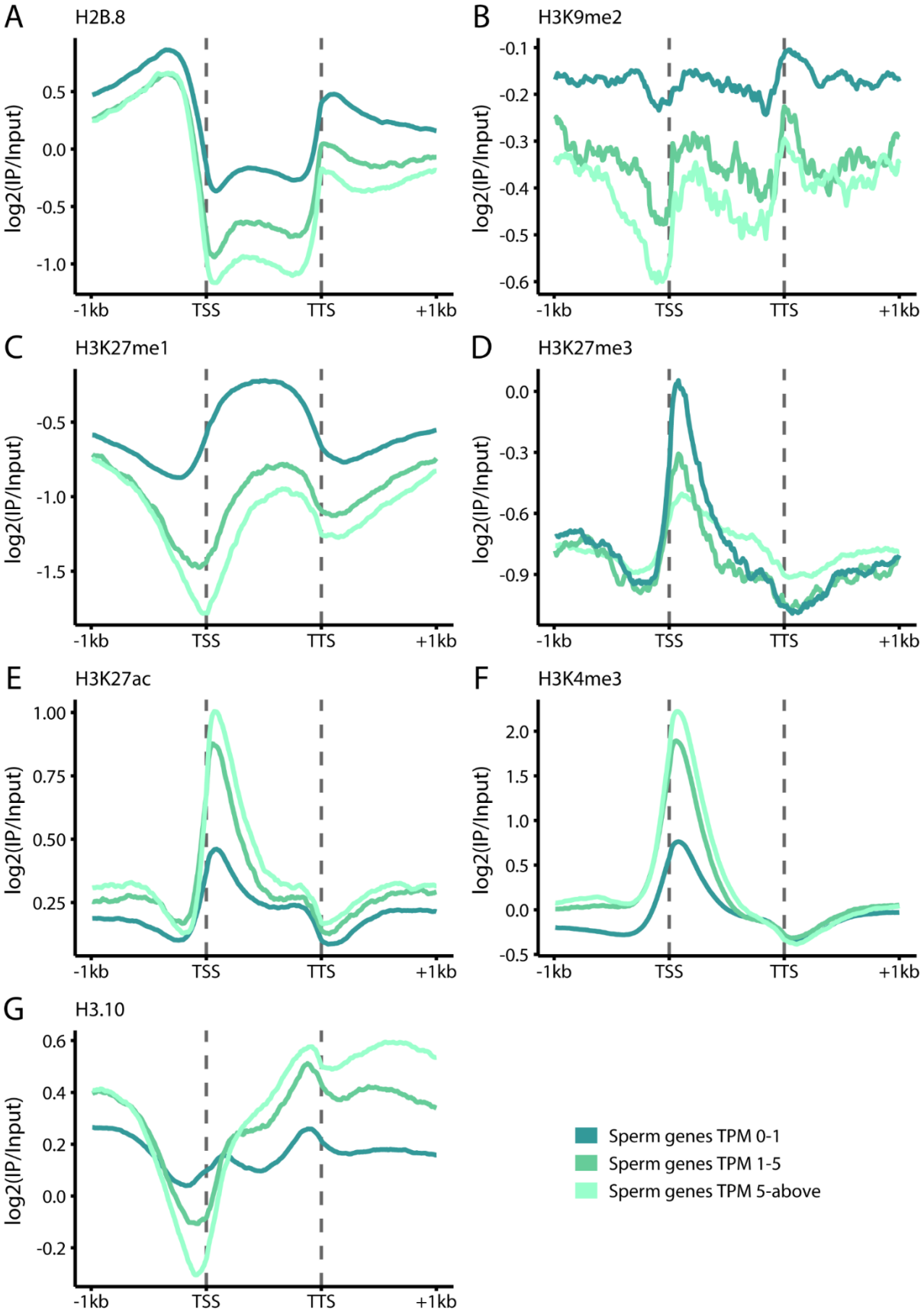
Supplemental Figure 2.4: Statistical significance of chimeric H2B proteins tested for complementation of *htb8* #7 sperm nuclear size. Heatmap of pairwise Tukey test *P* values (red – $P < 0.001$; orange – $P > 0.001$ and < 0.05; yellow – $P > 0.05$) comparing sperm nuclear size of WT, *htb8* #7 and *htb8* #3 along with complementation by H2B.8, H2B.8T-2B, H2B.2T-8B, H2B.8-N234K and H2B.2-K145N lines.



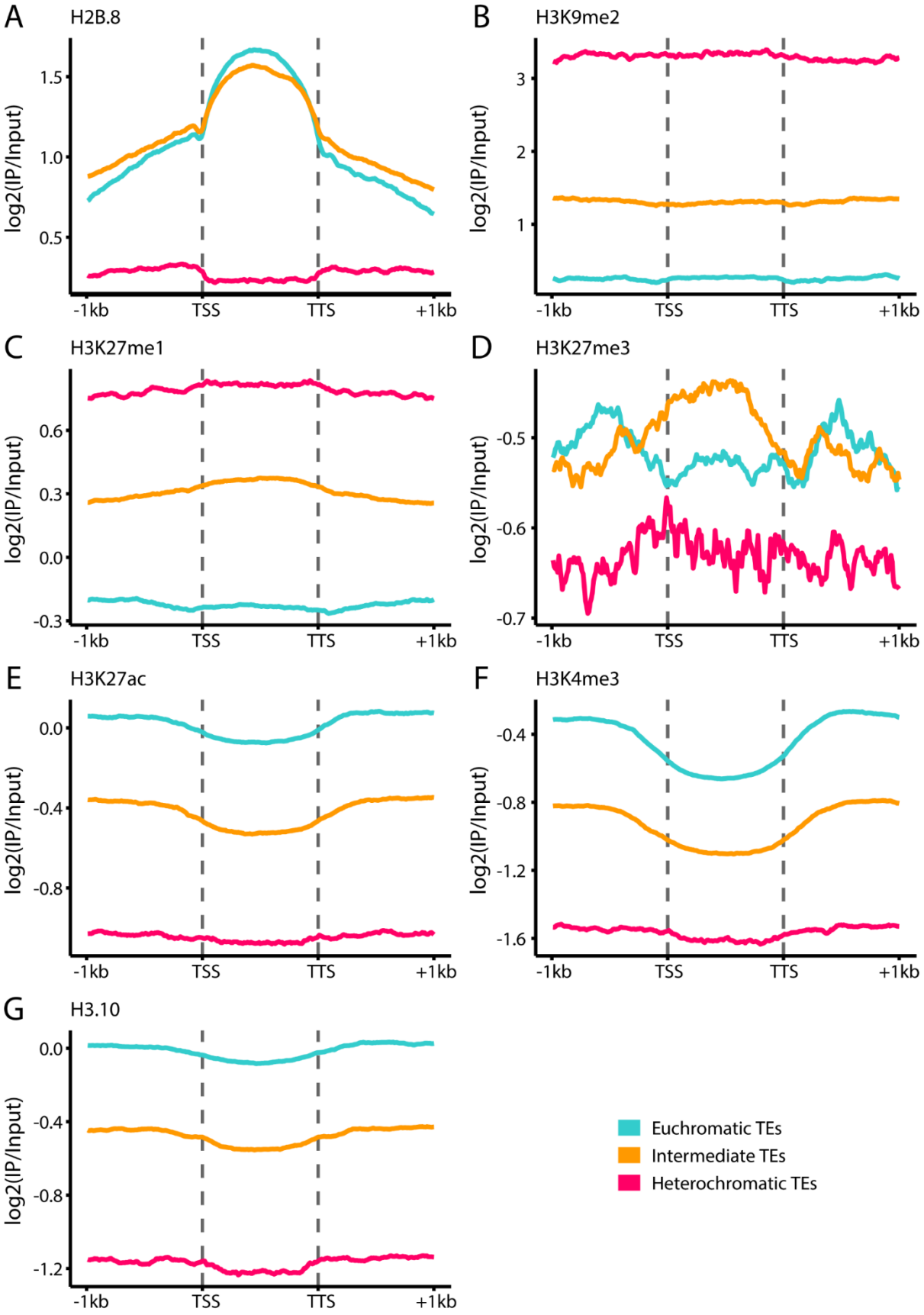
Supplemental Figure 2.5: Sperm and seedling H2B.8-eGFP native ChIP-seq libraries are highly correlated. (**A** to **C**) Native ChIP-seq replicates of H2B.8 are highly correlated for both sperm (A) and seedling (B). Genome-wide, H2B.8 deposition is decidedly similar between sperm and seedling (C). Scatterplots (A to C) show enrichment (log_2(IP/input)) over 1 kb windows across the genome. Linear regression models are plotted in red for each comparison. Spearman's rank correlation coefficients (*R*) were used to account for abnormal distribution of the data; *R* values are labelled on the figure.



Supplemental Figure 2.6: Enrichments of histone marks/variants over genomic features in sperm. (**A** to **E**) Boxplots of H2B.8 and other histone marks/variants enrichments (log_2(IP/input)) over genomic features in sperm. H2B.8 is largely depleted from genes (A) and enriched at TEs (B). Grouping TEs by chromatin environment reveals that H2B.8 is preferentially enriched at intermediate (D) and euchromatic (E) TEs, rather than heterochromatic TEs (C).


A
H2B.8
log2(IP/Input)
0.5
0.0
-0.5
-1.0
-1kb
TSS
TTS
+1kb
B
H3K9me2
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
C
H3K27me1
-1.5
D
H3K27me3
E
H3K27ac
1.00
0.75
0.50
0.25
F
H3K4me3
2.0
1.5
1.0
G
H3.10
0.6
0.4
0.2
Sperm genes TPM 0-1
Sperm genes TPM 1-5
Sperm genes TPM 5-above

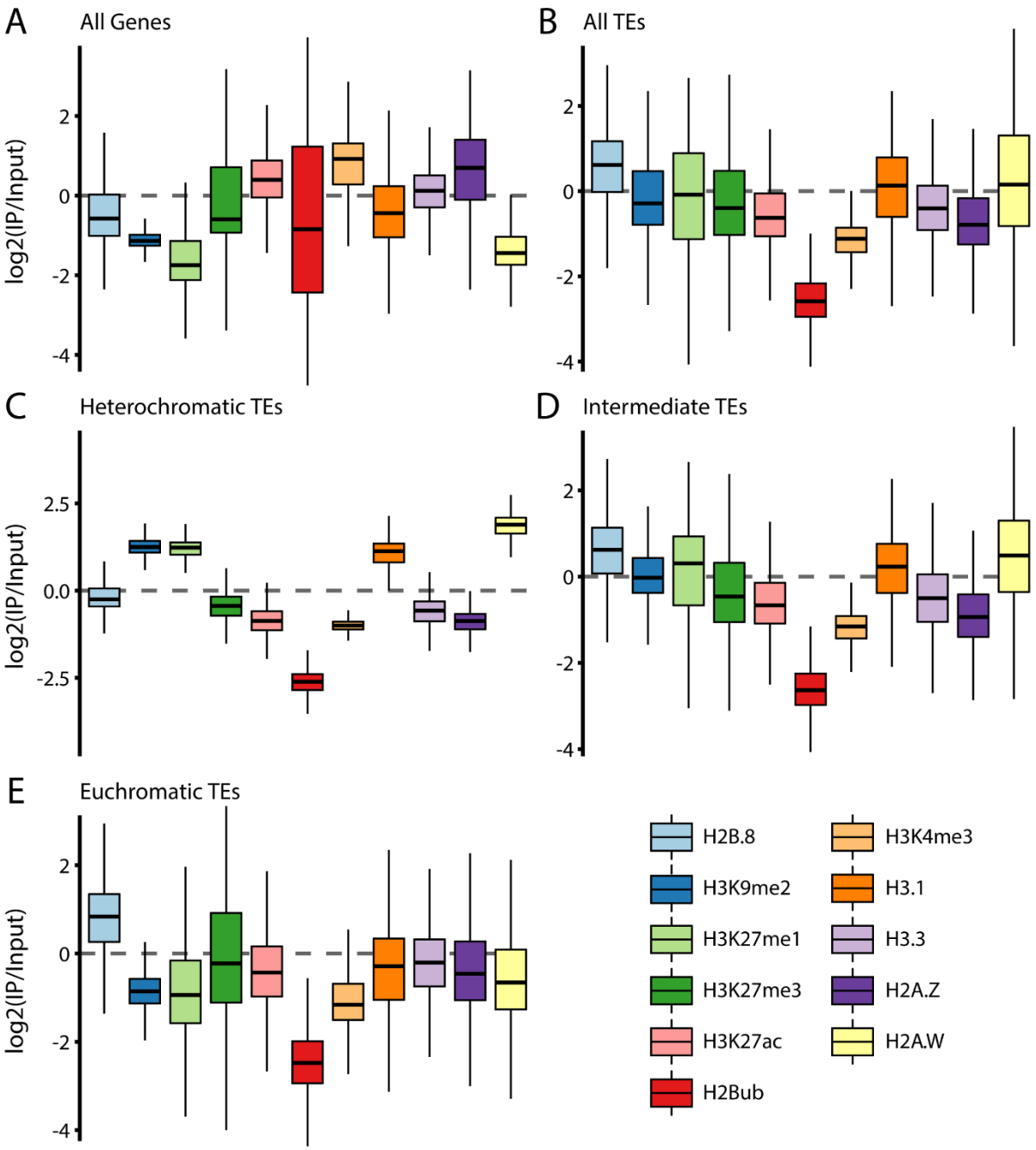
Supplemental Figure 2.7: Profiles of sperm histone marks/variants enrichments over genes grouped by expression. (**A** to **G**) Average enrichments (log_2(IP/input)) of H2B.8 (A), H3K9me2 (B), H3K27me1 (C), H3K27me3 (D), H3K27ac (E), H3K4me3 (F) and H3.10 (G) over genes grouped by expression (light green – TPM 5-above; mid green – TPM 1-5; dark green – TPM 0-1). Genes are scaled to 1 kb lengths; 1 kb is shown upstream/downstream of the TSS and TTS, respectively.


A
H2B.8
log2(IP/Input)
1.5
1.0
0.5
-1kb
TSS
TTS
+1kb
B
H3K9me2
log2(IP/Input)
3
2
1
-1kb
TSS
TTS
+1kb
C
H3K27me1
log2(IP/Input)
0.6
0.3
0.0
-0.3
-1kb
TSS
TTS
+1kb
D
H3K27me3
log2(IP/Input)
-0.5
-0.6
-0.7
-1kb
TSS
TTS
+1kb
E
H3K27ac
log2(IP/Input)
0.0
-0.4
-0.8
-1kb
TSS
TTS
+1kb
F
H3K4me3
log2(IP/Input)
-0.4
-0.8
-1.2
-1.6
-1kb
TSS
TTS
+1kb
G
H3.10
log2(IP/Input)
0.0
-0.4
-0.8
-1.2
-1kb
TSS
TTS
+1kb
Euchromatic TEs
Intermediate TEs
Heterochromatic TEs

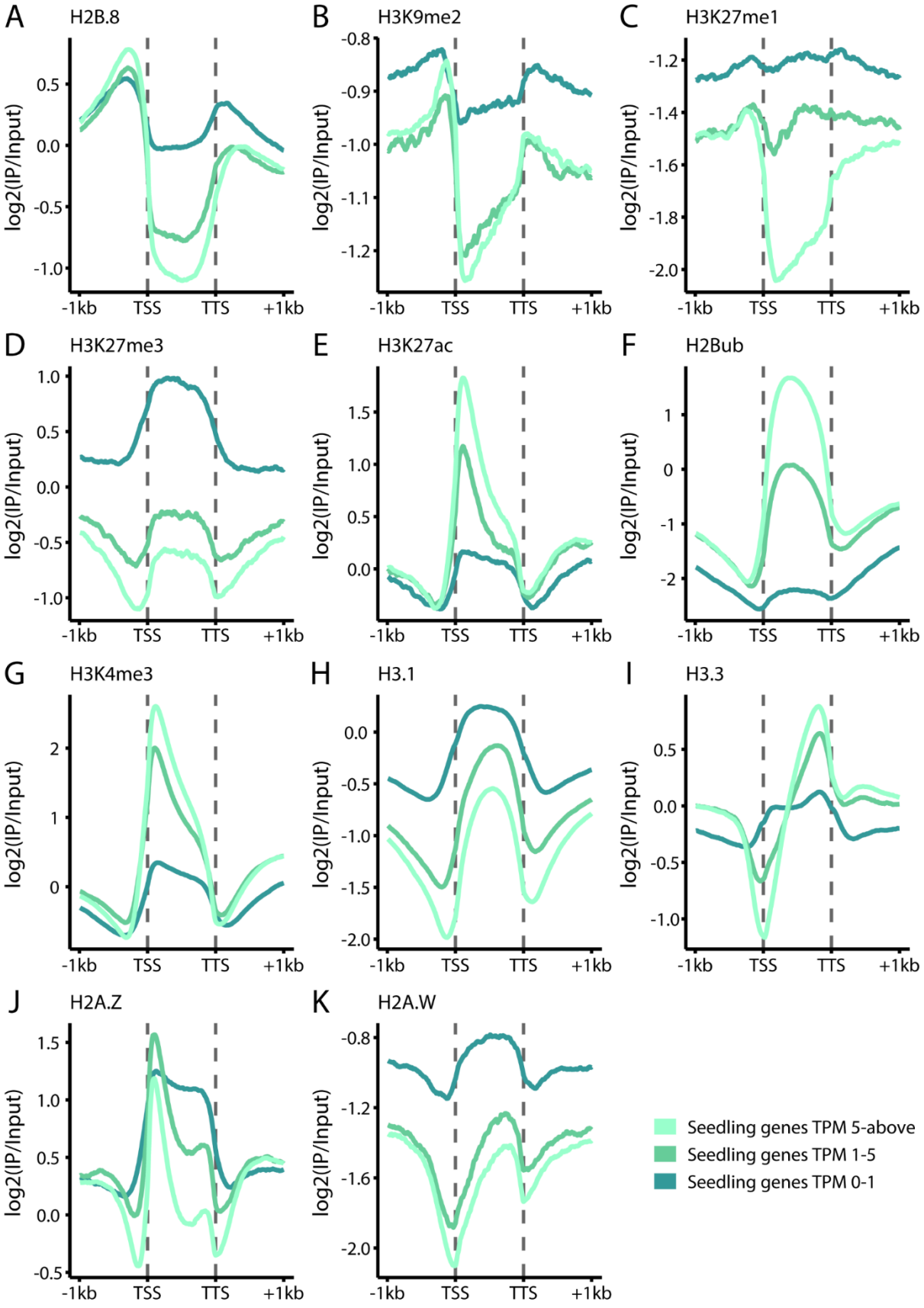
Supplemental Figure 2.8: Profiles of sperm histone marks/variants enrichments over TEs grouped by chromatin environment. (**A** to **G**) Average enrichments (log_2(IP/input)) of H2B.8 (A), H3K9me2 (B), H3K27me1 (C), H3K27me3 (D), H3K27ac (E), H3K4me3 (F) and H3.10 (G) over TEs grouped by chromatin environment (blue – euchromatic TEs; orange – intermediate TEs; pink – heterochromatic TEs). TEs are scaled to 1 kb lengths, 1 kb is shown upstream/downstream of the TSS and TTS, respectively.

H2B.8 peptides	Canonical H2B peptides	H2B.8 size (aa)	Average canonical H2B size (aa)	Proportion of H2B.8 peptides (%)	Proportion of H2B.8 peptides normalised by size (%)
9	56	243	151	13.8	9.1

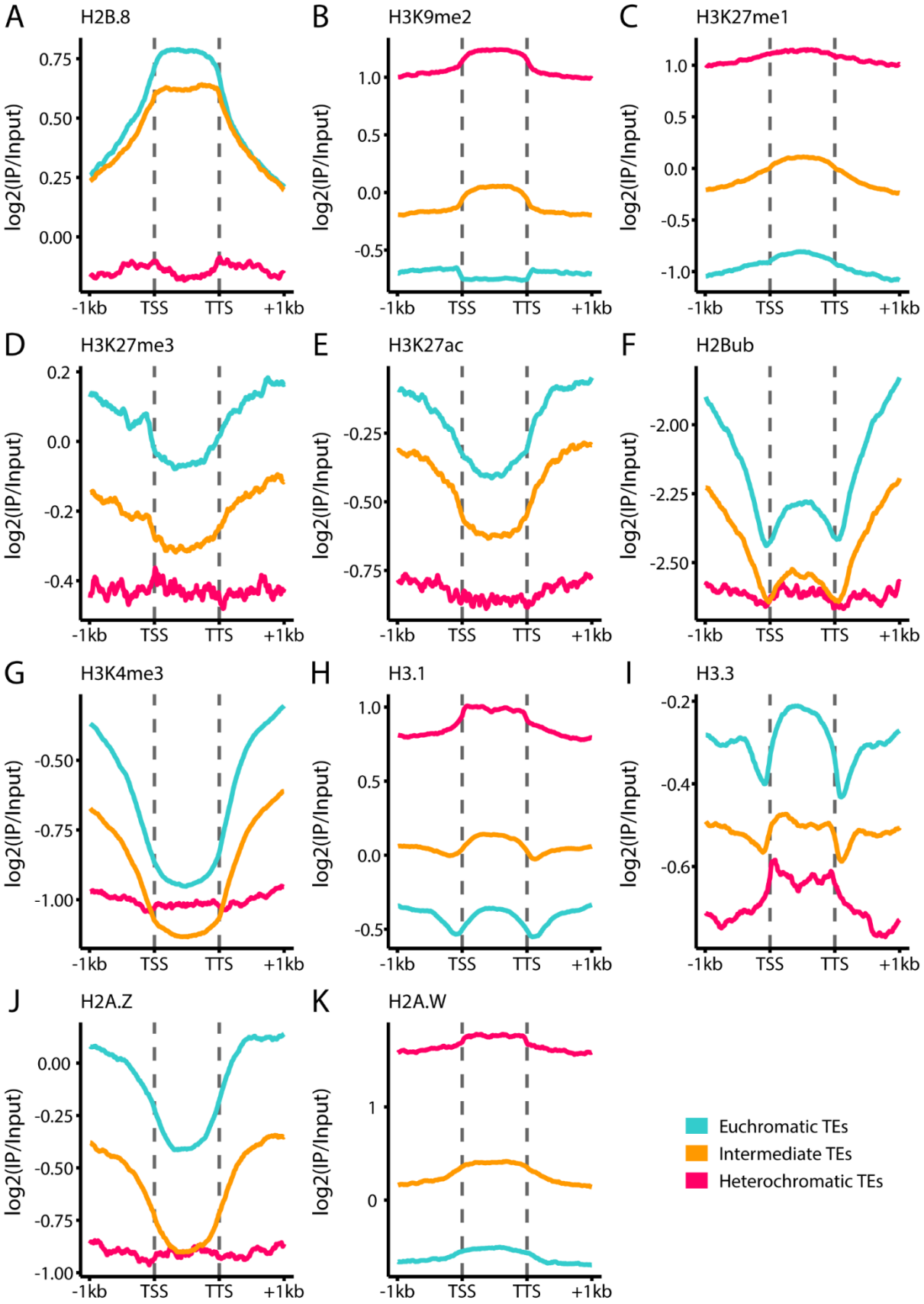
Supplemental Figure 2.9: H2B.8 can be ectopically expressed in seedlings. Ectopic H2B.8 protein comprises 9.1% of total histone H2B in *p35S::HTB8-eGFP* seedlings, adjusting for the larger size of H2B.8 compared to canonical H2Bs (243 amino acid versus 151).



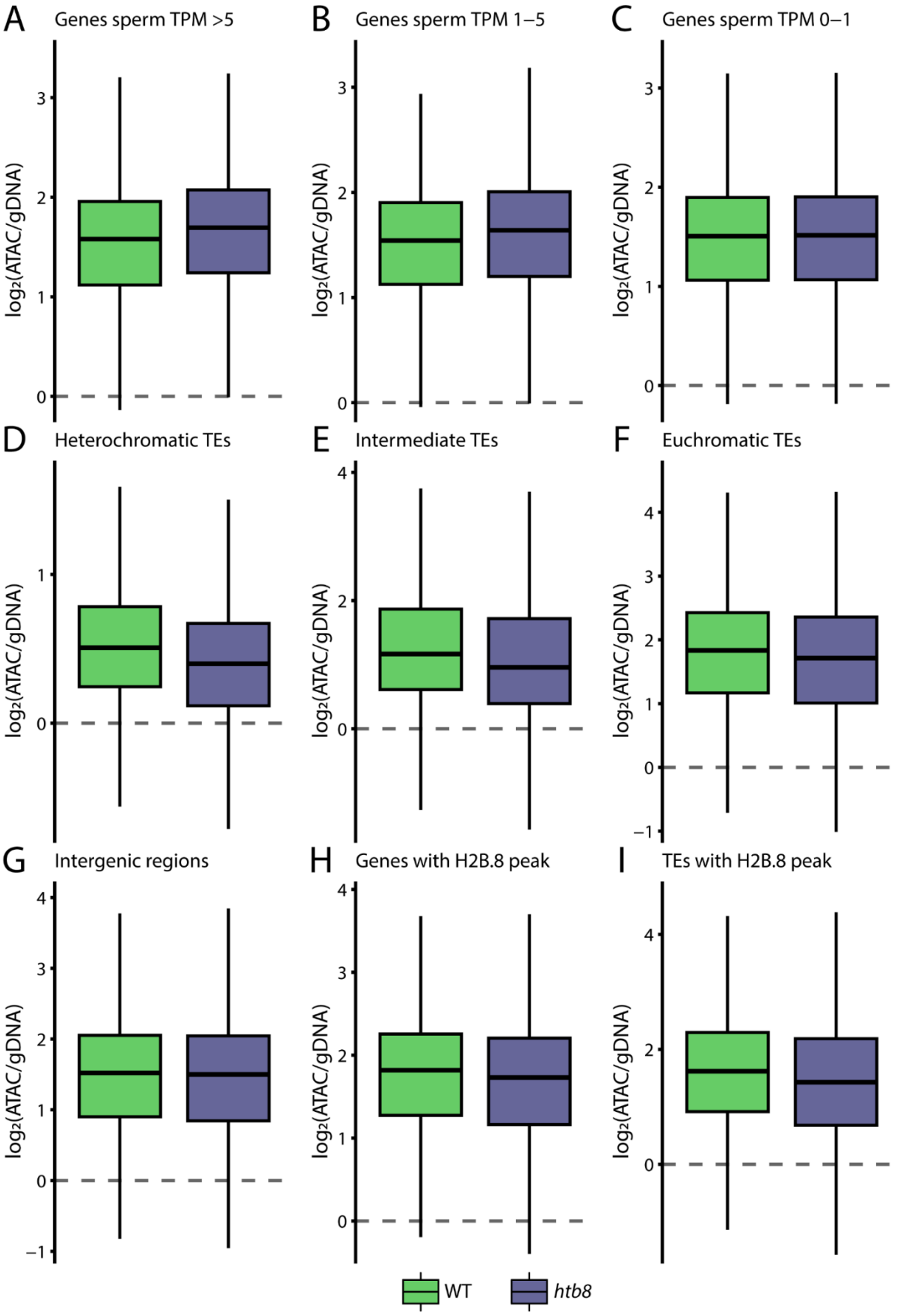
Supplemental Figure 2.10: Enrichments of histone marks/variants over genomic features in seedlings. (**A** to **E**) Boxplots of H2B.8 and other histone marks/variants enrichments (log_2(IP/input)) over genomic features in seedlings. H2B.8 is largely depleted from genes (A) and enriched at TEs (B). Grouping TEs by chromatin environment reveals that H2B.8 is preferentially enriched at intermediate (D) and euchromatic (E) TEs, rather than heterochromatic TEs (C). Therefore, H2B.8 deposition is broadly similar in sperm and soma.


A
H2B.8
B
H3K9me2
C
H3K27me1
D
H3K27me3
E
H3K27ac
F
H2Bub
G
H3K4me3
H
H3.1
I
H3.3
J
H2A.Z
K
H2A.W
log2(IP/Input)
-1kb
TSS
TTS
+1kb
Seedling genes TPM 5-above
Seedling genes TPM 1-5
Seedling genes TPM 0-1

Supplemental Figure 2.11: Profiles of seedling histone marks/variants enrichments over genes grouped by expression. (**A** to **K**) Average enrichments (log_2(IP/input)) of H2B.8 (A), H3K9me2 (B), H3K27me1 (C), H3K27me3 (D), H3K27ac (E), H2Bub (F), H3K4me3 (G), H3.1 (H), H3.3 (I), H2A.Z (J) and H2A.W (K) over genes grouped by expression (light green – TPM 5-above; mid green – TPM 1-5; dark green – TPM 0-1). Genes are scaled to 1 kb lengths; 1 kb is shown upstream/downstream of the TSS and TTS, respectively.


A
H2B.8
B
H3K9me2
C
H3K27me1
D
H3K27me3
E
H3K27ac
F
H2Bub
G
H3K4me3
H
H3.1
I
H3.3
J
H2A.Z
K
H2A.W
log2(IP/Input)
-1kb
TSS
TTS
+1kb
Euchromatic TEs
Intermediate TEs
Heterochromatic TEs

Supplemental Figure 2.12: Profiles of seedling histone marks/variants enrichments over TEs grouped by chromatin environment. (**A** to **K**) Average enrichments (log_2(IP/input)) of H2B.8 (A), H3K9me2 (B), H3K27me1 (C), H3K27me3 (D), H3K27ac (E), H2Bub (F), H3K4me3 (G), H3.1 (H), H3.3 (I), H2A.Z (J) and H2A.W (K) over TEs grouped by chromatin environment (blue – euchromatic TEs; orange – intermediate TEs; pink – heterochromatic TEs). TEs are scaled to 1 kb lengths, 1 kb is shown upstream/downstream of the TSS and TTS, respectively.


A
Genes sperm TPM >5
B
Genes sperm TPM 1–5
C
Genes sperm TPM 0–1
D
Heterochromatic TEs
E
Intermediate TEs
F
Euchromatic TEs
G
Intergenic regions
H
Genes with H2B.8 peak
I
TEs with H2B.8 peak
log2(ATAC/gDNA)
WT
htb8

Supplemental Figure 2.13: Local chromatin accessibility is very similar between WT and *htb8* over a range of genomic features. (**A** to **I**) Local chromatin accessibility was profiled by ATAC-seq (log_2(ATAC/gDNA)) between WT (green) and *htb8* (blue) sperm nuclei. Accessibility is largely similar at genes grouped by expression (A to C), TEs grouped by chromatin environment (D to F), intergenic regions (G) and genes (H) and TEs (I) marked by H2B.8 peaks.

2.16.2 Supplemental tables

Supplemental Table 2.1: List of primers used in this chapter.

Name	Sequence (5' – 3')	Purpose
***HTB8* CRIPSR guide RNAs**		
TB201	tgtggtctcaattgacaatcaaagtaaccgtcagtttaagagctatgctggaa	*HTB8* sgRNA 1
TB202	tgtggtctcaattgaagtaaccgtcacggaagagtttaagagctatgctggaa	*HTB8* sgRNA 2
TB203	tgtggtctcaattgggatctccttcttccgtgagtttaagagctatgctggaa	*HTB8* sgRNA 3
TB204	tgtggtctcaattgctgtctcggtgatcacacagtttaagagctatgctggaa	*HTB8* sgRNA 4
***HTB8* CRIPSR screening**		
PSH335	ccaccgcgtagtagacag	*HTB8* forward
PSH336	aagcgggagtttccggtg	*HTB8* reverse
PSH337	gctaaggtattcgaacgac	*HTB8* CRISPR sequencing primer
TB296	aggtcgtcgtcgtaggtgtc	Cas9 forward
TB297	gtgcagacctacaaccagct	Cas9 reverse
***htb8* #7 and *htb8* #3 genotyping**		
TB298	acagatacgaatcaccggat	*htb8* #7 nested PCR forward
TB299	gatacaacgattagggaatc	*htb8* #7 nested PCR reverse

TB293	agtagacagttaatcaccaatgctaaggta	*htb8* #7 dCAPS genotyping forward
TB294	tgatcacacacggatctccttcctccgtga	*htb8* #7 dCAPS genotyping reverse
TB335	gtgaaagggtgatcgtggtg	*htb8* #3 genotyping forward
TB336	tggcgggagatgagtatagg	*htb8* #3 genotyping reverse
***HTB8* qRT-PCR primers**		
TB239	tgggagatatgttcgagagga	*HTB8* qRT-PCR forward
TB240	cagaaccaacctaaccgctg	*HTB8* qRT-PCR reverse
H2B.8 Multisite Gateway cloning		
TB137	ggggacaactttgtatagaaaagttgatcagtggaggatgacatggc	*HTB8* promoter forward with attB4 site
TB138	ggggactgctttttgtacaaacttgattcttcgttagaaataaccg	*HTB8* promoter reverse with attB1r site
TB139	ggggacaagtttgtacaaaaaagcaggctatggcgccgagaaaaccaaaggt	*HTB8* CDS forward with attB1 site
TB140	ggggaccactttgtacaagaaagctgggtaacgtttcctactatcataacca	*HTB8* CDS reverse with attB2 site

H2B.8, H2B.8ΔIDR and H2B.2 In-Fusion cloning		
TB332	tctatctctctcgaggtaccatggcgccgagaaaaccaaag	*HTB8* In-Fusion forward
TB334	tgtcgactccgaattcacgtttcctactatcataaccaacg	*HTB8* In-Fusion reverse
PSH469	tatgaccatgattacgaattatcagtggaggatgacatggcg	*HTB8ΔIDR* In-Fusion overlapping forward N-terminus
PSH475	tcaaagtgccgacggttactttgattgtctcttcc	*HTB8ΔIDR* In-Fusion overlapping reverse N-terminus
PSH476	agtaaccgtcggcactttgaagaaaacagataagg	*HTB8ΔIDR* In-Fusion forward C-terminus
PSH474	ctcgcccttgctcacggatcccagatcttcctcagagatgagc	*HTB8ΔIDR* In-Fusion reverse C-terminus
TB333	tctatctctctcgaggtaccatggcgaaggcagataagaaa	*HTB2* In-Fusion forward
TB335	tgtcgactccgaattcagaactcgtaaacttcgtaaccgc	*HTB2* In-Fusion reverse
H2B chimeric line Multisite Gateway cloning		
PSH326	ggggaccactttgtacaagaaagctgggtaacgtttcctactatcataac	*HTB8* N-to-K point mutation reverse with attB2 site

PSH295	ggggaccactttgtacaagaaagctgggtaagaactcgtaaacttcgtaacagctttagagccttc	*HTB2* with H2B.8 C-terminal motif reverse with attB2 site
PSH296	ggggacaagtttgtacaaaaaagcaggctatggcgaaggcagataagaaac	*HTB2* forward with attB1 site
PSH297	ggggaccactttgtacaagaaagctgggtaagaactcgtaaagttcgtaaccgccttagtc	*HTB2* K-to-N point mutation reverse with attB2 site
PSH298	tgatctcttcttcttttttcttcttcttcttattc	*HTB8* and *HTB2* domain swap overlapping PCR reverse
PSH299	aagaagaaaaagaagaagagatcaaagaagaac	*HTB8* and *HTB2* domain swap overlapping PCR forward
PSH300	ggggaccactttgtacaagaaagctgggtaagaactcgtaaacttcgtaac	*HTB2* reverse with attB2 site
TB139	ggggacaagtttgtacaaaaaagcaggctatggcgccgagaaaaccaaaggt	*HTB8* CDS forward with attB1 site
TB140	ggggaccactttgtacaagaaagctgggtaacgtttcctactatcataacca	*HTB8* CDS reverse with attB2 site

Supplemental Table 2.2: Antibodies and ChIP beads used in this chapter.

Antigen	Application	Company	Name	Description
H3K9me2	Immunostaining / CUT&Tag	abcam	ab1220	Mouse monoclonal, IgG, ChIP grade
Myc	Immunostaining	abcam	ab32	Mouse monoclonal, IgG, ChIP grade
GFP	Native ChIP-seq	ChromoTek	gtma	GFP-Trap Magnetic Agarose beads, ChIP grade

Supplemental Table 2.3: Publicly available ChIP-seq data used in this chapter.

Dataset	IP or input	GEO dataset	GEO sample	Publication
Sperm				
H3K27me1	IP	GSE120669	GSM3407994 / GSM3407995	(Borg *et al.*, 2020)
	Input	GSE120669	GSM3407988 / GSM3407989	
H3K27me3	IP	GSE120669	GSM3408002 / GSM3408003	(Borg *et al.*, 2020)
	Input	GSE120669	GSM3407996 / GSM3407997	
H3K27ac	IP	GSE120669	GSM3407991 / GSM3407992 / GSM3407993	(Borg *et al.*, 2020)
	Input	GSE120669	GSM3407988 / GSM3407989 / GSM3407990	

H3K4me3	IP	GSE120669	GSM3408004 / GSM3408005 / GSM3408006	(Borg *et al.*, 2020)
	Input	GSE120669	GSM3407996 / GSM3407997 / GSM3407998	
H3.10	IP	GSE120669	GSM3407999 / GSM3408000 / GSM3408001	(Borg *et al.*, 2020)
	Input	GSE120669	GSM3407996 / GSM3407997 / GSM3407998	
Seedling				
H3K9me2	IP	GSE51304	GSM1242393 / GSM1667165	(Stroud, Do, Du, Zhong, Feng, Patel, *et al.*, 2014)
	Input	GSE51304	GSM1242392 / GSM1667164	
H3K27me1	IP	GSE111814	GSM3040071 / GSM3040072	(Ma *et al.*, 2018)
	Input	GSE111814	GSM3040051 / GSM3040052	
H3K27me3	IP	GSE72735	GSM2065751 / GSM2065752	(Li *et al.*, 2016)
	Input	GSE72735	GSM2065750	

H3K27ac	IP	GSE79524	GSM2475161 / GSM2096920	(Chen *et al.*, 2017)
	Input	GSE79524	GSM2096938	
H2Bub	IP	GSE112952	GSM3092016 / GSM3092017	(Nassrallah *et al.*, 2018)
	Input	GSE112952	GSM3092008 / GSM3092009	
H3K4me3	IP	GSE124319	GSM3528367 / GSM3528368	(Fiorucci *et al.*, 2019)
	Input	GSE124319	GSM3528363 / GSM3528364	
H3.1	IP	GSE34840	GSM856055	(Stroud *et al.*, 2012)
	Input	GSE34840	GSM856057	
H3.3	IP	GSE34840	GSM856054	(Stroud *et al.*, 2012)
	Input	GSE34840	GSM856056	
H2A.Z	IP	GSE50942	GSM1232782	(Yelagandula *et al.*, 2014)
	Input	GSE50942	GSM1232779	
H2A.W	IP	GSE50942	GSM1232780	(Yelagandula *et al.*, 2014)
	Input	GSE50942	GSM1232779	

Supplemental Table 2.4: Publicly available RNA-seq data used in this chapter.

Cell type	GEO dataset	GEO sample	Publication
Root	GSE122772	GSM3484754 / GSM3484755 / GSM3484756	(Tannenbaum *et al.*, 2018)
Leaf	GSE86583	GSM2306324 / GSM2306325 / GSM2306326	(Walker *et al.*, 2018)
Pollen	GSE120519	GSM3402476 / GSM3402477 / GSM3402478	(He *et al.*, 2019)
Seed	GSE94459	GSM2476064 / GSM2476065 / GSM2476066	(Narsai *et al.*, 2017)
Meiocyte	GSE86583	GSM2306313 / GSM2306314 / GSM2306315	(Walker *et al.*, 2018)
Embryo epidermis	GSE98176	GSM2588913 / GSM2588914 / GSM2588915 / GSM2588916	(Sakai *et al.*, 2018)
Embryo mesophyll	GSE98176	GSM2588917 / GSM2588918 / GSM2588919 / GSM2588920	(Sakai *et al.*, 2018)
Root endodermis	GSE79710	GSM2101451 / GSM2101452	(Kawakatsu *et al.*, 2016)

Root epidermis	GSE79710	GSM2101449	(Kawakatsu *et al.*, 2016)
Root cortex	GSE79710	GSM2101450	(Kawakatsu *et al.*, 2016)
Stele	GSE79710	GSM2101453	(Kawakatsu *et al.*, 2016)
Columella root cap	GSE79710	GSM2101454	(Kawakatsu *et al.*, 2016)

Supplemental Table 2.5: H2B.8 homologs and shared features.

Species	**Angiosperm lineage**	**Locus**	**Size (# amino acid)**	**Arginine (# amino acid)**	**K-to-N substitution in C-terminal motif (Y/N)**
Arabidopsis thaliana	Eudicot	AT1G08170	243	16	Y
Arabidopsis lyrata	Eudicot	AL1G18580	204	13	Y
Brassica oleracea	Eudicot	Bol041191	248	14	Y
Brassica rapa	Eudicot	Brara.F00506	257	14	Y
Trifolium pratense	Eudicot	Tp57577	222	15	Y
Medicago truncatula	Eudicot	Medtr2g084480	227	13	Y
Prunus persica	Eudicot	Prupe.6G114300	305	21	Y
Malus domestica	Eudicot	MDP0000126901	321	19	Y
Vitis vinifera	Eudicot	GSVIVG01031447001	216	20	Y

Crocus sativus	Eudicot	Cucsa.200320	166	13	Y
Kadua laxiflora	Eudicot	Kalax.0027s0129	231	21	Y
Kalanchoe fedtschenkoi	Eudicot	Kaladp0039s0686	204	18	Y
Glycine max	Eudicot	Glyma.11G191600	199	10	Y
Glycine max	Eudicot	Glyma.12G082700	225	13	Y
Phaseolus vulgaris	Eudicot	Phvul.011G087500	227	12	Y
Daucus carota	Eudicot	DCAR_006133	196	6	Y
Amaranthus hypochondriacus	Eudicot	AHYPO_016191	205	10	Y
Mimulus guttatus	Eudicot	Migut.N02535	270	16	Y
Solanum tuberosum	Eudicot	PGSC0003DMG400007131	273	16	Y
Solanum lycopersicum	Eudicot	Solyc06g074750	283	18	Y
Boechera stricta	Eudicot	Bostr.25219s0128	207	14	Y
Eutrema salsugineum	Eudicot	Thhalv10008775m	210	13	Y
Capsella grandiflora	Eudicot	Cagra.4395s0008	208	13	Y
Capsella rubella	Eudicot	Carubv10011530m	208	13	Y

Carica papaya	Eudicot	evm.model.supercontig_30.62	251	17	Y
Spirodela polyrhiza	Eudicot	Spipo0G0102400	287	27	Y
Linum usitatissimum	Eudicot	Lus10036579	243	19	Y
Linum usitatissimum	Eudicot	Lus10041351	227	16	Y
Manihot esculenta	Eudicot	Manes.11G153400	233	20	Y
Populus trichocarpa	Eudicot	Potri.001G212200	163	16	Y
Citrus clementina	Eudicot	Ciclev10029880m	260	14	Y
Citrus sinensis	Eudicot	orange1.1g048093m	261	14	Y
Fragaria vesca	Eudicot	mrna01119.1-v1.0-hybrid	327	17	Y
Gossypium raimondii	Eudicot	Gorai.006G103900	219	14	Y
Aquilegia coerulea	Eudicot	Aqcoe1G121500	220	8	Y
Cinnamomum micranthum	Magnoliids	RWR72246.1	253	11	Y
Lilium davidii	Monocot	A0A0U5KQX1	197	7	Y

Lilium longiflorum	Monocot	p22.5	158	11	Y
Zostera marina	Monocot	Zosma42g00130	192	9	Y
Ananas comosus	Monocot	Aco009323	266	28	Y
Musa acuminata	Monocot	randomP19680	210	15	Y
Oryza sativa	Monocot	Os09g39730	194	14	Y
Brachypodium distachyon	Monocot	Bradi4g38800	293	17	Y
Brachypodium stacei	Monocot	Brast05G224600	284	15	Y
Zea mays	Monocot	GRMZM2G442555	262	22	N
Sorghum bicolor	Monocot	Sobic.002G138000	299	20	N
Oropetium thomaeum	Monocot	06309A	268	21	N
Oropetium thomaeum	Monocot	27269A	267	19	N
Panicum hallii	Monocot	Pahal.B01951	305	25	N
Panicum virgatum	Monocot	Pavir.Ba02673	303	30	N
Setaria viridis	Monocot	Sevir.2G141900	297	30	N
Setaria italica	Monocot	Seita.2G136700	298	30	N
Nymphaea colorata	Nymphaeales	GWHPAAYW018457	216	20	Y

2.16.3 Supplemental movies

Supplemental movies can be accessed using the following link at the time of submission:

https://www.dropbox.com/sh/jtvdjzzndn7bz93/AACofgTFhNAw1TVhGIQYO_TOa?dl=0

Captions for supplemental movies are as follows:

Supplemental Movie 2.1: Movie of 3D-SIM images of WT and *htb8* sperm from Fig. 2.8.

Supplemental Movie 2.2: Movie of 3D-SIM images of WT sperm, vegetative and leaf nuclei from Fig. 2.10.

Supplemental Movie 2.3: Movie of 3D-SIM images of *pHTB8::HTB8-eGFP* sperm from Fig. 2.11.

Supplemental Movie 2.4: Movie of 3D-SIM images of *p35S::HTB8*-eGFP, *p35S::HTB8ΔIDR-YFP* and *p35S::HTB2-YFP* leaf nuclei from Fig. 2.14.

Supplemental Movie 2.5: Movie of FRAP images from Fig. 2.15A.

Supplemental Movie 2.6: Movie of droplet fusion time-lapse images from Supp. Fig. 2.15B.

Supplemental Movie 2.7: Movie of FRAP images from Fig. 2.15C.

Chapter 3 – HAM proteins are required for heterochromatic DNA demethylation in pollen vegetative nuclei

3.1 Introduction

DNA methylation is extensively reprogrammed through the course of male germline development in the model plant *Arabidopsis thaliana*. Previous work has characterised the presence of sexual-lineage-specific methylation in meiocytes that persists to the sperm (Walker *et al.*, 2018). Such methylation affects gene expression and is required for the correct splicing of a gene essential for meiosis. Furthermore, DNA methylation levels in the CG context has been shown to be more robustly maintained throughout the male sexual lineage to sperm cells (SC) than comparative somatic cell types (Hsieh *et al.*, 2016). This increased efficiency of maintenance is purported to enable accurate inheritance of methylation patterns across generations.

However, this chapter focuses on active DNA demethylation of the vegetative cell (VC) by DEMETER (DME), a DNA glycosylase (Choi *et al.*, 2002). DME expression is confined to the companion cells of both the male and female gametes (Schoft *et al.*, 2011). Active demethylation by DME facilitates the establishment of imprinted genes, with parentally biased expression in the endosperm of the seed (Hsieh *et al.*, 2011; Rodrigues and Zilberman, 2015). DME also functions to demethylate transposable elements (TEs), releasing them from transcriptional silencing (Slotkin *et al.*, 2009; Calarco *et al.*, 2012; Ibarra *et al.*, 2012). Such activated transposons are proposed to serve as templates for small interfering RNA (siRNA) biogenesis. In turn, siRNAs guide RNA-directed DNA methylation (RdDM) machinery in the sperm to reinforce transposon silencing for the next generation (Feng, Zilberman and Dickinson, 2013).

How DME is recruited to specific genomic locations and then gains access for enzymatic activity remains unknown. Recent work has demonstrated that the C-terminal catalytic core of DME is required for active demethylation at endogenous targets (C. Zhang *et al.*, 2019). However, without the N-terminus, the protein fails to fully demethylate heterochromatic loci whilst also gaining off-target effects (C. Zhang *et al.*, 2019). Therefore, both termini of DME are required for correct recruitment and function.

Epigenetic reprogramming of the VC is not limited to DNA methylation – global higher order chromatin structure is also extensively reconfigured. Centromeres are disassembled by active removal of the centromeric histone CenH3 (Mérai *et al.*, 2014). Additionally, heterochromatin is lost, with H3K9me2 and linker histone H1 absent (Schoft *et al.*, 2009; Hsieh *et al.*, 2016).

Recently, developmental depletion of H1 has been implicated in facilitating DME access to putative heterochromatic loci. Ectopic expression of H1 in VC caused a moderate increase of DNA methylation at DME-target sites and silenced a subset of pollen expressed TEs (He *et al.*, 2019). Therefore, H1 absence from VC can partially explain DME access to heterochromatin sites. Additionally, DME access to heterochromatic loci in the central cell, the female equivalent of the VC, has been demonstrated to be dependent on the FACT (Facilitates Chromatin Transcription) complex (Ikeda *et al.*, 2011; Frost *et al.*, 2018). Unlike the VC, H1 is present in the central cell. It is suggested that FACT enhances DNA accessibility at heterochromatic and H1-rich DME-target sites (Ikeda *et al.*, 2011; Frost *et al.*, 2018). Owing to the absence of H1 from VC, no such reliance on FACT is observed in the male germline companion cell. Besides these findings, knowledge on how DME is recruited to the majority of sites and gains access to chromatin is currently limited (C. Zhang *et al.*, 2019).

This chapter seeks to identify and characterise novel components of VC chromatin that could facilitate DME function. Employing mass spectrometry, Histone Acetyltransferase of the MYST family (HAM) proteins were identified as VC-specific in pollen. The following work explores the role of HAM proteins in permitting DME access to VC chromatin for active DNA demethylation and subsequent TE activation.

3.2 HAM proteins are present in VC of pollen

Chromatin is extensively reprogrammed through male germline development, resulting in highly dimorphic nuclei within mature pollen (Kawashima and Berger, 2014). Sperm chromatin is highly compacted whilst the vegetative nuclei has very loose chromatin conformation (Schoft *et al.*, 2009). The previous chapter described the use of low-input proteomics of pollen nuclei to identify novel regulators of sperm chromatin compaction. The same sperm and vegetative nuclear proteomes were also explored for potential contributors to the decondensed chromatin state in vegetative nuclei.

Mass spectrometry data suggested the presence of HAM proteins in VC but not SC (Fig. 3.1A). Peptides are identified for both HAM1 and HAM2 (AT5G64610 and AT5G09740, respectively) in VC across two pollen nuclear proteome replicates. Previous promoter GUS fusion experiments indicated the presence of HAM1 and HAM2 in pollen, along with actively dividing vegetative tissues (Latrasse *et al.*, 2008). To confirm HAM1 and HAM2 presence in VC chromatin, eGFP-tagged proteins were generated under the respective native promoters. Confocal microscopy of *pHAM1::HAM1-eGFP* and *pHAM2::HAM2-eGFP* verified the specificity of HAM proteins to the VC in pollen (Fig. 3.1B), agreeing with proteomic and transcriptomic analysis.

Together, mass spectrometry enabled the detection of HAM proteins as VC-specific in pollen. Confocal imaging of tagged lines confirmed the proteomic data. Given their absence from the highly condensed SC, HAM proteins were considered as potential determinants of VC chromatin decondensation.

A

MS rep	HAM1 (AT5G64610)		HAM2 (AT5G09740)	
	Sperm nuclei	Vegetative nuclei	Sperm nuclei	Vegetative nuclei
1	0	3	0	2
2	0	1	0	0
Both	0	4	0	2

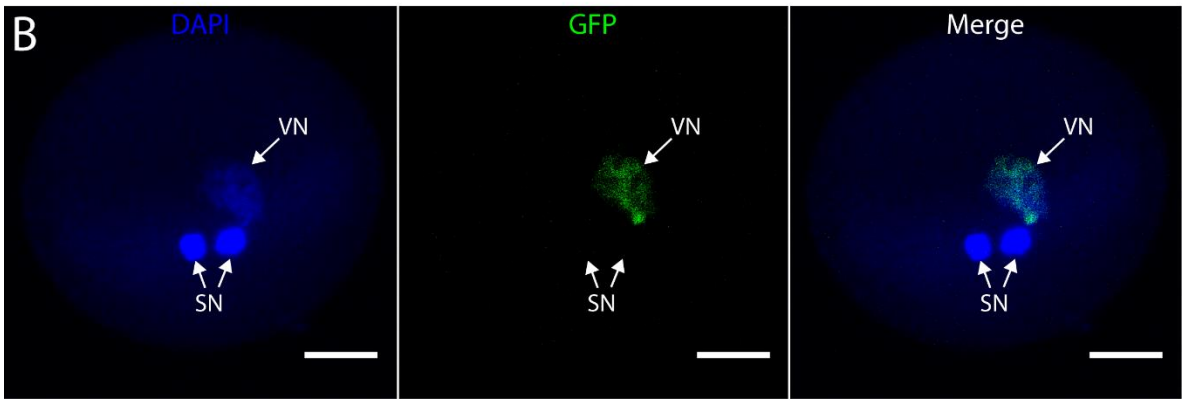


Figure 3.1: HAM proteins are VC-specific in pollen. (**A**) Mass spectrometry identifies peptides specific to HAM1 and HAM2 in VC but not SC of pollen. (**B**) GFP-tagged HAM1 (*pHAM1::HAM1-eGFP*) confirms protein specificity to VC in pollen. Scale bars are 5 µm in (C).

3.3 *ham* mutants phenocopy *dme* fertility defects

Previous characterisation of HAM proteins revealed the two alleles to be functionally redundant as no phenotypes were observed for single mutants (Latrasse *et al.*, 2008). However, the genes are necessary for development as double homozygous mutant plants could not be obtained (Latrasse *et al.*, 2008). Sesquimutant (homozygous mutant at one allele and heterozygous at the other; *ham1* +/- *ham2* -/- or *ham1* -/- *ham2* +/-; hereafter as *ham*) individuals display defects in both male and female post-meiotic germline development and fertilisation (Latrasse *et al.*, 2008). Allele transmission assays between WT and mutant indicate female defects are more severe, although male defects are likely masked by half of pollen containing a functional *HAM* allele (Latrasse *et al.*, 2008). Male fertility defects were further examined by Alexander staining for pollen viability and found half of pollen in the *ham* background to be defective (Latrasse *et al.*, 2008).

HAM proteins are chromatin modifiers that preferentially acetylate lysine 5 of histone H4 (H4K5ac), with a lesser ability to target histone H3 (Earley *et al.*, 2007). *In vitro* experiments suggest that HAM proteins are the only acetyltransferases in *Arabidopsis* able to deposit H4K5ac (Earley *et al.*, 2007). Acetylated histones are associated with permissive chromatin environments and active transcription (Roudier *et al.*, 2011). In somatic tissues, H4K5ac is targeted to genes and implicated in facilitating high gene expression; exampled by the developmental transitioning to flowering and the targeting to FLC (Xiao *et al.*, 2013; Bu *et al.*, 2014; An *et al.*, 2020). Histone H4 acetylation has been demonstrated to physically reduce nucleosome packaging and facilitate DNA accessibility (Shogren-Knaak *et al.*, 2006).

Excitingly, the *ham* phenotype strongly resembles that observed in *dme* mutants (Choi *et al.*, 2002; Latrasse *et al.*, 2008; Schoft *et al.*, 2011), suggesting that the histone acetyltransferase activity of HAM proteins may act to allow DME access to heterochromatic sites.

Collectively, the histone acetyltransferase activity of HAM proteins and the similar mutant phenotype to *dme*, suggested that the two could be linked. It was hypothesised that VC-expressed HAM proteins enable a permissive chromatin state by depositing H4K5ac to putative heterochromatic sites. The increased accessibility is proposed to enable DME access to transposons for active DNA demethylation and subsequent transcriptional activation. The work in this chapter tests this hypothesis, first examining DNA methylation profiles between WT and mutant VCs.

3.4 HAM proteins are required for transposon DNA demethylation in VC

To assess whether HAM proteins facilitate DME access to chromatin for active DNA demethylation, the DNA methylation profiles of WT, *ham* and *dme* mutant VCs were assessed. The mutants *ham1-1* (SALK_027726) and *ham2* (SALK_106046) were crossed to generate *ham1* +/- *ham2* -/- as described by Latrasse *et al*. (2008), hereafter referred to as *ham*. To note, owing to the sesquimutant nature of *ham* and the redundancy of HAM1 and HAM2, half of profiled VCs are WT. Therefore, potential methylation effects are diluted. However, this data is comparable to downloaded *dme* VC bisulfite-seq data used herein, which was also obtained from a heterozygous line and showed notable methylation recovery compared to wild-type (Ibarra *et al.*, 2012).

This section compares DNA methylation profiles between WT SCs, WT VCs, *ham* VCs and *dme* VCs. Analysis seeks to determine whether absence of HAM proteins can lead to improper DNA demethylation, indicative of failure of DME to access chromatin.

3.4.1 Transposon DNA methylation profile of *ham* VC mimics *dme* VC

Differences in CG DNA methylation profiles between SCs and VCs are caused by DME (Ibarra *et al.*, 2012). In transposon-rich, heterochromatic pericentromeric regions, differences are readily observed (Fig. 3.2A). Cytosines with negative methylation scores when WT SC is subtracted from WT VC indicate likely DME targets (Fig. 3.2A).

Comparison of *ham* and *dme* VC methylation profiles to WT VC indicates regions (positive scores) where DME has failed to correctly demethylate DNA (Fig. 3.2A). A representative browser view of such regions indicates profiles are similar between *ham* and *dme*; thus, indicating that HAM is required for DME access to putative heterochromatin.

Focusing on specific TE examples, DNA methylation is greater in *dme* VC compared to WT VC, with *ham* VC at intermediate levels (Fig. 3.2B). As previously reported, DME principally functions in the non-CG contexts. Here, increased methylation is more readily observed in the CHG and CHH tracks (Fig. 3.2B). Together suggesting DNA methylation differences are owing to DME absence (*dme* VC) or impeded DME access (*ham* VC).

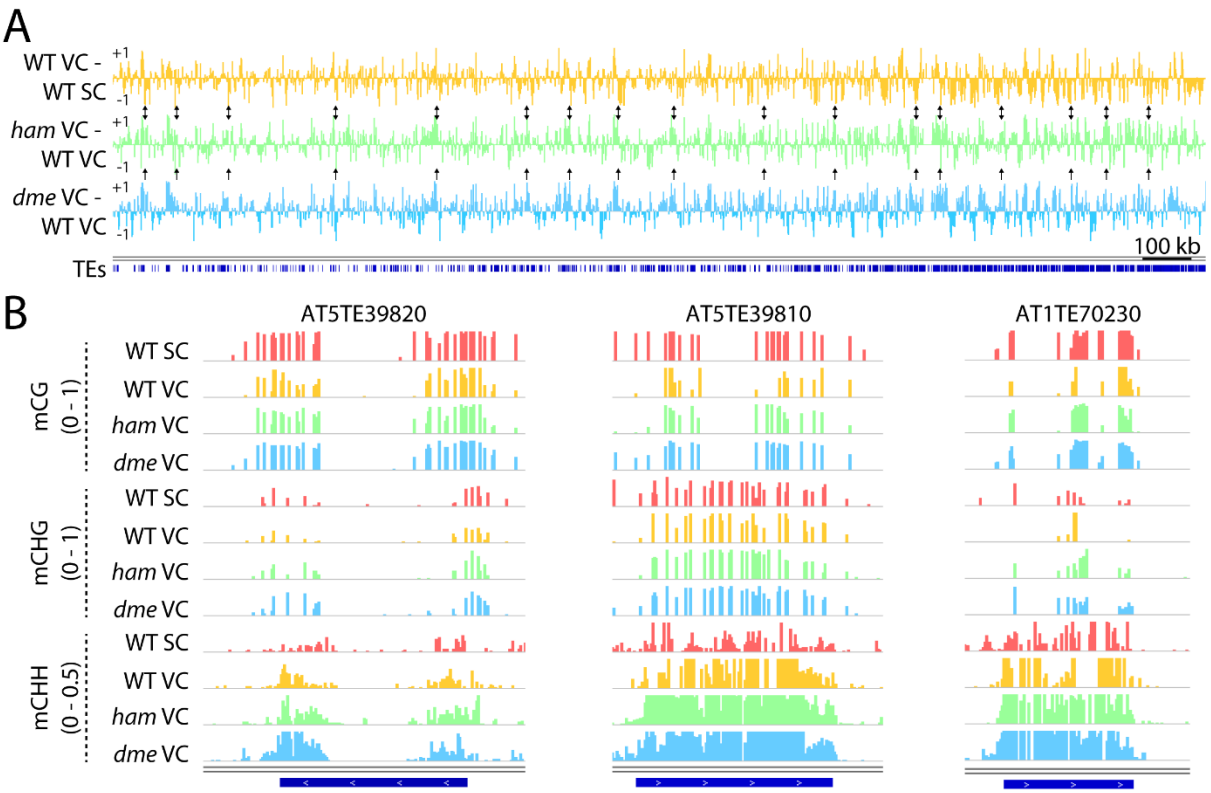


Figure 3.2: Transposon DNA demethylation is impaired in *ham* mutant VC. (**A**) Browser view of a transposon-rich (dark blue) genomic region of chromosome 1 (1:9,379,852-14,003,770). Tracks show CG methylation at single cytosine resolution. Negative methylation values in WT VC – WT SC (yellow) demonstrate regions that are demethylated in VC (arrows). Positive methylation scores of *ham* VC – WT VC (green) and *dme* VC – WT VC (blue) indicate failure to undergo demethylation. (**B**) Browser views of example TEs targeted by DME for active DNA demethylation in VC. CG (top), CHG (middle) and CHH (bottom) methylation values for WT SC (red), WT VC (yellow), *ham* VC (green) and *dme* VC (blue).

Next, to determine whether such trends were a more general feature of *ham* VC, DNA methylation profiles across all TEs was considered. Ends analysis plots show TEs aligned at the 5′ and 3′ ends (dashed lines) and average methylation levels for each 100-bp interval are plotted. In the CG context, *ham* VC DNA methylation at TEs is slightly higher than WT VC but not to the extent of *dme* VC (Fig. 3.3A). However, in the non-CG contexts, CHG and CHH, DNA methylation profiles averaged across all TEs is very similar between *ham* and *dme* VC, far greater than that observed for WT VC (Fig. 3.3, B and C). Gene body methylation in the CG context is unchanged in *ham* and *dme* VC compared to WT (Supp. Fig. 3.1), agreeing with previous reports that DME does not target such methylation (Ibarra *et al.*, 2012).

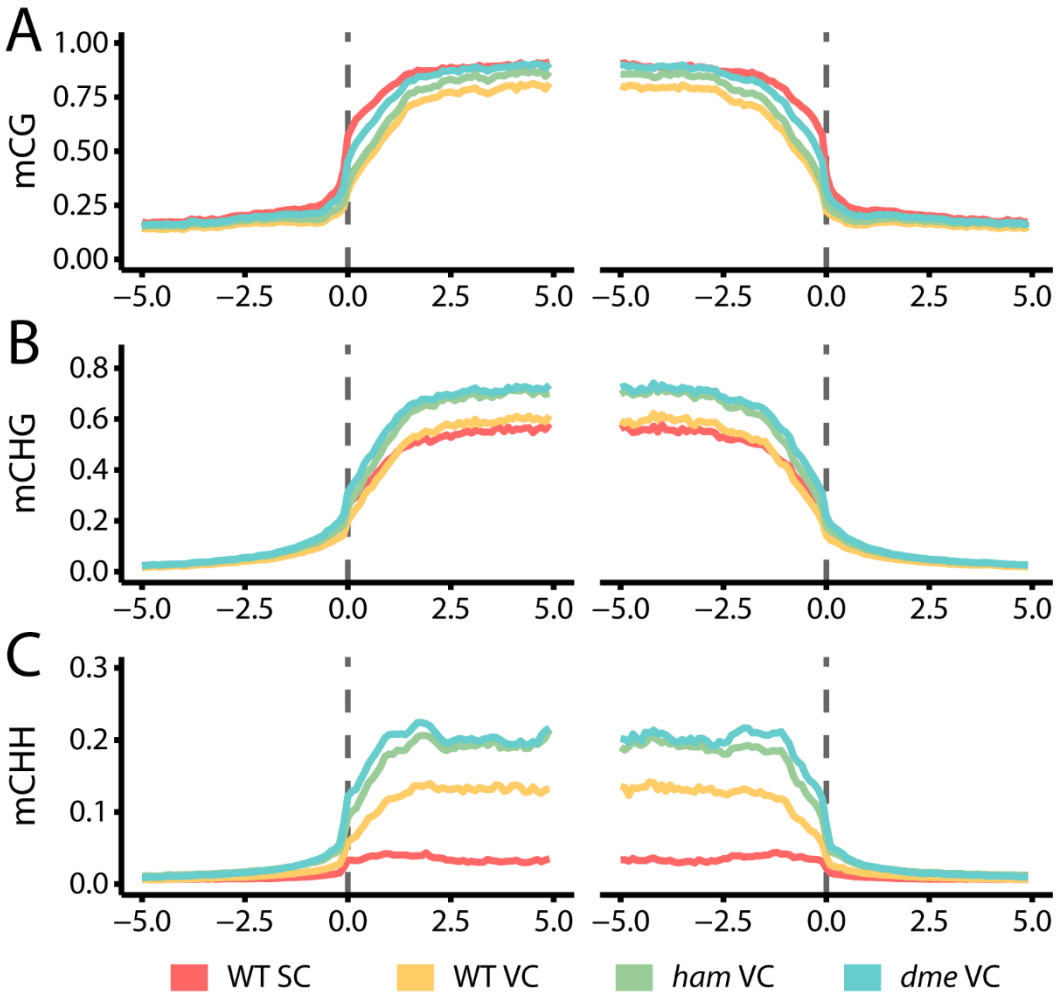


Figure 3.3: Transposon DNA methylation profiles are similar between *ham* and *dme*. (**A** to **C**) Ends analysis plots of DNA methylation at all TEs in CG (A), CHG (B) and CHH (C) contexts for WT SC (red), WT VC (yellow), *ham* VC (green) and *dme* VC (blue). Average CG/CHG/CHH methylation levels for 100 bp windows were plotted for 5 kb up and downstream of TEs aligned at the TSS (left) and TTS (right).

Initial observations of DNA methylation profiles of *ham* VC compared to *dme* and WT VC suggest that HAM proteins affect DME access to TEs, impeding subsequent demethylation activity. However, TEs exist within different chromatin environments. The following section addresses the requirement of HAM proteins for DME function at various classes of TEs.

3.4.2 Heterochromatic transposons are preferentially demethylated by HAM-facilitated DME

TEs are situated throughout the genome and can be categorised depending on the chromatin environment in which they are found. The most heterochromatic TEs are primarily found at pericentromeric regions and are densely methylated and strongly marked by H3K9me2 in somatic tissues (Simon *et al.*, 2015). TEs also occur along euchromatic chromosome arms and are less heavily methylated and can be somewhat depleted of H3K9me2. Analysis of DNA methylation profiles of TEs grouped by somatic H3K9me2 levels can determine the preferential activity of DME and HAM proteins in VC.

Consideration of all TEs shows great variability in the CG methylation level (Fig. 3.4A). However, grouping TEs by chromatin environment shows that heterochromatic TEs are the most heavily methylated and euchromatic TEs are the least, with intermediate TEs between the two (Fig. 3.4, B to D; Fig. 3.5, A to C). Absence of DME leads to greater CG methylation at all three TE groups, with the greatest difference occurring at intermediate TEs (Fig. 3.4, B to D; Fig. 3.5, A to C). Assessment of *ham* VC suggests that the protein is not required for DME function at euchromatic and intermediate TEs in the CG context, as little increase is observed compared to WT VC (Fig. 3.4, C and D; Fig. 3.5, B and C). However, profiles at heterochromatic TEs between *ham* VC and *dme* VC are very similar, indicating that DME requires HAM for function at such sites (Fig. 3.4B; Fig. 3.5A).

TEs specifically activated in the VC of pollen have been determined (He *et al.*, 2019). Such TEs exhibit far higher CG methylation in *dme* VC compared to WT, with *ham* VC intermediate (Fig. 3.4E). This indicates that HAM is required for DME access to a subset of VC-activated TEs.

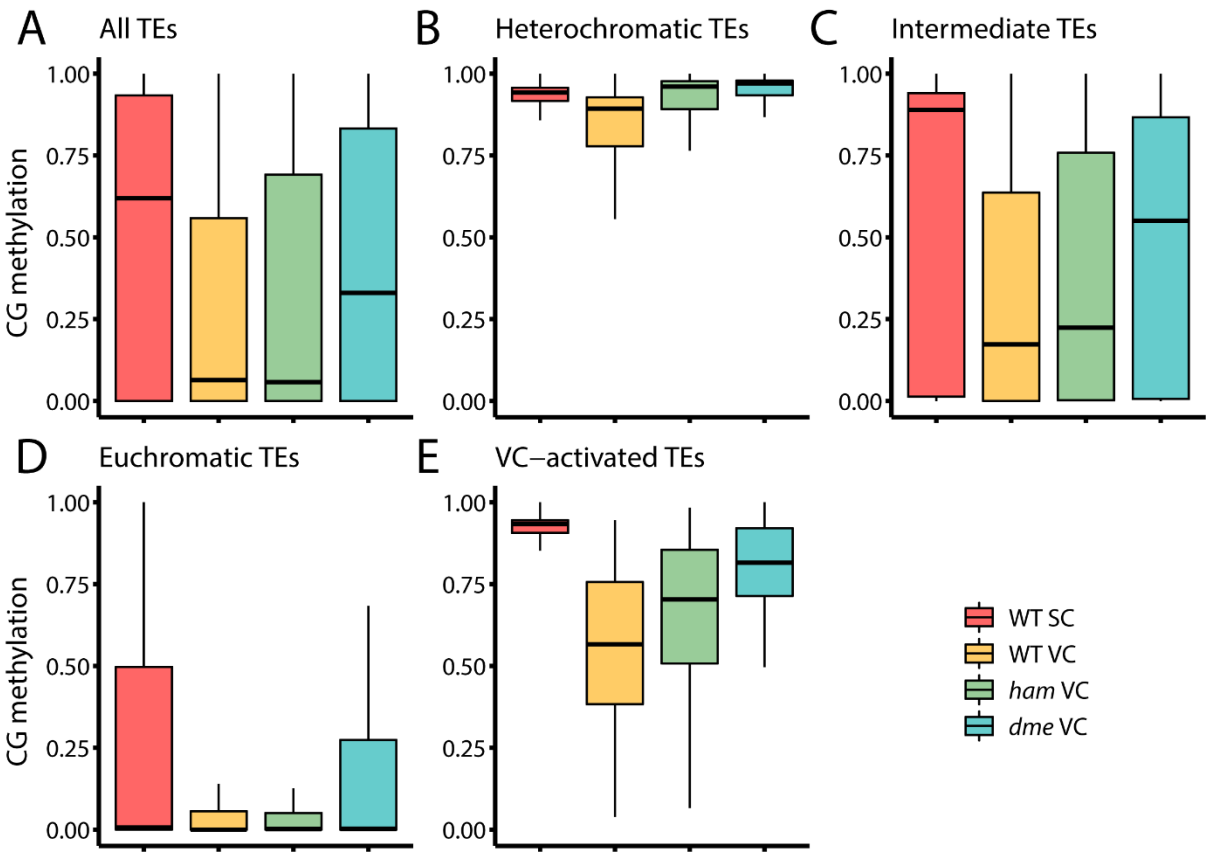


Figure 3.4: Heterochromatic transposons are preferentially targeted by HAM and DME in the CG context. (**A** to **E**) DNA methylation levels in the CG context for WT SC (red), WT VC (yellow), *ham* VC (green) and *dme* VC (blue) at all TEs (A), heterochromatic TEs (B), intermediate TEs (C), euchromatic TEs (D) and VC-activated TEs (E). Boxplots shows median (thick black bar) and first and third quartiles, with lower and upper whiskers extending to 1.5 times the interquartile range of the first and third quartiles, respectively.

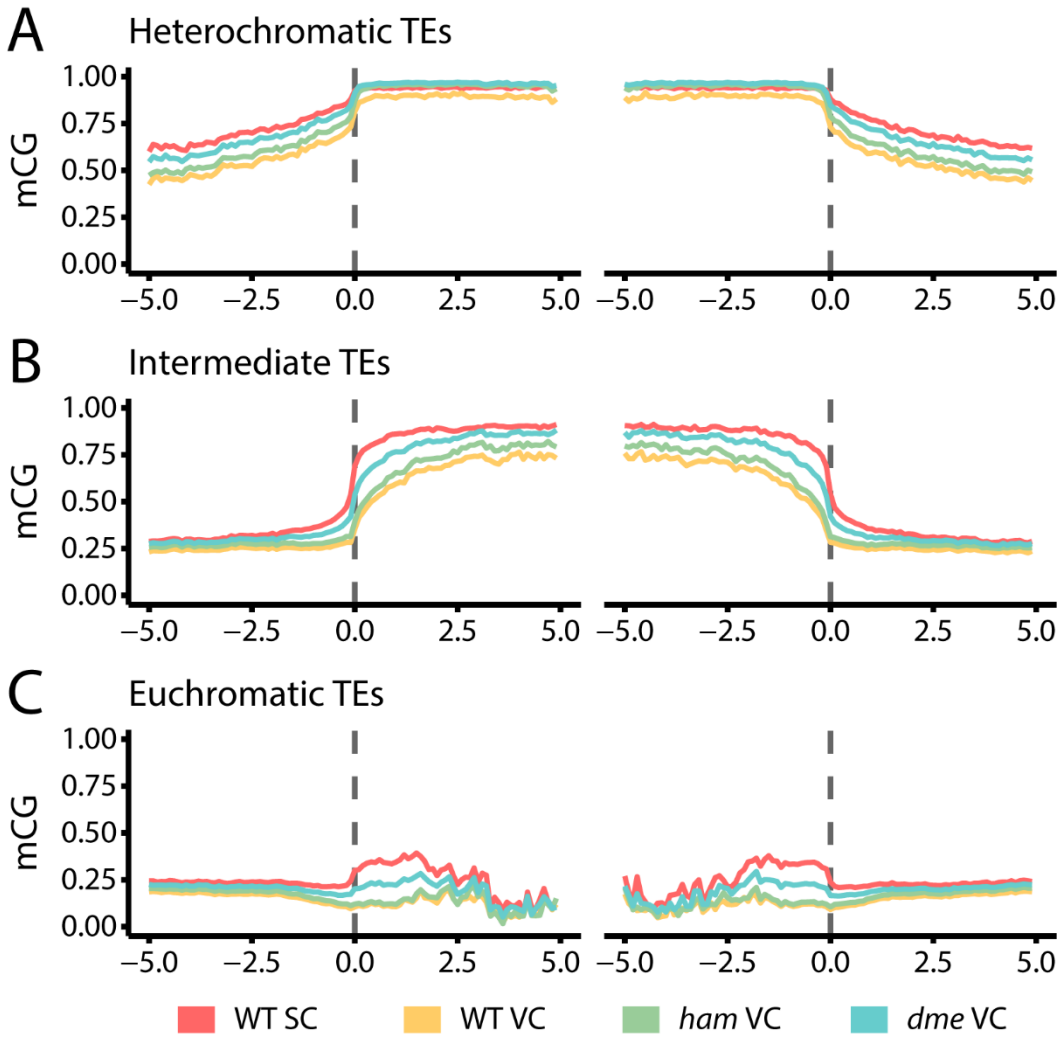


Figure 3.5: DNA methylation profiles in the CG context are most similar at heterochromatic transposons in *ham* and *dme* VC. (**A** to **C**) Ends analysis plots of DNA methylation in the CG context at heterochromatic (A), intermediate (B) and euchromatic (C) TEs for WT SC (red), WT VC (yellow), *ham* VC (green) and *dme* VC (blue). Average CG methylation levels for 100 bp windows were plotted for 5 kb up and downstream of TEs aligned at the TSS (left) and TTS (right).

DME is reported to preferentially function in non-CG contexts (Ibarra *et al.*, 2012). Clustering of TEs by chromatin environment reveals that heterochromatic and intermediate TEs are primarily undergo CHG demethylation by DME (Fig. 3.6, A to D; Fig. 3.7, A to C). Strikingly, CHG methylation profiles at heterochromatic TEs is very similar for *ham* VC and *dme* VC (Fig. 3.6B; Fig. 3.7A). This strongly indicates that HAM facilitates DME function at such TEs. To a lesser extent, a HAM requirement is also observed for intermediate TEs, but not for euchromatic TEs (Fig. 3.6, C and D; Fig. 3.7, B and C). VC-activated TEs exhibit an intermediate CHG methylation level in *ham* VC, between *dme* and WT (Fig. 3.6E). This suggests that HAM is required for a subset of DME-activated TEs.

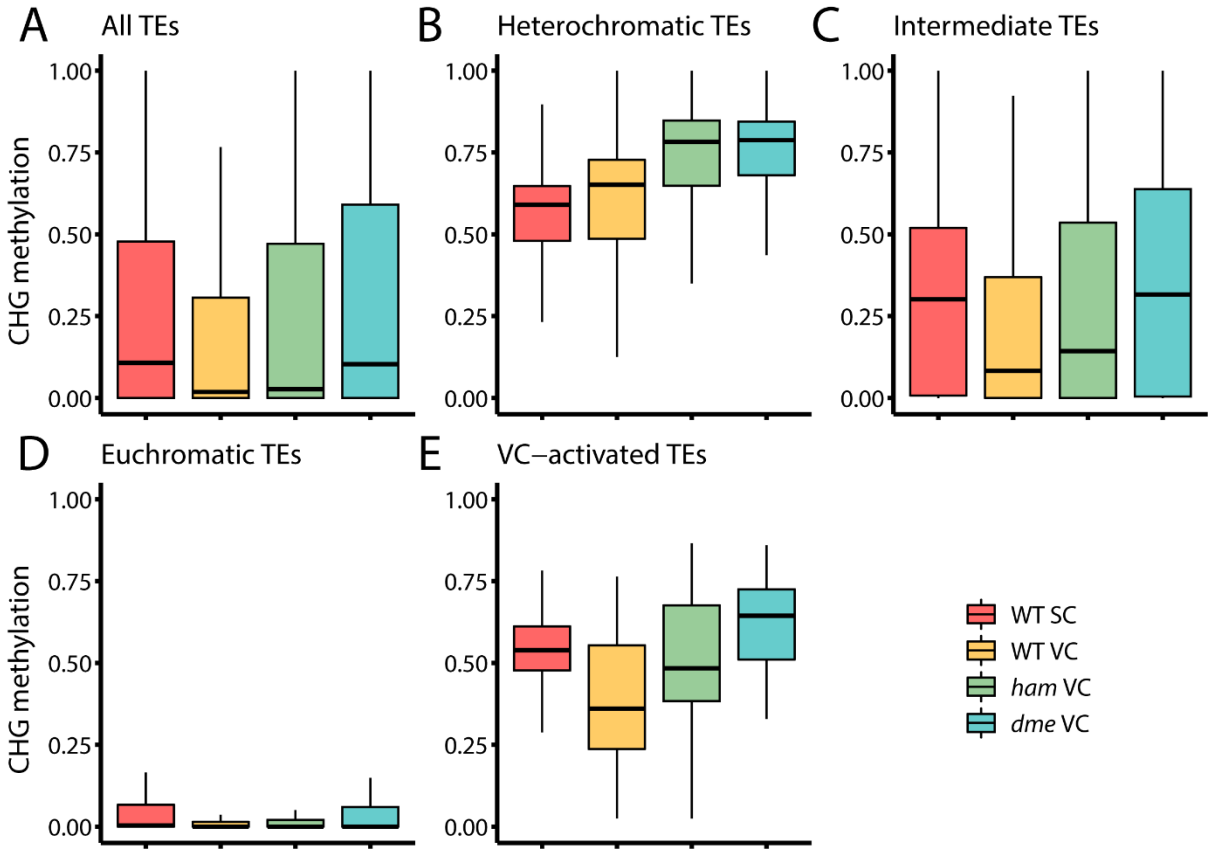


Figure 3.6: Heterochromatic transposons are preferentially targeted by HAM and DME in the CHG context. (**A** to **E**) DNA methylation levels in the CHG context for WT SC (red), WT VC (yellow), *ham* VC (green) and *dme* VC (blue) at all TEs (A), heterochromatic TEs (B), intermediate TEs (C), euchromatic TEs (D) and VC-activated TEs (E). Boxplots shows median (thick black bar) and first and third quartiles, with lower and upper whiskers extending to 1.5 times the interquartile range of the first and third quartiles, respectively.

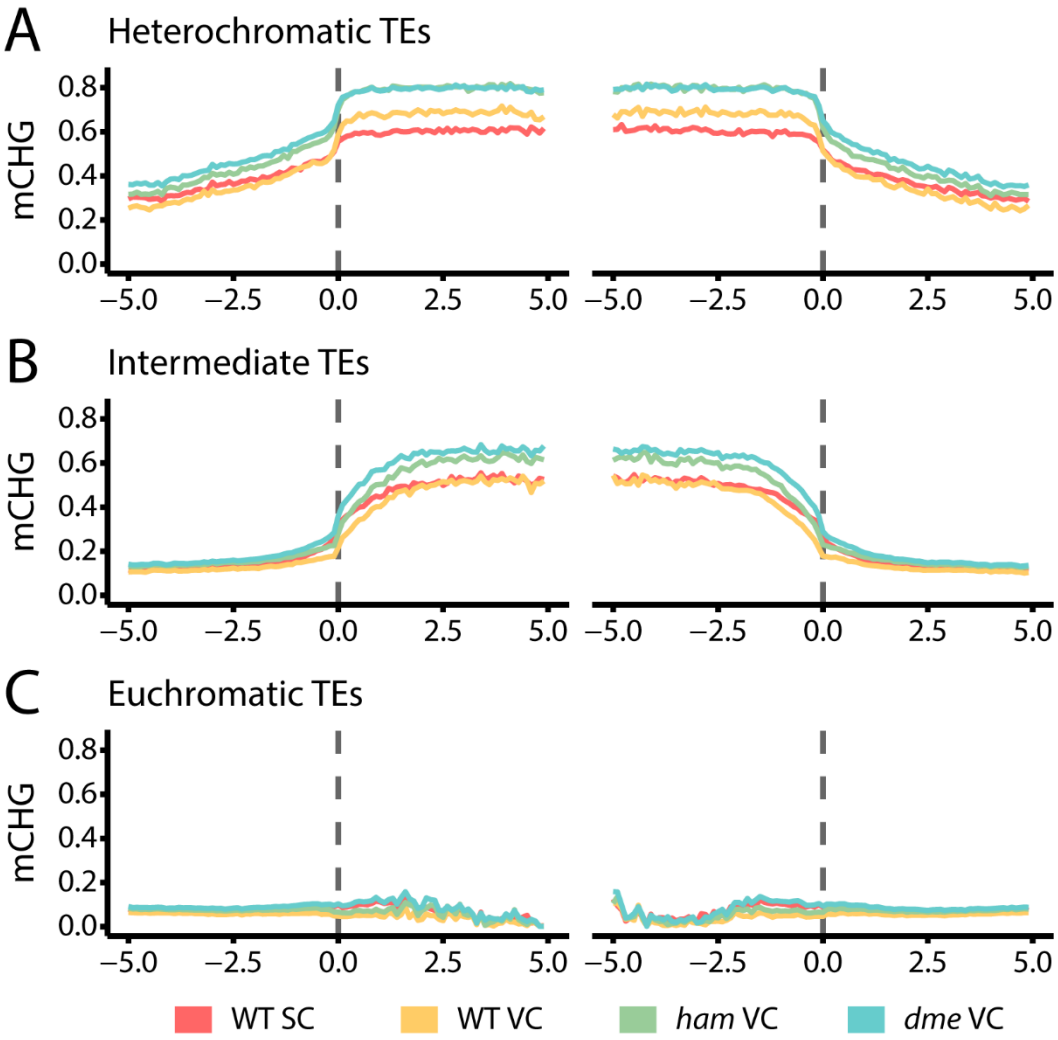


Figure 3.7: DNA methylation profiles in the CHG context are most similar at heterochromatic transposons in *ham* and *dme* VC. (**A** to **C**) Ends analysis plots of DNA methylation in the CHG context at heterochromatic (A), intermediate (B) and euchromatic (C) TEs for WT SC (red), WT VC (yellow), *ham* VC (green) and *dme* VC (blue). Average CHG methylation levels for 100 bp windows were plotted for 5 kb up and downstream of TEs aligned at the TSS (left) and TTS (right).

Similar to CHG methylation, CHH methylation at heterochromatic TEs is very similar between *ham* and *dme* VCs and considerably higher than WT VC (Fig. 3.8B; Fig. 3.9A). The effect is also observed at intermediate TEs, although to a slightly lesser degree (Fig. 3.8C; Fig. 3.9B). Euchromatic TEs gain a small increase of CHH methylation in the *ham* and *dme* mutant background (Fig. 3.8D; Fig. 3.9C). Additionally, VC-activated TEs largely fail to undergo CHH demethylation in the *ham* VC, almost to *dme* VC levels (Fig. 3.8E).

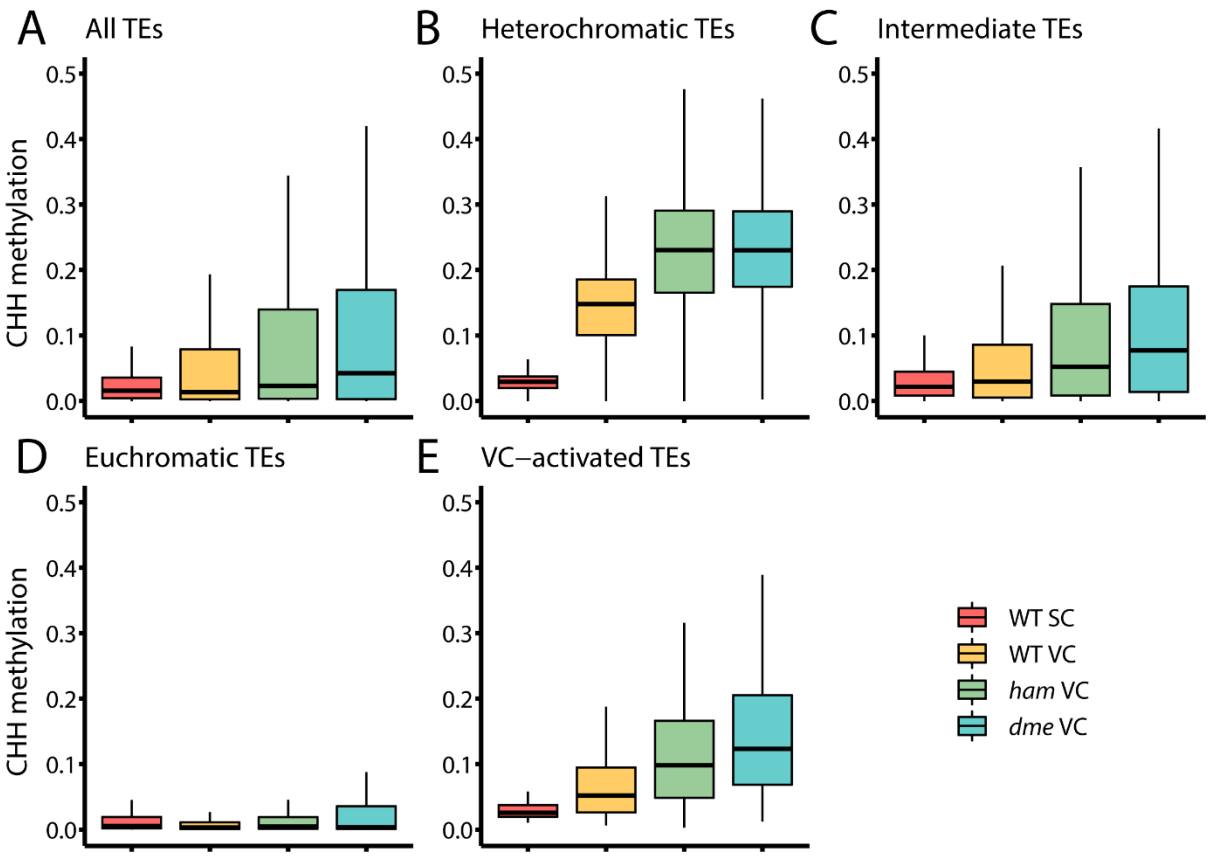


Figure 3.8: Heterochromatic transposons are preferentially targeted by HAM and DME in the CHH context. (**A** to **E**) DNA methylation levels in the CHH context for WT SC (red), WT VC (yellow), *ham* VC (green) and *dme* VC (blue) at all TEs (A), heterochromatic TEs (B), intermediate TEs (C), euchromatic TEs (D) and VC-activated TEs (E). Boxplots shows median (thick black bar) and first and third quartiles, with lower and upper whiskers extending to 1.5 times the interquartile range of the first and third quartiles, respectively.

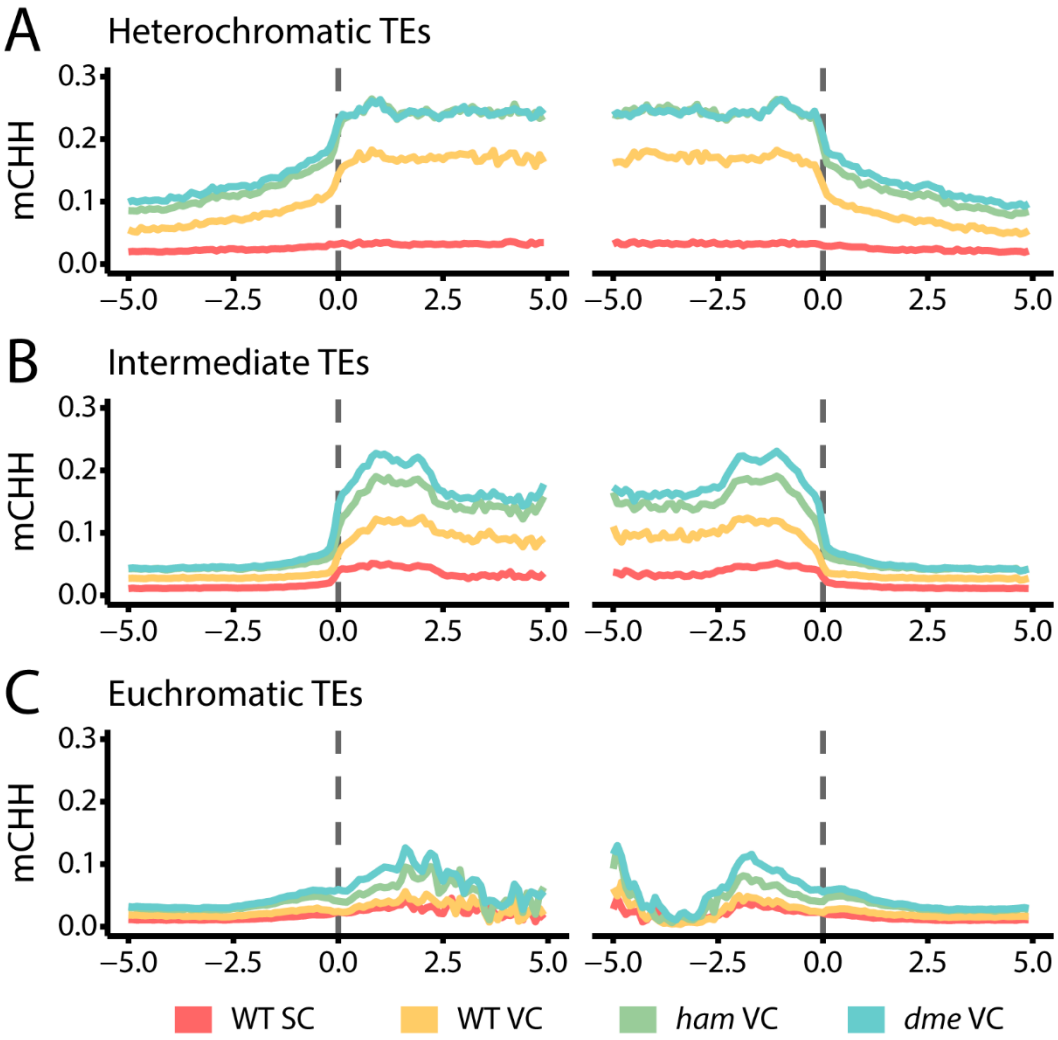


Figure 3.9: DNA methylation profiles in the CHH context are most similar at heterochromatic transposons in *ham* and *dme* VC. (**A** to **C**) Ends analysis plots of DNA methylation in the CHH context at heterochromatic (A), intermediate (B) and euchromatic (C) TEs for WT SC (red), WT VC (yellow), *ham* VC (green) and *dme* VC (blue). Average CHH methylation levels for 100 bp windows were plotted for 5 kb up and downstream of TEs aligned at the TSS (left) and TTS (right).

Overall, this section has demonstrated the preference of DME to target heterochromatic TEs across all three sequence contexts, consistent with previous reports (He *et al.*, 2019). DME also targets intermediate and euchromatic TEs, although demethylates to a lesser extent. This section has presented evidence for HAM proteins as novel determinants of DME function at heterochromatic TEs, particularly in non-CG contexts. Absence of HAM proteins causes increased methylation at such TEs, indicating that DME is unable to undertake its demethylation activity.

This analysis has considered methylation across TEs. However, DME is known to function at specific loci, primarily at TE edges. The following section assesses whether methylation in *ham* VC is similar to *dme* VC at reported DME-target loci.

3.5 DME-target loci DNA demethylation is impeded by HAM absence

DME targets have been extensively characterised by comparison of SC and VC DNA methylation data, as all CG hypomethylation in the latter is caused by DME (Ibarra *et al.*, 2012). He *et al*. (2019) identified approximately ten thousand differentially methylated regions (DMRs) between the pollen nuclei types. Such DMRs are defined as DME target loci.

DME target loci exhibit greater DNA methylation in the *dme* mutant compared to WT VC across all three sequence contexts (Fig. 3.10, A to C). Interestingly, *ham* VC methylation data is intermediate between WT and *dme* VC levels (Fig. 3.10, A to C). Analysis of loci of the same number and size to DME targets, randomised throughout the genome, shows no such trend (Supp. Fig. 3.2, A to C). This suggests that the increased methylation is specific to DME targets, and not a global effect. Considering individual DME-target loci, the vast majority of sites are methylated in both *dme* and *ham* VC, although to a slightly lesser extent in the latter (Fig. 3.11, A to C). However, a subset of loci are demethylated in *ham* VC but not in *dme* VC, particularly evident in the CG context (Fig. 3.11, A to C). Such loci are therefore demethylated independently of HAM function.

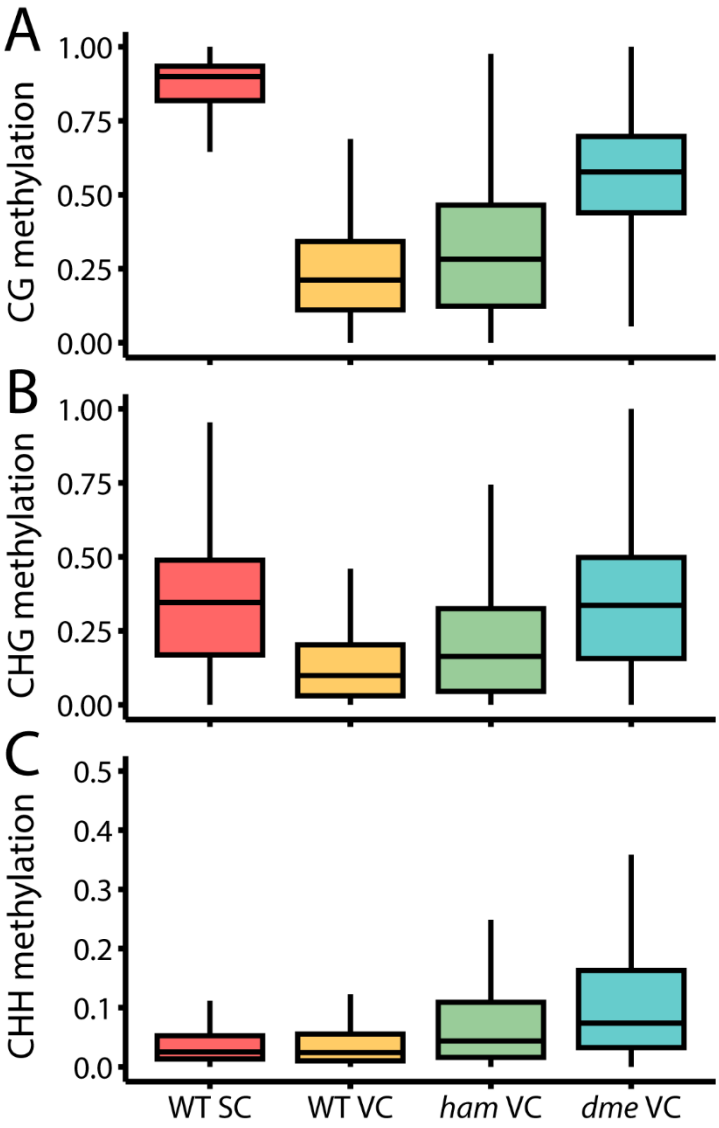


Figure 3.10: DME-target loci fail to fully undergo DNA demethylation in the *ham* mutant. (**A** to **C**) DNA methylation levels in the CG (A), CHG (B) and CHH (C) contexts for WT SC (red), WT VC (yellow), *ham* VC (green) and *dme* VC (blue) at DME-target loci. Boxplots shows median (thick black bar) and first and third quartiles, with lower and upper whiskers extending to 1.5 times the interquartile range of the first and third quartiles, respectively.

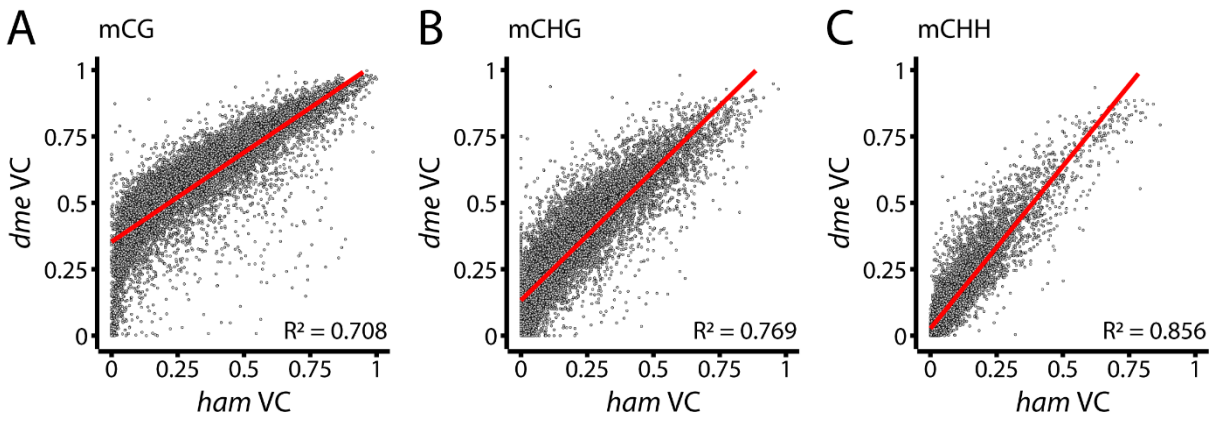


Figure 3.11: The majority of DME-target loci are dependent on HAM proteins. (**A** to **C**) DNA methylation in CG (A), CHG (B) and CHH (C) contexts at DME-target loci for *ham* VC and *dme* VC. Linear models at plotted in red, with R^2 values indicated on the figure.

Kernel density estimation plots further evidence the reliance of DME-target loci on HAM proteins for correct DNA demethylation. DME-target loci are highly hypomethylated in VC compared to SC in the CG context, whereas non-DME-target sites exhibit no such bias (Fig. 3.12A). CHG methylation is also much lower at DME-target sites, with a slight bias for higher methylation at non-DME-target sites (Fig. 3.12A). Distribution of methylation differences is bimodal for CHH contexts, with sites overlapping DME-target loci are lower in VC than SC (Fig. 3.12A). This is expected due to the obvious reinforcement of RdDM at these loci in sperm (Ibarra *et al.*, 2012; Hsieh *et al.*, 2016).

Comparing methylation differences for DME and non-DME targets between *dme* and *ham* mutant VC with WT VC assess whether demethylation effects are specific, or a more global effect. In all three sequence contexts, DNA methylation is lower for DME-target sites for both *ham* VC and *dme* VC when compared to WT VC (Fig. 3.12, B and C). Non-DME-target sites exhibit no bias for more or less methylation between mutants and WT (Fig. 3.12, B and C).

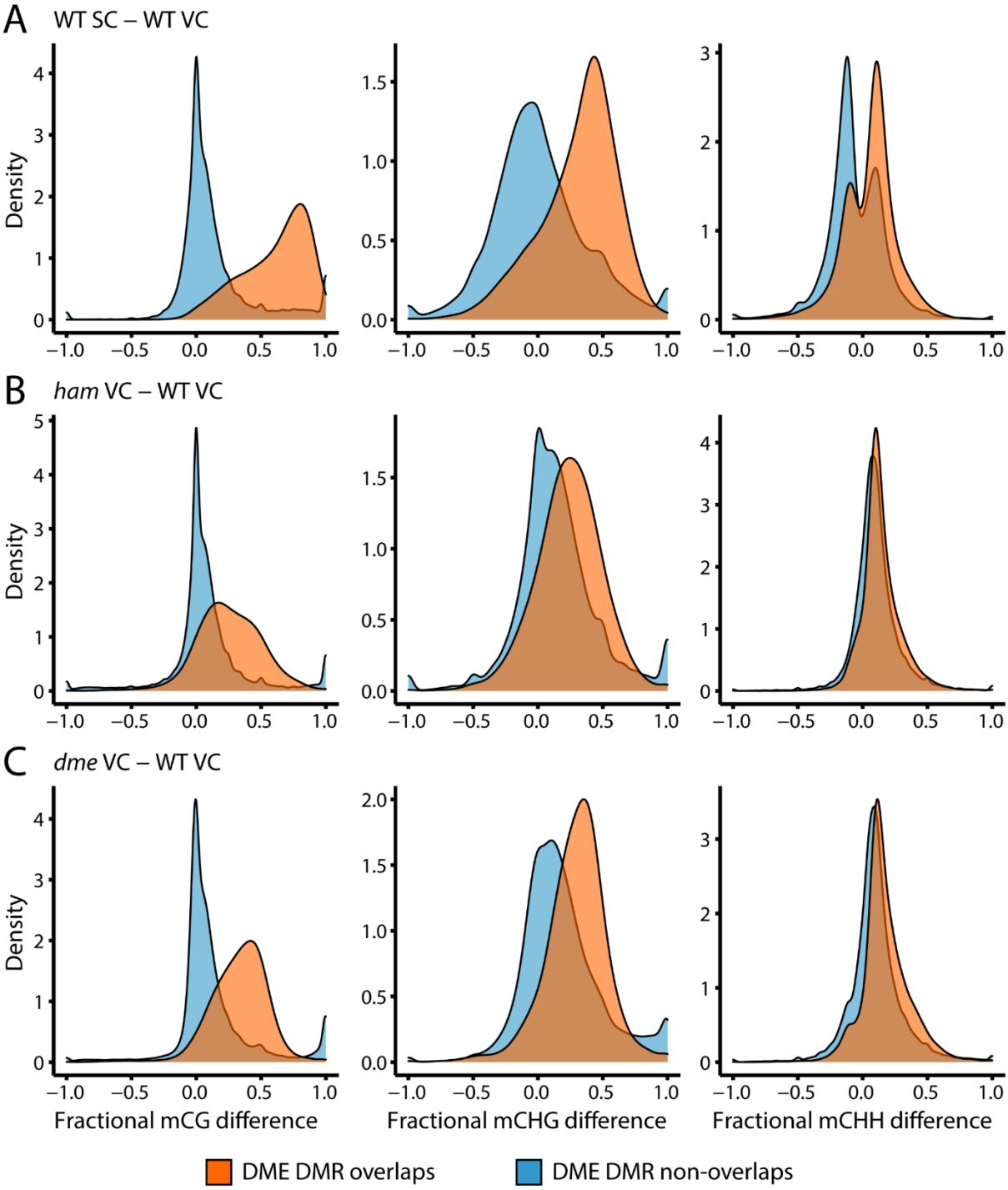


Figure 3.12: HAM-facilitated DNA methylation effect is specific to DME-target loci. (**A** to **C**) Kernel density estimation plots of CG (left), CHG (middle) and CHH (right) methylation in 50 bp windows overlapping (orange) or not overlapping (blue) DME-target loci for WT SC – WT VC (A), *ham* VC – WT VC (B) and *dme* VC – WT VC (C).

Collectively, these data describe specific effects of the HAM proteins upon DME function in VC, rather than global methylation differences. The trends between *ham* and *dme* VC are very similar, indicating that HAM is required for DME function at the majority of sites throughout the genome.

This section has provided evidence for HAM proteins as mediators of DME demethylation activity in VC. However, the mechanism by which HAM enables DME function remains unknown. The following section seeks to address whether HAM proteins could facilitate DME access to chromatin.

3.6 Chromatin accessibility is impaired at transposons in *ham* mutant VC

ATAC-seq (Assay for Transposase-Accessible Chromatin followed by sequencing) has emerged as an important tool for profiling local chromatin accessibility (Buenrostro *et al.*, 2013). Here, ATAC-seq was used to determine whether HAM proteins could facilitate DME access to target loci.

The VC is known to be absent of noted heterochromatic features, such as H3K9me2, CenH3 and linker histone H1 (Schoft *et al.*, 2009; Mérai *et al.*, 2014; Hsieh *et al.*, 2016; He *et al.*, 2019). Despite this absence, WT VC heterochromatic TEs remain less accessible than intermediate and euchromatic counterparts (Fig. 3.13, A to C). As such, accessibility could still represent a challenge for DME function at such loci.

Profiling chromatin accessibility of TEs in *ham* mutant VC revealed dramatic differences to WT. Heterochromatic TEs are far less accessible in the absence of HAM proteins (Fig. 3.13A; Fig. 3.14, A and B). Such a result is consistent with DNA methylation data, whereby heterochromatic TEs were least susceptible to DME function in the *ham* mutant VC. In agreement, intermediate TEs are marginally less accessible in *ham* VC compared to WT (Fig. 3.13B; Fig. 3.14, A and B). Such TEs fail to fully undergo demethylation in *ham*, although not to the same extent as heterochromatic TEs. Thus, the chromatin accessibility and DNA methylation effects of *ham* mutation are proportional. Euchromatic TE accessibility is unchanged in *ham* VC from WT (Fig. 3.13C; Fig. 3.14, A and B). Consistently, loss of HAM proteins does not affect DME demethylation at such TEs.

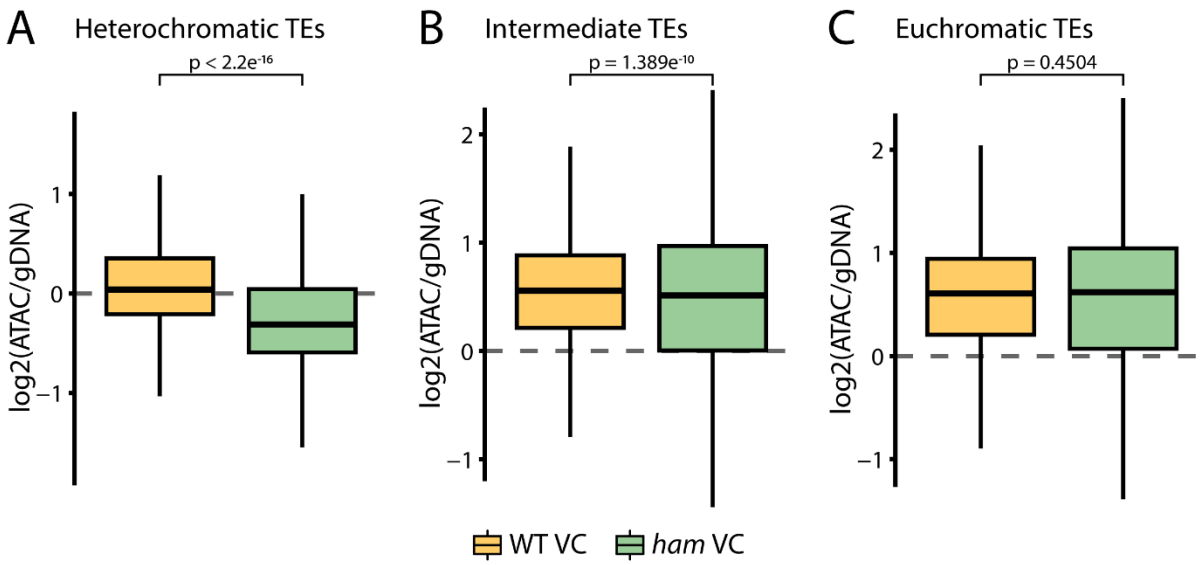


Figure 3.13: Heterochromatic, and not intermediate or euchromatic, transposons are less accessible in *ham* VC compared to WT VC. (**A** to **C**) Normalised ATAC-seq enrichment (log_2(ATAC/gDNA)) of WT VC (yellow) and *ham* VC (green) at heterochromatic (A), intermediate (B) and euchromatic (C) TEs. Boxplots shows median (thick black bar) and first and third quartiles, with lower and upper whiskers extending to 1.5 times the interquartile range of the first and third quartiles, respectively. Statistical significance was determined by Wilcoxon signed-rank test, *P* values are denoted on the figure.

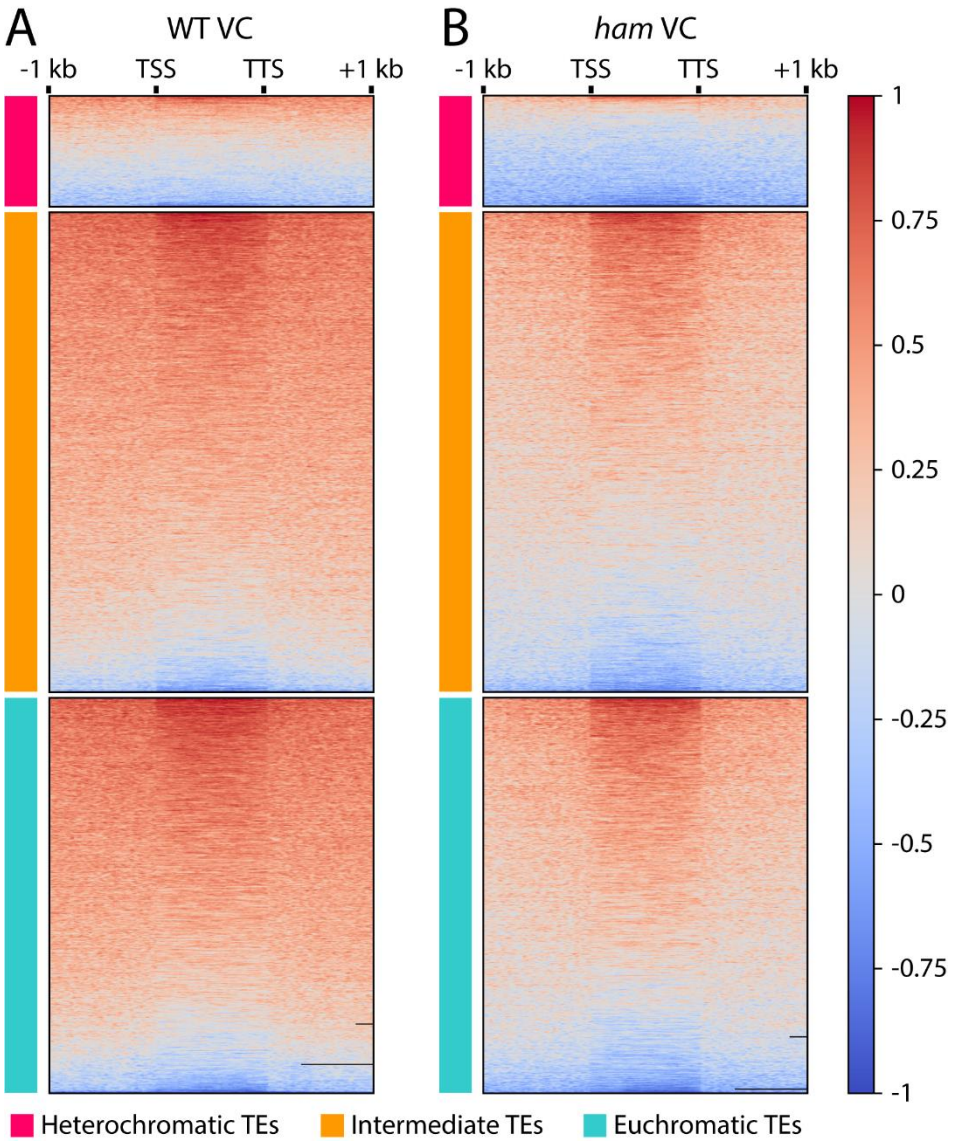


Figure 3.14: Heterochromatic transposons are less accessible in *ham* VC compared to WT VC. (**A** and **B**) Heatmaps of normalised ATAC-seq enrichment (log_2(ATAC/gDNA)) of WT VC (A) and *ham* VC (B) at heterochromatic (pink), intermediate (orange) and euchromatic (blue) TEs. TEs are scaled to 1 kb lengths; 1 kb is shown upstream/downstream of the TSS and TTS, respectively. Heatmaps are sorted from high to low accessibility in WT VC.

Altogether, HAM proteins are required for a large increase of chromatin accessibility at heterochromatic TEs and a minor increase of chromatin accessibility at intermediate TEs. HAM proteins are not required for chromatin accessibility at euchromatic TEs. These data are consistent with DNA methylation dependencies, whereby loss of HAM proteins most affected DNA demethylation at heterochromatic TEs and to a lesser extent, intermediate TEs. HAM proteins were not required for demethylation of euchromatic TEs.

Overall, these findings support the hypothesis that HAM proteins facilitate DME access to heterochromatic regions for active DNA demethylation. It was next asked whether pollen expressed TEs require HAM proteins for activation.

3.7 A subset of pollen-expressed transposons are downregulated in *ham* mutant

DME mediates demethylation of numerous TEs, although comparatively few are released from silencing (Slotkin *et al.*, 2009; Calarco *et al.*, 2012; Ibarra *et al.*, 2012). Of the TEs expressed in pollen, 114 are specifically activated in VC and are primarily situated in heterochromatic regions (He *et al.*, 2019). Previous sections have established that HAM proteins facilitate DME access to heterochromatic TEs for DNA demethylation. It was next asked whether HAM proteins are required for TE activation in pollen VC.

Transcriptome data was generated for WT and *ham* mutant pollen. Analysis of transcripts revealed 9508 TEs with expression either in either genotype (TPM > 0). 39 of which are downregulated in *ham* pollen, indicating HAM is required for activation (Fig. 3.15A). Curiously, 20 TEs are upregulated in the absence of HAM proteins, suggesting disruption of other TE regulatory mechanisms. However, such TEs upregulation could also be due to statistical noise. Overall, a slight bias exists towards lower TE expression in *ham* VC compared to WT VC (Fig. 3.15B).

This slight bias is also observed at genes, but to a much lesser extent (Fig. 3.15C). Furthermore, *ham* mutation does not lead to substantial gene misregulation, with only 6 and 7 genes up or downregulated, respectively, of 25981 genes with expression in either genotype (TPM > 0; Supp. Fig. 3.3). As such, it seems that HAM function in pollen is primarily not gene targeted. This in contrast to reported function in somatic cell types, where HAM-mediated H4K5ac can principally serves to upregulate gene expression, for example at FLC (Xiao *et al.*, 2013; Bu *et al.*, 2014; An *et al.*, 2020).

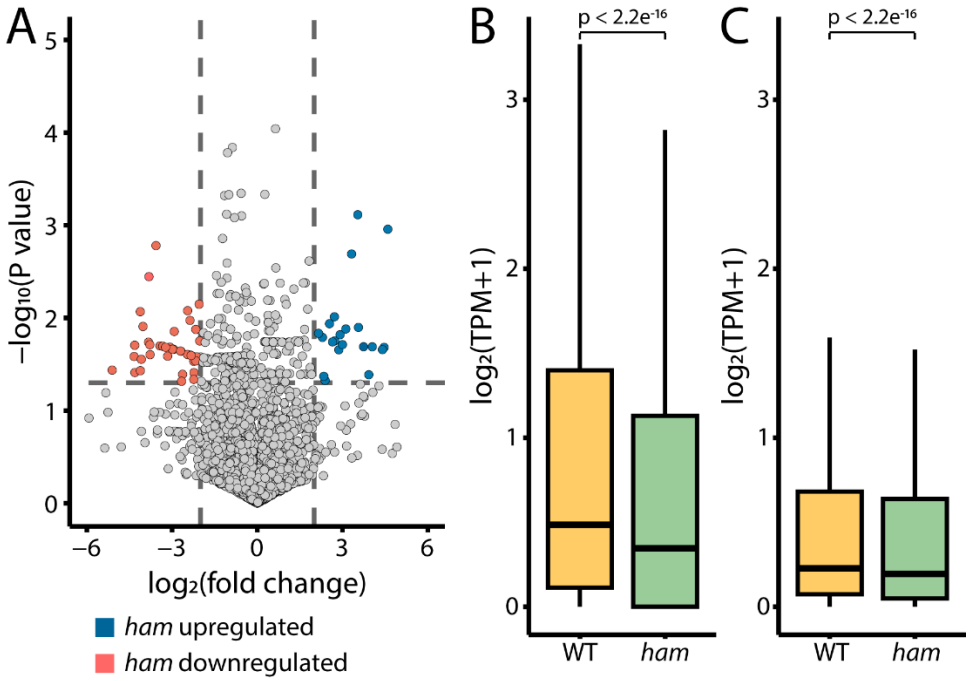


Figure 3.15: A subset of pollen-expressed transposons are downregulated in *ham* mutant. (**A**) Volcano plot of TE expression between WT and *ham* mutant pollen. TEs are defined as differentially expressed should they meet the following criteria: > 2 or < -2 $\log_2$(fold change) and P < 0.05 (likelihood ratio test; -$\log_{10}$ transformed in figure). TEs are coloured according to upregulation (blue) or downregulation (red) in *ham* mutant compared to WT. (**B**) Average expression of TEs ($\log_2$(TPM+1) between WT (yellow) and *ham* (green) pollen. Overall, TE expression in *ham* is significantly lower than WT. (**C**) Average expression of genes ($\log_2$(TPM+1) between WT (yellow) and *ham* (green) pollen. Overall, gene expression in *ham* is significantly lower than WT but with a substantially lesser difference than at TEs. Boxplots (B, C) show median (thick black bar) and first and third quartiles, with lower and upper whiskers extending to 1.5 times the interquartile range of the first and third quartiles, respectively. Statistical significance in (B, C) was determined by Wilcoxon signed-rank test, *P* values are denoted on the figure.

The majority of *ham* downregulated TEs are categorised as heterochromatic (29 of 39), with the remainder being deemed intermediate in chromatin state. Accordingly, such TEs exhibit greater methylation in *ham* VC compared to WT, particularly in the non-CG contexts (Fig. 3.16).

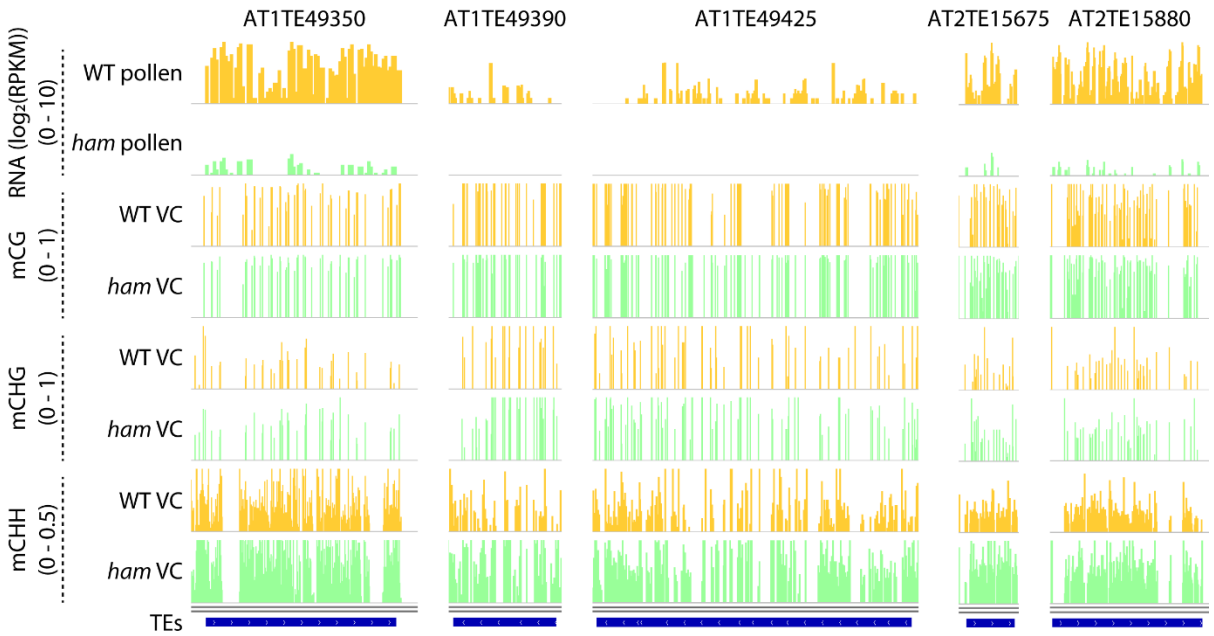


Figure 3.16: DNA methylation levels are higher at *ham* downregulated transposons. Browser view of HAM-regulated pollen expressed transposons (dark blue). Tracks show RNA expression data (log_2(RPKM)) for WT (yellow) and *ham* mutant pollen (green). DNA methylation data in CG, CHG and CHH contexts is displayed between WT and *ham* VC.

Together, analysis of WT and *ham* pollen transcriptomes has revealed that HAM proteins are required for expression of a subset of VC-expressed TEs. Absence of HAM proteins causes a negligible effect on gene transcription, indicating the function is primarily directed to TEs. HAM-regulated TEs are mostly heterochromatic in nature, in agreement with preferences observed in DNA methylation and chromatin accessibility data.

3.8 Conclusion

This chapter has implicated HAM proteins as novel determinants of active DNA demethylation in pollen VC, providing greater mechanistic insight to DME access to heterochromatic TEs.

Proteomes of pollen nuclei revealed differential proteins that could contribute to the highly dimorphic chromatin states of SC and VC. HAM proteins were identified as VC-specific in pollen and confirmed with GFP-tagging. Comparisons of WT, *ham* and *dme* VC BS-seq data revealed that DNA demethylation was impaired in the absence of HAM, particularly at heterochromatic TEs. Such TEs also exhibited lower chromatin accessibility in the *ham* mutant when profiled with ATAC-seq. Finally, transcriptome analysis revealed the downregulation of a subset of pollen expressed TEs in *ham* mutant pollen, whilst gene expression was largely unperturbed.

Collectively, this work has provided evidence that HAM proteins facilitate greater chromatin accessibility at heterochromatic TEs, likely through acetyltransferase activity targeting histone H4 lysine 5 (Fig. 3.17). The greater chromatin accessibility is suggested to enable efficient DME-mediated DNA demethylation and subsequent TE activation. Ultimately, HAM-mediated TE activation in the VC may serve to enable siRNA biogenesis and reinforce TE silencing in SC by RdDM (Fig. 3.17).

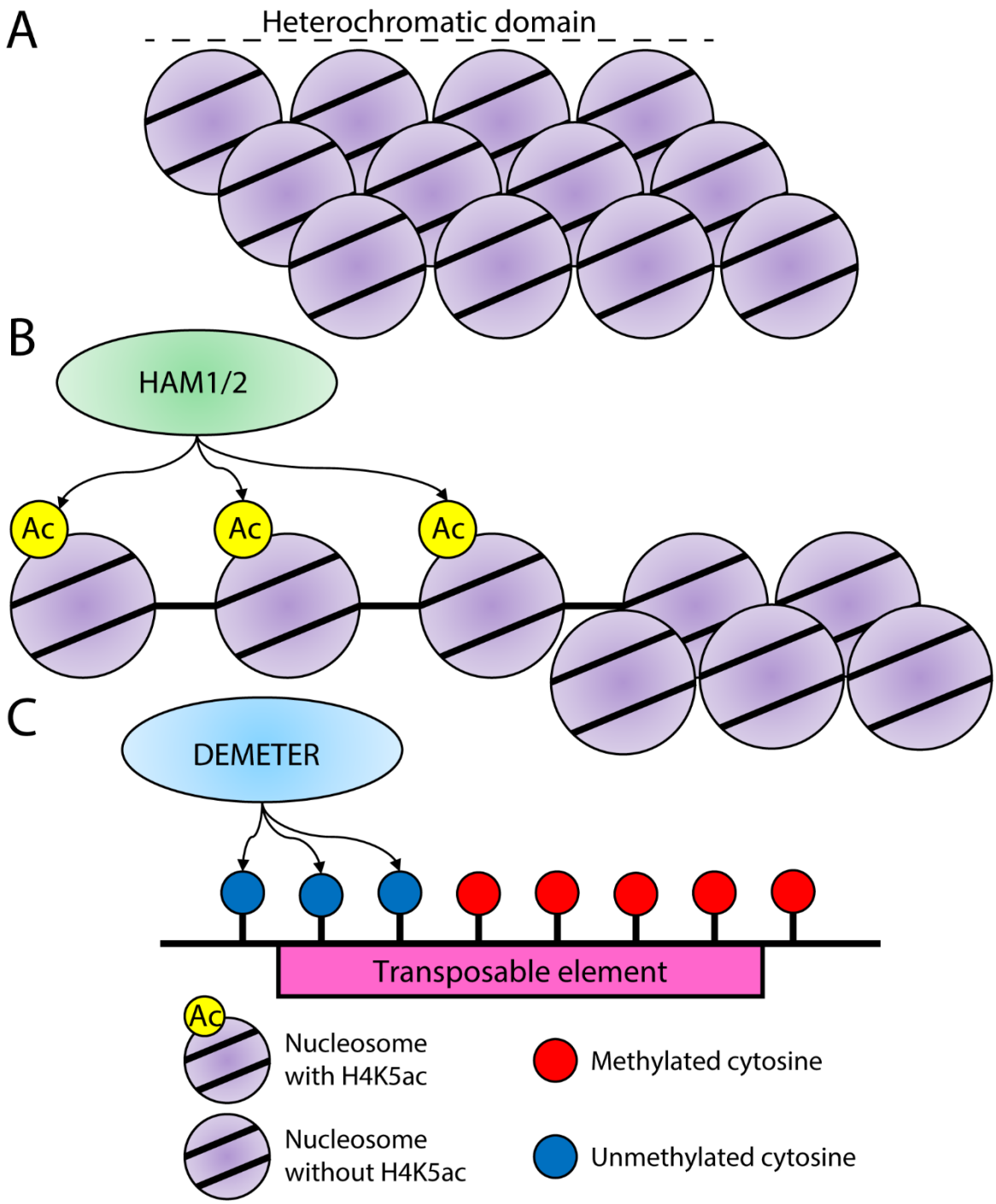


Figure 3.17: Model of HAM-mediated DME access to heterochromatic TEs in VC. (**A**) Heterochromatic domains remain locally inaccessible in VC chromatin. (**B**) HAM-mediated H4K5ac relaxes chromatin structure, increasing DNA accessibility. (**C**) DME gains entry to heterochromatin targets, such as TEs, and performs active DNA demethylation. Decreased methylation activates TEs in VC. Expressed TEs serve as templates for siRNA biogenesis which, in turn, guide RdDM machinery in SC to reinforce TE silencing for the next generation.

3.9 Discussion

Histone acetylation has long been implicated in promoting gene expression (Roudier *et al.*, 2011). Specifically, MYST-type acetyltransferases, conserved across eukaryotes, are widely reported to target genes for activation through development (Carrozza *et al.*, 2003). However, histone acetylation localised at TEs in heterochromatic regions is comparatively unexplored.

ROS1, a DNA glycosylase (Gong *et al.*, 2002), functions in somatic tissues and is reported to target TEs close to active genes in order to prevent aberrant DNA methylation spread (Zhu *et al.*, 2007). ROS1 sites are enriched for histone H3 lysine 18 acetylation (H3K18ac) but depleted of heterochromatic marks such as H3K9me2 and H3K27me1 (Tang *et al.*, 2016). Therefore, a precedent exists for active demethylation at acetylated histones but within largely euchromatic genomic environments. HAM-mediated H4K5ac at heterochromatic DME targets represents a novel approach to providing DNA glycosylase access.

Future work will seek to further validate the relationship between HAM-mediated H4K5ac and DME access by performing ChIP-seq against the histone mark in VC. Given the permissive nature of H4K5ac and its noted association with increased gene expression in somatic tissues (Xiao *et al.*, 2013; Bu *et al.*, 2014; An *et al.*, 2020), it will be interesting to note whether a dual role exists for the modification in VC. It is possible that H4K5ac deposited to heterochromatic TEs by HAM proteins both permits DME-mediated demethylation and also promotes transcriptional machinery access for TE expression.

How a euchromatic mark (H4K5ac) is recruited to normally heterochromatic regions remains unknown. VC chromatin is remarkably different to other reported global chromatin states in *Arabidopsis*. Cytologically detectable chromocenters are not observed, along with characteristic heterochromatic marks, such as H1, CenH3 and H3K9me2 (Schoft *et al.*, 2009; Mérai *et al.*, 2014; Hsieh *et al.*, 2016). Despite this, heterochromatic TEs remain less accessible than intermediate or euchromatic counterparts (Fig. 3.12, A to C). The highly compacted nature of heterochromatic regions in other tissues could occlude active chromatin modifiers from entering and fulfilling function. However, in VC, where such compaction is lost, a mechanism can be envisaged whereby active modifiers can gain entry and enact a further chromatin relaxation.

However, the question remains as to how DME is recruited to specific loci. Analysis of DME domains suggested the C-terminal catalytic core is sufficient for targeting to euchromatic sites, but the N-terminal region is required for recruitment to heterochromatic loci (C. Zhang *et al.*, 2019). The authors posit that the N-terminus serves to enlist other factors for heterochromatic targeting (C. Zhang *et al.*, 2019). Evidence presented in this chapter suggests that HAM proteins are required for DME function at heterochromatic sites; therefore, the two may function together in a complex.

In fact, it is well-documented that MYST family histone acetyltransferases function as multi-protein complexes in animal systems (Carrozza *et al.*, 2003; Avvakumov and Côté, 2007). Such complexes remain largely unexplored in plants. Future experiments would include IP-MS (immunoprecipitation followed by mass spectrometry) using the HAM1/2-GFP tagged lines. Employing the technique in pollen versus somatic tissues could identify complex components specific to each VC chromatin. Furthermore, IP-MS would indicate whether HAM and DME directly interact with one another *in vivo*.

It is also possible that HAM proteins are responsible for DME targeting to heterochromatic sites. HAM1/2 are defined by the MYST acetyltransferase domain, although the protein also contains a putative chromodomain and Zinc-finger C2H2-type (Znf-C2H2) domains (Latrasse *et al.*, 2008). Chromodomains bind methylated H3 histone tails (Eissenberg, 2012). Evidence from somatic tissues suggests a crosstalk of H3K36me3 with HAM proteins (Bu *et al.*, 2014; An *et al.*, 2020), it is possible that this permissive mark is also required for DME function in VC heterochromatin. Znf-C2H2 domains are well characterised as sequence-specific DNA binding motifs (Englbrecht, Schoof and Böhm, 2004). The Znf-C2H2 domain of HAM proteins could guide acetyltransferase activity to specific targets encoded by the DNA sequence. Overall, this chapter has provided evidence that HAM and DME function together at heterochromatic TEs to release silencing, but how this targeting occurs remains unknown.

DME also functions in the female companion cell, known as the central cell (Choi *et al.*, 2002; Schoft *et al.*, 2011). Similar to *dme* mutant, the *ham* fertility defect is more severe in female sexual lineage development than male (Choi *et al.*, 2002; Latrasse *et al.*, 2008). In addition to the TE demethylation activity, DME establishes imprinting at a number of central cell genes which result in maternally biased expression in the endosperm of seeds (Hsieh *et al.*, 2011; Rodrigues and Zilberman, 2015). It would be interesting to explore whether HAM proteins are also required for DME function in female sexual lineage development.

3.10 Materials and methods

3.10.1 Plant growth conditions

Plants were grown under long day (16 hr light; 8 hr dark) conditions at 22 °C. Seedlings were grown on germination medium (GM) plates without glucose under the same conditions.

3.10.2 Genotyping of T-DNA mutants

T-DNA insertion mutants, *ham1-1* (SALK_027726) and *ham2* (SALK_106046) were obtained from NASC (Nottingham Arabidopsis Stock Centre). Heterozygous *ham1-1* was crossed to homozygous *ham2* to generate the sesquimutant line *ham1* +/- *ham2* -/- (*ham*), as reported by Latrasse *et al*. (2008). Self-fertilised progeny were genotyped following each sowing to ensure the correct genetic background for experiments. The same primers were used as per Latrasse *et al*. (2008).

3.10.3 Molecular cloning and plant transformation

HAM1 and *HAM2* genomic DNA sequences were amplified, along with ~2 kb upstream for the respective promoters. PCR products were ligated to the pCAMBIA1300 vector backbone containing a C-terminal eGFP tag using the In-Fusion cloning system (Takara Bio). Constructs were transformed to GV3101 *Agrobacterium tumefaciens* and subsequently used for floral dip *Arabidopsis* transformation (Clough and Bent, 1998). T1 individuals were selected on GM plates with Hygromycin.

3.10.4 Confocal microscopy

Pollen was obtained from *pHAM1::HAM1-eGFP* and *pHAM2::HAM2-eGFP* by collecting open flowers and vertexing in Galbraith buffer (45 mM $MgCl_2$, 30 mM sodium citrate, 20 mm MOPS, 0.1 % Triton X-100, pH 7.0). DAPI (2 µg/ml) was added to the pollen suspension before spinning down. The pellet was pipetted to a slide and covered with a coverslip. Images were obtained using a Leica SP5 confocal microscope.

3.10.5 Bisulfite-seq library preparation and sequencing

FACS was used to isolate vegetative nuclei from pollen, as described in Chapter 2 section 2.15.2. Libraries were constructed using the Nugen Ovation Ultralow Methyl-Seq Library Systems (0336) kit, incorporating two rounds of bisulfite conversion with the Qiagen EpiTect Fast Bisulfite Conversion (59802) kit.

Library quality was assessed by Bioanalyzer (Agilent). The Nextseq 500 (Illumina), at the John Innes Centre, was used to sequence libraries single-end with 76 bp reads.

3.10.6 Analysis of bisulfite-seq data

Bisufite-seq reads were mapped using Nightfury, a mapper currently under development by Dr Martin Vickers (Feng Lab, John Innes Centre). Methylation was called using MethylDackel (https://github.com/dpryan79/MethylDackel). Information was extracted for all three sequence contexts using the following options: MethylDackel extract --CHG --CHH.

Average methylation across TEs and DME-target loci for boxplots and ends analysis was determined using bssequel-0.0.1 and xftools-0.0.1, in-house packages maintained by the Zilberman and Feng labs, respectively (John Innes Centre). Data was plotted using ggplot2 in R-3.6.0 (Wickham *et al.*, 2019). DME DMRs were downloaded from He *et al*. (2019). Control loci for DME-target sites, termed pseudo-DME-target loci, were generated using the BEDtools-2.28.0 (Quinlan and Hall, 2010) shuffle command.

Kernel density plots utilise methylation data over 50 bp windows. Data was filtered such that at least one of the samples tested had a minimum of 20 informative sequenced cytosines and a fractional methylation of >0.5 for CG, >0.4 for CHG or >0.1 for CHH. Windows were overlapped with DME-target sites using the BEDtools intersect command. Windows were deemed to overlap DME-target sites if ≥ 25 bp intersected (-f 0.5). Overlaps and non-overlaps were called using -wa and -v options, respectively. Kernel density plots were generated in R using ggplot2.

3.10.7 Vegetative nuclei ATAC-seq library preparation, sequencing and analysis

Vegetative nuclei were isolated by FACS, as described in Chapter 2 section 2.15.2. ATAC-seq libraries were prepared and sequenced as per Chapter 2 section 2.15.20. Two replicates for each genotype were confirmed to be highly correlated and were then analysed in the same way as Chapter 2 section 2.15.21.

3.10.8 Pollen RNA-seq library preparation, sequencing and analysis

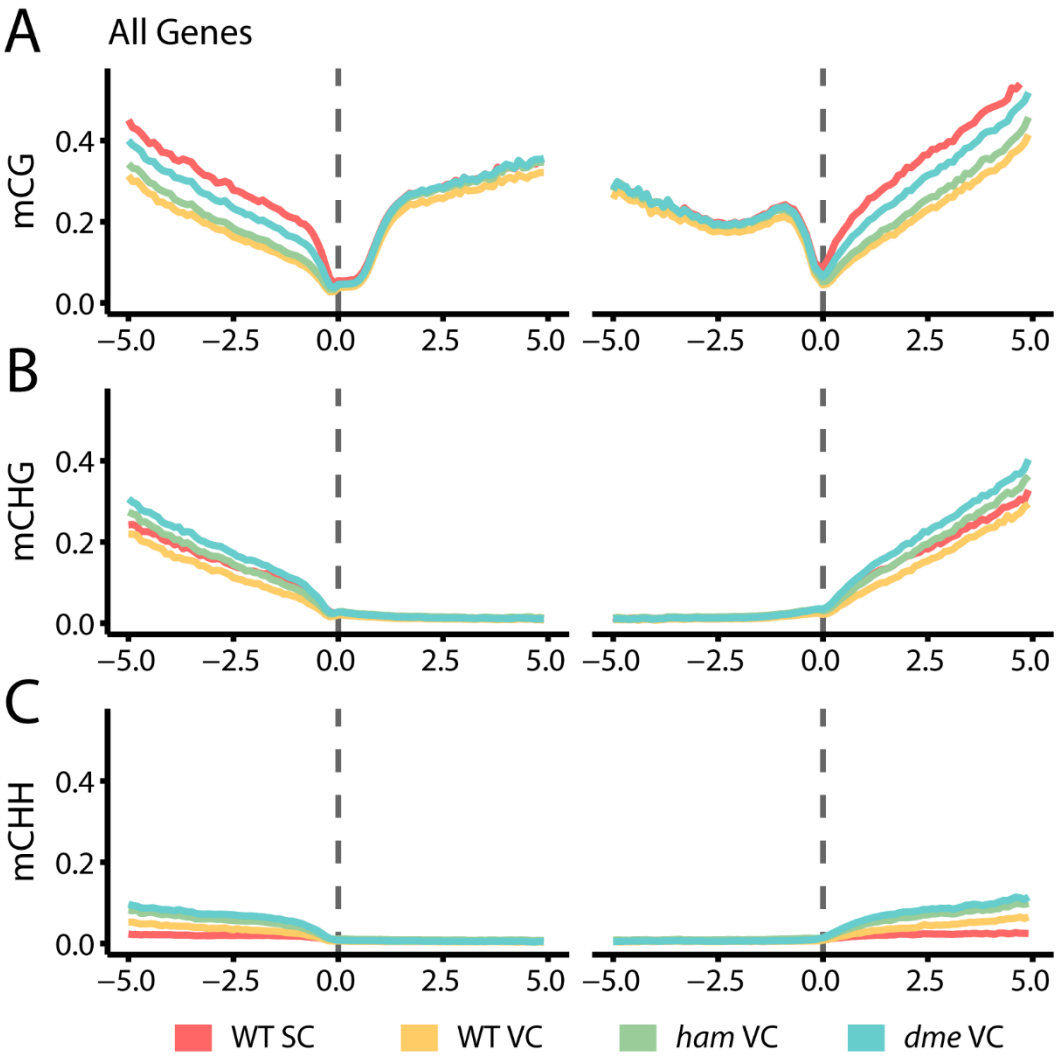
Open flowers were collected from WT and *ham* mutant plants. Flowers were vortexed in Galbraith buffer to extract pollen. RNA was extracted using the RNeasy plant mini kit (QIAGEN). DNA was removed by digestion with the DNase digestion kit (QIAGEN). Libraries were prepared from 100 ng input RNA using the Nugen Universal RNA-seq kit (0364).

Library concentrations and fragment size distributions were determined by Bioanalyzer (Agilent). Libraries were sequenced single-end, 76 bp reads, at the John Innes Centre, using the NextSeq 500 (Illumina).

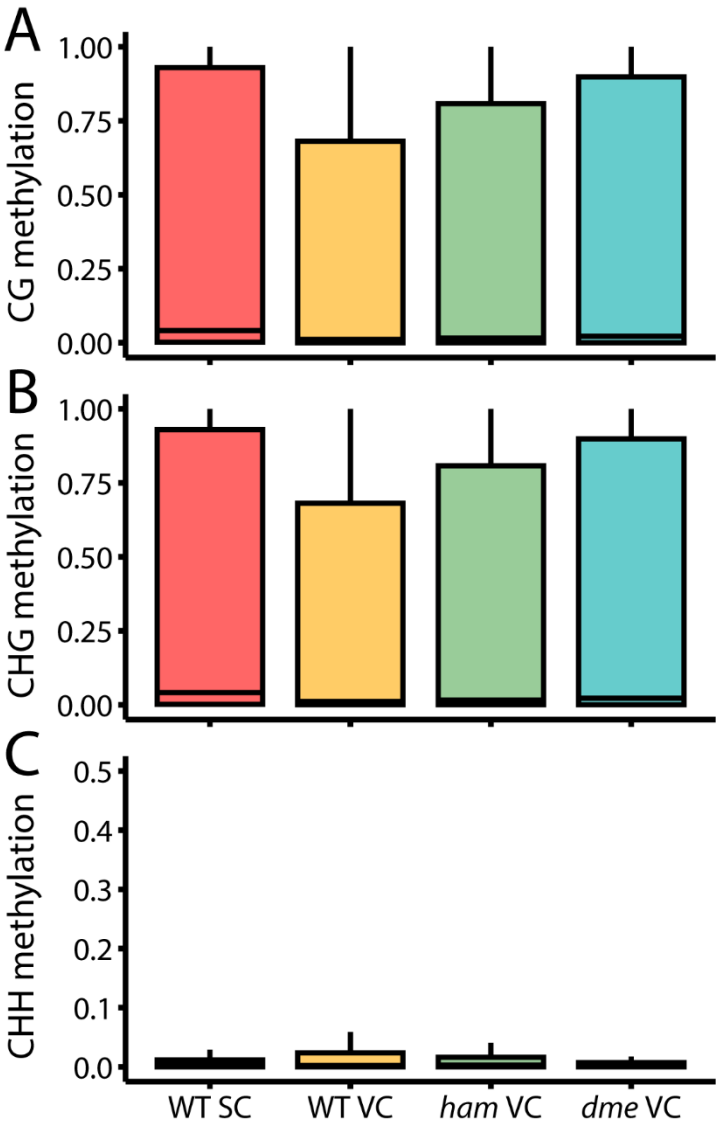
Reads were mapped to TAIR10 using TopHat-2.0.10 (Kim *et al.*, 2013). Two replicates were used for each genotype. Kallisto-0.43.0 (Bray *et al.*, 2016) and Sleuth-0.30.0 (Pimentel *et al.*, 2017) were used to obtain TPM and *P* values, respectively. Genes and TEs were required to have ≥ 2-fold change and $P < 0.05$ between samples to be categorised as differentially expressed. Volcano plots of fold change versus *P* values were generated using a custom script in R.

3.11 Supplemental data

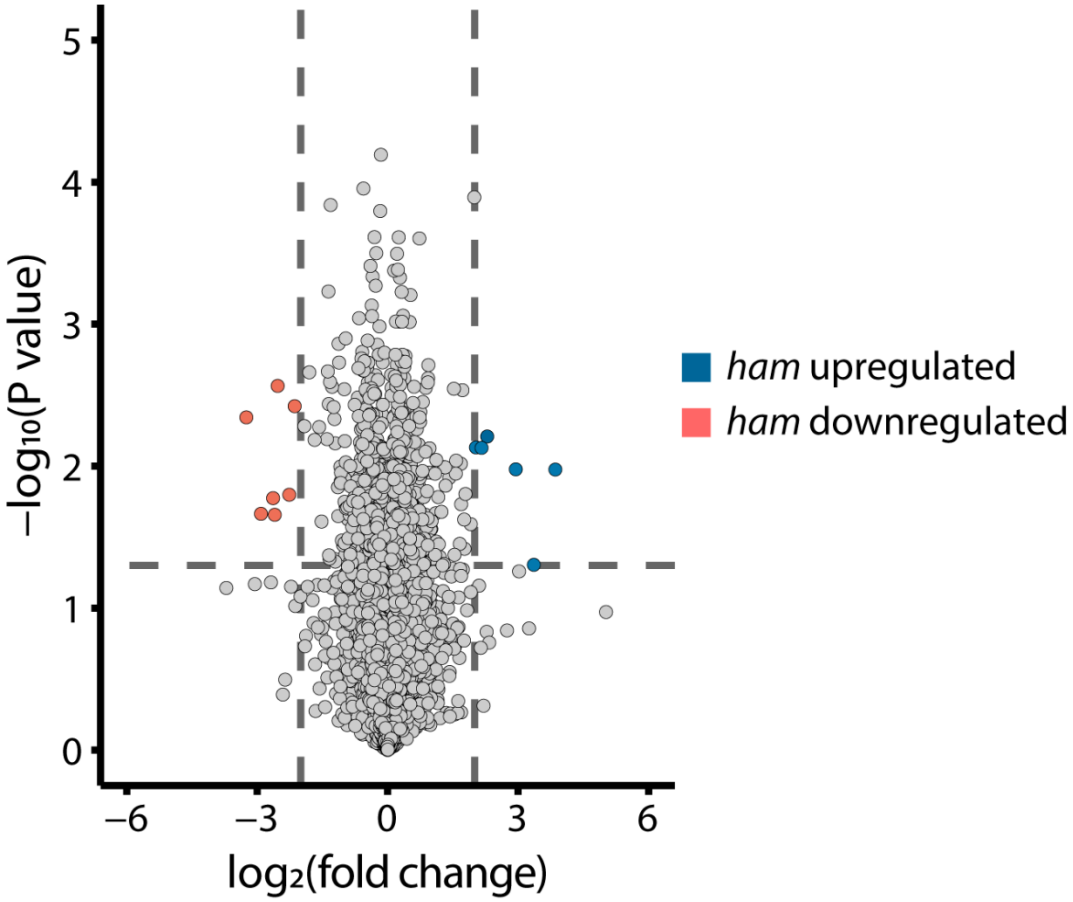
3.11.1 Supplemental figures



Supplemental Figure 3.1: DNA methylation profiles are unchanged at genes in *ham* and *dme* VC. (**A** to **C**) Ends analysis plots of DNA methylation at all genes in CG (A), CHG (B) and CHH (C) contexts for WT SC (red), WT VC (yellow), *ham* VC (green) and *dme* VC (blue). Average CG/CHG/CHH methylation levels for 100 bp windows were plotted for 5 kb up and downstream of genes aligned at the TSS (left) and TTS (right).



Supplemental Figure 3.2: Pseudo-DME-target loci DNA methylation is the same between WT and mutant VC. (**A** to **C**) DNA methylation levels in the CG (A), CHG (B) and CHH (C) contexts for WT SC (red), WT VC (yellow), *ham* VC (green) and *dme* VC (blue) at pseudo-DME-target loci. Pseudo-DME-target loci are of the same number and size to DME targets but randomised in location throughout the genome. Boxplots shows median (thick black bar) and first and third quartiles, with lower and upper whiskers extending to 1.5 times the interquartile range of the first and third quartiles, respectively.



Supplemental Figure 3.3: Gene expression is largely unchanged between WT and *ham* pollen. Volcano plot of gene expression between WT and *ham* mutant pollen. Genes are defined as differentially expressed should they meet the following criteria: > 2 or < -2 $\log_2$(fold change) and P < 0.05 (likelihood ratio test; -$\log_{10}$ transformed in figure). Genes are coloured according to upregulation (blue) or downregulation (red) in *ham* mutant compared to WT.

3.11.2 Supplemental tables

Supplemental Table 3.1: List of primers used in this chapter.

Name	Sequence (5'-3')	Purpose
***ham* mutant genotyping**		
TB080	atggtgtgcgaatctatgacc	*HAM1* forward primer
TB081	tcaaggtcaagctgttcaagc	*HAM1* reverse primer
TB082	gtcgaagaagaggaaaatggg	*HAM2* forward primer
TB083	catatgcctttgaagctgctc	*HAM2* reverse primer
LBb1.3	attttgccgatttcggaac	SALK T-DNA forward primer
***HAM1* and *HAM2* In-Fusion cloning**		
TB324	tatgaccatgattacgaattcagtcataagcttcatcatcaagatc	*HAM1* promoter forward primer
TB325	acgatcccatttctttagtcgggtcggag	*HAM1* promoter reverse primer
TB326	gactaaagaaatgggatcgtctgcggatacag	*HAM1* forward primer
TB327	ctcgcccttgctcacggatccgctctgctctttgtaaggagtcc	*HAM1* reverse primer
TB328	acgatcccatttctcggtcgggtcggag	*HAM2* promoter forward primer
TB329	cgaccgagaaatgggatcgtcagcgaatacag	*HAM2* promoter reverse primer
TB330	ctcgcccttgctcacggatccactctggtccttgtaaggtgtcc	*HAM2* forward primer
TB331	tatgaccatgattacgaattctgctcgatgaagctatacgc	*HAM2* reverse primer

Supplemental Table 3.2: Publicly available BS-seq data used in this chapter.

Genotype	Cell type	GEO dataset	GEO sample	Publication
WT	SC	GSE120519	GSM3402473	(He *et al.*, 2019)
WT	VC	GSE120519	GSM3402474	(He *et al.*, 2019)
dme	VC	GSE38935	GSM952446	(Ibarra *et al.*, 2012)

Chapter 4 – Discussion

4.1 Introduction

This chapter will discuss the results presented in this thesis and their importance in the wider field of chromatin biology. Consideration will be given to the method used to identify novel candidate chromatin modifying proteins and the *Arabidopsis* male sexual lineage as a model for studying chromatin reprogramming. Further sections will discuss H2B.8 and the lessons that could be applicable in a broader sense. The evolution of sperm chromatin compaction mechanisms across eukaryotes will be reviewed. Finally, the interplay between epigenetic modifications will be discussed in the context of HAM protein function in the vegetative nucleus of pollen.

4.2 Proteomics as a tool to identify chromatin modifying proteins

Optimisation of a novel protein extraction protocol enabled the profiling of sperm and vegetative nuclear proteomes. Owing to the low input nature of the method, a balance was achieved between performing FACS and obtaining mass spectrometry results. Given the broad scope of proteins identified and quantifiable data generated, this approach could be applied more broadly to profile proteomes of other rare nuclear types.

This thesis has focused on a single candidate from each of the proteomes generated. However, many more proteins likely contribute to the chromatin dimorphism observed in nuclei of pollen. As such, the data generated could serve as a useful resource for further characterisation of chromatin reprogramming in male germline development.

Furthermore, rather than enriching for chromatin associated proteins (Kustatscher *et al.*, 2014), the proteomes generated utilise whole nuclei. As such, the proteomes could be mined for alternate purposes, such as nuclear envelope constituents or protein machinery involved in other nuclear processes.

4.3 The *Arabidopsis* male sexual lineage as a model for chromatin reprogramming

The *Arabidopsis* male sexual lineage is an excellent model for the study of chromatin in plants. Extensive epigenetic reprogramming occurs across a small number of well-defined cell divisions. Thus, alterations in chromatin structure can be easily traced. This is most clearly exampled by the sperm and vegetative nuclei of pollen, which represent extremes of chromatin compaction but are only separated by two mitotic events.

The highly compacted structure of sperm nuclei offers the opportunity to discover novel determinants that can condense chromatin. As the sperm contributes genetic material to the next generation, it is critical to understand epigenetic regulation and how this could impact upon reproduction. Chapter 2 presented evidence of the importance of H2B.8 to germline reprogramming, but additional factors could contribute to sperm compaction. For example, sperm nuclei are very small in size and become larger in the absence of H2B.8; however, the increase is moderate. As such, other unknown chromatin factors are likely also required.

In contrast to sperm, the vegetative nucleus undergoes extensive decompaction of chromatin. Besides the vegetative nucleus, heterochromatic domains are maintained throughout the nuclei of the plant. Therefore, this cell type offers the opportunity to study how heterochromatin domains are maintained and subsequently depleted in a natural setting. Such heterochromatin loss and the associated active DNA demethylation leads to the release of TE silencing in vegetative nuclei. Natural TE expression can reveal the multiple levels of epigenetic mechanisms that normally lead to silencing.

Recent technological advancements have also enabled a greater understanding of epigenetic reprogramming in the *Arabidopsis* male sexual lineage. The advent of Fluorescence-Activated Cell Sorting (FACS) techniques has enabled the isolation of pure cell types from plant tissues. Much of the work presented here is reliant on FACS and the ability to isolate pure populations of sperm and vegetative nuclei from pollen. Furthermore, a greater understanding of chromatin reprogramming has been enabled by the development of methods to profile DNA methylation, histone variants/modifications and transcription from low input materials. Additionally, the use of super-resolution microscopy offers greater insight to nuclear dynamics.

Collectively, such technological developments could extend our understanding of the reproductive epigenetics in plants by examining the female side and the dynamics immediately post-fertilisation.

4.4 Histone variants and phase separation

H2B.8 is the first example of a core histone variant preferentially undergoing phase separation compared to canonical histones of the same family. Variants of the core histones are found throughout eukaryotes from the highly conserved H2A.Z, found in all lineages (Kawashima *et al.*, 2015), to the highly specialised transitional variants in metazoan spermatogenesis (Hao, Ni and Yang, 2019). This great diversity leads to the possibility that phase separation could be a more general feature of histone variants.

It is possible that other eukaryotic lineages have evolved core histone variants that are expressed in particular cell types and undergo phase separation to regulate chromatin structure. Alternatively, histone variants could have evolved to become refractory to entering phase separated domains, such as heterochromatin. Moreover, specific modifications of histones could also alter phase separation properties. Indeed, chromatin has been demonstrated to have liquid-like properties *in vivo* (Gibson *et al.*, 2019). The deposition of histone variants or post-translational modifications at specific genomic locations could enable the fine-tuning of chromatin behaviour and thus enable physical separation of nuclear processes. Owing to the nature of core histones being imbedded within chromatin, such properties could be comparatively stable, in contrast to chromatin binding proteins. Although, the ability of chaperones to swap core histones means the properties would be reversible if required.

The discovery of H2B.8 in flowering plants and its highly specialised ability to form euchromatic foci in sperm will hopefully serve as a precedent for the exploration of histone variants as mediators of chromatin liquid behaviour across eukaryotes.

4.5 Alternate H2B.8 mechanisms in chromatin compaction

H2B.8 has three key features that distinguish the variant; the N-terminal tail IDR, increased histone body surface arginine and the K-to-N substitution in the C-terminal motif. This work has primarily focused on the contribution of the IDR to H2B.8 function, however roles likely exist for the other features.

The histone body arginine residues are proposed to increase DNA-nucleosome binding affinity. Indeed, the arginine of protamines strongly bind DNA in contrast to synthetic lysine substituted versions (Derouchey, Hoover and Rau, 2013). Furthermore, accumulation of arginine residues in H2A variants correlates with increased chromatin compaction (Macadangdang *et al.*, 2014). Work in Chapter 2 section 2.8.2 demonstrated that chimeric H2Bs carrying the IDR alone was insufficient to fully complement sperm nuclear size and that the histone body was required. As such, the histone body arginine residues are important for *in vivo* H2B.8 function. Future work could utilise comparative *in vitro* studies of nucleosome stability, similar to experiments in (Osakabe *et al.*, 2018), to test whether the arginine residues confer tighter DNA association.

The K-to-N substitution at the conserved ubiquitylation modification site could also be further explored. H2B.8 and H2Bub are anticorrelated, although further *in vitro* experiments are required to determine whether the lysine replacement truly prevents ubiquitylation. Irrespective of H2B.8 presence, the transcriptome is unchanged in both sperm and seedling, therefore it seems likely that H2Bub freely occurs at expressed genes and can evict H2B.8 if required. Further work should seek to address this interplay and more broadly how H2B.8 is incorporated to chromatin.

Regarding H2B.8 deposition, it is interesting that the IDR is critical for localisation to euchromatin. Work presented in Chapter 2 section 2.7.3 shows that IDR deletion entirely inverts localisation to heterochromatic domains. It was expected that loss of the IDR would lead to similar localisation throughout chromatin, as per H2B.2.This finding suggests that the IDR prevents deposition to heterochromatic regions, likely through differential phase separation properties. Whereas without the IDR, the H2B.8 body preferentially localises to compacted chromatin regions, perhaps owing to the increased arginine and inability to undergo ubiquitylation. Altogether, exciting questions remain regarding H2B.8 behaviour.

4.6 Evolution of sperm chromatin compaction mechanisms

This thesis has presented evidence that H2B.8 is a key driver of sperm chromatin compaction in flowering plants. H2B.8 represents an answer to the long-standing mystery of how sperm chromatin is compacted in the absence of protamines (Southworth and Cresti, 1997). From metazoans to non-seed plants, the mechanism has been long established and with a clear purpose for swimming sperm (Rathke *et al.*, 2014). H2B.8 provides a mechanism that explains sperm chromatin compaction, although the purpose of small nuclear size remains unclear for immotile sperm.

Future work should explore the *htb8* fertility defect in greater detail to ascertain the requirement for sperm chromatin compaction. A possibility is that male chromatin must be highly condensed to enable correct fusion with the female gamete. Furthermore, an evolutionary perspective should be considered. Gymnosperms do not have condensed sperm chromatin and have lineages with both pollen tubes and flagella (such as *Ginkgo biloba*) (Southworth and Cresti, 1997). Exploration of the requirement for compaction across

gymnosperms and angiosperms could answer questions surrounding eukaryotic reproduction more widely.

Interestingly, H2B.8 employs phase separation to compact sperm chromatin. Phase separation is also a feature of protamine function in metazoans (Gou *et al.*, 2020). Both condensates exhibit gel-like behaviour. Perhaps stable phase separation in sperm is a convergently evolved feature across eukaryotes.

4.7 Interplay between epigenetic marks

Chapter 3 presents evidence for HAM proteins as mediators of DME access to heterochromatic TEs for active DNA demethylation. HAM proteins likely deposit H4K5ac to relax local chromatin structure, thus enabling further epigenetic activity at otherwise inaccessible genomic locations. This work has highlighted an interplay between epigenetic mechanisms at different levels, whereby DNA methylation levels are dependent upon histone modification mediated chromatin packaging.

The somatic DNA demethylase ROS1 has been shown to associate with particular histone modifications (Tang *et al.*, 2016). However, direct evidence linking the two levels of epigenetic regulation is lacking. Furthermore, ROS1 sites are comparatively few when compared to the activity of DME in VC (Ibarra *et al.*, 2012; He *et al.*, 2019). Therefore, the prevalence of DME-target sites and reliance on HAM at heterochromatic loci offers the opportunity to uncover a direct interplay of DNA demethylation and histone modification. Furthermore, lessons garnered from the HAM/DME pathway could be applicable to DNA methylation activity and how methyltransferases gain access to heterochromatin.

4.8 Conclusions

To conclude, this thesis has explored the extreme chromatin dimorphism observed in *Arabidopsis* pollen. Proteomics was used to identify candidates that could contribute to sperm chromatin compaction and vegetative nuclear decondensation. H2B.8 was demonstrated to compact sperm chromatin by localising to inactive euchromatin and forming IDR-dependent phase separated foci. HAM proteins are required for DME demethylation in heterochromatin by facilitating accessibility, likely through H4K5ac deposition.

References

An, Z., Yin, L., Liu, Y., Peng, M., Shen, W. H. and Dong, A. (2020) 'The histone methylation readers MRG1/MRG2 and the histone chaperones NRP1/NRP2 associate in fine-tuning Arabidopsis flowering time', *Plant Journal*, 103(3), pp. 1010–1024. doi: 10.1111/tpj.14780.

Anderson, S. N., Johnson, C. S., Jones, D. S., Conrad, L. J., Gou, X., Russell, S. D. and Sundaresan, V. (2013) 'Transcriptomes of isolated Oryza sativa gametes characterized by deep sequencing: Evidence for distinct sex-dependent chromatin and epigenetic states before fertilization', *Plant Journal*, 76(5), pp. 729–741. doi: 10.1111/tpj.12336.

Avvakumov, N. and Côté, J. (2007) 'The MYST family of histone acetyltransferases and their intimate links to cancer', *Oncogene*, 26(37), pp. 5395–5407. doi: 10.1038/sj.onc.1210608.

Baldi, S., Korber, P. and Becker, P. B. (2020) 'Beads on a string—nucleosome array arrangements and folding of the chromatin fiber', *Nature Structural and Molecular Biology*, 27(2), pp. 109–118. doi: 10.1038/s41594-019-0368-x.

Balhorn, R. (2007) 'The protamine family of sperm nuclear proteins', *Genome Biology*, 8(9), pp. 1–8. doi: 10.1186/gb-2007-8-9-227.

Berger, F. (2019) 'Emil Heitz, a true epigenetics pioneer', *Nature Reviews Molecular Cell Biology*, 20(10), p. 572. doi: 10.1038/s41580-019-0170-y.

Berger, F., Hamamura, Y., Ingouff, M. and Higashiyama, T. (2008) 'Double fertilization - caught in the act', *Trends in Plant Science*, 13(8), pp. 437–443. doi: 10.1016/j.tplants.2008.05.011.

Berger, F. and Twell, D. (2011) 'Germline specification and function in plants.', *Annual review of plant biology*, 62, pp. 461–484. doi: 10.1146/annurev-arplant-042110-103824.

Boehning, M., Dugast-Darzacq, C., Rankovic, M., Hansen, A. S., Yu, T., Marie-Nelly, H., McSwiggen, D. T., Kokic, G., Dailey, G. M., Cramer, P., Darzacq, X. and Zweckstetter, M. (2018) 'RNA polymerase II clustering through carboxy-terminal domain phase separation', *Nature Structural and Molecular Biology*, 25(9), pp. 833–840. doi: 10.1038/s41594-018-0112-y.

Borg, M. and Berger, F. (2015) 'Chromatin remodelling during male gametophyte development', *Plant Journal*, 83(1), pp. 177–188. doi: 10.1111/tpj.12856.

Borg, M., Brownfield, L., Khatab, H., Sidorova, A., Lingaya, M. and Twell, D. (2011) 'The R2R3

MYB transcription factor DUO1 activates a male germline-specific regulon essential for sperm cell differentiation in Arabidopsis.', *The Plant Cell*, 23(2), pp. 534–549. doi: 10.1105/tpc.110.081059.

Borg, M., Buendía, D. and Berger, F. (2019) 'A simple and robust protocol for immunostaining Arabidopsis pollen nuclei', *Plant Reproduction*, 32(1), pp. 39–43. doi: 10.1007/s00497-018-00360-7.

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), pp. 621–629. doi: 10.1038/s41556-020-0515-y.

Borges, F., Gardner, R., Lopes, T., Calarco, J. P., Boavida, L. C., Keith Slotkin, R., Martienssen, R. A. and Becker, J. D. (2012) 'FACS-based purification of Arabidopsis microspores, sperm cells and vegetative nuclei', *Plant Methods*, 8(44), pp. 1–8. doi: 10.1186/1746-4811-8-44.

Borges, F., Gomes, G., Gardner, R., Moreno, N., McCormick, S., Feijó, J. a and Becker, J. D. (2008) 'Comparative transcriptomics of Arabidopsis sperm cells.', *Plant Physiology*, 148(2), pp. 1168–1181. doi: 10.1104/pp.108.125229.

Bourguet, P., Picard, C., Yelagandula, R., Pélissier, T., Lorković, Z., Pouch-Pélissier, M.-N., Jacobsen, S., Berger, F. and Mathieu, O. (2020) 'The histone variant H2A.W promotes heterochromatin accessibility for efficient DNA methylation in Arabidopsis', *bioRxiv*. doi: 10.1101/2020.03.19.998609.

Braun, R. E. (2001) 'Packaging paternal chromosomes with protamine', *Nature Genetics*, 28(1), pp. 10–12. doi: 10.1038/88194.

Bray, N. L., Pimentel, H., Melsted, P. and Pachter, L. (2016) 'Near-optimal probabilistic RNA-seq quantification', *Nature Biotechnology*, 34(5), pp. 525–527. doi: 10.1038/nbt.3519.

Brownfield, L., Hafidh, S., Borg, M., Sidorova, A., Mori, T. and Twell, D. (2009) 'A plant germline-specific integrator of sperm specification and cell cycle progression', *PLoS Genetics*, 5(3). doi: 10.1371/journal.pgen.1000430.

Brownfield, L., Hafidh, S., Durbarry, A., Khatab, H., Sidorova, A., Doerner, P. and Twell, D.

(2009) ‘Arabidopsis DUO POLLEN3 is a key regulator of male germline development and embryogenesis’, *The Plant Cell*, 21(7), pp. 1940–1956. doi: 10.1105/tpc.109.066373.

Bu, Z., Yu, Y., Li, Z., Liu, Y., Jiang, W., Huang, Y. and Dong, A. W. (2014) ‘Regulation of Arabidopsis Flowering by the Histone Mark Readers MRG1/2 via Interaction with CONSTANS to Modulate FT Expression’, *PLoS Genetics*, 10(9), pp. 1–11. doi: 10.1371/journal.pgen.1004617.

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. and Greenleaf, W. J. (2013) ‘Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position.’, *Nature Methods*, 10(12), pp. 1213–8. doi: 10.1038/nmeth.2688.

Calarco, J. P., Borges, F., Donoghue, M. T. A., Van Ex, F., Jullien, P. E., Lopes, T., Gardner, R., Berger, F., Feijó, J. A., Becker, J. D. and Martienssen, R. A. (2012) ‘Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA’, *Cell*, 151(1), pp. 194–205. doi: 10.1016/j.cell.2012.09.001.

Carrozza, M. J., Utley, R. T., Workman, J. L. and Côté, J. (2003) ‘The diverse functions of histone acetyltransferase complexes’, *Trends in Genetics*, 19(6), pp. 321–329. doi: 10.1016/S0168-9525(03)00115-X.

Castel, B., Tomlinson, L., Locci, F., Yang, Y. and Jones, J. D. G. (2019) ‘Optimization of T-DNA architecture for Cas9-mediated mutagenesis in Arabidopsis’, *PLoS ONE*, 14(1), p. e0204778. doi: 10.1101/419952.

Chan, S. W. L., Henderson, I. R. and Jacobsen, S. E. (2005) ‘Gardening the genome: DNA methylation in Arabidopsis thaliana.’, *Nature reviews Genetics*, 6(5), pp. 351–360. doi: 10.1038/nrg1664.

Chaw, S. M., Liu, Y. C., Wu, Y. W., Wang, H. Y., Lin, C. Y. I., Wu, C. S., Ke, H. M., Chang, L. Y., Hsu, C. Y., Yang, H. T., Sudianto, E., Hsu, M. H., Wu, K. P., Wang, L. N., Leebens-Mack, J. H. and Tsai, I. J. (2019) ‘Stout camphor tree genome fills gaps in understanding of flowering plant genome evolution’, *Nature Plants*, 5(1), pp. 63–73. doi: 10.1038/s41477-018-0337-0.

Chen, C., Li, C., Wang, Y., Renaud, J., Tian, G., Kambhampati, S., Saatian, B., Nguyen, V., Hannoufa, A., Marsolais, F., Yuan, Z. C., Yu, K., Austin, R. S., Liu, J., Kohalmi, S. E., Wu, K.,

Huang, S. and Cui, Y. (2017) 'Cytosolic acetyl-CoA promotes histone acetylation predominantly at H3K27 in Arabidopsis', *Nature Plants*, 3(10), pp. 814–824. doi: 10.1038/s41477-017-0023-7.

Chen, J., Hao, Z., Guang, X., Zhao, C., Wang, P., Xue, L., Zhu, Qihui, Yang, Linfeng, Sheng, Y., Zhou, Y., Xu, H., Xie, H., Long, X., Zhang, J., Wang, Z., Shi, M., Lu, Y., Liu, S., Guan, L., Zhu, Qianhua, Yang, Liming, Ge, S., Cheng, T., Laux, T., Gao, Q., Peng, Y., Liu, N., Yang, S. and Shi, J. (2019) 'Liriodendron genome sheds light on angiosperm phylogeny and species–pair differentiation', *Nature Plants*, 5(1), pp. 18–25. doi: 10.1038/s41477-018-0323-6.

Chen, Z., Hafidh, S., Poh, S. H., Twell, D. and Berger, F. (2009) 'Proliferation and cell fate establishment during Arabidopsis male gametogenesis depends on the Retinoblastoma protein', *Proceedings of the National Academy of Sciences*, 106(17), pp. 7257–7262. doi: 10.1073/pnas.0810992106.

Cho, C., Willis, W. D., Goulding, E. H., Jung-Ha, H., Choi, Y. C., Hecht, N. B. and Eddy, E. M. (2001) 'Haploinsufficiency of protamine-1 or -2 causes infertility in mice', *Nature Genetics*, 28(1), pp. 82–86. doi: 10.1038/88313.

Choi, J., Lyons, D. B., Kim, M. Y., Moore, J. D. and Zilberman, D. (2020) 'DNA Methylation and Histone H1 Jointly Repress Transposable Elements and Aberrant Intragenic Transcripts', *Molecular Cell*, 77(2), pp. 310–323. doi: 10.1016/j.molcel.2019.10.011.

Choi, Y., Gehring, M., Johnson, L., Hannon, M., Harada, J. J., Goldberg, R. B., Jacobsen, S. E. and Fischer, R. L. (2002) 'DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in Arabidopsis', *Cell*, 110(1), pp. 33–42. doi: 10.1016/S0092-8674(02)00807-3.

Clapier, C. R. and Cairns, B. R. (2009) 'The biology of chromatin remodeling complexes', *Annual review of biochemistry*, 78, pp. 273–304. doi: 10.1146/annurev.biochem.77.062706.153223.

Clough, S. J. and Bent, A. F. (1998) 'Floral dip: A simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana', *The Plant Journal*, 16(6), pp. 735–743. doi: 10.1046/j.1365-313X.1998.00343.x.

Cremer, M., Schmid, V. J., Kraus, F., Markaki, Y., Hellmann, I., Maiser, A., Leonhardt, H., John,

S., Stamatoyannopoulos, J. and Cremer, T. (2017) 'Initial high-resolution microscopic mapping of active and inactive regulatory sequences proves non-random 3D arrangements in chromatin domain clusters', *Epigenetics and Chromatin*, 10(1), pp. 1–17. doi: 10.1186/s13072-017-0146-0.

D'Ippolito, R. A., Minamino, N., Rivera-Casas, C., Cheema, M. S., Bai, D. L., Kasinsky, H. E., Shabanowitz, J., Eirin-Lopez, J. M., Ueda, T., Hunt, D. F. and Ausió, J. (2019) 'Protamines from liverwort are produced by post-translational cleavage and C-terminal di-aminopropanelation of several male germ-specific H1 histones', *Journal of Biological Chemistry*, 294(44), pp. 16364–16373. doi: 10.1074/jbc.RA119.010316.

Derouchey, J., Hoover, B. and Rau, D. C. (2013) 'A comparison of DNA compaction by arginine and lysine peptides: A physical basis for arginine rich protamines', *Biochemistry*, 52(17), pp. 3000–3009. doi: 10.1021/bi4001408.

Doğan, E. S. and Liu, C. (2018) 'Three-dimensional chromatin packing and positioning of plant genomes', *Nature Plants*, 4(8), pp. 521–529. doi: 10.1038/s41477-018-0199-5.

Du, G., Drexler, G. A., Friedland, W., Greubel, C., Hable, V., Krücken, R., Kugler, A., Tonelli, L., Friedl, A. A. and Dollinger, G. (2011) 'Spatial dynamics of DNA damage response protein foci along the ion trajectory of high-LET particles', *Radiation Research*, 176(6), pp. 706–715. doi: 10.1667/RR2592.1.

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

Earley, K. W., Shook, M. S., Brower-Toland, B., Hicks, L. and Pikaard, C. S. (2007) 'In vitro specificities of Arabidopsis co-activator histone acetyltransferases: Implications for histone hyperacetylation in gene activation', *The Plant Journal*, 52(4), pp. 615–626. doi: 10.1111/j.1365-313X.2007.03264.x.

Eissenberg, J. C. (2012) 'Structural biology of the chromodomain: Form and function', *Gene*, 496(2), pp. 69–78. doi: 10.1016/j.gene.2012.01.003.

Emre, N. C. T. and Berger, S. L. (2004) 'Histone H2B ubiquitylation and deubiquitylation in

genomic regulation', *Cold Spring Harbor Symposia on Quantitative Biology*, 69, pp. 289–299. doi: 10.1101/sqb.2004.69.289.

Englbrecht, C. C., Schoof, H. and Böhm, S. (2004) 'Conservation, diversification and expansion of C2H2 zinc finger proteins in the Arabidopsis thaliana genome', *BMC Genomics*, 5, pp. 1–17. doi: 10.1186/1471-2164-5-39.

Erdel, F., Rademacher, A., Vlijm, R., Tünnermann, J., Frank, L., Weinmann, R., Schweigert, E., Yserentant, K., Hummert, J., Bauer, C., Schumacher, S., Al Alwash, A., Normand, C., Herten, D. P., Engelhardt, J. and Rippe, K. (2020) 'Mouse Heterochromatin Adopts Digital Compaction States without Showing Hallmarks of HP1-Driven Liquid-Liquid Phase Separation', *Molecular Cell*, 78(2), pp. 236-249.e7. doi: 10.1016/j.molcel.2020.02.005.

Erdel, F. and Rippe, K. (2018) 'Formation of Chromatin Subcompartments by Phase Separation', *Biophysical Journal*, 114(10), pp. 2262–2270. doi: 10.1016/j.bpj.2018.03.011.

Fang, X., Wang, L., Ishikawa, R., Li, Y., Fiedler, M., Liu, F., Calder, G., Rowan, B., Weigel, D., Li, P. and Dean, C. (2019) 'Arabidopsis FLL2 promotes liquid–liquid phase separation of polyadenylation complexes', *Nature*, 569(7755), pp. 265–269. doi: 10.1038/s41586-019-1165-8.

Feng, S., Cokus, S., Schubert, V., Zhai, J., Pellegrini, M. and Jacobsen, S. (2014) 'Genome-wide Hi-C Analyses in Wild-Type and Mutants Reveal High-Resolution Chromatin Interactions in Arabidopsis', *Molecular Cell*, 55(5), pp. 694–707. doi: 10.1016/j.molcel.2014.07.008.

Feng, X., Zilberman, D. and Dickinson, H. (2013) 'A Conversation across Generations: Soma-Germ Cell Crosstalk in Plants', *Developmental Cell*, 24(3), p. 224. doi: 10.1016/j.devcel.2013.01.014.

Feric, M., Vaidya, N., Harmon, T. S., Mitrea, D. M., Zhu, L., Richardson, T. M., Kriwacki, R. W., Pappu, R. V. and Brangwynne, C. P. (2016) 'Coexisting Liquid Phases Underlie Nucleolar Subcompartments', *Cell*, 165(7), pp. 1686–1697. doi: 10.1016/j.cell.2016.04.047.

Fierz, B., Chatterjee, C., Mcginty, R. K., Bar-dagan, M., Daniel, P. and Muir, T. W. (2011) 'Histone H2B ubiquitylation disrupts local and higher order chromatin compaction', *Nature Chemical Biology*, 7(2), pp. 113–119. doi: 10.1038/nchembio.501.Histone.

Fiorucci, A. S., Bourbousse, C., Concia, L., Rougée, M., Deton-Cabanillas, A. F., Zabulon, G., Layat, E., Latrasse, D., Kim, S. K., Chaumont, N., Lombard, B., Stroebel, D., Lemoine, S., Mohammad, A., Blugeon, C., Loew, D., Bailly, C., Bowler, C., Benhamed, M. and Barneche, F. (2019) 'Arabidopsis S2Lb links AtCOMPASS-like and SDG2 activity in H3K4me3 independently from histone H2B monoubiquitination', *Genome Biology*, 20(1). doi: 10.1186/s13059-019-1705-4.

Fisher, R. S. and Elbaum-Garfinkle, S. (2020) 'Tunable multiphase dynamics of arginine and lysine liquid condensates', *Nature Communications*, 11(1). doi: 10.1038/s41467-020-18224-y.

Frost, J. M., Kim, M. Y., Park, G. T., Hsieh, P. H., Nakamura, M., Lin, S. J. H., Yoo, H., Choi, J., Ikeda, Y., Kinoshita, T., Choi, Y., Zilberman, D. and Fischer, R. L. (2018) 'FACT complex is required for DNA demethylation at heterochromatin during reproduction in Arabidopsis', *Proceedings of the National Academy of Sciences*, 115(20), pp. E4720–E4729. doi: 10.1073/pnas.1713333115.

Frottin, F., Schueder, F., Tiwary, S., Gupta, R., Körner, R., Schlichthaerle, T., Cox, J., Jungmann, R., Hartl, F. U. and Hipp, M. S. (2019) 'The nucleolus functions as a phase-separated protein quality control compartment', *Science*, 365(6451), pp. 342–347. doi: 10.1126/science.aaw9157.

Fyodorov, D. V., Zhou, B. R., Skoultchi, A. I. and Bai, Y. (2018) 'Emerging roles of linker histones in regulating chromatin structure and function', *Nature Reviews Molecular Cell Biology*, 19(3), pp. 192–206. doi: 10.1038/nrm.2017.94.

Gallego, L. D., Schneider, M., Mittal, C., Romanauska, A., Gudino Carrillo, R. M., Schubert, T., Pugh, B. F. and Köhler, A. (2020) 'Phase separation directs ubiquitination of gene-body nucleosomes', *Nature*, 579(7800), pp. 592–597. doi: 10.1038/s41586-020-2097-z.

Gentry, M. and Hennig, L. (2014) 'Remodelling chromatin to shape development of plants', *Experimental Cell Research*, 321(1), pp. 40–46. doi: 10.1016/j.yexcr.2013.11.010.

Gibson, B. A., Doolittle, L. K., Schneider, M. W. G., Jensen, L. E., Gamarra, N., Henry, L., Gerlich, D. W., Redding, S. and Rosen, M. K. (2019) 'Organization of Chromatin by Intrinsic and Regulated Phase Separation', *Cell*, 179(2), pp. 470–484. doi: 10.1016/j.cell.2019.08.037.

Gong, Z., Morales-Ruiz, T., Ariza, R. R., Roldán-Arjona, T., David, L. and Zhu, J. K. (2002) 'ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA glycosylase/lyase', *Cell*, 111(6), pp. 803–814. doi: 10.1016/S0092-8674(02)01133-9.

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N. and Rokhsar, D. S. (2012) 'Phytozome: A comparative platform for green plant genomics', *Nucleic Acids Research*, 40(1), pp. 1178–1186. doi: 10.1093/nar/gkr944.

Gou, L. T., Lim, D. H., Ma, W., Aubol, B. E., Hao, Y., Wang, X., Zhao, J., Liang, Z., Shao, C., Zhang, Xuan, Meng, F., Li, H., Zhang, Xiaorong, Xu, R., Li, D., Rosenfeld, M. G., Mellon, P. L., Adams, J. A., Liu, M. F. and Fu, X. D. (2020) 'Initiation of Parental Genome Reprogramming in Fertilized Oocyte by Splicing Kinase SRPK1-Catalyzed Protamine Phosphorylation', *Cell*, 180(6), pp. 1212–1227. doi: 10.1016/j.cell.2020.02.020.

Grob, S., Schmid, M. W. and Grossniklaus, U. (2014) 'Hi-C Analysis in Arabidopsis Identifies the KNOT, a Structure with Similarities to the flamenco Locus of Drosophila', *Molecular Cell*, 55(5), pp. 678–693. doi: 10.1016/j.molcel.2014.07.009.

Grunewald, S., Paasch, U., Glander, H. J. and Anderegg, U. (2005) 'Mature human spermatozoa do not transcribe novel RNA', *Andrologia*, 37(2), pp. 69–71. doi: 10.1111/j.1439-0272.2005.00656.x.

Hackenberg, D. and Twell, D. (2019) *The evolution and patterning of male gametophyte development*, *Current Topics in Developmental Biology*. doi: 10.1016/bs.ctdb.2018.10.008.

Hamamura, Y., Saito, C., Awai, C., Kurihara, D., Miyawaki, A., Nakagawa, T., Kanaoka, M. M., Sasaki, N., Nakano, A., Berger, F. and Higashiyama, T. (2011) 'Live-cell imaging reveals the dynamics of two sperm cells during double fertilization in Arabidopsis thaliana', *Current Biology*, 21(6), pp. 497–502. doi: 10.1016/j.cub.2011.02.013.

Hammoud, S. S., Nix, D. A., Zhang, H., Purwar, J., Carrell, D. T. and Cairns, B. R. (2009) 'Distinctive chromatin in human sperm packages genes for embryo development', *Nature*, 460(7254), pp. 473–478. doi: 10.1038/nature08162.

Han, S. K., Wu, M. F., Cui, S. and Wagner, D. (2015) 'Roles and activities of chromatin remodeling ATPases in plants', *The Plant Journal*, 83(1), pp. 62–77. doi: 10.1111/tpj.12877.

Hao, S. L., Ni, F. Da and Yang, W. X. (2019) 'The dynamics and regulation of chromatin remodeling during spermiogenesis', *Gene*, 706(March), pp. 201–210. doi: 10.1016/j.gene.2019.05.027.

He, S., Vickers, M., Zhang, J. and Feng, X. (2019) 'Natural depletion of histone H1 in sex cells causes DNA demethylation, heterochromatin decondensation and transposon activation', *eLife*, 8, p. e42530. doi: 10.7554/eLife.42530.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H. and Glass, C. K. (2010) 'Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities', *Molecular Cell*, 38(4), pp. 576–589. doi: 10.1016/j.molcel.2010.05.004.

Henderson, I. R. and Jacobsen, S. E. (2007) 'Epigenetic inheritance in plants.', *Nature*, 447(7143), pp. 418–424. doi: 10.1038/nature05917.

Higo, A., Kawashima, T., Borg, M., Zhao, M., López-Vidriero, I., Sakayama, H., Montgomery, S. A., Sekimoto, H., Hackenberg, D., Shimamura, M., Nishiyama, T., Sakakibara, K., Tomita, Y., Togawa, T., Kunimoto, K., Osakabe, A., Suzuki, Y., Yamato, K. T., Ishizaki, K., Nishihama, R., Kohchi, T., Franco-Zorrilla, J. M., Twell, D., Berger, F. and Araki, T. (2018) 'Transcription factor DUO1 generated by neo-functionalization is associated with evolution of sperm differentiation in plants', *Nature Communications*, 9(1), pp. 1–13. doi: 10.1038/s41467-018-07728-3.

Hsieh, P. H., He, S., Buttress, T., Gao, H., Couchman, M., Fischer, R. L., Zilberman, D. and Feng, X. (2016) 'Arabidopsis male sexual lineage exhibits more robust maintenance of CG methylation than somatic tissues', *Proceedings of the National Academy of Sciences*, 113(52), pp. 15132–15137. doi: 10.1073/pnas.1619074114.

Hsieh, T.-F., Shin, J., Uzawa, R., Silva, P., Cohen, S., Bauer, M. J., Hashimoto, M., Kirkbride, R. C., Harada, J. J., Zilberman, D. and Fischer, R. L. (2011) 'Regulation of imprinted gene expression in Arabidopsis endosperm.', *Proceedings of the National Academy of Sciences of the United States of America*, 108(5), pp. 1755–62. doi: 10.1073/pnas.1019273108.

Huettel, B., Kanno, T., Daxinger, L., Aufsatz, W., Matzke, A. J. M. and Matzke, M. (2006) 'Endogenous targets of RNA-directed DNA methylation and Pol IV in Arabidopsis.', *The EMBO*

journal, 25(12), pp. 2828–2836. doi: 10.1038/sj.emboj.7601150.

Hyman, A. A., Weber, C. A. and Jülicher, F. (2014) 'Liquid-liquid phase separation in biology', *Annual review of cell and developmental biology*, 30, pp. 39–58. doi: 10.1146/annurev-cellbio-100913-013325.

Ibarra, C. A., Feng, X., Schoft, V. K., Hsieh, T. F., Uzawa, R., Rodrigues, J. A., Zemach, A., Chumak, N., Machlicova, A., Nishimura, T., Rojas, D., Fischer, R. L., Tamaru, H. and Zilberman, D. (2012) 'Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes', *Science*, 337(6100), pp. 1360–1364. doi: 10.1126/science.1224839.

Ikeda, Yoko, Kinoshita, Y., Susaki, D., Ikeda, Yuriko, Iwano, M., Takayama, S., Higashiyama, T., Kakutani, T. and Kinoshita, T. (2011) 'HMG Domain containing SSRP1 is required for DNA demethylation and genomic imprinting in arabidopsis', *Developmental Cell*, 21(3), pp. 589–596. doi: 10.1016/j.devcel.2011.08.013.

Ingouff, M., Hamamura, Y., Gourgues, M., Higashiyama, T. and Berger, F. (2007) 'Distinct Dynamics of HISTONE3 Variants between the Two Fertilization Products in Plants', *Current Biology*, 17(12), pp. 1032–1037. doi: 10.1016/j.cub.2007.05.019.

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

Jacob, Y., Bergamin, E., Donoghue, M. T. A., Mongeon, V., LeBlanc, C., Voigt, P., Underwood, C. J., Brunzelle, J. S., Michaels, S. D., Reinberg, D., Couture, J.-F. and Martienssen, R. A. (2014) 'Selective Methylation of Histone H3 Variant H3.1 Regulates Heterochromatin Replication', *Science*, 343(6176), pp. 1249–1253. doi: DOI: 10.1126/science.1248357.

Jiang, D. and Berger, F. (2017) 'DNA replication–coupled histone modification maintains Polycomb gene silencing in plants', *Science*, 357(6356), pp. 1146–1149.

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), pp. 1–22. doi: 10.1371/journal.pgen.1008964.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S. and Madden, T. L. (2008) 'NCBI BLAST: a better web interface', *Nucleic acids research*, 36(1), pp. 5–9. doi: 10.1093/nar/gkn201.

Kalyanikrishna, Mikulski, P. and Schubert, D. (2020) 'Measurement of Arabidopsis thaliana Nuclear Size and Shape', in Spillane, C. and McKeown, P. (eds) *Plant Epigenetics and Epigenomics*. 1st edn. New York, NY: Springer US, pp. 107–113. doi: 10.1007/978-1-0716-0179-2_8.

Kaneshiro, K. R., Rechtsteiner, A. and Strome, S. (2019) 'Sperm-inherited H3K27me3 impacts offspring transcription and development in C. elegans', *Nature Communications*, 10(1), pp. 1–9. doi: 10.1038/s41467-019-09141-w.

Kanno, T., M. Florian Mette, 2, David P. Kreil, 3, Werner Aufsatz, 1, Marjori Matzke, 1 and Matzke, and A. J. M. (2004) 'Involvement of Putative SNF2 Chromatin Remodeling Protein DRD1 in RNA-Directed DNA Methylation', *Current Biology*, 14, pp. 801–805. doi: 10.1016/j.

Kawakatsu, T., Stuart, T., Valdes, M., Breakfield, N., Schmitz, R. J., Nery, J. R., Urich, M. A., Han, X., Lister, R., Benfey, P. N. and Ecker, J. R. (2016) 'Unique cell-type-specific patterns of DNA methylation in the root meristem', *Nature Plants*, 2(5), pp. 1–8. doi: 10.1038/NPLANTS.2016.58.

Kawashima, T. and Berger, F. (2014) 'Epigenetic reprogramming in plant sexual reproduction', *Nature Reviews Genetics*, 15(9), pp. 613–624. doi: 10.1038/nrg3685.

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), pp. 419–425. doi: 10.1016/j.tplants.2015.04.005.

Kaya-Okur, H. S., Janssens, D. H., Henikoff, J. G., Ahmad, K. and Henikoff, S. (2020) 'Efficient low-cost chromatin profiling with CUT&Tag', *Nature Protocols*, 15(10), pp. 3264–3283. doi: 10.1038/s41596-020-0373-x.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. and Sternberg, M. J. (2015) 'The Phyre2 web portal for protein modeling, prediction and analysis', *Nature Protocols*, 10(6), pp. 845–858. doi: 10.1038/nprot.2015-053.

Khadka, J., Pesok, A. and Grafi, G. (2020) 'Plant histone htb (H2b) variants in regulating chromatin structure and function', *Plants*, 9(11), pp. 1–14. doi: 10.3390/plants9111435.

Kim, D., Paggi, J. M., Park, C., Bennett, C. and Salzberg, S. L. (2019) 'Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype', *Nature Biotechnology*, 37(8), pp. 907–915. doi: 10.1038/s41587-019-0201-4.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. and Salzberg, S. L. (2013) 'TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions', *Genome Biology*, 14(36), pp. 1–13.

Kimmins, S. and Sassone-Corsi, P. (2005) 'Chromatin remodelling and epigenetic features of germ cells', *Nature*, 434(7033), pp. 583–589. doi: 10.1038/nature03368.

Kumar, S., Stecher, G., Li, M., Knyaz, C. and Tamura, K. (2018) 'MEGA X: Molecular evolutionary genetics analysis across computing platforms', *Molecular Biology and Evolution*, 35(6), pp. 1547–1549. doi: 10.1093/molbev/msy096.

Kustatscher, G., Wills, K. L. H., Furlan, C. and Rappsilber, J. (2014) 'Chromatin enrichment for proteomics', *Nature Protocols*, 9(9), pp. 2090–9. doi: 10.1038/nprot.2014.142.

Labun, K., Montague, T. G., Krause, M., Torres Cleuren, Y. N., Tjeldnes, H. and Valen, E. (2019) 'CHOPCHOP v3: Expanding the CRISPR web toolbox beyond genome editing', *Nucleic acids research*, 47(1), pp. 171–174. doi: 10.1093/nar/gkz365.

Lakadamyali, M. and Cosma, M. P. (2015) 'Advanced microscopy methods for visualizing chromatin structure', *FEBS Letters*, 589(20), pp. 3023–3030. doi: 10.1016/j.febslet.2015.04.012.

Langmead, B. and Salzberg, S. L. (2012) 'Fast gapped-read alignment with Bowtie 2', *Nature Methods*, 9(4), pp. 357–359. doi: 10.1038/nmeth.1923.

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

Latrasse, D., Benhamed, M., Henry, Y., Domenichini, S., Kim, W., Zhou, D. X. and Delarue, M.

(2008) ‘The MYST histone acetyltransferases are essential for gametophyte development in arabidopsis’, *BMC Plant Biology*, 8, pp. 1–16. doi: 10.1186/1471-2229-8-121.

Law, J. A. and Jacobsen, S. E. (2010) ‘Establishing, maintaining and modifying DNA methylation patterns in plants and animals’, *Nature Reviews Genetics*, 11(3), pp. 204–20. doi: 10.1038/nrg2719.

Lermontova, I., Rutten, T. and Schubert, I. (2011) ‘Deposition, turnover, and release of CENH3 at Arabidopsis centromeres’, *Chromosoma*, 120(6), pp. 633–640. doi: 10.1007/s00412-011-0338-5.

Lermontova, I., Sandmann, M., Mascher, M., Schmit, A. C. and Chabouté, M. E. (2015) ‘Centromeric chromatin and its dynamics in plants’, *The Plant Journal*, 83(1), pp. 4–17. doi: 10.1111/tpj.12875.

Li, C., Gu, L., Gao, L., Chen, C., Wei, C. Q., Qiu, Q., Chien, C. W., Wang, S., Jiang, L., Ai, L. F., Chen, C. Y., Yang, S., Nguyen, V., Qi, Y., Snyder, M. P., Burlingame, A. L., Kohalmi, S. E., Huang, S., Cao, X., Wang, Z. Y., Wu, K., Chen, X. and Cui, Y. (2016) ‘Concerted genomic targeting of H3K27 demethylase REF6 and chromatin-remodeling ATPase BRM in Arabidopsis’, *Nature Genetics*, 48(6), pp. 687–693. doi: 10.1038/ng.3555.

Liu, C., Cheng, Y. J., Wang, J. W. and Weigel, D. (2017) ‘Prominent topologically associated domains differentiate global chromatin packing in rice from Arabidopsis’, *Nature Plants*, 3(9), pp. 742–748. doi: 10.1038/s41477-017-0005-9.

Liu, C., Wang, C., Wang, G., Becker, C., Zaidem, M. and Weigel, D. (2016) ‘Genome-wide analysis of chromatin packing in Arabidopsis thaliana at single-gene resolution’, *Genome Research*, pp. 1057–1068. doi: 10.1101/gr.204032.116.

Luo, X., Ou, Y., Li, R. and He, Y. (2020) ‘Maternal transmission of the epigenetic “memory of winter cold” in Arabidopsis’, *Nature Plants*, 6(10), pp. 1211–1218. doi: 10.1038/s41477-020-00774-0.

Ma, Z., Castillo-González, C., Wang, Z., Sun, D., Hu, X., Shen, X., Potok, M. E. and Zhang, X. (2018) ‘Arabidopsis Serrate Coordinates Histone Methyltransferases ATXR5/6 and RNA Processing Factor RDR6 to Regulate Transposon Expression’, *Developmental Cell*, 45(6), pp. 769–784. doi: 10.1016/j.devcel.2018.05.023.

Macadangdang, B. R., Oberai, A., Spektor, T., Campos, O. A., Sheng, F., Carey, M. F., Vogelauer, M. and Kurdistani, S. K. (2014) 'Evolution of histone 2A for chromatin compaction in eukaryotes', *eLife*, 2014(3), pp. 1–22. doi: 10.7554/eLife.02792.

Maeshima, K., Ide, S., Hibino, K. and Sasai, M. (2016) 'Liquid-like behavior of chromatin', *Current Opinion in Genetics and Development*, 37, pp. 36–45. doi: 10.1016/j.gde.2015.11.006.

Malo, A. F., Gomendio, M., Garde, J., Lang-Lenton, B., Soler, A. J. and Roldan, E. R. S. (2006) 'Sperm design and sperm function', *Biology Letters*, 2(2), pp. 246–249. doi: 10.1098/rsbl.2006.0449.

March-Díaz, R. and Reyes, J. C. (2009) 'The beauty of being a variant: H2A.Z and the SWR1 complex in plants', *Molecular Plant*, 2(4), pp. 565–577. doi: 10.1093/mp/ssp019.

Martínez, G., Panda, K., Köhler, C. and Slotkin, R. K. (2016) 'Silencing in sperm cells is directed by RNA movement from the surrounding nurse cell', *Nature Plants*, 2(4). doi: 10.1038/NPLANTS.2016.30.

Matsuda, A., Chikashige, Y., Ding, D. Q., Ohtsuki, C., Mori, C., Asakawa, H., Kimura, H., Haraguchi, T. and Hiraoka, Y. (2015) 'Highly condensed chromatins are formed adjacent to subtelomeric and decondensed silent chromatin in fission yeast', *Nature Communications*, 6, pp. 1–3. doi: 10.1038/ncomms8753.

Matzke, M. A. and Mosher, R. A. (2014) 'RNA-directed DNA methylation: an epigenetic pathway of increasing complexity', *Nature Reviews Genetics*, 15(6), pp. 394–408. doi: 10.1038/nrg3683.

Matzke, M., Kanno, T., Daxinger, L., Huettel, B. and Matzke, A. J. (2009) 'RNA-mediated chromatin-based silencing in plants', *Current Opinion in Cell Biology*, 21(3), pp. 367–376. doi: 10.1016/j.ceb.2009.01.025.

Maximiliano, T., Montserrat, G. and Roldan, E. R. S. (2011) 'Sperm competition and the evolution of sperm design in mammals', *BMC Evolutionary Biology*, 11(12), pp. 14–21. doi: 10.1111/rda.13552.

McSwiggen, D. T., Mir, M., Darzacq, X. and Tjian, R. (2019) 'Evaluating phase separation in live cells: diagnosis, caveats, and functional consequences', *Genes & development*, 33(23–24), pp.

1619–1634. doi: 10.1101/gad.331520.119.

Meistrich, M. L., Trostle-Weige, P. K., Lin, R., Allis, C. D. and Bhatnagar, Y. M. (1992) 'Highly acetylated H4 is associated with histone displacement in rat spermatids', *Molecular Reproduction and Development*, 31(3), pp. 170–181. doi: 10.1002/mrd.1080310303.

Mérai, Z., Chumak, N., García-Aguilar, M., Hsieh, T.-F., Nishimura, T., Schoft, V. K., Bindics, J., Slusarz, L., Arnoux, S., Opravil, S., Mechtler, K., Zilberman, D., Fischer, R. L. and Tamaru, H. (2014) 'The AAA-ATPase molecular chaperone Cdc48/p97 disassembles sumoylated centromeres, decondenses heterochromatin, and activates ribosomal RNA genes', *Proceedings of the National Academy of Sciences*, 111(45), pp. 16166–16171. doi: 10.1073/pnas.1418564111.

Mergner, J., Frejno, M., List, M., Papacek, M., Chen, X., Chaudhary, A., Samaras, P., Richter, S., Shikata, H., Messerer, M., Lang, D., Altmann, S., Cyprys, P., Zolg, D. P., Mathieson, T., Bantscheff, M., Hazarika, R. R., Schmidt, T., Dawid, C., Dunkel, A., Hofmann, T., Sprunck, S., Falter-Braun, P., Johannes, F., Mayer, K. F. X., Jürgens, G., Wilhelm, M., Baumbach, J., Grill, E., Schneitz, K., Schwechheimer, C. and Kuster, B. (2020) 'Mass-spectrometry-based draft of the Arabidopsis proteome', *Nature*, 579(7799), pp. 409–414. doi: 10.1038/s41586-020-2094-2.

Montellier, E., Boussouar, F., Rousseaux, S., Zhang, K., Buchou, T., Fenaille, F., Shiota, H., Debernardi, A., Héry, P., Curtet, S., Jamshidikia, M., Barral, S., Holota, H., Bergon, A., Lopez, F., Guardiola, P., Pernet, K., Imbert, J., Petosa, C., Tan, M., Zhao, Y., Gérard, M. and Khochbin, S. (2013) 'Chromatin-to-nucleoprotamine transition is controlled by the histone H2B variant TH2B', *Genes and Development*, 27(15), pp. 1680–1692. doi: 10.1101/gad.220095.113.

Morgan, C., Zhang, H., Henry, C. E., Franklin, C. F. H. and Bomblies, K. (2020) 'Derived alleles of two axis proteins affect meiotic traits in autotetraploid Arabidopsis arenosa', *Proceedings of the National Academy of Sciences*, 117(16), pp. 8980–8988. doi: 10.1073/pnas.1919459117.

Mozgova, I., Köhler, C. and Hennig, L. (2015) 'Keeping the gate closed: Functions of the polycomb repressive complex PRC2 in development', *The Plant Journal*, 83(1), pp. 121–132. doi: 10.1111/tpj.12828.

Murphy, P. J., Wu, S. F., James, C. R., Wike, C. L. and Cairns, B. R. (2018) 'Placeholder

Nucleosomes Underlie Germline-to-Embryo DNA Methylation Reprogramming', *Cell*, 172(5), pp. 993–1006. doi: 10.1016/j.cell.2018.01.022.

Narsai, R., Gouil, Q., Secco, D., Srivastava, A., Karpievitch, Y. V., Liew, L. C., Lister, R., Lewsey, M. G. and Whelan, J. (2017) 'Extensive transcriptomic and epigenomic remodelling occurs during Arabidopsis thaliana germination', *Genome Biology*, 18(1), pp. 1–18. doi: 10.1186/s13059-017-1302-3.

Nassrallah, A., Rougée, M., Bourbousse, C., Drevensek, S., Fonseca, S., Iniesto, E., Ait-Mohamed, O., Deton-Cabanillas, A. F., Zabulon, G., Ahmed, I., Stroebel, D., Masson, V., Lombard, B., Eeckhout, D., Gevaert, K., Loew, D., Genovesio, A., Breyton, C., de Jaeger, G., Bowler, C., Rubio, V. and Barneche, F. (2018) 'DET1-mediated degradation of a SAGA-like deubiquitination module controls H2Bub homeostasis', *eLife*, 7, p. e37892. doi: 10.7554/eLife.37892.

Norstog, K. J., Gifford, E. M. and Stevenson, D. W. (2004) 'Comparative development of the spermatozoids of cycads and Ginkgo biloba', *Botanical Review*, 70(1), pp. 5–15. doi: 10.1663/0006-8101(2004)070[0005:CDOTSO]2.0.CO;2.

Ochs, F., Karemore, G., Miron, E., Brown, J., Sedlackova, H., Rask, M. B., Lampe, M., Buckle, V., Schermelleh, L., Lukas, J. and Lukas, C. (2019) 'Stabilization of chromatin topology safeguards genome integrity', *Nature*, 574(7779), pp. 571–574. doi: 10.1038/s41586-019-1659-4.

Okada, T., Endo, M., Singh, M. B. and Bhalla, P. L. (2005) 'Analysis of the histone H3 gene family in Arabidopsis and identification of the male-gamete-specific variant AtMGH3', *Plant Journal*, 44(4), pp. 557–568. doi: 10.1111/j.1365-313X.2005.02554.x.

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), pp. 7675–7685. doi: 10.1093/nar/gky540.

Ostermeier, G. C., Dix, D. J., Miller, D., Khatri, P. and Krawetz, S. A. (2002) 'Spermatozoal RNA profiles of normal fertile men', *The Lancet*, 360(9335), pp. 772–777. doi: 10.1016/S0140-6736(02)09899-9.

Ostermeier, G. C., Miller, D., Huntriss, J. D., Diamond, M. P. and Krawetz, S. (2004) 'Delivering spermatozoan RNA to the oocyte', *Nature*, 429(6988), pp. 1–1. doi: 10.1038/nature02602.

Pérez-Montero, S., Carbonell, A. and Azorín, F. (2016) 'Germline-specific H1 variants: the "sexy" linker histones', *Chromosoma*, 125(1), pp. 1–13. doi: 10.1007/s00412-015-0517-x.

Pertea, M., Kim, D., Pertea, G. M., Leek, J. T. and Salzberg, S. L. (2016) 'Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown', *Nature Protocols*, 11(9), pp. 1650–1667. doi: 10.1038/nprot.2016.095.

Pillot, M., Baroux, C., Vazquez, M. A., Autran, D., Leblanc, O., Vielle-Calzada, J. P., Grossniklaus, U. and Grimanelli, D. (2010) 'Embryo and endosperm inherit distinct chromatin and transcriptional states from the female gametes in Arabidopsis', *The Plant Cell*, 22(2), pp. 307–320. doi: 10.1105/tpc.109.071647.

Pimentel, H., Bray, N. L., Puente, S., Melsted, P. and Pachter, L. (2017) 'Differential analysis of RNA-seq incorporating quantification uncertainty', *Nature Methods*, 14(7), pp. 687–690. doi: 10.1038/nmeth.4324.

Pinon, V., Yao, X., Dong, A. and Shen, W.-H. (2017) 'SDG2-Mediated H3K4me3 Is Crucial for Chromatin Condensation and Mitotic Division during Male Gametogenesis in Arabidopsis', *Plant Physiology*, 174(2), pp. 1205–1215. doi: 10.1104/pp.17.00306.

Pontvianne, F. and Liu, C. (2020) 'Chromatin domains in space and their functional implications', *Current Opinion in Plant Biology*, 54, pp. 1–10. doi: 10.1016/j.pbi.2019.11.005.

Quinlan, A. R. and Hall, I. M. (2010) 'BEDTools: A flexible suite of utilities for comparing genomic features', *Bioinformatics*, 26(6), pp. 841–842. doi: 10.1093/bioinformatics/btq033.

Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Dündar, F. and Manke, T. (2016) 'deepTools2: a next generation web server for deep-sequencing data analysis', *Nucleic acids research*, 44(1), pp. 160–165. doi: 10.1093/nar/gkw257.

Rathke, C., Baarends, W. M., Awe, S. and Renkawitz-Pohl, R. (2014) 'Chromatin dynamics during spermiogenesis', *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1839(3), pp. 155–168. doi: 10.1016/j.bbagrm.2013.08.004.

Ravi, M., Kwong, P. N., Menorca, R. M. G., Valencia, J. T., Ramahi, J. S., Stewart, J. L., Tran, R.

K., Sundaresan, V., Comai, L. and Chan, S. W. L. (2010) 'The rapidly evolving centromere-specific histone has stringent functional requirements in Arabidopsis thaliana', *Genetics*, 186(2), pp. 461–471. doi: 10.1534/genetics.110.120337.

Renzaglia, K. S. and Garbary, D. J. (2001) 'Motile gametes of land plants: Diversity, development, and evolution', *Critical Reviews in Plant Sciences*, 20(2), pp. 107–213. doi: 10.1080/20013591099209.

Reynolds, W. F. and Wolfe, S. L. (1984) 'Protamines in Plant Sperm', *Experimental Cell Research*, 152, pp. 443–448.

Rodrigues, J. A. and Zilberman, D. (2015) 'Evolution and function of genomic imprinting in plants', *Genes and Development*, 29(24), pp. 2517–2531. doi: 10.1101/gad.269902.115.

Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J. and Dunker, A. K. (2001) 'Sequence complexity of disordered protein', *Proteins: Structure, Function and Genetics*, 42(1), pp. 38–48. doi: 10.1002/1097-0134.

Rotman, N., Durbarry, A., Wardle, A., Yang, W. C., Chaboud, A., Faure, J.-E., Berger, F. and Twell, D. (2005) 'A Novel Class of MYB Factors Controls Sperm-Cell Formation in Plants', *Current Biology*, 15, pp. 244–248. doi: 10.1016/j.

Roudier, F., Ahmed, I., Bérard, C., Sarazin, A., Mary-Huard, T., Cortijo, S., Bouyer, D., Caillieux, E., Duvernois-Berthet, E., Al-Shikhley, L., Giraut, L., Després, B., Drevensek, S., Barneche, F., Dèrozier, S., Brunaud, V., Aubourg, S., Schnittger, A., Bowler, C., Martin-Magniette, 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), pp. 1928–1938. doi: 10.1038/emboj.2011.103.

Rutowicz, K., Lirski, M., Mermaz, B., Teano, G., Schubert, J., Mestiri, I., Kroteń, M. A., Fabrice, T. N., Fritz, S., Grob, S., Ringli, C., Cherkezyan, L., Barneche, F., Jerzmanowski, A. and Baroux, C. (2019) 'Linker histones are fine-scale chromatin architects modulating developmental decisions in Arabidopsis', *Genome Biology*, 20(1), pp. 1–22. doi: 10.1186/s13059-019-1767-3.

Sabari, B. R., Dall'Agnese, A., Boija, A., Klein, I. A., Coffey, E. L., Shrinivas, K., Abraham, B. J., Hannett, N. M., Zamudio, A. V., Manteiga, J. C., Li, C. H., Guo, Y. E., Day, D. S., Schuijers, J., Vasile, E., Malik, S., Hnisz, D., Lee, T. I., Cisse, I. I., Roeder, R. G., Sharp, P. A., Chakraborty, A.

K. and Young, R. A. (2018) ‘Coactivator condensation at super-enhancers links phase separation and gene control’, *Science*, 361(6400), p. eaar3958. doi: 10.1126/science.aar3958.

Sakai, K., Taconnat, L., Borrega, N., Yansouni, J., Brunaud, V., Paysant-Le Roux, C., Delannoy, E., Martin Magniette, M. L., Lepiniec, L., Faure, J. D., Balzergue, S. and Dubreucq, B. (2018) ‘Combining laser-assisted microdissection (LAM) and RNA-seq allows to perform a comprehensive transcriptomic analysis of epidermal cells of Arabidopsis embryo’, *Plant Methods*, 14(10), pp. 1–12. doi: 10.1186/s13007-018-0275-x.

Santos, M. R., Bispo, C. and Becker, J. D. (2017) ‘Isolation of Arabidopsis Pollen, Sperm Cells, and Vegetative Nuclei by Fluorescence-Activated Cell Sorting (FACS)’, in Schmidt, A. (ed.) *Plant Germline Development*. 1st edn. New York, NY: Springer US, pp. 193–210. doi: 10.1007/978-1-4939-7286-9_16.

Sassone-Corsi, P. (2002) ‘Unique chromatin remodeling and transcriptional regulation in spermatogenesis.’, *Science*, 296(5576), pp. 2176–2178. doi: 10.1126/science.1070963.

Schagdarsurengin, U., Paradowska, A. and Steger, K. (2012) ‘Analysing the sperm epigenome: Roles in early embryogenesis and assisted reproduction’, *Nature Reviews Urology*, 9(11), pp. 609–619. doi: 10.1038/nrurol.2012.183.

Schoft, V. K., Chumak, N., Choi, Y., Hannon, M., Garcia-Aguilar, M., Machlicova, A., Slusarz, L., Mosiolek, M., Park, J.-S., Park, G. T., Fischer, R. L. and Tamaru, H. (2011) ‘Function of the DEMETER DNA glycosylase in the Arabidopsis thaliana male gametophyte.’, *Proceedings of the National Academy of Sciences*, 108(19), pp. 8042–7. doi: 10.1073/pnas.1105117108.

Schoft, V. K., Chumak, N., Mosiolek, M., Slusarz, L., Komnenovic, V., Brownfield, L., Twell, D., Kakutani, T. and Tamaru, H. (2009) ‘Induction of RNA-directed DNA methylation upon decondensation of constitutive heterochromatin.’, *EMBO reports*, 10(9), pp. 1015–21. doi: 10.1038/embor.2009.152.

Schrödinger, L. (2015) *The {PyMOL} Molecular Graphics System, Version~1.8*.

Shakya, A., Park, S., Rana, N. and King, J. T. (2020) ‘Liquid-Liquid Phase Separation of Histone Proteins in Cells: Role in Chromatin Organization’, *Biophysical Journal*, 118(3), pp. 753–764. doi: 10.1016/j.bpj.2019.12.022.

Shogren-Knaak, M., Ishii, H., Sun, J.-M., Pazin, M. J., Davie, J. R. and Peterson, C. (2006) 'Histone H4-K16 Acetylation Controls Chromatin Structure and Protein Interactions', *Science*, 16(2), pp. 844–848. doi: https://doi.org/10.1126/science.1124000.

Simon, L., Voisin, M., Tatout, C. and Probst, A. V. (2015) 'Structure and function of centromeric and pericentromeric heterochromatin in arabidopsis thaliana', *Frontiers in Plant Science*, 6(11), pp. 1–8. doi: 10.3389/fpls.2015.01049.

Slotkin, R. K., Vaughn, M., Borges, F., Tanurdžić, M., Becker, J. D., Feijó, J. A. and Martienssen, R. A. (2009) 'Epigenetic Reprogramming and Small RNA Silencing of Transposable Elements in Pollen', *Cell*, 136(3), pp. 461–472. doi: 10.1016/j.cell.2008.12.038.

Soares, L. M. and Buratowski, S. (2013) 'Histone Crosstalk: H2Bub and H3K4 Methylation', *Molecular Cell*, 49(6), pp. 1019–1020. doi: 10.1016/j.molcel.2013.03.012.

Southworth, D. and Cresti, M. (1997) 'Comparison of flagellated and nonflagellated sperm in plants', *American Journal of Botany*, 84(9), pp. 1301–1311. doi: 10.2307/2446056.

Sparkes, I. A., Runions, J., Kearns, A. and Hawes, C. (2006) 'Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants', *Nature Protocols*, 1(4), pp. 2019–2025. doi: 10.1038/nprot.2006.286.

Starmer, J. and Magnuson, T. (2016) 'Detecting broad domains and narrow peaks in ChIP-seq data with hiddenDomains', *BMC Bioinformatics*, 17(1), pp. 1–10. doi: 10.1186/s12859-016-0991-z.

Stovner, E. B. and Sætrom, P. (2019) 'Epic2 efficiently finds diffuse domains in ChIP-seq data', *Bioinformatics*, 35(21), pp. 4392–4393. doi: 10.1093/bioinformatics/btz232.

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

Stroud, H., Do, T., Du, J., Zhong, X., Feng, S., Johnson, L., Patel, D. J. and Jacobsen, S. E. (2014) 'Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis', *Nature Structural and Molecular Biology*, 21(1), pp. 64–72. doi: 10.1038/nsmb.2735.

Stroud, H., Do, T., Du, J., Zhong, X., Feng, S., Patel, D. J. and Jacobsen, S. E. (2014) 'The roles

of non-CG methylation in Arabidopsis', *Nature Structural and Molecular Biology*, 21(1), pp. 64–72. doi: 10.1038/nsmb.2735.

Stroud, H., Otero, S., Desvoyes, B., Ramírez-Parra, E., Jacobsen, S. E. and Gutierrez, C. (2012) 'Genome-wide analysis of histone H3.1 and H3.3 variants in Arabidopsis thaliana', *Proceedings of the National Academy of Sciences*, 109(14), pp. 5370–5375. doi: 10.1073/pnas.1203145109.

Sundell, D., Mannapperuma, C., Netotea, S., Delhomme, N., Lin, Y. C., Sjödin, A., Van de Peer, Y., Jansson, S., Hvidsten, T. R. and Street, N. R. (2015) 'The Plant Genome Integrative Explorer Resource: PlantGenIE.org', *New Phytologist*, 208(4), pp. 1149–1156. doi: 10.1111/nph.13557.

Tabuchi, T. M., Rechtsteiner, A., Jeffers, T. E., Egelhofer, T. A., Murphy, C. T. and Strome, S. (2018) 'Caenorhabditis elegans sperm carry a histone-based epigenetic memory of both spermatogenesis and oogenesis', *Nature Communications*, 9(1), pp. 1–11. doi: 10.1038/s41467-018-06236-8.

Tachiwana, H., Kagawa, W., Osakabe, A., Kawaguchi, K., Shiga, T., Hayashi-Takanaka, Y., Kimura, H. and Kurumizaka, H. (2010) 'Structural basis of instability of the nucleosome containing a testis-specific histone variant, human H3T', *Proceedings of the National Academy of Sciences*, 107(23), pp. 10454–10459. doi: 10.1073/pnas.1003064107.

Talbert, P. B. and Henikoff, S. (2010) 'Histone variants--ancient wrap artists of the epigenome', *Nature Reviews Molecular Cell Biology*, 11(4), pp. 264–275. doi: nrm2861 [pii]\n10.1038/nrm2861.

Talbert, P. B. and Henikoff, S. (2016) 'Histone variants on the move: substrates for chromatin dynamics', *Nature Reviews Molecular Cell Biology*, 18(2), pp. 115–126. doi: 10.1038/nrm.2016.148.

Tang, K., Lang, Z., Zhang, H. and Zhu, J. K. (2016) 'The DNA demethylase ROS1 targets genomic regions with distinct chromatin modifications', *Nature Plants*, 2(November). doi: 10.1038/nplants.2016.169.

Tannenbaum, M., Sarusi-Portuguez, A., Krispil, R., Schwartz, M., Loza, O., Benichou, J. I. C., Mosquna, A. and Hakim, O. (2018) 'Regulatory chromatin landscape in Arabidopsis thaliana roots uncovered by coupling INTACT and ATAC-seq', *Plant Methods*, 14(113), pp. 1–12. doi:

10.1186/s13007-018-0381-9.

Teperek, M., Simeone, A., Gaggioli, V., Miyamoto, K., Allen, G. E., Erkek, S., Kwon, T., Marcotte, E. M., Zegerman, P., Bradshaw, C. R., Peters, A. H. F. M., Gurdon, J. B. and Jullien, J. (2016) 'Sperm is epigenetically programmed to regulate gene transcription in embryos', *Genome Research*, 26(8), pp. 1034–1046. doi: 10.1101/gr.201541.115.

The UniProt Consortium (2019) 'UniProt: A worldwide hub of protein knowledge', *Nucleic acids research*, 47(1), pp. 506–515. doi: 10.1093/nar/gky1049.

Thorvaldsdóttir, H., Robinson, J. T. and Mesirov, J. P. (2012) 'Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration', *Briefings in Bioinformatics*, 14(2), pp. 178–192. doi: 10.1093/bib/bbs017.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pimentel, H., Salzberg, S. L., Rinn, J. L. and Pachter, L. (2012) 'Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks', *Nature Protocols*, 7(3), pp. 562–578. doi: 10.1038/nprot.2012.016.

Turner, A. L., Watson, M., Wilkins, O. G., Cato, L., Travers, A., Thomas, J. O. and Stott, K. (2018) 'Highly disordered histone H1−DNA model complexes and their condensates', *Proceedings of the National Academy of Sciences*, 115(47), pp. 11964–11969. doi: 10.1073/pnas.1805943115.

Twell, D. (2011) 'Male gametogenesis and germline specification in flowering plants', *Sexual Plant Reproduction*, 24(2), pp. 149–160. doi: 10.1007/s00497-010-0157-5.

Tyanova, S., Temu, T. and Cox, J. (2016) 'The MaxQuant computational platform for mass spectrometry-based shotgun proteomics', *Nature Protocols*, 11(12), pp. 2301–2319. doi: 10.1038/nprot.2016.136.

Ueda, K., Kinoshita, Y., Xu, Z., Ide, N., Ono, M., Akahori, Y., Tanaka, I. and Inoue, M. (2000) 'Unusual core histones specifically expressed in male gametic cells of Lilium longiflorum', *Chromosoma*, 108, pp. 491–500. doi: 10.1007/s004120050401.

Ueda, K. and Tanaka, I. (1995) 'The appearance of male gamete-specific histones gH2B and gH3 during pollen development in Lilium longiflorum', *Developmental Biology*, pp. 210–217.

doi: 10.1006/dbio.1995.1138.

Uversky, V. N. (2017) 'Intrinsically disordered proteins in overcrowded milieu: Membrane-less organelles, phase separation, and intrinsic disorder', *Current Opinion in Structural Biology*, 44, pp. 18–30. doi: 10.1016/j.sbi.2016.10.015.

Vavrdová, T., Šamajová, O., Křenek, P., Ovečka, M., Floková, P., Šnaurová, R., Šamaj, J. and Komis, G. (2019) 'Multicolour three dimensional structured illumination microscopy of immunolabeled plant microtubules and associated proteins', *Plant Methods*, 15(1), pp. 1–17. doi: 10.1186/s13007-019-0406-z.

Walker, J., Gao, H., Zhang, J., Aldridge, B., Vickers, M., Higgins, J. D. and Feng, X. (2018) 'Sexual-lineage-specific DNA methylation regulates meiosis in Arabidopsis', *Nature Genetics*, 50(1), pp. 130–137. doi: 10.1038/s41588-017-0008-5.

Wang, L., Gao, Y., Zheng, X., Liu, C., Dong, S., Li, R., Zhang, G., Wei, Y., Qu, H., Li, Y., Allis, C. D., Li, G., Li, H. and Li, P. (2019) 'Histone Modifications Regulate Chromatin Compartmentalization by Contributing to a Phase Separation Mechanism', *Molecular Cell*, 76(4), pp. 646-659.e6. doi: 10.1016/j.molcel.2019.08.019.

West, P. T., Li, Q., Ji, L., Eichten, S. R., Song, J., Vaughn, M. W., Schmitz, R. J. and Springer, N. M. (2014) 'Genomic distribution of H3K9me2 and DNA methylation in a maize genome', *PLoS ONE*, 9(8), pp. 1–10. doi: 10.1371/journal.pone.0105267.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K. and Yutani, H. (2019) 'Welcome to the Tidyverse', *Journal of Open Source Software*, 4(43), p. 1686. doi: 10.21105/joss.01686.

Wu, S. F., Zhang, H. and Cairns, B. R. (2011) 'Genes for embryo development are packaged in blocks of multivalent chromatin in zebrafish sperm', *Genome Research*, 21(4), pp. 578–589. doi: 10.1101/gr.113167.110.

Xiao, J., Lee, U. S. and Wagner, D. (2016) 'Tug of war: adding and removing histone lysine methylation in Arabidopsis', *Current Opinion in Plant Biology*, 34, pp. 41–53. doi: 10.1016/j.pbi.2016.08.002.

Xiao, J., Zhang, H., Xing, L., Xu, S., Liu, H., Chong, K. and Xu, Y. (2013) 'Requirement of histone acetyltransferases HAM1 and HAM2 for epigenetic modification of FLC in regulating flowering in Arabidopsis', *Journal of Plant Physiology*, 170(4), pp. 444–451. doi: 10.1016/j.jplph.2012.11.007.

Yan, W., Chen, D., Schumacher, J., Durantini, D., Engelhorn, J., Chen, M., Carles, C. C. and Kaufmann, K. (2019) 'Dynamic control of enhancer activity drives stage-specific gene expression during flower morphogenesis', *Nature Communications*, 10(1), pp. 1–16. doi: 10.1038/s41467-019-09513-2.

Yang, H., Yang, N. and Wang, T. (2016) 'Proteomic analysis reveals the differential histone programs between male germline cells and vegetative cells in Lilium davidii', *The Plant Journal*, 85(5), pp. 660–674. doi: 10.1111/tpj.13133.

Yao, X., Yang, H., Zhu, Y., Xue, J., Wang, T., Song, T., Yang, Z. and Wang, S. (2018) 'The canonical E2Fs are required for germline development in Arabidopsis', *Frontiers in Plant Science*, 9(5), pp. 1–10. doi: 10.3389/fpls.2018.00638.

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), pp. 98–109. doi: 10.1016/j.cell.2014.06.006.

Van Zanten, M., Koini, M. A., Geyer, R., Liu, Y., Brambilla, V., Bartels, D., Koornneef, M., Fransz, P. and Soppe, W. J. J. (2011) 'Seed maturation in Arabidopsis thaliana is characterized by nuclear size reduction and increased chromatin condensation', *Proceedings of the National Academy of Sciences*, 108(50), pp. 20219–20224. doi: 10.1073/pnas.1117726108.

Zemach, A., Kim, M. Y., Hsieh, P. H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., Harmer, S. L. and Zilberman, D. (2013) 'The arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin', *Cell*, 153(1), pp. 193–205. doi: 10.1016/j.cell.2013.02.033.

Zhang, C., Du, X., Tang, K., Yang, Z., Pan, L., Zhu, P., Luo, J., Jiang, Y., Zhang, Hui, Wan, H., Wang, X., Wu, F., Tao, W. A., He, X. J., Zhang, Heng, Bressan, R. A., Du, J. and Zhu, J. K. (2018) 'Arabidopsis AGDP1 links H3K9me2 to DNA methylation in heterochromatin', *Nature*

Communications, 9(1). doi: 10.1038/s41467-018-06965-w.

Zhang, C., Hung, Y. H., Rim, H. J., Zhang, D., Frost, J. M., Shin, H., Jang, H., Liu, F., Xiao, W., Iyer, L. M., Aravind, L., Zhang, X. Q., Fischer, R. L., Huh, J. H. and Hsieh, T. F. (2019) 'The catalytic core of DEMETER guides active DNA demethylation in Arabidopsis', *Proceedings of the National Academy of Sciences*, 116(35), pp. 17563–17571. doi: 10.1073/pnas.1907290116.

Zhang, K., Sridhar, V. V., Zhu, J., Kapoor, A. and Zhu, J. K. (2007) 'Distinctive core histone post-translational modification patterns in Arabidopsis thaliana', *PLoS ONE*, 2(11). doi: 10.1371/journal.pone.0001210.

Zhang, L., Chen, Fei, Zhang, X., Li, Z., Zhao, Y., Lohaus, R., Chang, X., Dong, W., Ho, S. Y. W., Liu, X., Song, A., Chen, J., Guo, W., Wang, Z., Zhuang, Y., Wang, H., Chen, Xuequn, Hu, J., Liu, Yanhui, Qin, Y., Wang, K., Dong, S., Liu, Yang, Zhang, S., Yu, X., Wu, Q., Wang, L., Yan, X., Jiao, Y., Kong, H., Zhou, X., Yu, C., Chen, Y., Li, F., Wang, J., Chen, W., Chen, Xinlu, Jia, Q., Zhang, C., Jiang, Y., Zhang, W., Liu, G., Fu, J., Chen, Feng, Ma, H., Van de Peer, Y. and Tang, H. (2019) 'The water lily genome and the early evolution of flowering plants', *Nature*, 577(7788), pp. 79–84. doi: 10.1038/s41586-019-1852-5.

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nussbaum, C., Myers, R. M., Brown, M., Li, W. and Shirley, X. S. (2008) 'Model-based analysis of ChIP-Seq (MACS)', *Genome Biology*, 9(9). doi: 10.1186/gb-2008-9-9-r137.

Zhao, S., Cheng, L., Gao, Y., Zhang, B., Zheng, X., Wang, L., Li, P., Sun, Q. and Li, H. (2019) 'Plant HP1 protein ADCP1 links multivalent H3K9 methylation readout to heterochromatin formation', *Cell Research*, 29(1), pp. 54–66. doi: 10.1038/s41422-018-0104-9.

Zhu, J., Kapoor, A., Sridhar, V. V., Agius, F. and Zhu, J. K. (2007) 'The DNA Glycosylase/Lyase ROS1 Functions in Pruning DNA Methylation Patterns in Arabidopsis', *Current Biology*, 17(1), pp. 54–59. doi: 10.1016/j.cub.2006.10.059.

Zilberman, D. (2017) 'An evolutionary case for functional gene body methylation in plants and animals', *Genome Biology*, 18(1), pp. 17–19. doi: 10.1186/s13059-017-1230-2.