Title
A role for genic DNA methylation in the governance of H2A.Z enrichment within gene bodies and the transcriptional regulation of responsive genes

Author
Coleman-Derr, Devin A.

Publication Date
2012

Peer reviewed|Thesis/dissertation

A role for genic DNA methylation in the governance of H2A.Z enrichment within gene bodies and the transcriptional regulation of responsive genes

By

Devin A. Coleman-Derr

A dissertation submitted in partial satisfaction of the
requirement for the degree of
Doctor of Philosophy
in
Plant Biology
in the
Graduate Division
of the
University of California, Berkeley

Committee in charge:
Professor Daniel Zilberman, Chair
Professor Robert Fischer
Professor Michael Freeling
Professor Barbara Meyer

Spring 2012

A role for genic DNA methylation in the governance of H2A.Z enrichment within gene bodies and the transcriptional regulation of responsive genes

By

Devin A. Coleman-Derr

Abstract

A role for genic methylation in the governance of H2A.Z enrichment within gene bodies

and

the transcriptional regulation of responsive genes

by

Devin A. Coleman-Derr

Doctor of Philosophy in Plant Biology

University of California, Berkeley

Professor Daniel Zilberman, Chair

One remarkable property of the eukaryotic cell is its ability to orchestrate the activities of thousands of genes in a complex temporal symphony of transcriptional expression. Development in multicellular species often requires that many genes lie dormant in early undifferentiated cellular lineages, awakening only in the tissues that they help define. Even single cellular species, such as the yeast *Saccharomyces cerevisiae*, need to keep some genes in a temporary, transcriptionally-repressed state until the onset of particular environmental conditions. This is no easy feat to accomplish, and cells use many different molecular mechanisms to do so; this includes the intricate interplay of many epigenetic regulatory systems, such as the post-translational modification of histones, the incorporation of histone variants, and a covalent but reversible modification of the DNA itself, DNA methylation.

In this dissertation, I describe a series of experiments designed to help understand the interaction between two of these epigenetic factors, DNA methylation and the histone variant H2A.Z, within the context of gene regulation. This work was conceived after initial mapping experiments in the model plant *Arabidopsis thaliana* revealed that the genome-wide distributions of H2A.Z and DNA methylation are strikingly anticorrelated. Additionally experiments have revealed that the basis for this

relationship is the exclusion of H2A.Z from chromatin by the presence of DNA methylation, an epigenetic principle that appears likely to be an ancient invention conserved among both plants and animals.

To better understand what purpose this relationship might hold in eukaryotes, I developed an Arabidopsis partial loss-of-function *h2a.z* mutant, and surveyed its genome-wide RNA expression profile. These experiments revealed strong correlations between transcriptional misregulation in the *h2a.z* mutant, the presence of H2A.Z within gene bodies, and levels of gene responsiveness, a measure of the degree to which a gene's expression varies across tissue types or environmental conditions. As we have shown that the presence of DNA methylation antagonizes H2A.Z incorporation across the genome, we propose that one basal function of gene-body methylation, an ancient and yet mysterious chromatin feature found in many eukaryotes, may be the prevention of H2A.Z incorporation within the bodies of genes that need to be constitutively expressed.

How gene body methylation is targeted in the first place remains unclear. The fact that genic methylation in all species is almost exclusively limited to CG sites, even in plants which have two other contexts of methylation that they use for the silencing of transposons, suggests that the various methylation targeting machineries are somehow able to distinguish between gene sequences and their other heterochromatic targets. Recently, several mutants in Arabidopsis have been shown to accumulate non-CG methylation within gene bodies. In order to understand the mechanisms responsible for this hypermethylation of genes, we examined the methylation profiles of these mutants. We discovered that the hypermethylation phenotypes of these mutants are quite different from one another in several respects, including their correlation with normal genic methylation, their distribution patterns across the gene body, and their dependence on the endogenous RNAi machinery. This suggests that multiple mechanisms may be responsible for controlling genic methylation patterns.

Contents

List of Figures

Chapter 2

Chapter 3

Chapter 4

List of Tables

Chapter 3

Acknowledgments

This dissertation would not exist without my advisor, Dr. Daniel Zilberman. It is primarily thanks to his support and mentoring over the past five years that this project has been a success. In particular, I want to thank him for relating to me many stories from his own years as a graduate student, which have helped me place my own experiences within a broader context and have encouraged me to persevere.

I am also indebted to the other members of the Zilberman lab, for their continued assistance with technical issues during my experiments, as well as for their camaraderie and good cheer, which have helped make my graduate career an enjoyable experience. In particular, I would thank Toshiro Nishimura and Pedro Silva, our bioinformaticists, who have provided the fantastic tools that my analyses rely heavily on. Members of the Fischer lab, with whom we frequently collaborate, have also been very gracious with their time, energy, and resources, and have my many thanks.

My thesis committee has been very supportive of me, through the ups and downs of my various projects. Prof. Bob Fischer, Prof. Mike Freeling, and Prof. Barbara Meyer have offered me the advice and encouragement that I needed, and I thank them for generously giving me so much of their time.

My graduate experience has been strongly influenced by my relationships with my undergraduate mentees, Anna Lee, Jimmy T. Le, and Maggie Wang, who have been an inspiration to me through their incredible patience, positive attitudes, and admirable work ethics. I could never have completed this work without their help.

Of course, graduate school can at times be challenging and overwhelming, and I want to extend a special thanks to my family. In particular, I would like to thank my mom, my dad, and my sister for the many hours of their time they shared with me in conversation about my project and for their continued support.

Above all, I want to thank my wife, Dawn, for inspiring me to embark on this rewarding journey in the first place. I could not have asked for a better companion in this half-decade adventure in plant biology.

Chapter 1

A role for genic methylation in the governance of H2A.Z enrichment within gene bodies and the transcriptional control of responsive genes

1.1 Introduction

1.1.1 Histone variant H2A.Z's relationship with transcription variability

In addition to the canonical histones, there exist a variety of non-canonical histone variants that play roles in a diverse array of cellular processes, including meiotic recombination, chromosome segregation, DNA repair, transcription initiation, and sex chromosome condensation (Talbert and Henikoff 2010). H2A.Z, which replaces canonical H2A, is perhaps the most well-conserved of all histone variants, and considerable progress has been made in recent decades towards understanding H2A.Z's role in eukaryotic biology (Zlatanova and Thakar 2008). Unlike some other histone variants, which are lineage specific (Talbert and Henikoff 2010), H2A.Z appears to have had a single origin at the root of the eukaryotic tree, and shares an approximate 90% amino acid sequence identity between species as diverse as *Saccharomyces cerevisiae, Drosophila melanogaster, Homo sapiens* and *Arabidopsis thaliana (Redon, Pilch et al. 2002; Zlatanova and Thakar 2008)*. Mutant analyses have revealed that, unlike most other species, the unicellular yeasts *S. cerevisiae* and *Schizosaccharomyces pombe* can tolerate a loss of H2A.Z, though they exhibit environmental sensitivities and slower growth (Jackson and Gorovsky 2000; Redon, Pilch et al. 2002). By contrast, the presence of H2A.Z in animals is strictly required for viability, with mutants exhibiting early developmental arrest and death (van Daal and van der Leij 1992; Liu, Li et al. 1996; Clarkson, Wells et al. 1999; Faast, Thonglairoam et al. 2001; Ridgway, Brown et al. 2004). Unlike many of these other model systems, which have only a single copy of H2A.Z, the Arabidopsis genome contains three. In Arabidopsis, a loss of two of the three *H2A.Z* genes leads to a variety of developmental phenotypes, including a loss of apical dominance, smaller and fewer rosette leaves, decreased fertility, and early flowering (March-Diaz, Garcia-Dominguez et al. 2008). A strong loss-of-function *h2a.z* mutant with insertions in all three *H2A.Z* genes exhibits a similar, but considerably stronger, phenotype (Coleman-Derr et. al., 2012, in press) while remaining viable, potentially making Arabidopsis an unusual and useful model system among multicellular species for understanding H2A.Z biology.

The high degree of conservation in H2A.Z among eukaryotes, taken in conjunction with the phenotypes of *h2a.z* mutants, suggests the existence of one or more ancient and important functions for this variant. A considerable proportion of H2A.Z research has focused on analyzing differences between the protein structures of H2A.Z and H2A in order to reveal these core, H2A.Z-specific functions (Billon and Cote 2011). Despite considerable similarities between the two molecules, the crystal structure of this molecule has revealed key amino acid residue differences that affect the manner in which H2A.Z molecules interact with each other and with the H3/H4 core, potentially altering nucleosome stability (Suto, Clarkson et al. 2000). Like many aspects of H2A.Z biology, the nature of the effect that H2A.Z has on nucleosome stability remains controversial, and experiments have found evidence for both stabilizing (Fan, Gordon et al. 2002; Park, Dyer et al. 2004; Thambirajah, Dryhurst et al. 2006; Ishibashi,

Dryhurst et al. 2009), and destabilizing (Suto, Clarkson et al. 2000; Placek, Harrison et al. 2005; Zhang, Roberts et al. 2005) effects.

Other amino acid residues have been discovered which are required for interaction with a specific SWR1 chromatin remodeler and the subsequent deposition of H2A.Z into chromatin (Wu, Alami et al. 2005; Jensen, Santisteban et al. 2011; Wang, Aristizabal et al. 2011). This multi-subunit, ATP-dependent chromatin remodeling complex was first discovered in yeast, where it was shown to be capable of replacing H2A/H2B dimers with H2A.Z/H2B *in vitro* and *in vivo* (Kobor, Venkatasubrahmanyam et al. 2004; Krogan, Baetz et al. 2004; Mizuguchi, Shen et al. 2004; Li, Pattenden et al. 2005). More recently, homologous remodelers with similar functional capabilities have been discovered in animals (Kusch, Florens et al. 2004; Zhang, Roberts et al. 2005; Ruhl, Jin et al. 2006; Gevry, Chan et al. 2007; Wong, Cox et al. 2007; Cuadrado, Corrado et al. 2011). In Arabidopsis, the *SWR1* homolog *PIE1* has been demonstrated to interact with H2A.Z and to be required for the deposition of H2A.Z (Choi, Park et al. 2007; Deal, Topp et al. 2007; March-Diaz, Garcia-Dominguez et al. 2008). Mutations in the catalytic subunit *PIE1*, as well as in other members of the PIE1 complex, produce phenotypes that are broadly similar to those found in the Arabidopsis *h2a.z* mutants, though the presence of several phenotypic differences suggests that mutations in PIE1 and H2A.Z may not be completely functionally redundant (Noh and Amasino 2003; Deal, Kandasamy et al. 2005; Choi, Park et al. 2007; March-Diaz, Garcia-Dominguez et al. 2007; March-Diaz, Garcia-Dominguez et al. 2008). Whether PIE1 has H2A.Z-independent functions, which has been shown for eukaryotic SWR1 homologs in animals, remains unclear (Kobor, Venkatasubrahmanyam et al. 2004; Auger, Galarneau et al. 2008; Bowman, Wong et al. 2011). Possibly indicative of such non-redundant roles, we have recently shown that double mutants of *pie1* and *h2a.z* are not viable, exhibiting early developmental arrest shortly after germination (Coleman-Derr et. al., 2012 in press). This is quite different from results reported in yeast, which showed that simultaneous mutation of *SWR1* and *HTZ1* (yeast H2A.Z) lead to an amelioration of the phenotypes found in *htz1* mutants alone (Morillo-Huesca, Clemente-Ruiz et al. 2010); this result was hypothesized to be caused by SWR1's continued remodeling and consequent destabilization of chromatin in the absence of its substrate, *HTZ1*. Though the cause of the more severe phenotype in *h2a.z;pie1* Arabidopsis plants remains unclear, these results suggest that there may be important differences in H2A.Z-related function between eukaryotic species.

It was recently shown in yeast that INO80, a chromatin remodeler belonging to the same subfamily as SWR1 (Morrison and Shen 2009), can regulate the genome-wide distribution of H2A.Z by promoting the eviction of H2A.Z from promoters during transcriptional induction (Papamichos-Chronakis, Watanabe et al. 2011). This discovery, along with the fact that H2A.Z is incorporated into nucleosomes at low levels in both *pie1* and *swr1* mutants (Kobor, Venkatasubrahmanyam et al. 2004; Wu, Alami et al. 2005; Deal, Topp et al. 2007), suggests that similar remodelers may also be capable of depositing H2A.Z into chromatin in Arabidopsis. The fact that in both *S. cerevisiae* and Arabidopsis the sets of genes that are misregulated in H2A.Z and SWR1-related

mutants only partially overlap (Kobor, Venkatasubrahmanyam et al. 2004; March-Diaz, Garcia-Dominguez et al. 2008; Morillo-Huesca, Clemente-Ruiz et al. 2010) further supports this hypothesis. Homologs of INO80 exist in Arabidopsis (Fritsch, Benvenuto et al. 2004), and an important question that remains to be answered is whether any of these remodelers possess similar H2A.Z remodeling capabilities and assist in shaping the characteristic genomic distribution pattern of H2A.Z.

Genome-wide localization experiments in fungi, animals, and plants have all shown that H2A.Z is preferentially found within the few nucleosomes surrounding the transcription start sites (TSS) of genes (Guillemette and Gaudreau 2006; Creyghton, Markoulaki et al. 2008; Zilberman, Coleman-Derr et al. 2008), though which genes show enrichment seems to vary depending on species. In yeast, the majority of genes have a peak of H2A.Z at their 5' end, regardless of whether they are active or inactive (Guillemette, Bataille et al. 2005; Li, Pattenden et al. 2005). By contrast, in animals, the 5' enrichment of H2A.Z appears more frequently on active than inactive genes (Whittle, McClinic et al. 2008; Hardy, Jacques et al. 2009; Jin, Zang et al. 2009). Interestingly, it was found that in mammalian ES cells the genes enriched by H2A.Z change during the process of differentiation (Creyghton, Markoulaki et al. 2008). Contrary to the general consensus that active genes are enriched for H2A.Z in animals, genes showing enrichment in the ES cell lineage were overrepresented for silent developmental genes required in later cell types (Creyghton, Markoulaki et al. 2008). Similarly, in *C. elegans*, it was found that the 23% of genes that show H2A.Z enrichment by ChIP-chip were enriched for GO terms related to embryonic development, larval development, and gamete generation (Whittle, McClinic et al. 2008). The enrichment of H2A.Z at specific classes of gene, as well as the changes in enrichment patterns during differentiation, suggest potential innovations for H2A.Z function in multicellular organisms that might not be found in yeast. In Arabidopsis, we have shown that the majority of genes exhibit a strong 5' peak of H2A.Z enrichment, as well as a smaller peak at their 3' end (Zilberman, Coleman-Derr et al. 2008) and (Coleman-Derr et. al., 2012, in press). The presence of this 3' peak in Arabidopsis genes is supported by previous reports from ChIP-PCR of specific loci (Deal, Topp et al. 2007). Taken together, the localization data demonstrate that while the primary feature of H2A.Z distribution across eukaryotes is a consistent and strong presence at the 5' ends of genes, there are lineage-specific differences to enrichment patterns that may well have functional significance.

Not surprisingly, most research on H2A.Z has focused on the 5' enrichment of H2A.Z at the TSS of genes, and only a few studies have reported the presence of H2A.Z in other parts of the gene (Fujimoto, Seebart et al. 2012). Recently, Fujimoto et. al. reported that a reexamination of the available genome-wide mapping data from animals and fungi, however, demonstrate the presence of H2A.Z within gene bodies as well (Li, Pattenden et al. 2005; Zhang, Roberts et al. 2005; Millar, Xu et al. 2006; Albert, Mavrich et al. 2007; Zlatanova and Thakar 2008; Tolstorukov, Kharchenko et al. 2009; Zemach, McDaniel et al. 2010). Another study suggests that perhaps as much as 40% of all H2A.Z containing nucleosomes may reside downstream of the promoter in human genes (Tolstorukov, Kharchenko et al. 2009). In Arabidopsis, we have recently reported

that in addition to H2A.Z enrichment at the TSS, many genes exhibit considerable levels of H2A.Z across their coding regions (Coleman-Derr et. al., 2012, in press). Though the significance of this enrichment remains unexplored, it has been hypothesized that the H2A.Z-mediated functions in gene bodies and promoters may well be different (Fujimoto, Seebart et al. 2012).

Given the conserved H2A.Z distribution pattern at transcriptional start sites (TSS), many studies have focused on dissecting the relationship between H2A.Z and transcription. Counter intuitively, H2A.Z enrichment at promoters in yeast was found to be simultaneously required for, and inversely correlated with, transcription (Guillemette, Bataille et al. 2005; Zhang, Roberts et al. 2005; Millar, Xu et al. 2006). *PHO5*, a gene that is repressed on phosphate rich media and active under low phosphate conditions, shows a loss of H2A.Z upon transcriptional induction under phosphate deprivation (Millar, Xu et al. 2006). A return to phosphate rich conditions resulted in transcriptional repression and increased H2A.Z occupancy (Millar, Xu et al. 2006). Similar results were found for the *GAL1* and *GAL10* loci, which are induced upon addition of galactose, but it was also demonstrated that *GAL1* and *GAL10* require H2A.Z in order for transcriptional activation (Adam, Robert et al. 2001). Overall a number of genes required for growth under specific conditions showed defects in transcriptional activation in an *h2a.z* mutant in yeast (Santisteban, Kalashnikova et al. 2000; Larochelle and Gaudreau 2003; Dhillon, Oki et al. 2006; Wan, Saleem et al. 2009).

In contrast to this pattern in yeast, most studies in animals report that H2A.Z exhibits a positive correlation with transcription. In human and *Drosophila*, H2A.Z enrichment at promoters correlates with the level of gene expression (Barski, Cuddapah et al. 2007; Mavrich, Jiang et al. 2008) and in humans H2A.Z has been shown to associate with gene promoters upon their induction and aid in the recruitment of transcriptional machinery (Hardy, Jacques et al. 2009). However, some studies suggest a more complex relationship; studies with mice ES cells have demonstrated that H2A.Z is present at silent Retinoic Acid Response (RAR) genes and is removed upon the addition of Retinoic Acid and the transcriptional activation of these genes (Amat and Gudas 2011). In both *C. elegans* and the pufferfish *Tetraodon nigridiris*, the relationship appears parabolic; H2A.Z enrichment at promoters are positively correlated with transcription up to a point, after which the correlation becomes negative (Whittle, McClinic et al. 2008; Zemach, McDaniel et al. 2010).

Similar to the results from *C. elegans* and pufferfish, we have shown that in Arabidopsis promoter H2A.Z enrichment and transcription share a roughly parabolic relationship, with H2A.Z levels at their highest in moderately transcribed genes, and at their lowest at either transcriptional extreme. The presence of this correlation between H2A.Z and transcription level in both plants and some animals suggests that it may well be the ancestral relationship for eukaryotes. Interestingly, we have found that the relationship between transcription and H2A.Z enrichment within gene bodies in Arabidopsis shows a negative correlation, with the lowest expressed genes showing the greatest gene body enrichment of H2A.Z (Zilberman, Coleman-Derr et al. 2008;

Coleman-Derr 2012). Indeed, several other studies in animals have noted that, unlike promoter enrichment of H2A.Z, the presence of H2A.Z within gene bodies appears to be negatively correlated with transcription (Barski, Cuddapah et al. 2007; Hardy, Jacques et al. 2009; Zemach, McDaniel et al. 2010). Given that H2A.Z enrichment at promoters and gene bodies have different relationships with transcription, it is possible that some of the discrepancies in the many transcriptional studies from various species may be due to differences in how H2A.Z enrichment has been measured within genes.

1.1.2 DNA methylation guides the global distribution of H2A.Z

Recently, we and others have found that H2A.Z and DNA methylation are anticorrelated on a genome-wide basis in both plants and animals (Zilberman, Coleman-Derr et al. 2008; Conerly, Teves et al. 2010; Edwards, O'Donnell et al. 2010; Zemach, McDaniel et al. 2010). Methylation, in the form of 5-methylcytosine, is a well-conserved feature of eukaryotic chromatin, present in all vertebrates examined to date, as well as in many invertebrates, fungi, and plants (Cokus, Feng et al. 2008; Feng, Cokus et al. 2010; Zemach, McDaniel et al. 2010; Glastad, Hunt et al. 2011). In plants, DNA methylation exists in three different sequence contexts, CG, CHG, and CHH (H= A, T, or C), which are catalyzed by distinct families of DNA methyltransferase enzymes that differ in their mechanisms of targeting (Zemach and Zilberman 2010). Broadly speaking, the distribution patterns of DNA methylation and H2A.Z are quite different; while H2A.Z is primarily found at the 5' ends of genes, DNA methylation is typically associated with TEs and other repetitive sequences and is generally absent from gene promoters (Zilberman, Coleman-Derr et al. 2008). Not surprisingly then, the global distribution patterns of DNA methylation and H2A.Z are strongly anticorrelated (Zilberman, Coleman-Derr et al. 2008). Interestingly, however, this anticorrelation exists at the local level as well, irrespective of the type of sequence examined; transposons (TEs) that have escaped DNA methylation targeting in Arabidopsis are enriched for H2A.Z and genes with DNA methylation are depleted of H2A.Z (Zilberman, Coleman-Derr et al. 2008). The opposing distributions of these two chromatin marks suggest that one or both of the marks may be acting to prevent the incorporation of the other.

Notably, we found that upon changes in DNA methylation patterning in Arabidopsis, induced by a loss of the primary maintenance methyltransferases MET1, opposite changes in H2A.Z deposition were detected (Zilberman, Coleman-Derr et al. 2008). While almost all TEs show a strong loss of CG DNA methylation in the *met1* background, only about half show transcriptional derepression. An examination of these two populations of TE (i.e. those that remain silenced, and those that become transcriptionally active in *met1*) revealed that they were equally enriched in H2A.Z in *met1*, which strongly argues that it is changes in DNA methylation rather than changes in transcription that cause the redistribution of H2A.Z (Zilberman, Coleman-Derr et al. 2008). In support of these results, *in vitro* experiments conducted with human tumor cells demonstrated that changes in DNA methylation associated with cancer progression were associated with opposite changes in H2A.Z levels (Conerly, Teves et al. 2010).

To test whether H2A.Z also acts to inhibit DNA methylation distribution we used two experiments which lead to results that on the surface seem contradictory. First, we used a mutant line deficient for PIE1 to mimic a loss of H2A.Z in chromatin and examined changes in DNA methylation by MeDIP-chip. Though the limitations of this technique (including its inability to distinguish between methylation sequence contexts and reliance on normalization of signal intensities) made these discoveries more difficult to interpret than the changes in H2A.Z distribution observed in the *met1* experiments, we did find small increases and decrease in genic and TE methylation, respectively (Zilberman, Coleman-Derr et al. 2008). More recently, we used whole-genome shotgun bisulfite sequencing to examine changes in methylation in an *h2a.z* loss-of-function line, and found that a loss of H2A.Z does not affect genic methylation (Coleman-derr et. al., 2012, in press). In support of this, Conerly et al. found almost no change in methylation at active gene promoters in the transition from WT to Em-MYC cells, despite a loss of H2A.Z from the promoters of these genes (Conerly, Teves et al. 2010). Taken together, these studies strongly suggest that the genome-wide anticorrelation between H2A.Z and DNA methylation is the result of DNA methylation acting to prevent the incorporation of H2A.Z into chromatin, and that this is an ancient feature of eukaryotes.

1.1.3 Gene body methylation is associated with constitutive expression

For the past two decades, DNA methylation was thought to primarily act as a mediator of transcription silencing, particularly at TEs (Suzuki and Bird 2008); more recently DNA methylation was also discovered in the bodies of genes in many organisms (Jones and Laird 1999; Tran, Henikoff et al. 2005). Genome-wide analyses in a variety of eukaryotic species have revealed that gene body methylation most likely existed in the very first eukaryotes, and remains well-conserved among both plants and animals today (Lister, Pelizzola et al. 2009; Feng, Cokus et al. 2010; Xiang, Zhu et al. 2010; Zemach, Kim et al. 2010; Zemach, McDaniel et al. 2010). The experiments conducted in plant species have demonstrated that although plants heavily methylate their TEs in all three contexts (CG, CHG, and CHH), the methylation found in genes is at lower levels and limited to CG sites (Lippman, Gendrel et al. 2004; Zhang, Yazaki et al. 2006; Zilberman, Gehring et al. 2007; Cokus, Feng et al. 2008; Lister, O'Malley et al. 2008; Zemach, McDaniel et al. 2010). Interestingly, most animals except vertebrates only have DNA methylation within gene bodies and have lost the ability to methylate TE sequences (Suzuki, Kerr et al. 2007; Feng, Cokus et al. 2010). This discovery has challenged the standard belief that DNA methylation exists primarily as a genomic surveillance system, and has opened the door to a broader evolutionary understanding of epigenetics in the context of gene regulation.

Despite the prevalence of gene body methylation in both plants and animals, the mechanisms for its establishment and the role it plays in gene regulation, if any, remain mysterious (Zemach and Zilberman 2010). Originally it was proposed that genic methylation may act to prevent cryptic promoters from firing within the bodies of actively

transcribed genes (Zilberman, Gehring et al. 2007). More recently, it was discovered that genic methylation is preferentially targeted to exons in both animals and plants, leading to the suggestion that gene-body methylation may play a role in exon definition and accurate/alternative splicing (Lorincz, Dickerson et al. 2004; Chodavarapu, Feng et al. 2010; Luco, Pan et al. 2010). At present, the evidence for this is still somewhat limited. Some hint to genic methylation's purpose may come from the now well-established relationship between DNA methylation and transcription.

Unlike DNA methylation at promoters, gene body methylation does not appear to antagonize transcription; on the contrary, targets of genic methylation tend to be expressed (Suzuki and Bird 2008). Gene body methylation is highest in moderately transcribed genes in both plants and animals, with the lowest levels of genic methylation in either transcriptional extreme (Zhang, Yazaki et al. 2006; Zilberman, Gehring et al. 2007; Zemach, McDaniel et al. 2010). In Arabidopsis, genic methylation is positively correlated with transcription up to the 70th percentile of transcription, after which the relationship becomes strongly negative (Zilberman, Coleman-Derr et al. 2008). Similar results were found in pufferfish, as well as in several invertebrates, including honeybee, silkworm, sea anemone and sea squirt (Zemach, McDaniel et al. 2010).

In addition, genic methylation also appears to correlate with constitutive gene expression. This relationship was first discovered in Arabidopsis, where it was shown that body methylated genes are more constituitively expressed (as measured by Shannon's entropy) than other genes (Zhang, Yazaki et al. 2006). Recent work has corroborated this; further analysis of Arabidopsis methylation data has shown that genes with body methylation are enriched for housekeeping functions, whereas those with no gene body methylation are more likely to exhibit high tissue and environmental condition specificity (Aceituno, Moseyko et al. 2008). Functional analysis of body methylated genes from another plant species (*populus trichocarpa*) revealed significant enrichment for gene categories which are typically constituitively expressed, including translation/protein metabolism, nucleic acid binding and RNA metabolism (Vining, Pomraning et al. 2012). Similar results were found in four invertebrates species, with body methylated genes showing enrichment for housekeeping functions, such as translation, ribosome biogenesis, RNA splicing, and protein localization (Sarda, Zeng et al. 2012). These data suggest that in addition to the ancestral relationship between gene body methylation and transcription level, the correlation of genic methylation with constitutive expression is an invention that predates the divergence of the plant and animal kingdoms.

1.1.4 Gene body enrichment of H2A.Z antagonizes transcriptional consistency

Quite the opposite of the relationship between DNA methylation and constituitively expressed genes, H2A.Z studies in many species have revealed an association between H2A.Z and various inducible classes of genes. Specifically, upon a loss of H2A.Z in yeast or animals, it is genes required only in specific environmental

conditions or tissue types, respectively, that are reported to be misregulated (Adam, Robert et al. 2001; Millar, Xu et al. 2006; Updike and Mango 2006; Creyghton, Markoulaki et al. 2008; Whittle, McClinic et al. 2008; Amat and Gudas 2011; Petter, Lee et al. 2011; Sadeghi, Bonilla et al. 2011). In Arabidopsis, plants lacking *PIE1* exhibit misregulation of many genes involved in the innate immune response (March-Diaz, Garcia-Dominguez et al. 2008). Recent work has shown that Arabidopsis plants with a mutated copy of *ARP6*, another component of the PIE1 complex, inappropriately express temperature response genes, which lead to the proposal that H2A.Z may act as a thermosensor in plants (Kumar and Wigge 2010). Other experiments done with ARP6 have demonstrated that H2A.Z is deposited at a number of Pi starvation response genes and that a loss of H2A.Z leads to their derepression (Smith, Jain et al. 2009).

We have recently found that in a strong loss-of-function *h2a.z* mutant there are three-fold more genes up-regulated than down regulated; similarly ratios of up- to downregulated genes were reported for *pie1* and another *h2a.z* mutant (March-Diaz, Garcia-Dominguez et al. 2008). Functional analyses of the 1200 genes upregulated in the *h2a.z* mutant revealed a significant over-representation for response genes in general (Coleman-Derr et. al., 2012, in press). In fact, all 19 of the top overrepresented GO terms (P-values less than $1x10^{-5}$) are specifically "response" categories (Coleman-Derr et. al., 2012, in press). Importantly, while all three response-related categories previously reported to be misregulated in H2A.Z-related mutants (innate immune response, response to temperature stimulus, and response to phosphate starvation) are also overrepresented in our *h2a.z* upregulated gene list, there are also many examples of other, unrelated response categories, including response to wounding, water, abiotic stimulus, carbohydrate stimulus, endogenous stimulus, salt stress, and hormone stimulus (Coleman-Derr et. al., 2012, in press). This result demonstrates that a loss of H2A.Z in plants likely leads to misregulation of response genes in general, rather than a specific category of response gene, and may help to reconcile the previous, seemingly disparate results from Arabidopsis.

In a similar vein, we found that the degree of misregulation in our loss-of-function *h2a.z* mutant correlated with gene responsiveness (where high gene responsiveness scores indicate greater differential expression across a variety of tissues or environmental conditions, as described in Aceituno 2009) (Aceituno, Moseyko et al. 2008) and (Coleman-Derr et. al., 2012, in press). Furthermore, the misregulation in this mutant appears to be directly related to the presence of H2A.Z within gene bodies, as the degree of misregulation directly correlated with the level of H2A.Z body enrichment (Coleman-Derr et. al., 2012, in press). Considering these two results, it is unsurprising that we have also found that the presence of H2A.Z within gene bodies is positively associated with measures of gene responsiveness (Coleman-Derr et. al., 2012, in press). Similarly, in yeast analyses of H2A.Z enrichment across gene coding sequences demonstrated significant overlap with genes that are differentially expressed after environmental stresses (Sadeghi, Bonilla et al. 2011).

Taken together, these findings point towards the conclusion that the presence of H2A.Z within gene bodies either antagonizes constitutive and strong gene expression, or promotes variability of levels and patterns of expression, in both plants and animals. Considering that evidence from both of these kingdoms suggests that the genome-wide anticorrelation between DNA methylation and H2A.Z is likely established by the prevention of H2A.Z deposition in chromatin by DNA methylation, we propose that one basal function of genic DNA methylation may be the establishment of constitutive expression patterns within housekeeping genes by preventing H2A.Z from becoming incorporated within their bodies. As H2A.Z has been linked to the expression of inducible genes in many species (Adam, Robert et al. 2001; Millar, Xu et al. 2006; Zanton and Pugh 2006; Creyghton, Markoulaki et al. 2008; Whittle, McClinic et al. 2008; Smith, Jain et al. 2009; Kumar and Wigge 2010; Amat and Gudas 2011; Petter, Lee et al. 2011; Sadeghi, Bonilla et al. 2011), including species such as *S. cerevisiae* and *C. elegans* which lack DNA methylation all together, it strikes us a stronger candidate for a direct link to gene responsiveness than DNA methylation.

1.1.5 Genic methylation targeting and hypermethylation mutants

Despite considerable knowledge about the targeting of DNA methylation to TEs and repeats, the mechanism for the establishment of gene body methylation remains mysterious (Suzuki and Bird 2008). In flowering plants, the absence of CHG and CHH methylation in gene bodies demonstrates that genes and transposons are differentially recognized by the various methylation machineries (Cokus, Feng et al. 2008; Lister, O'Malley et al. 2008; Feng, Cokus et al. 2010; Zemach, McDaniel et al. 2010). In many eukaryotes, TEs are interspersed amongst genes, which necessitates special mechanisms for the accurate and precise targeting of non-CG methylation found in these sequences.

One mechanism for establishing the presence of non-CG methylation in TEs but not genes would be to specifically target TEs with the methylation machinery. In plants, the RNA-directed DNA methylation pathway (RdDM) works in this way. This system, which results in methylation primarily in the CHH context, targets repeats and TEs through small regions of siRNA/DNA sequence homology (Furner and Matzke 2011). These siRNAs are the degradation products of double-stranded RNAs derived from plant specific RNA polymerase IV transcripts derived from repetitive loci. These transcripts are copied into dsRNA by RNA-dependent RNA polymerase 2 (Meyer 2010), and cleaved into smaller siRNAs by Dicer nucleases. Each siRNA strand is then incorporated into ARGONOTE complexes that can target specific regions of siRNA/DNA homology within the genome for DNA methylation and silencing (Chan, Zilberman et al. 2004). This CHH methylation is mediated by members of the DRM family of methyltransferase, and is classified as *de novo* methylation, as it doesn't require the presence of an initial epigenetic signal at the methylated locus for its targeting (Mathieu, Reinders et al. 2007). While precise and accurate, *de novo* methylation requires active targeting in each new cell after replication. Fortunately, plants also have a mechanism for propagating established DNA methylation marks directly during replication.

This mechanism, referred to as maintenance methylation, is responsible for perpetuating the presence of DNA methylation on both chromatids after DNA replication. To achieve this, the MET1 methyltransferase targets sites that have a methylatable cytosine on both strands of the DNA, specifically at CG sites (Kankel, Ramsey et al. 2003). Virtually all CG methylation within the Arabidopsis genome is catalyzed by MET1, including sites within both TEs and genes (Cokus, Feng et al. 2008; Lister, O'Malley et al. 2008). As discussed previously, a loss of MET1 leads to a complete genome-wide loss of CG methylation. Other factors have been shown to act as positive regulators of DNA methylation. One such factor, DDM1, has been shown to be responsible for helping establish DNA methylation at TEs; loss of the DDM1 chromatin remodeling complex causes a strong decrease in DNA methylation in TEs and repeats (Lippman, Gendrel et al. 2004; Tao, Xi et al. 2011).

Plants also contain a third methyltransferase family, CMT3, which is responsible for DNA methylation in the CHG context (Lindroth, Cao et al. 2001). This enzyme does not strictly fit the *de novo* or maintenance classification, but shares elements of both. Unlike, CHH sites, the presence of cytosines on either strand in CHG sites allows for the potential propagation of this mark across cell divisions (Furner and Matzke 2011). Similar to the RdDM pathway, however, CMT3 can be actively targeted to previously unmethylated sequences; it achieves this through interaction with methylation of H3 lysine 9 (H3K9me2), a second epigenetic hallmark of silencing that is conserved among animals, plants and fungi (Enke2011). CMT3 contains a chromodomain that allows it to bind H3K9me2, potentially offering a distinct method from RdDM for the targeting of new DNA methylation (Lindroth, Shultis et al. 2004). Additionally, one of the enzymes responsible for catalyzing H3K9me2 in Arabidopsis, KRYPTONITE (KYP), contains an SRA domain that is capable of binding to methylated cytosines (Johnson, Bostick et al. 2007). This mutual binding by KYP and CMT3 of their reciprocal enzymatic products creates an autocatalytic feedback loop that allows for the propagation and reinforcement of both CHG methylation and H3K9me2 across targeted sequences (Furner and Matzke 2011).

An alternative method for achieving the differential distribution of non-CG methylation in TEs and genes involves the active removal of methylation or associated marks from genes. Recent work has shown that, whereas H3K9 methylation is known to be enriched over TEs in many eukaryotic species, in plants H3K9 methylation can be targeted to genes as well. A histone demethylase, IBM1, was recently reported to be responsible for removing H3K9me2 from genic sequences (Saze, Shiraishi et al. 2008; Miura, Nakamura et al. 2009; Inagaki and Kakutani 2010). IBM1 is a member of the JHDM2/KDM3 demethylases that are conserved from plants to animals (Inagaki and Kakutani 2010). Presumably, it is the removal of H3K9me2 from these genic sequences by IBM1 that prevents the spread of non-CG methylation.

Consistent with this, loss-of-function *ibm1* mutants cause an increase in both H3K9me2 and non-CG methylation in genes (Saze, Shiraishi et al. 2008; Miura, Nakamura et al. 2009; Inagaki and Kakutani 2010). Interestingly, both *met1* and *ddm1* mutants have also been shown to exhibit increases in non-CG methylation in the bodies

of some genes (Jacobsen and Meyerowitz 1997; Mathieu, Reinders et al. 2007; Saze and Kakutani 2007; Cokus, Feng et al. 2008; Lister, O'Malley et al. 2008). These discoveries present two important questions in need of answering. First, are these three examples of genic hypermethylation conveyed by similar mechanisms? Second, do they bear any relation to the gene body methylation normally found in WT? To that end, we have examined the methylation profiles in these three mutant lines by whole-genome bisulfite sequencing. We have found that the CHG and CHH hypermethylation in *ibm1* correlate well with one another and, perhaps more importantly, with the presence of WT CG methylation (Zemach et. al., 2012, in press). Similar to *ibm1*, mutants in *ddm1* exhibit CHG hypermethylation of genes that correlates with the presence of genic CG methylation in WT (Zemach et. al., 2012, in press). By contrast, the hypermethylation found in the *met1* mutant showed no such correlation, appearing in many genes without WT genic methylation (Zemach et. al., 2012, in press).

Despite the strong correlation between WT CG methylation and both *ibm1* and *ddm1* non-CG hypermethylation, the hypermethylations in these two mutants show different patterns of distribution across the gene body. *ibm1* CHG and CHH hypermethylation exhibited a 5' bias, with greater levels towards the start of the gene and less methylation at the three prime end (Zemach et. al., 2012, in press). By contrast, *ddm1* CHG hypermethylation exhibited the opposite trend, reaching its highest levels at the 3' end of the gene (Zemach et. al., 2012, in press). Furthermore, crosses between *ibm1* and a mutation in a gene required for RdDM (*drd1*) do not abolish the CHG and CHH hypermethylation caused by *ibm1*, strongly suggesting that this hypermethylation is not mediated by the RdDM pathway (Zemach et. al., 2012, in press). By contrast, double mutants of *ddm1* and mutants in various RNAi components lost nearly all of the hypermethylation seen in *ddm1*; taken together, these results strongly support the hypothesis that although the hypermethylation in *ibm1* and *ddm1* both appear to follow the distribution of WT CG, the genic hypermethylation caused by *ibm1* and *ddm1* are mediated by distinct processes.

It has been hypothesized that the hypermethylations in *ddm1* and *met1* might be induced by a similar mechanism as the hypermethylation in *ibm1*; specifically the decreases in DNA methylation in and depression of TEs in both *met1* and *ddm1* might lead to a titration of IBM1 protein from genic to TE sequences, leaving genes without a means of adequately pruning H3K9me2 sequences (Miura, Nakamura et al. 2009; Inagaki and Kakutani 2010). However, the differences in their correlations with WT CG strongly suggest that the hypermethylation found in *met1* and *ddm1* are likely caused by separate mechanisms. Furthermore, the idea that the CHG hypermethylation seen in *met1* is a compensatory mechanism for the loss of genic CG methylation (Mathieu, Reinders et al. 2007) seems unlikely, as its presence does not appear solely in the genes typically targeted by gene body methylation. Taken together, these results argue for the presence of at least three distinct processes responsible for hypermethylation in these mutants, and the need for future studies to help refine our knowledge of the complex systems governing DNA methylation within genes.

Chapter 2

Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks

2.1 Abstract

Eukaryotic chromatin is separated into functional domains differentiated by posttranslational histone modifications, histone variants, and DNA methylation(Malik and Henikoff 2003; Goll and Bestor 2005; Bernstein and Hake 2006; Klose and Bird 2006; Bhaumik, Smith et al. 2007; Gehring and Henikoff 2007). Methylation is associated with repression of transcriptional initiation in plants and animals, and is frequently found in transposable elements. Proper methylation patterns are critical for eukaryotic development (Goll and Bestor 2005; Gehring and Henikoff 2007), and aberrant methylation-induced silencing of tumor suppressor genes is a common feature of human cancer (Feinberg, Ohlsson et al. 2006). In contrast to methylation, the histone variant H2A.Z is preferentially deposited by the Swr1 ATPase complex near 5' ends of genes where it promotes transcriptional competence (Meneghini, Wu et al. 2003; Mizuguchi, Shen et al. 2004; Guillemette, Bataille et al. 2005; Li, Pattenden et al. 2005; Raisner, Hartley et al. 2005; Zhang, Roberts et al. 2005; Guillemette and Gaudreau 2006; Millar, Xu et al. 2006; Updike and Mango 2006; Barski, Cuddapah et al. 2007; Brickner, Cajigas et al. 2007; Deal, Topp et al. 2007; Venkatasubrahmanyam, Hwang et al. 2007). How DNA methylation and H2A.Z influence transcription remains largely unknown. Here we show that in the plant *Arabidopsis thaliana*, regions of DNA methylation are quantitatively deficient in H2A.Z. Exclusion of H2A.Z is seen at sites of DNA methylation in the bodies of actively transcribed genes and in methylated transposons. Mutation of the MET1 DNA methyltransferase, which causes both losses and gains of DNA methylation (Goll and Bestor 2005; Gehring and Henikoff 2007), engenders opposite changes in H2A.Z deposition. Our findings indicate that DNA methylation can influence chromatin structure and effect gene silencing by excluding H2A.Z.

2.2 Results

2.2.1 Genome-wide mapping in the model plant *Arabidopsis thaliana* reveals an anticorrelation between H2A.Z and DNA methylation

To investigate H2A.Z deposition in plant chromatin, we generated a high resolution genome-wide map of H2A.Z in Arabidopsis by adapting the *in vivo* biotinylation system we used to affinity-purify Arabidopsis chromatin (Mito, Henikoff et al. 2005). We tagged Arabidopsis H2A.Z with a peptide specifically recognized by the *E. coli* biotin ligase BirA (biotin ligase recognition peptide, BLRP), and created transgenic plants co-expressing BLRP-H2A.Z with BirA. Cytological localization revealed that BLRP-H2A.Z has a diffuse nuclear distribution, but is excluded from heterochromatic chromocenters (Fig. 1abc), the same pattern as that of endogenous H2A.Z (Deal, Topp et al. 2007). Following digestion with micrococcal nuclease to mostly mononucleosomes (Fig. 1de), we purified biotinylated chromatin from root tissue and co-hybridized the associated DNA with control DNA on high resolution microarrays

representing the entire Arabidopsis genome (Zilberman, Gehring et al. 2007). To ensure that our results were not influenced by potential tagging artifacts, we repeated the experiment with antibodies against endogenous H2A.Z(Deal, Topp et al. 2007). We also mapped DNA methylation in roots (we have previously published a dataset was from aerial tissues(Zilberman, Gehring et al. 2007)).

The maps generated by streptavidin pull-down and immunoprecipitation were virtually the same (Fig. 2bc). The most striking feature was a strong, quantitative anticorrelation with DNA methylation (Pearson's r = -0.81). Distinct peaks of H2A.Z around the 5' ends of genes were also evident (Fig. 2b). To better visualize the H2A.Z distribution, we aligned all Arabidopsis annotated sequences, which include genes, pseudogenes, and transposable elements, at the 5' end, and stacked them from the top of chromosome 1 to the bottom of chromosome 5 (Fig. 3a). An obvious feature of this alignment is a vertical strip of high H2A.Z that roughly corresponds to the first nucleosome following the start of transcription. This pattern of H2A.Z deposition is consistent with those in yeast and humans (Guillemette, Bataille et al. 2005; Li, Pattenden et al. 2005; Raisner, Hartley et al. 2005; Zhang, Roberts et al. 2005; Millar, Xu et al. 2006; Barski, Cuddapah et al. 2007), indicating that this is a general feature of eukaryotic genes. There were also five conspicuous horizontal stripes of low H2A.Z incorporation. These correspond to transposon-rich, heavily methylated heterochromatin surrounding the five Arabidopsis centromeres. This pattern of incorporation is precisely the opposite of that of DNA methylation (Fig. 3b).

Methylation is not distributed evenly within the genome. Transposons are heavily and uniformly methylated, whereas some genes have short stretches of methylation, and most none at all (Zhang, Yazaki et al. 2006; Vaughn, Tanurd Ic et al. 2007; Zilberman, Gehring et al. 2007; Cokus, Feng et al. 2008; Lister, O'Malley et al. 2008). These three groups of sequences display a corresponding triphasic distribution of H2A.Z signal: low H2A.Z levels are found in transposons, intermediate levels in methylated genes, and high levels in unmethylated genes (Fig. 4). One possibility is that the low levels of H2A.Z in transposons are caused by intrinsic sequence preferences, rather than DNA methylation. To test this, we examined the tiny fraction (49) of Arabidopsis transposons that are not methylated (Table 1). Tellingly, all such transposons had high H2A.Z levels, indicating that low H2A.Z incorporation is not a feature of transposons per se (Fig. 2c and 3c-d). Unmethylated transposons also lacked any discernible H2A.Z peaks, suggesting that these are unique features of endogenous genes. Unsupervised *k*-means clustering of annotated Arabidopsis sequences based on H2A.Z patterns produced three groups that closely correspond to unmethylated genes, body-methylated genes and transposons (Fig. 5a, Table 2). Again, H2A.Z and DNA methylation levels showed a striking anticorrelation (Fig. 5b).

There are many histone modifications, such as acetylation and methylation of lysine 4 of H3, that are directly correlated with transcription (Bhaumik, Smith et al. 2007). Such modifications would be absent from transcriptionally inactive transposable elements, and therefore would be anticorrelated with DNA methylation. Similarly, there are histone modifications that mark silent chromatin, particularly dimethylation of lysine

9 of H3 (H3K9me2), which is highly enriched in Arabidopsis transposons (Lippman, Gendrel et al. 2004; Turck, Roudier et al. 2007) and is thus correlated with DNA methylation (Pearson's *r* = 0.63) and anticorrelated with H2A.Z (Pearson's *r* = -0.68, Supplementary Table 1). An important question is whether H2A.Z is simply excluded from silent chromatin, or has a more specific correlation with DNA methylation. Unlike H3K9me2, DNA methylation is abundant in bodies of actively transcribed genes in Arabidopsis, with methylated genes actually expressed at a higher level, on average, than unmethylated genes (Zhang, Yazaki et al. 2006; Zilberman, Gehring et al. 2007). Therefore, if H2A.Z is only anticorrelated with silent chromatin, the correlation with DNA methylation should break down in genes. However, our data show that this is not the case: H2A.Z is selectively depleted from methylated genes (Pearson's *r* = -0.63, Fig. 5ab). A scatter plot comparing H2A.Z and DNA methylation shows the same linear trend in genes and transposable elements (Fig. 6a). On the other hand, the correlation between DNA methylation and H3K9me2 in genes is, at best, very low (Pearson's *r* = 0.1), and there is no correlation in genes between H2A.Z and H3K9me2 (Pearson's *r* = -0.04). A scatter plot comparing H2A.Z and H3K9me2 shows two distinct groups: genes, which are low in H3K9me2 and have variable H2A.Z, and transposons, which are high in H3K9me2 and low in H2A.Z (Fig. 6b). Thus the overall anticorrelation between H2A.Z and H3K9me2 exists solely because they are found in active and silent chromatin, respectively, whereas DNA methylation and H2A.Z are mutually exclusive chromatin features independent of sequence context or transcription potential.

2.2.2 H2A.Z, DNA methylation and transcription

Because the likelihood of DNA methylation within a gene varies with transcription (Zhang, Yazaki et al. 2006; Zilberman, Gehring et al. 2007), as do H2A.Z levels in yeast and humans(Guillemette, Bataille et al. 2005; Li, Pattenden et al. 2005; Zhang, Roberts et al. 2005; Millar, Xu et al. 2006; Barski, Cuddapah et al. 2007), we examined the relationship between methylation, H2A.Z and transcription. To obtain average profiles, we aligned genes with annotated 5' and 3' UTRs at their 5' ends, grouped them into deciles by transcription level based on published root microarray data (Schmid, Davison et al. 2005), and further subdivided each decile into methylated and unmethylated genes (Fig. 7a). We also examined H2A.Z enrichment within each transcription percentile: averages were calculated for gene bodies starting from 1.5 kb downstream of the start of transcription (gene body; Fig. 7b), as well as for the first 500 bp of genes (H2A.Z peak; Fig.8 and 9). These analyses show that methylated genes have much less H2A.Z within the gene body than unmethylated genes (Fig.7ab). Both gene types show a robust negative correlation between H2A.Z and transcription with essentially identical rate of change (slope; Fig. 7b), indicating that the influences of methylation and transcription are independent and additive: methylation correlates with strong exclusion of H2A.Z, whereas transcription correlates with weaker exclusion. These findings are in agreement with earlier results showing that transcriptional activation of the *FLC* gene corresponds with decreased H2A.Z (Deal, Topp et al. 2007).

Most methylated and unmethylated genes have a robust 5' H2A.Z peak, regardless of expression level, with the exception of the least transcribed methylated

genes, which have much reduced peaks (Fig. 7a, 9ab). For unmethylated genes, H2A.Z enrichment within the first 500 bp follows a roughly parabolic distribution, with H2A.Z prevalence increasing up to about the 50th percentile, then gradually decreasing (Fig. 8a). This is consistent with the H2A.Z distribution observed in human promoters (Barski, Cuddapah et al. 2007). Below the 40th percentile, there is a strong positive correlation between expression and H2A.Z in both methylated and unmethylated genes, but the rate of change is much greater for methylated genes (Fig.8ab and 9ab). Because the first 500 bp represents a small subset of genic sequence, the overall trend is for H2A.Z to decline with increasing transcription (Fig. 8c). DNA methylation of the first 500 bp shows a pattern opposite to that of H2A.Z, with enrichment in the least transcribed genes declining linearly through the 40th percentile (Fig. 9c). One possible explanation is that, although only genes with annotated UTRs were included, some of the least transcribed methylated genes might be, or are in the process of becoming, pseudogenes. Another possibility is that some of these genes are generally repressed by methylation, but can become activated by demethylation in specific tissues. Such behavior is exhibited by the imprinted *FWA* gene, which is silenced in most tissues by methylation of the promoter and the start of transcription. Methylation is selectively removed in the endosperm, activating the gene(Kinoshita, Miura et al. 2004). *FWA* lacks a discernable 5' H2A.Z peak when repressed, but acquires one upon demethylation (Fig. 10a). Regardless of the reason for the relative methylation enrichment in the first 500 bp of the least transcribed genes, H2A.Z incorporation is apparently incompatible with DNA methylation, even in a region where H2A.Z is usually most abundant. Importantly, our data show that DNA methylation does not exclude H2A.Z through transcriptional repression, both because methylated genes have much less H2A.Z than unmethylated genes that are transcribed at the same rate (Fig. 7 and 9), and because such a model would predict that H2A.Z incorporation should increase with transcription, whereas the opposite is the case.

2.2.3 DNA methylation acts to exclude H2A.Z from chromatin

So far, our results indicate a strong anticorrelation between methylation and H2A.Z deposition, but we cannot distinguish which is causal. In order to address this issue, we took advantage of a line bearing a null mutation in the DNA methyltransferase *MET1, met1-6* (Xiao, Gehring et al. 2003; Goll and Bestor 2005). Mutations in *MET1* cause major reductions in overall DNA methylation, but also significant hypermethylation mediated by other methyltransferases (Lister, O'Malley et al. 2008). We reasoned that if DNA methylation influences H2A.Z deposition, changes in DNA methylation should be mirrored by changes in H2A.Z distribution. Notably, because *met1* causes both losses and gains of DNA methylation, we should see both gains and losses of H2A.Z. To test our hypothesis, we mapped H2A.Z, as well as DNA methylation and transcription, in *met1-6* plants.

Changes in DNA methylation indeed engendered changes in H2A.Z distribution (Fig. 10 and 11). To visualize these changes, we subtracted the wild type (WT) H2A.Z dataset from the *met1* H2A.Z dataset, so that high values represent increased H2A.Z incorporation in *met1* (Fig. 11). Examples of informative loci are shown in Fig. 10a-c.

As mentioned above, the *FWA* gene, which normally has 5' methylation and lacks an H2A.Z peak, loses promoter methylation and gains 5' H2A.Z in *met1* (Fig. 10a). The retrotransposon *At5g13205* is heavily methylated in WT, but loses methylation and gains H2A.Z in *met1* (Fig. 10b). Gene *At1g22000*, which encodes an F-box protein, is hypermethylated in *met1*, leading to loss of its 5' H2A.Z peak (Fig. 10c).

To get a comprehensive view of H2A.Z dynamics in *met1-6*, we aligned and arranged all annotated Arabidopsis sequences as in Fig. 3a. The same conspicuous pericentric stripes were evident in this profile (Fig. 12a) – H2A.Z levels are elevated in transposable elements, which lose most of their methylation and become reactivated in *met1*(Zhang, Yazaki et al. 2006; Zilberman, Gehring et al. 2007). Unbiased sorting of the data produced three clusters that roughly encompass unmethylated genes, methylated genes, and transposons, respectively (Fig. 12b, and Table 2, sequences are categorized as in(Zilberman, Gehring et al. 2007)). The changes in H2A.Z closely correspond to DNA methylation – sequences that gain H2A.Z are methylated in WT (Fig. 12c). Conversely, loci with decreased H2A.Z incorporation are unmethylated in WT, but methylated in *met1-6* (Fig. 13). There were relatively few such sequences (Fig. 11), reflecting the sporadic nature of *met1*-induced hypermethylation. Overall, changes in DNA methylation were mirrored by changes in H2A.Z in a manner that strongly argues that methylation inhibits H2A.Z incorporation.

Because some transposons and genes undergo transcriptional upregulation in *met1* plants (Zilberman, Gehring et al. 2007), we had an opportunity to test whether H2A.Z incorporation is negatively influenced by methylation or positively influenced by transcription. Within genes, there is a robust correlation between DNA methylation in WT and H2A.Z changes in *met1-6* (average Pearson's $r = 0.51$), but there is no correlation between transcriptional and H2A.Z changes (average Pearson's $r = 0.05$). *FWA*, which is strongly overexpressed in *met1*, has reduced levels of H2A.Z in the body of the gene, where it has no methylation in WT (Fig. 10a). Similarly, of the handful of transposons that are not methylated in wild type, two (*At4g10690* and *At5g35205*) are nevertheless upregulated in *met1* (Fig. 14). Both also have less H2A.Z in *met1* than in wild type, the opposite of other transposons.

Because only about half of all transposable elements are upregulated in *met1*, we could ask whether those elements preferentially gain H2A.Z, as would be expected if H2A.Z incorporation was associated with transcriptional activity. To ensure that the size of the datasets and methylation are not an issue, we compared 12,500 probes that represent each transposon class and have identical methylation profiles. We find that both transposon classes are equally enriched in H2A.Z (Fig. 15abc). Thus, changes in DNA methylation, rather than transcription, cause the redistribution of H2A.Z we observe in *met1*.

2.2.4 Loss of PIE1 leads to genome-wide changes in DNA methylation

Our results show that DNA methylation excludes H2A.Z. An intriguing question is whether H2A.Z can also exclude methylation. Some of our data suggest that this is indeed the case. The most striking feature of H2A.Z incorporation, the 5' genic peak, is

independent of DNA methylation, yet methylation is strongly excluded from precisely this area (Zhang, Yazaki et al. 2006; Zilberman, Gehring et al. 2007) (Fig. 2). Likewise, the higher H2A.Z levels in the bodies of less-transcribed genes might explain the puzzling observation that the chances of a gene becoming methylated increase with transcription (up to about the 70^{th} percentile) (Zhang, Yazaki et al. 2006; Zilberman, Gehring et al. 2007).

To address this issue, we mapped DNA methylation in plants with a strong loss-of-function allele of *PIE1* (the conserved catalytic component of Swr1) that disrupts proper deposition of H2A.Z (Deal, Topp et al. 2007). The overall methylation pattern in *pie1-5* plants remained similar to WT (Table 3), but there was a modest but consistent increase in DNA methylation (Fig. 16ab). To visualize the methylation changes in *pie1* we subtracted the methylation patterns of matched WT controls (F2 sibs) from *pie1* and displayed the resulting data as a heatmap (Fig. 16b). This analysis revealed genome-wide hypermethylation of gene bodies. Using the ChIPOTle algorithm (Buck, Nobel et al. 2005), we identified 1201 hypermethylated regions (corresponding to 1172 genes) for further analysis (threshold $p<10^{-7}$, Table 4).

In plants DNA methylation can occur at any cytosine (Gehring and Henikoff 2007). Most methylation is found in symmetric CG sites, like it is in animals, and is mediated by MET1, but there is also a substantial amount of methylation in other sequence contexts catalyzed by other methyltransferases (hence the hypermethylation observed in *met1*) (Cokus, Feng et al. 2008; Lister, O'Malley et al. 2008). To determine how the *pie1* mutation affects DNA methylation in different contexts, we used bisulfite sequencing to analyze the methylation of individual cytosines in five loci scored as hypermethylated by ChIPOTle: *At1g69850* (a nitrate transporter), *At3g22340* (a COPIA-like retrotransposon), *At4g03480* (an ankyrin repeat containing protein), *At4g38190* (a cellulose synthase) and *At5g37450* (a protein kinase). All five showed a modest but consistent gain of CG methylation (Fig. 17ab), confirming the microarray analysis. There was very little non-CG methylation at any of the loci in either WT or *pie1* (data not shown). Interestingly, all of the loci had some methylation in WT, so the overall genomic hypermethylation we observe in *pie1* is likely to be primarily caused by increased methylation of normally lightly methylated loci rather than *de novo* methylation of previously unmethylated loci.

Given the wide-spread hypermethylation caused by the *pie1* mutation, we asked whether the hypermethylated loci are representative of the genome as a whole. As might be expected, *pie1* hypermethylated genes have high levels of H2A.Z in WT (i.e. those generally found in unmethylated genes; Fig. 18a). They are also generally enriched in low transcribed genes, with greatest enrichment around the 30^{th} transcription percentile (Fig. 18b). This pattern is very different from that of normally methylated genes, which are most prevalent around the 70^{th} percentile (Fig. 18b), and is also unlike unmethylated genes, which are enriched in both low and highly expressed genes (Zilberman, Gehring et al. 2007). *pie1* hypermethylated genes do, however, closely parallel the overall distribution of H2A.Z (Fig. 18b). These loci also include 17 of the 49 transposons that are enriched in H2A.Z and unmethylated in WT (Table 1, 2), a

10 fold overrepresentation ($p=10^{-4}$, Fisher's exact test). Thus sequences that are generally preferred targets of DNA methylation (gene bodies and transposons) are hypermethylated in *pie1*, consistent with the presence of low levels of DNA methylation in these sequences in WT (Fig 17ab). As described in the following chapter, further experiments that utilized an *h2a.z* triple mutant yielded data that on the surface contradicts these results, and in retrospect we believe that the relative increases in genic methylation described above are in fact relative decreases in TE methylation, a phenomena also seen in the *h2a.z* mutant data.

2.3 Discussion

How methylation silences genes has been a vexing question for decades. A popular model is that proteins that bind to methylated DNA engender silencing by recruiting histone deacetylases (Klose and Bird 2006). However, careful gene disruption studies in mice have shown that these proteins are unlikely to fully account for methylation-induced repression (Guy, Hendrich et al. 2001; Hendrich, Guy et al. 2001). Previous work has provided strong evidence that H2A.Z contributes to promoter competence (Meneghini, Wu et al. 2003; Updike and Mango 2006; Brickner, Cajigas et al. 2007; Deal, Topp et al. 2007). Therefore, exclusion of H2A.Z would represent a novel mechanism of gene silencing by DNA methylation. Given that DNA methylation and H2A.Z are both ancient chromatin components, their interaction likely plays an important general role in regulating eukaryotic gene expression.

2.4 Materials and Methods

Transgenic lines. We adapted the biotin-mediated affinity purification system we developed in Drosophila tissue culture cells(Mito, Henikoff et al. 2005) to allow protein purification from Arabidopsis plants. We constructed a binary plasmid that contained the *E. coli* biotin ligase, BirA, driven by the Arabidopsis *ACTIN2* promoter, and the Arabidopsis H2A.Z gene *At1g52740* driven by its endogenous promoter and tagged at the N-terminus with the biotin ligase recognition peptide (BLRP). BLRP is a high affinity substrate for BirA, which biotinylates a lysine residue within the peptide. We sent the plasmid to the UC Riverside Plant Transformation Research Center (http://www.ptrc.ucr.edu), where transgenic Arabidopsis lines were created by vacuum infiltration in ecotype Columbia.

Affinity purification. About 100 seeds were sterilized in 20% bleach and 0.5% Tween-20 for 10 minutes. Seeds were germinated in 300 ml of Gamborg's B-5 medium supplemented with 5 mM biotin, and roots harvested after four weeks. 4 grams of roots were ground in liquid nitrogen to a fine powder, suspended in 20 ml of modified Honda buffer (25 mM Tris, pH 7.6, 0.44 M sucrose, 10 mM $MgCl_2$, 2 mM spermine, 0.1% Triton X-100, 10 mM beta-mercaptoethanol) and homogenized with a tissue homogenizer. The homogenate was filtered through miracloth, transferred to a 30 ml round bottom glass tube, and spun at 4000 rpm (2000g), 4C, in an SS-34 rotor for 10 min. The pellet was resuspended in Honda buffer B (Honda buffer minus spermine), spun in a

microcentrifuge at 1500 rpm (200g), 4C, for 2 min, and resuspended in 1 ml of TNE (10 mM Tris, pH 8.0, 100 mM NaCl, 1 mM EDTA). The suspension was warmed to 37C and digested with micrococcal nuclease in the presence of 4 mM $CaCl_2$ (Fig. 1) to liberate nucleosomes. The reaction was stopped with 25 mM EDTA and spun at high speed in a microcentrifuge for 5 min at 4C. Biotinylated proteins were purified from the supernatant as described(Mito, Henikoff et al. 2005). Endogenous H2A.Z was immunopurified as described(Deal, Topp et al. 2007), except the IP was performed in TNE. The antibodies are predicted to cross-react with all three Arabidopsis H2A.Z proteins(Deal, Topp et al. 2007).

Microarray analysis. Our microarray design is described in(Zilberman, Gehring et al. 2007). We analyzed DNA methylation in five independent samples from the Columbia ecotype: two from wild type roots (WT root-1 and root-2), two from *met1-6* roots (*met1-6* root-1 and root-2), and one from *met1-6* aerial tissues. We followed our protocol, as described in(Zilberman, Gehring et al. 2007), except we omitted the T7 RNA polymerase-mediated amplification step for all samples except aerial *met1-6*. Instead, sufficient amplification was achieved in the labeling step. We also utilized our WT aerial methylation data published in(Zilberman, Gehring et al. 2007).

For *pie1* methylation analysis, we mapped methylation in three *pie1* replicates and three matched WT controls (F2 sibs). DNA was extracted from tissue collected from >100 whole 10 day seedlings to eliminate the possibility of detecting random variations in DNA methylation(Vaughn, Tanurd Ic et al. 2007). The samples were amplified with the Sigma WGA2 kit before labeling. For ChIPOTle analysis(Buck, Nobel et al. 2005), outliers were removed from each dataset by median smoothing (3 probe window), the three *pie1* and WT datasets were averaged, WT was subtracted from *pie1*, and the resulting dataset was smoothed (triangular smoothing, $y = 0.25(x_{n-1}) + 0.5(x_n) + 0.25(x_{n+1})$) and normalized to a mean of zero. We removed the 270 kb mitochondrial DNA insertion on chromosome 2 before analysis. 1201 peaks were called with a conservative threshold of $p<10^{-7}$. As a control, we determined the number of 'negative' peaks that would represent hypomethylation: only 53 peaks were called. Even assuming the unlikely scenario that all the negative peaks are false positives, the false positive rate would be 4%.

We assayed H2A.Z in 8 samples: BLRP-1 and BLRP-2 were from one transgenic line, BLRP-3 and BLRP-4 from an independent transgenic line, WT Ab-1 and *met1-6* Ab-1 were paired immunoprecipitation experiments from WT and *met1-6* roots, respectively, and WT Ab-2 and *met1-6* Ab-2 were a second set of paired experiments. All samples except BLRP-2, BLRP-3 and BLRP-4 were T7 RNA polymerase-amplified(Zilberman, Gehring et al. 2007), the rest were sufficiently amplified in the labeling step.

Expression analysis of two independent *met1-6* RNA samples (paired with two independent wild type samples) was carried out as described in(Zilberman, Gehring et al. 2007), except random hexamers were used for cDNA synthesis instead of an oligo d(T) primer. All labeled samples were sent to NimbleGen Systems (Madison, WI) for

hybridization, except the *pie1* samples, which were hybridized at the FHCRC DNA array facility.

Bisulfite sequencing. 2 micrograms of genomic DNA for each sample were bisulfite-converted with the Qiagen EpiTect kit. PCR products were cloned with the Invitrogen PCR4 TOPO kit. Primer sequences are available upon request.

Figure 1

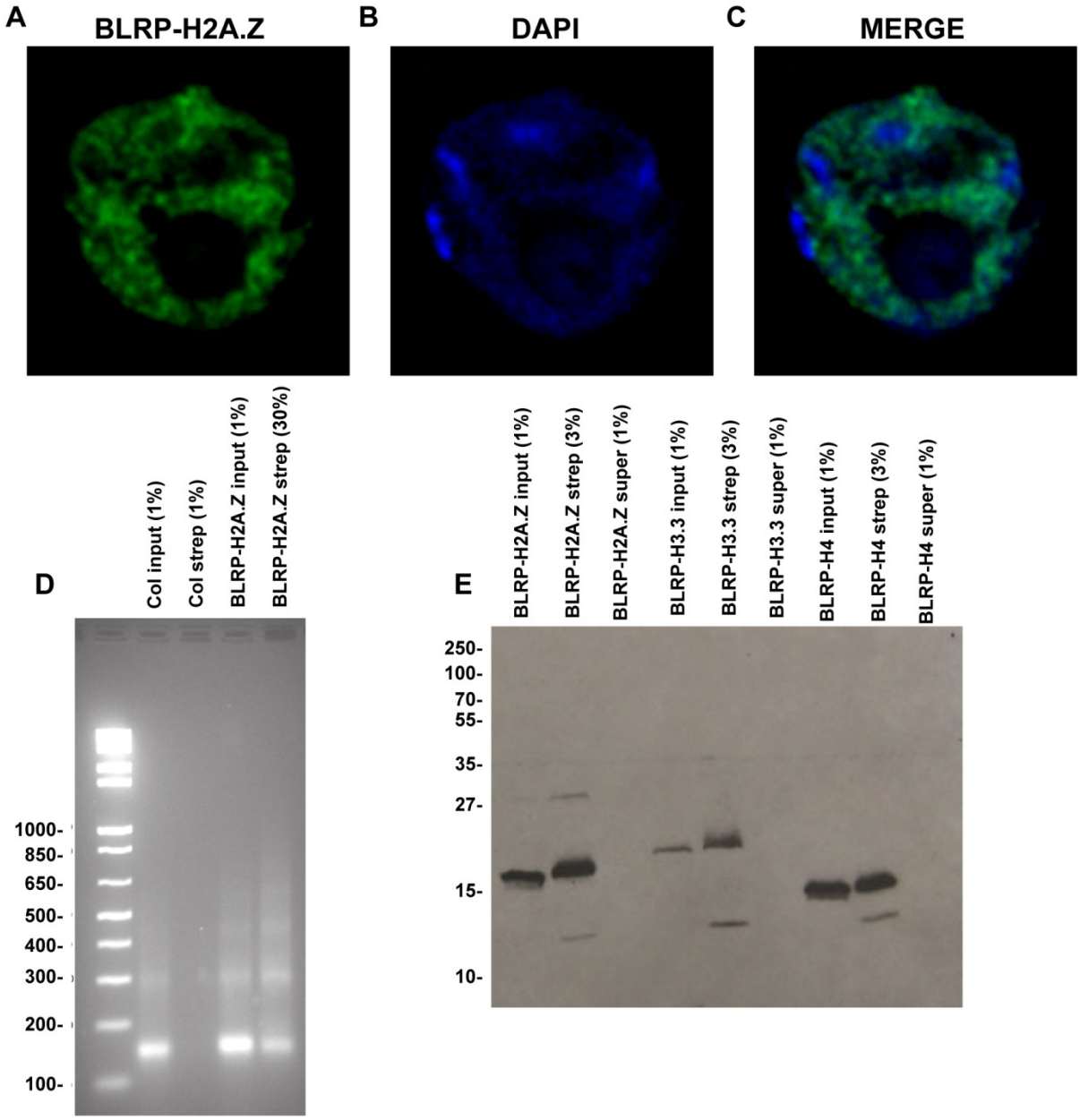

Figure 1. Purification of BLRP-tagged H2A.Z. a, BLRP-H2A.Z stained with streptavidin-conjugated Alexa Fluor 488. **b,** The same cell as in **a** stained with DAPI to visualize DNA. **c,** Merge of **a** and **b**. **d,** DNA from chromatin digested with micrococcal nuclease run on a 2% agarose gel stained with ethidium bromide. DNA markers are in the first lane, chromatin from wild type ecotype Columbia (Col) in the second, streptavidin-purified Col chromatin in the third, BLRP-H2A.Z chromatin in the fourth and streptavidin-purified BLRPH2A.Z chromatin in the fifth. **e,** Protein samples from isolated nuclei were run out on a 18% SDS-PAGE gel and treated with streptavidin-conjugated horse radish peroxidase to detect biotinylated proteins. BLRP-H2A.Z samples are in the first three lanes, BLRP-tagged histones H3.3 and H4 are shown as controls.

Figure 2

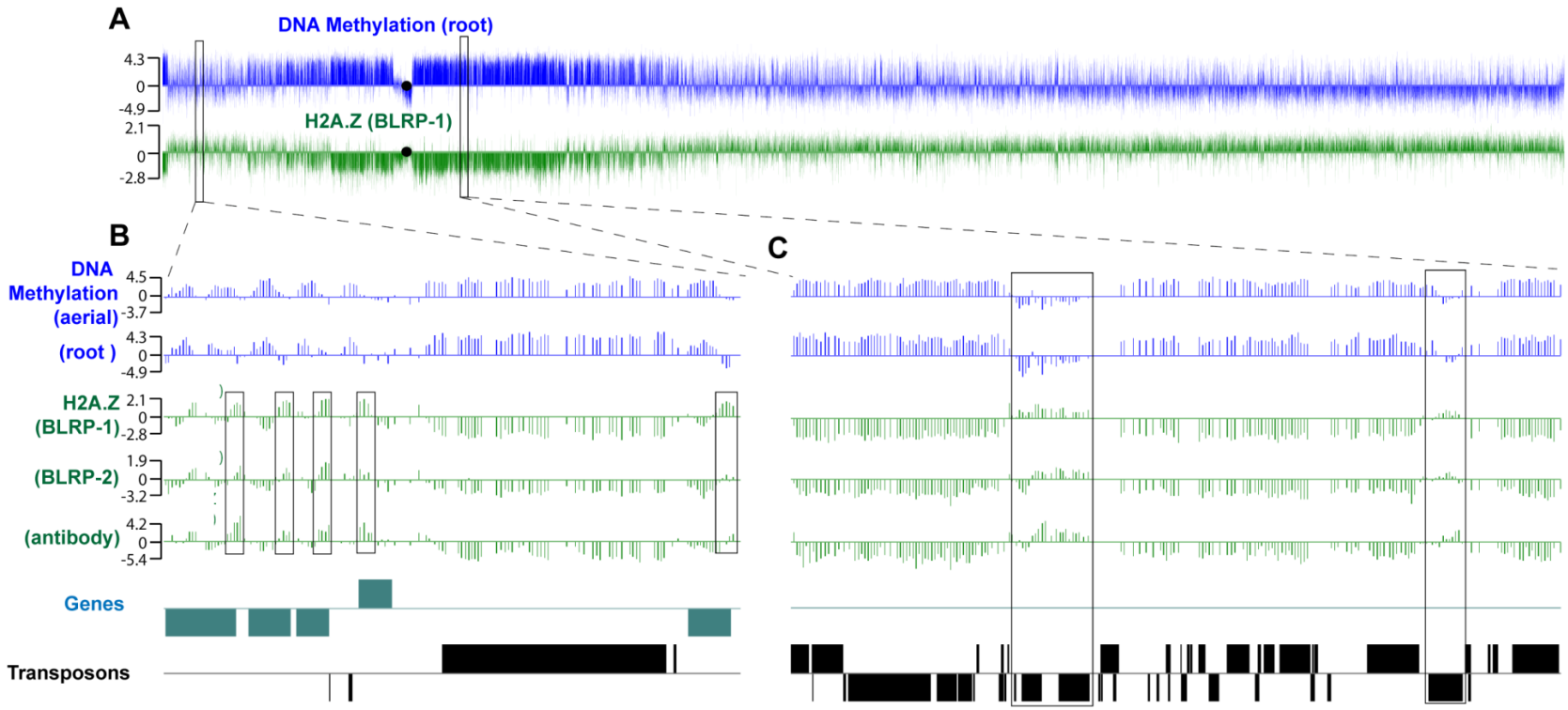

Figure 2. High resolution maps of Arabidopsis H2A.Z and DNA methylation. a, H2A.Z (green) and DNA methylation (blue) profiles of Arabidopsis chromosome 2. Each vertical bar represents the $\log_2$ signal ratio of the test sample signal divided by the input control signal ($\log_2$ (test/input)). The black circles denote the position of the centromeric sequence gap. **b-c,** More detailed views of a euchromatic (positions 547,000 – 587,000, **b**) and a heterochromatic (4,407,000 – 4,463,000, **c**) genomic region. DNA methylation from aerial tissues and roots is shown in blue; HA2.Z profiles obtained from two independent BLRP-H2A.Z transgenic lines and via immunoprecipitation of endogenous H2A.Z are shown in green. Genes and transposons on the top and the bottom strands are shown above and below the line, respectively. 5' peaks of H2A.Z in genes are emphasized by boxes in **b**. Unmethylated transposons with relatively high levels of H2A.Z are emphasized by boxes in **c**.

Figure 3

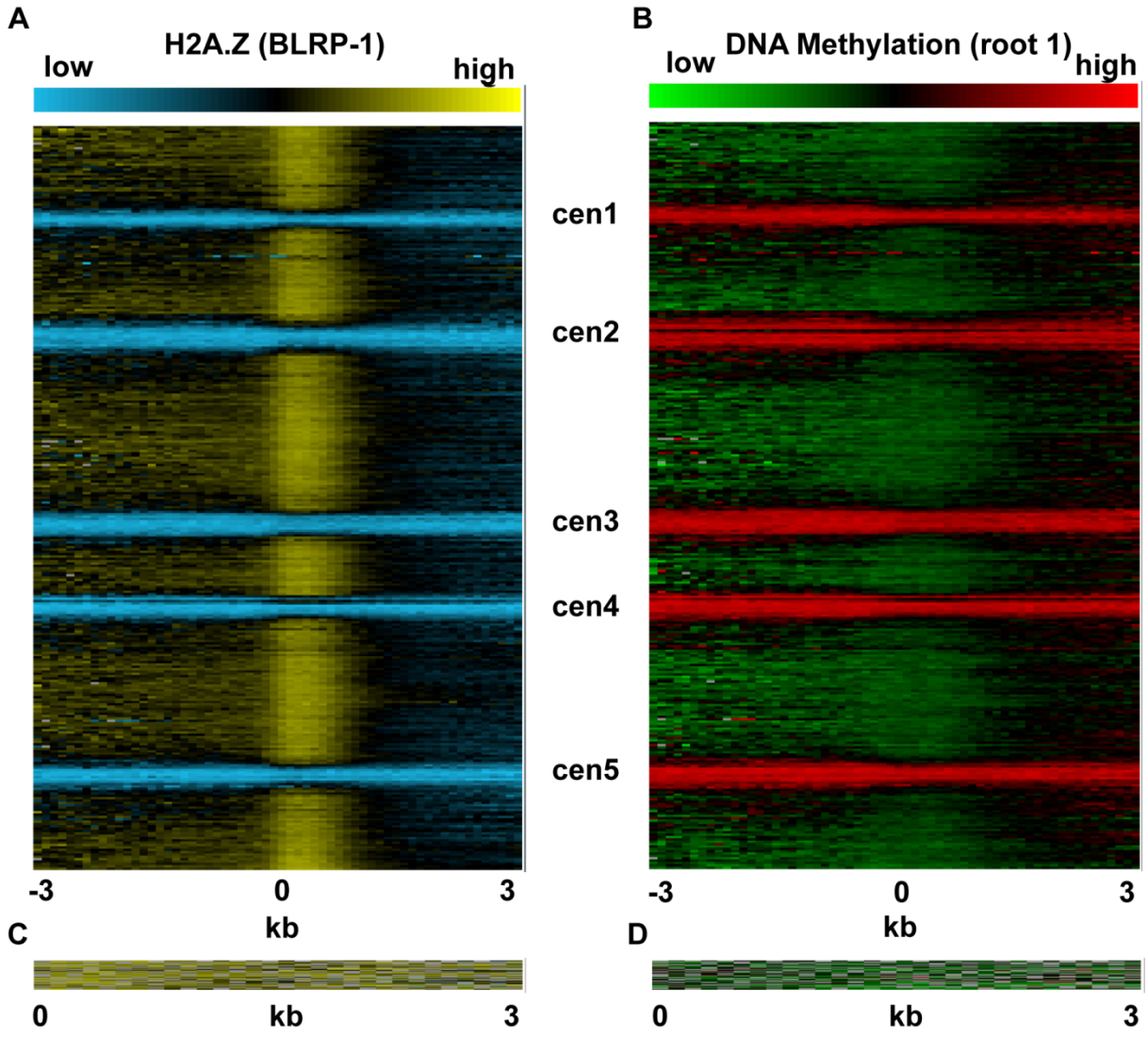

Figure 3. H2A.Z and DNA methylation heatmap distributions. a-b, All TAIR 7 annotated sequences (31,762) were aligned at the 5' end and stacked from the top of chromosome 1 to the bottom of chromosome 5. BLRP-H2A.Z is displayed as a heat map in **a**; root DNA methylation is displayed in **b**. Note the high degree of anticorrelation between H2A.Z and methylation. **c-d,** Unmethylated transposable elements (listed in Supplementary Table 3). BLRP-H2A.Z is displayed as a heat map in **c**; root DNA methylation is displayed in **d**.

Figure 4

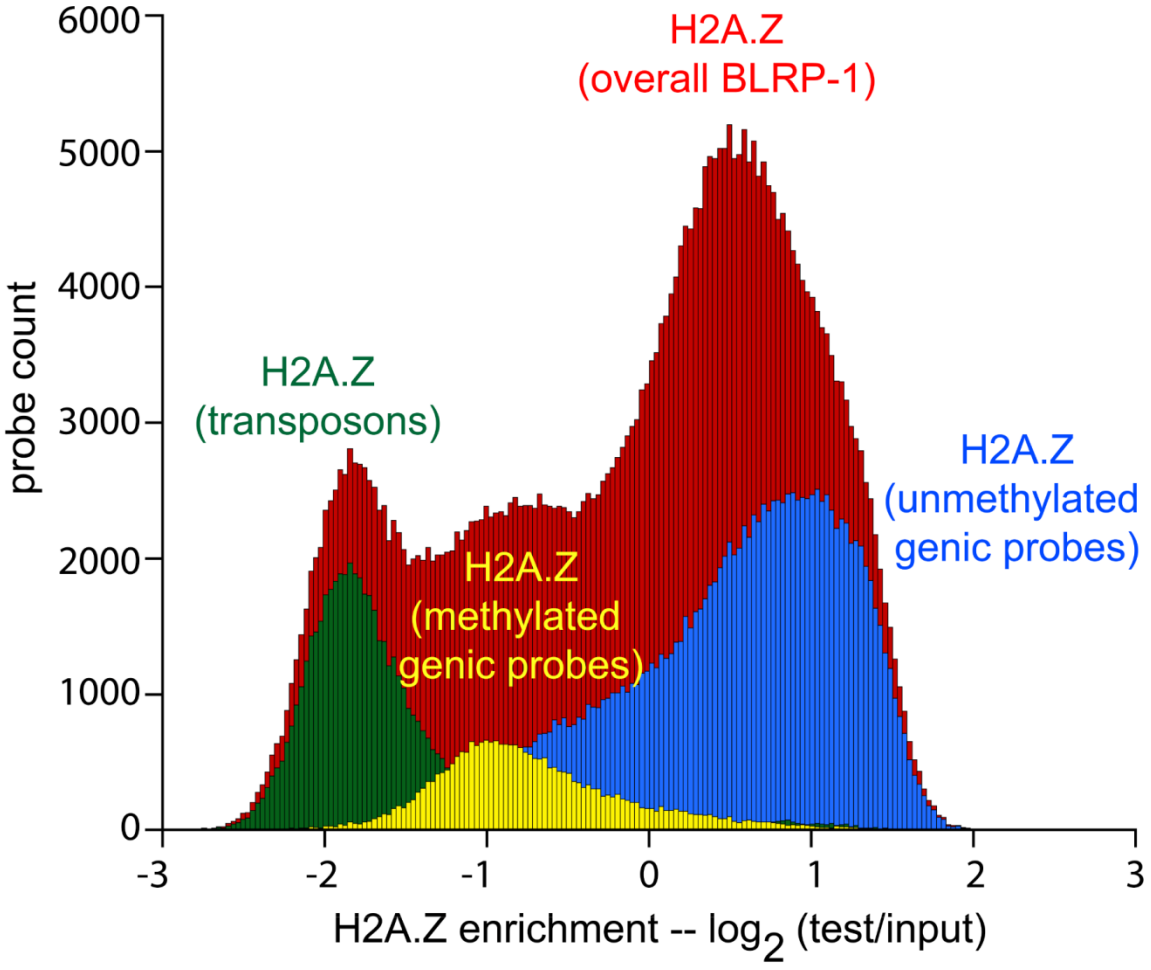

Figure 4. Histogram of H2A.Z by sequence context. Distribution of H2A.Z BLRP-1 signal scores: $\log_2$ (test/input) on the X-axis and probe counts on the Y-axis. Red bars represent the distribution of all probes on the array. Green bars represent the distribution of probes within transposable elements, yellow bars within methylated genes, and blue bars within unmethylated genes, as designated in[5].

Figure 5

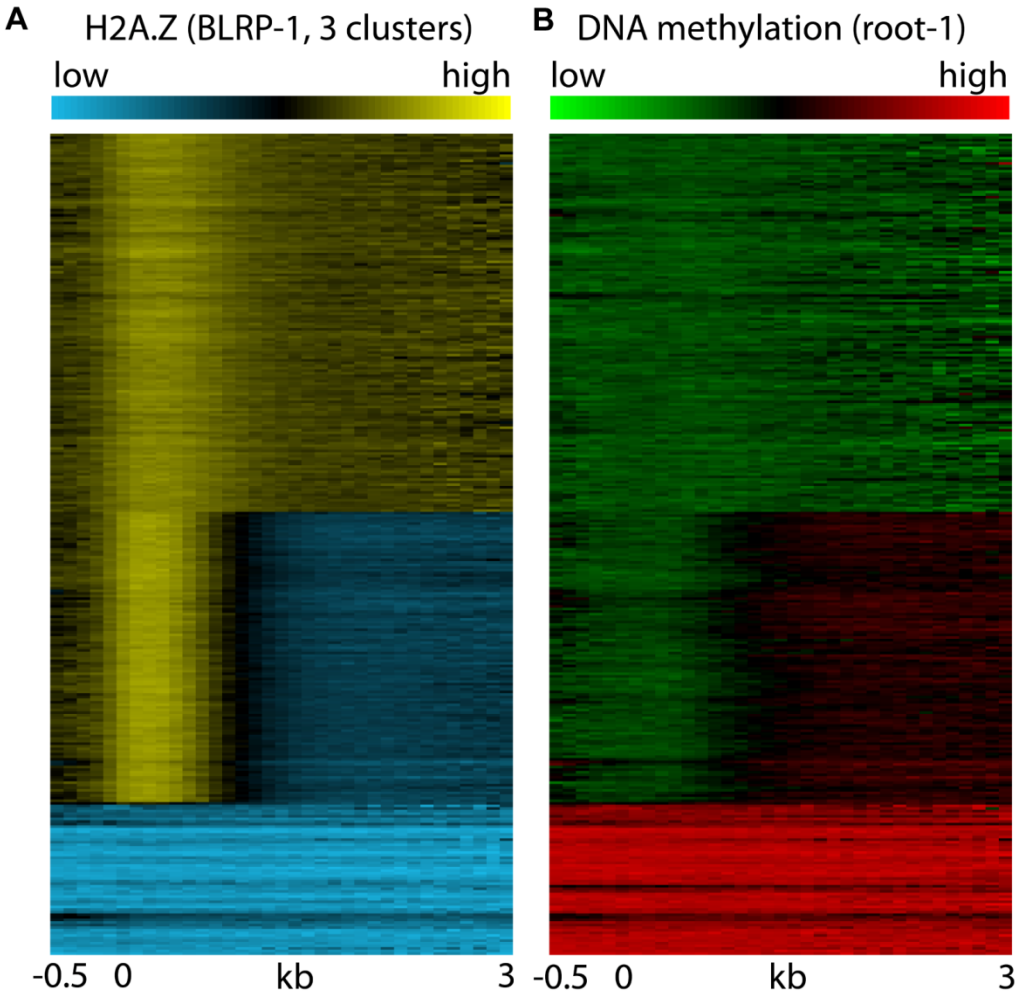

Figure 5. K-means clustering of H2A.Z and DNA methylation distribution. a, All TAIR 7 annotated sequences were *k*-means clustered (*k*=3) based on BLRP-H2A.Z patterns, and displayed as a heat map. For comparison, root DNA methylation of the same sequences is shown as a heat map in **b**.

Figure 6

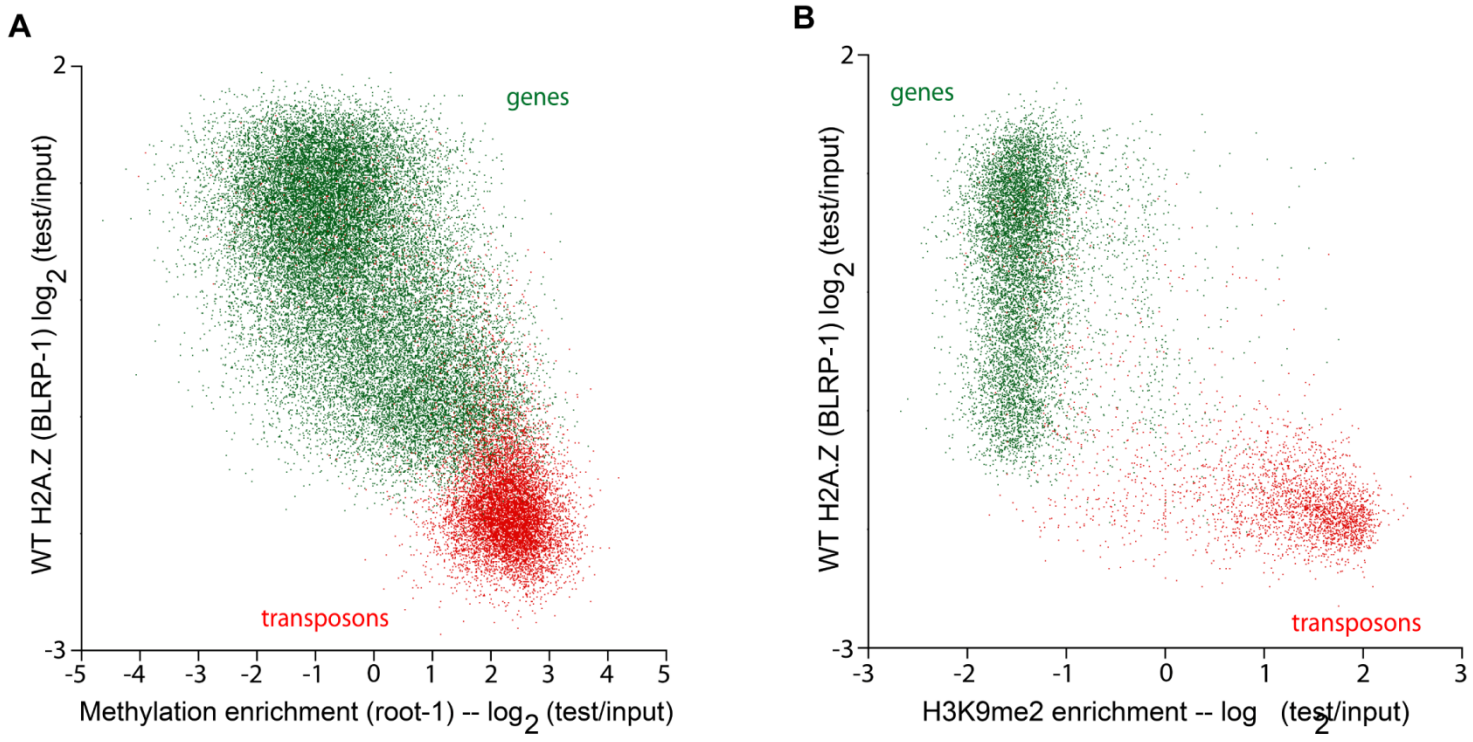

Figure 6. Scatter plots of H2A.Z, DNA methylation, and H3K9me2 in Genes and Transposons. a, Scatter plot showing the relationship between H2A.Z (BLRP-1, Y-axis) and DNA methylation (root-1, X-axis). Genes are shown in green and transposable elements in red. **b,** Scatter plot showing the relationship between H2A.Z (BLRP-1, Y-axis) and H3K9me2 data from the study of Turck et al. (X-axis). Genes are shown in green and transposable elements in red.

Figure 7

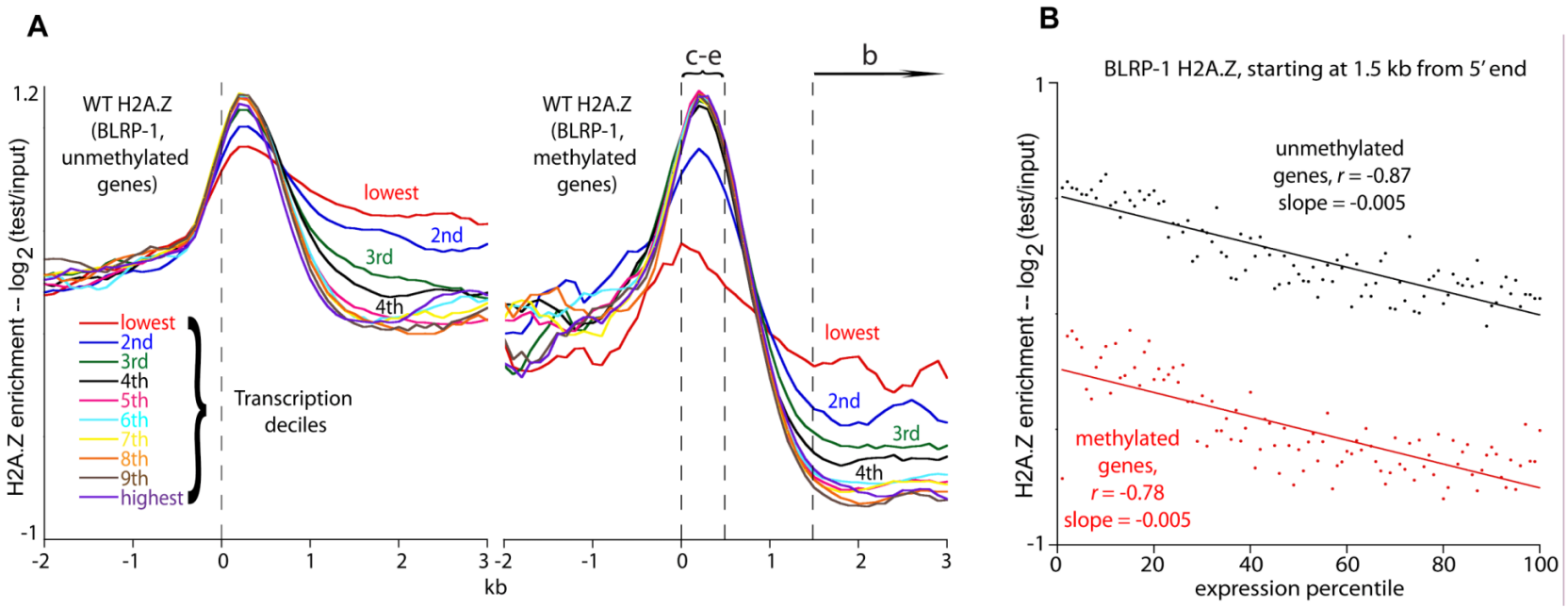

Figure 7. H2A.Z profiles across Genes binned by Transcription and Methylation status a, Genes with annotated 5' and 3' UTRs (15,743) were grouped into deciles based on transcription levels in roots, with each decile subdivided into unmethylated genes (left panel) and methylated genes (right panel). All genes were aligned at the 5' end and average scores for each 100-bp interval are plotted from 2 kb away from the gene (negative numbers) to 3 kb into the gene (positive numbers). The data were smoothed with a 5-point sliding window. The dashed line at zero represents the point of alignment. **b,** Genes were grouped into percentiles based on transcription level, with each percentile subdivided into unmethylated genes (black dots) and methylated genes (red dots). Average H2A.Z scores across gene bodies are plotted (starting at 1.5 kb from the 5' end.

Figure 8

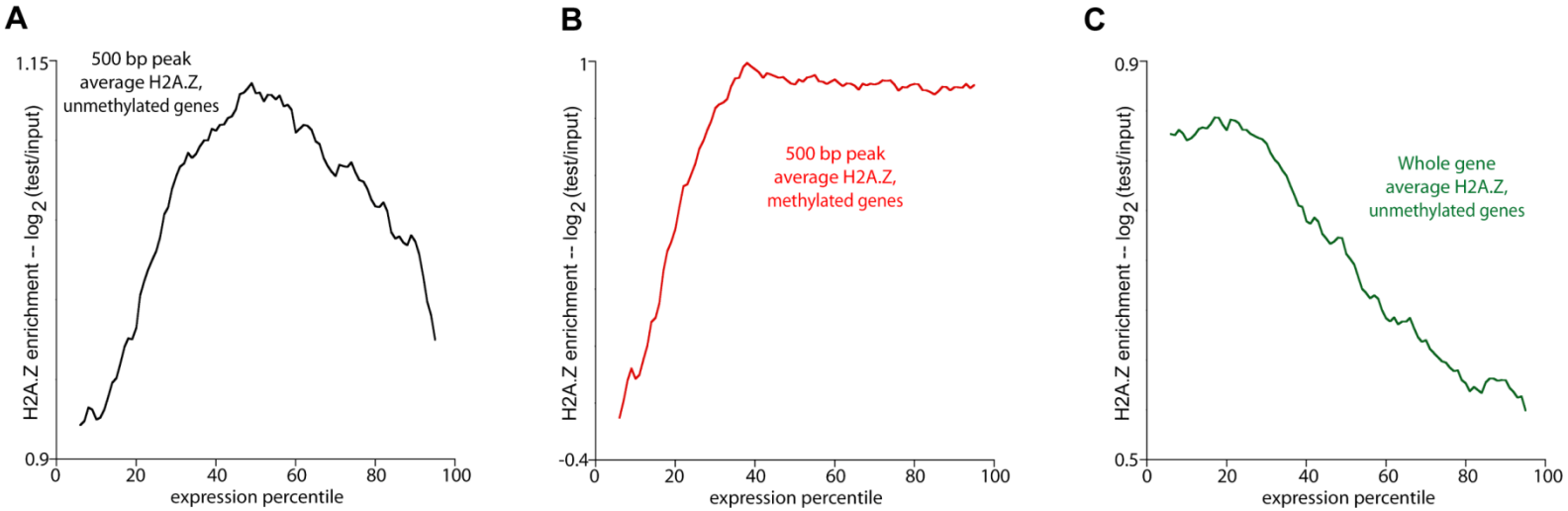

Figure 8. H2A.Z 5' peak profile across Transcription percentiles. a-c, All six H2A.Z datasets were averaged and the profiles for the first 500 bp of unmethylated genes (**a**), the first 500 bp of methylated genes (**b**) and the whole gene average of unmethylated genes (**c**) are plotted. The data were smoothed with a 10-point sliding window. Note the difference in scales between **a** and **b**.

Figure 9

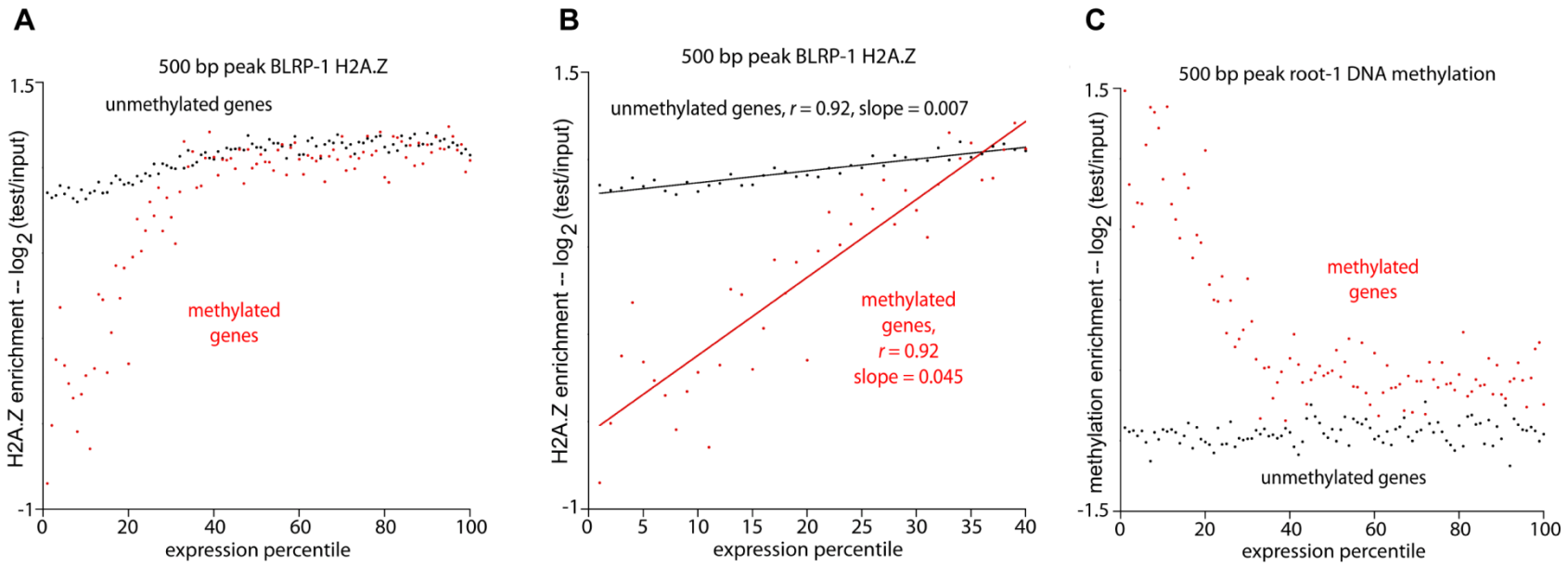

Figure 9. H2A.Z and DNA Methylation enrichment binned by Transcription percentile. a-c, Genes were grouped into percentiles based on transcription level, with each percentile subdivided into unmethylated genes (black dots) and methylated genes (red dots). Average H2A.Z (**a,b)** or DNA methylation(**c)** scores are plotted, as indicated in each panel, for the 500bp peak window indicated in Fig. 7a.

Figure 10

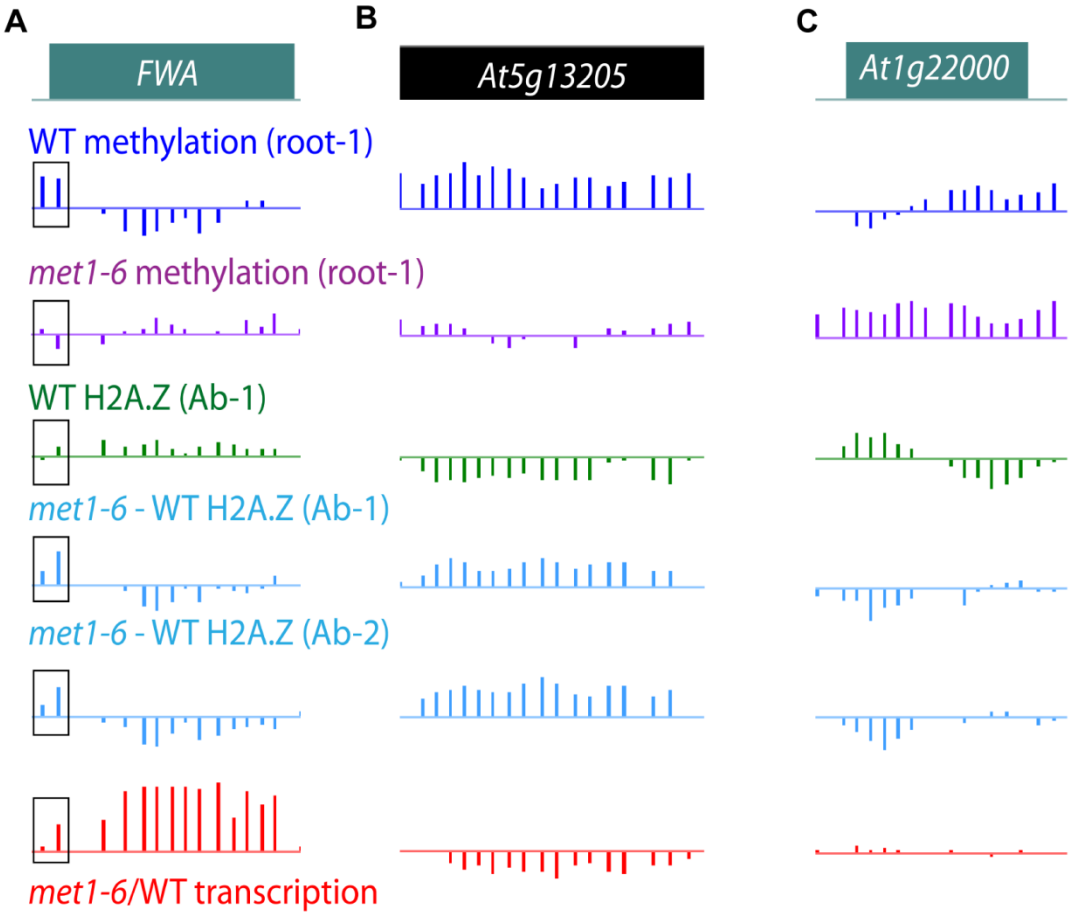

Figure 10. H2A.Z incorporation changes in *met1-6* for three loci. a-c, Wild type (WT) root DNA methylation (dark blue), *met1-6* root DNA methylation (purple), WT H2A.Z (antibody,green), WT H2A.Z profile subtracted from the *met1-6* H2A.Z profile (two sets of independent paired experiments, light blue), and *met1-6*/WT transcription (red) for *FWA* in **a**, copia-like transposable element *At5g13205* that loses methylation and gains H2A.Z in *met1-6* in **b**, and F box gene *At1g22000* that is hypermethylated and loses H2A.Z in *met1-6* in **c**. The 5'

Figure 11

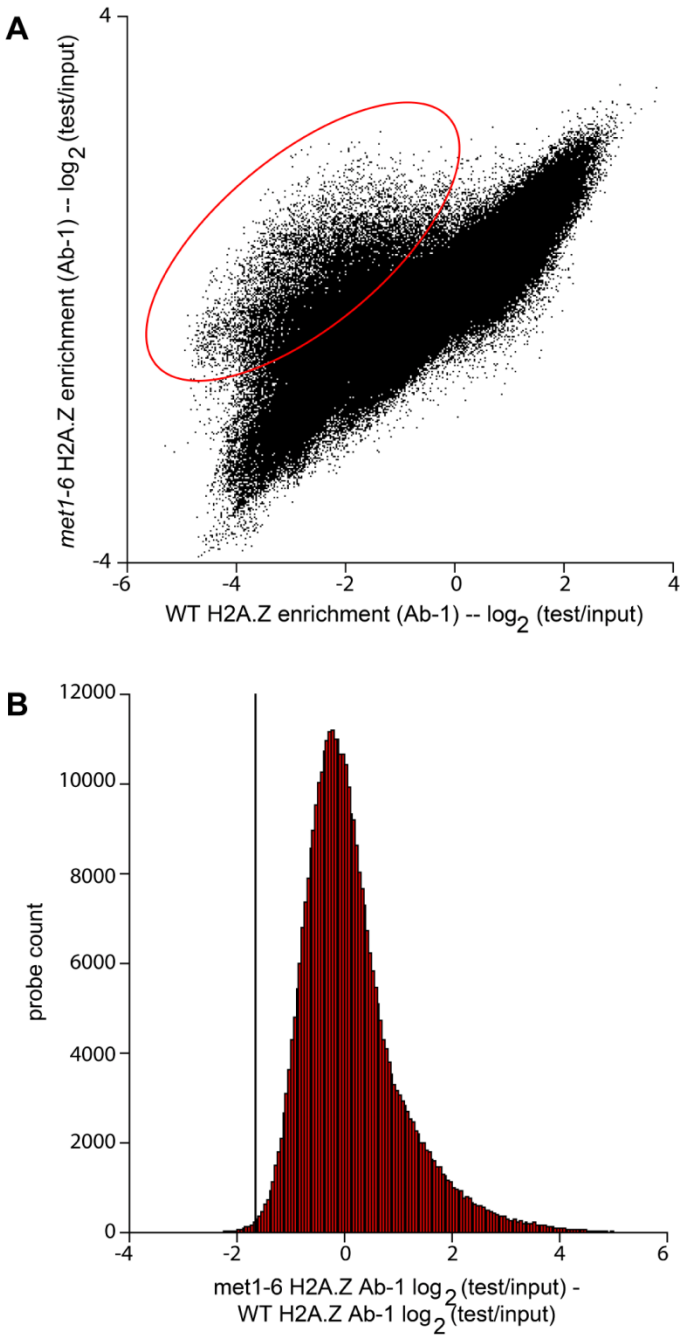

Figure 11. H2A.Z changes in *met1-6* mutant. a, Scatter plot showing the relationship between two paired microarray experiments: one assaying H2A.Z in wild type (WT) Arabidopsis roots (X-axis), and the other in *met1-6* roots (Y-axis). Note the large number of probes with low WT but high *met1-6* signal (red oval) representing increased H2A.Z incorporation in *met1-6* **b,** We subtracted the dataset represented on the X axis in (**A**) from the one on the Y-axis (we describe this procedure in detail in [41]). The histogram illustrates the distribution of the resulting signal scores. The shoulder on the right hand side indicates probes that have higher H2A.Z signal in *met1-6* than WT. The solid lines in **b** demarcate two standard deviations from the mean.

Figure 12

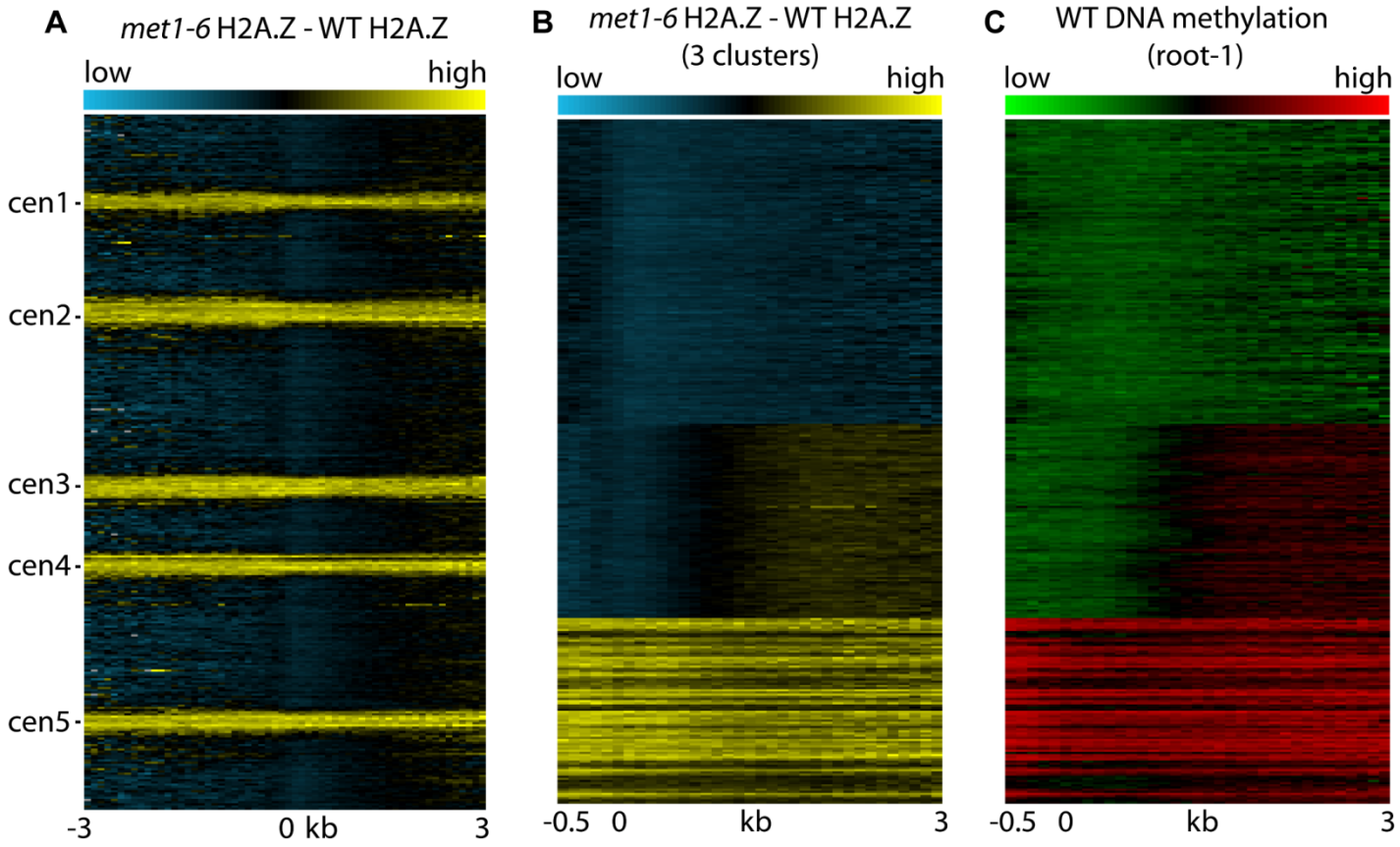

Figure 12. Differences in H2A.Z distribution in *met1-6* and WT. a, All TAIR 7 annotated sequences were aligned at the 5' end and stacked from the top of chromosome 1 to the bottom of chromosome 5. The WT H2A.Z pattern subtracted from the *met1-6* H2A.Z pattern is displayed as a heat map. The same data after *k*-means clustering (*k*=3) are shown in **b**. For comparison, root DNA methylation of sequences arranged as in **b** is shown as a heat map in **c**.

Figure 13

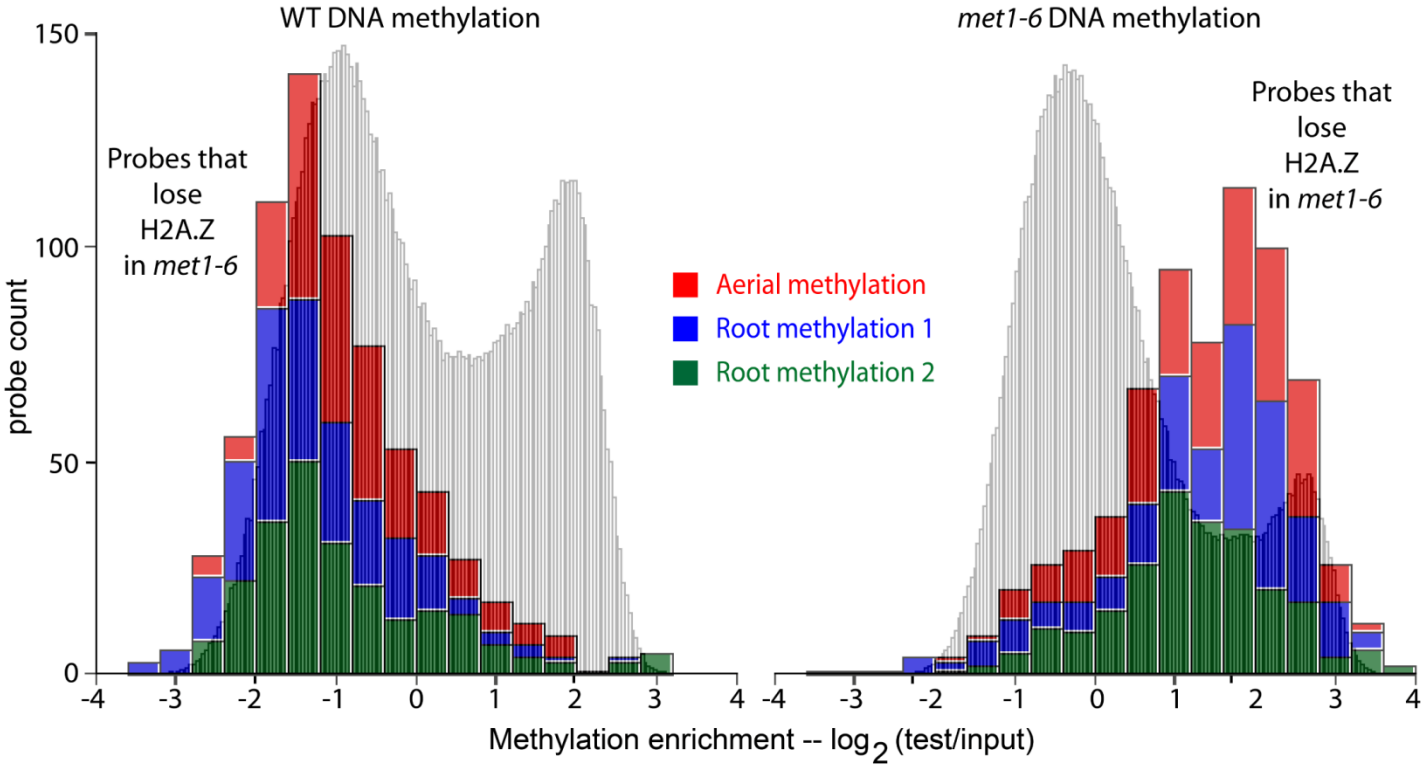

Figure 13. Histogram of *met1-6* induced changes in H2A.Z enrichment. WT methylation levels (left) and *met1-6* methylation levels (right) for probes representing a significant decrease of H2A.Z in *met1-6*. The histogram is cumulative for three independent methylation datasets. Grey histograms in the background show the signal distribution for all probes.

Figure 14

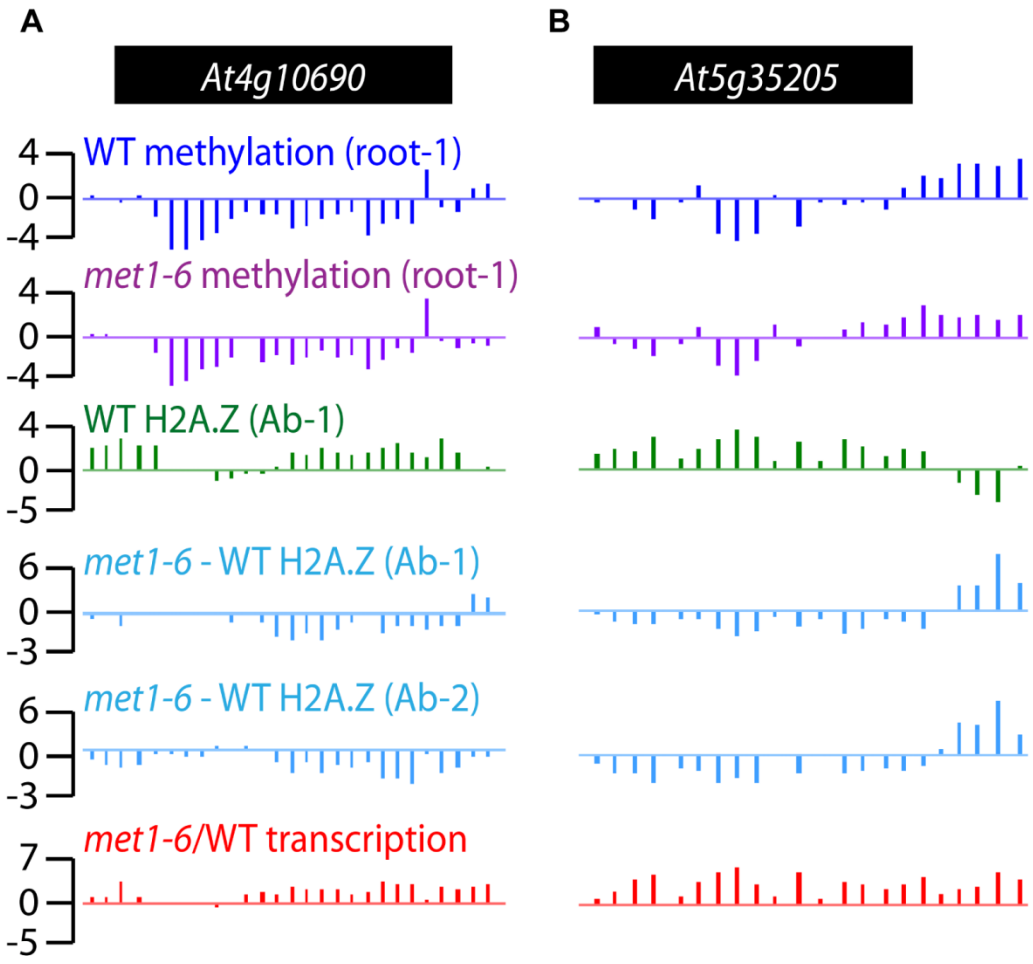

Figure 14. H2A.Z change in *met1-6* for two Transposons. Wild type (WT) root DNA methylation (dark blue), *met1-6* root DNA methylation (purple), WT H2A.Z (antibody, green), WT H2A.Z profile subtracted from the *met1-6* H2A.Z profile (two sets of independent paired experiments, light blue), and *met1-6*/WT transcription (red) for two transposable elements, *At4g10690* and *At5g35205*.

Figure 15

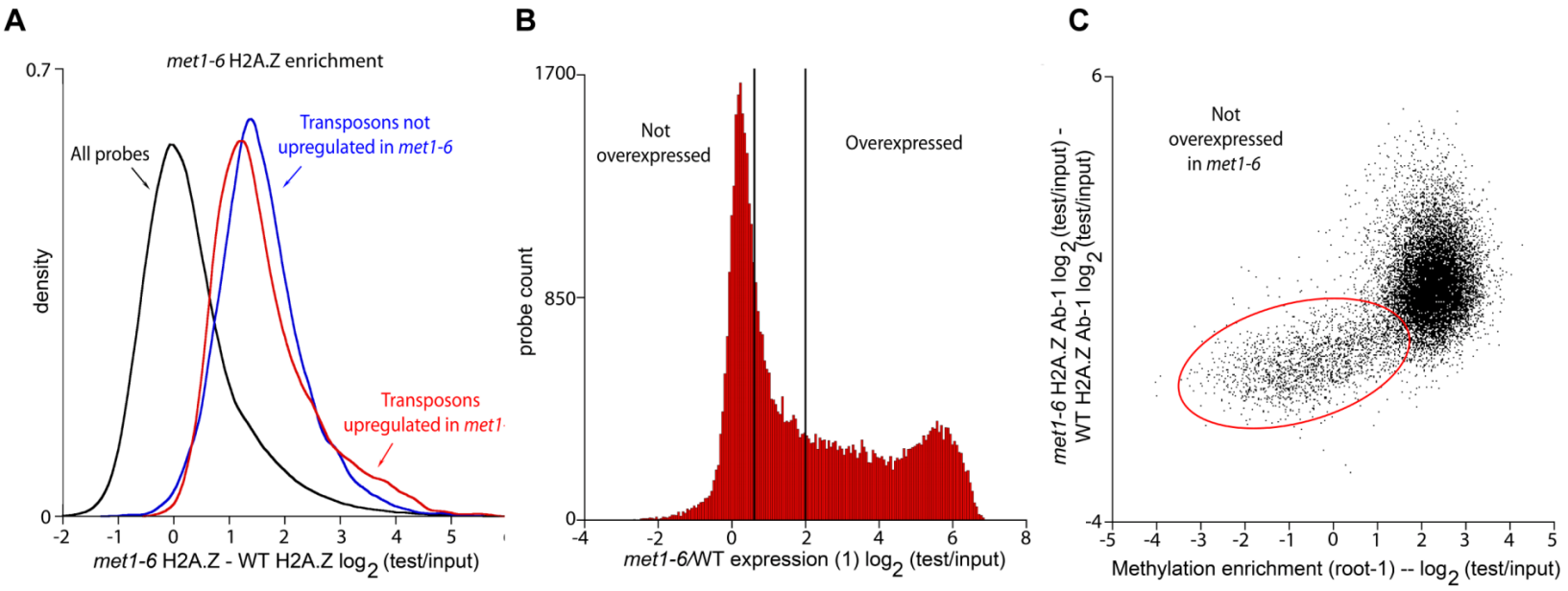

Figure 15 . Changes in H2A.Z Distribution in met1-6 for TEs. a, Kernel density plot, which has the effect of tracing the frequency distribution, of all probes in the dataset displayed in Fig. 12a (black trace), transposable elements upregulated in *met1-6* (red trace), and transposable elements not upregulated in *met1-6* (blue trace). **b,** The histograms illustrate the distribution of *met1-6*/WT transcription for transposable element probes. Probes with scores below 0.5 were considered not overexpressed, while probes with scores above 2 were considered overexpressed. **c,** Scatter plot showing the relationship between the *met1-6* H2A.Z - WT H2A.Z Ab-1 dataset (Y-axis) and root-1 WT DNA methylation (X-axis) for transposable element probes not overexpressed in *met1.*

Figure 16

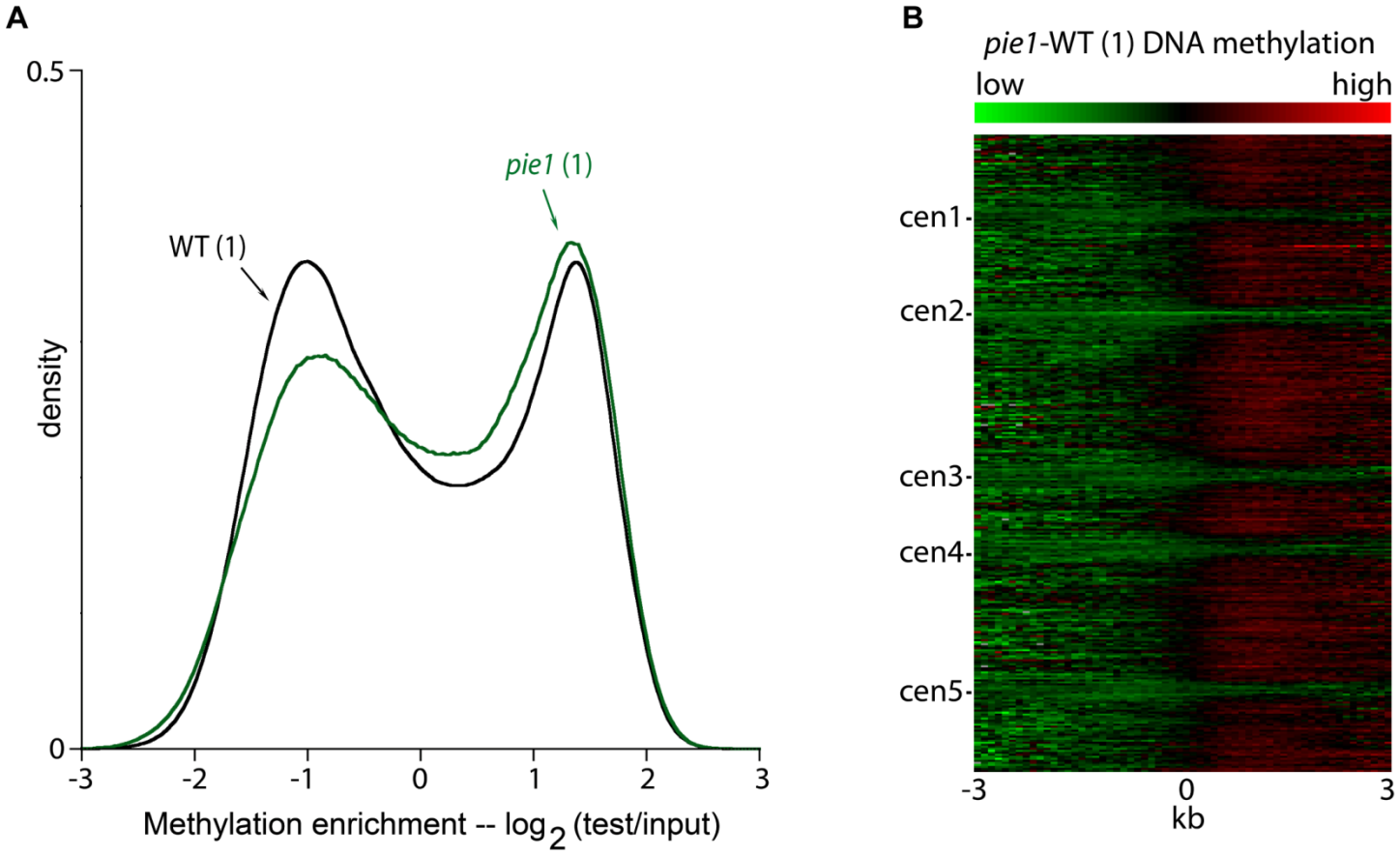

Figure 16. *pie1* induced changes in DNA methylation. a, Kernel density plots, which have the effect of tracing the frequency distribution, of DNA methylation enrichment for all WT probes (black trace) and all *pie1* probes (green trace). **b,** All TAIR 7 annotated sequences were aligned at the 5' end and stacked from the top of chromosome 1 to the bottom of chromosome 5. The WT methylation pattern subtracted from the *pie1* methylation pattern is displayed as a heat map.

Figure 17

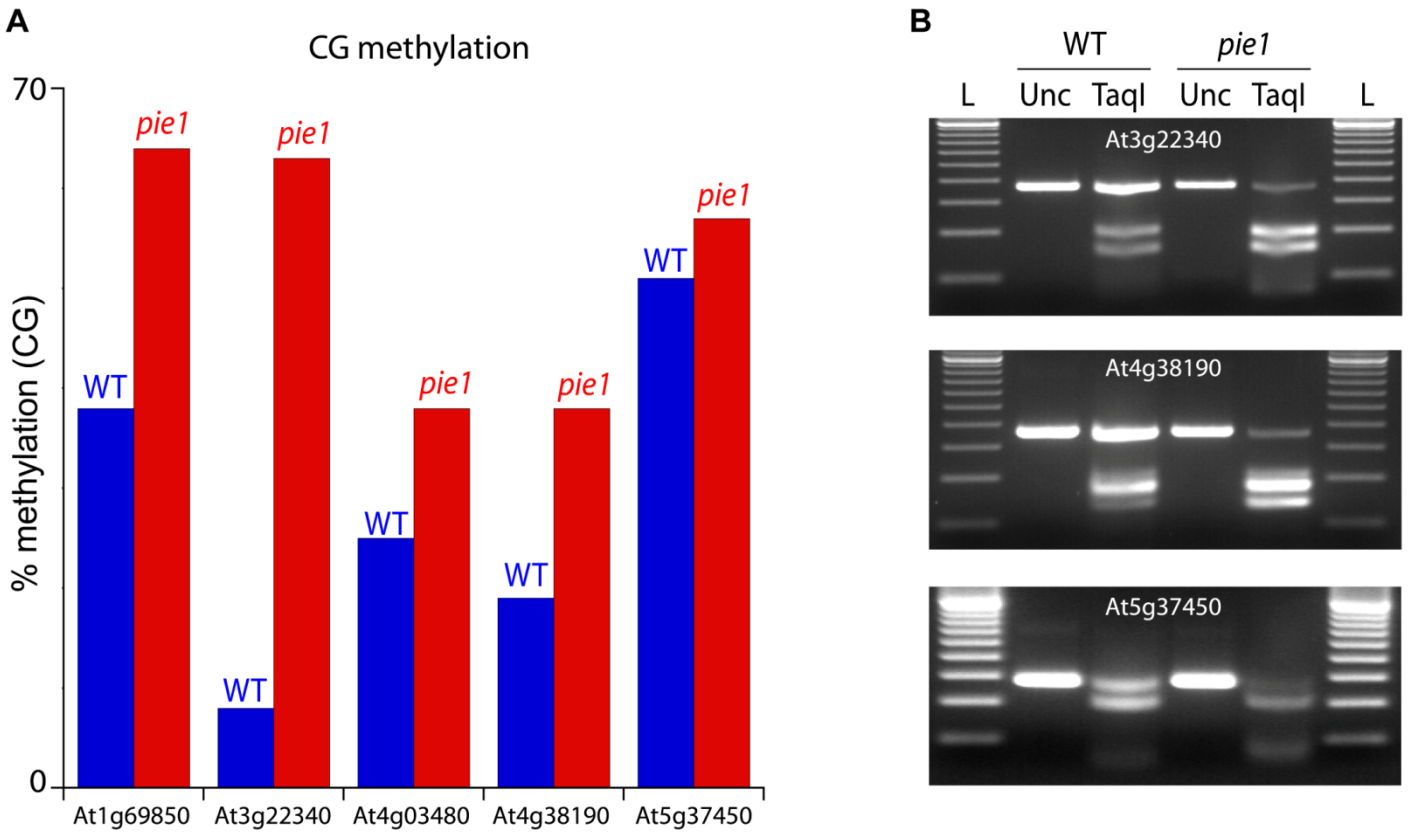

Figure 17. Increases in DNA methylation in *pie1* for specific loci. a, Bisulfite sequencing results for five loci. We sequenced 12 clones from each genotype, except for At1g69850 (10 clones in *pie1*) and At4g38190 (11 clones in *pie1*). **b,** PCR products from bisulfite-converted genomic DNA were digested with TaqI, which recognizes TCGA and will cut only if the C is unconverted (and therefore methylated). L = 100 bp ladder, Unc = uncut PCR product, TaqI = PCR product digested with TaqI. Note the greater digestion, which represents greater methylation, in *pie1* compared to WT.

Figure 18

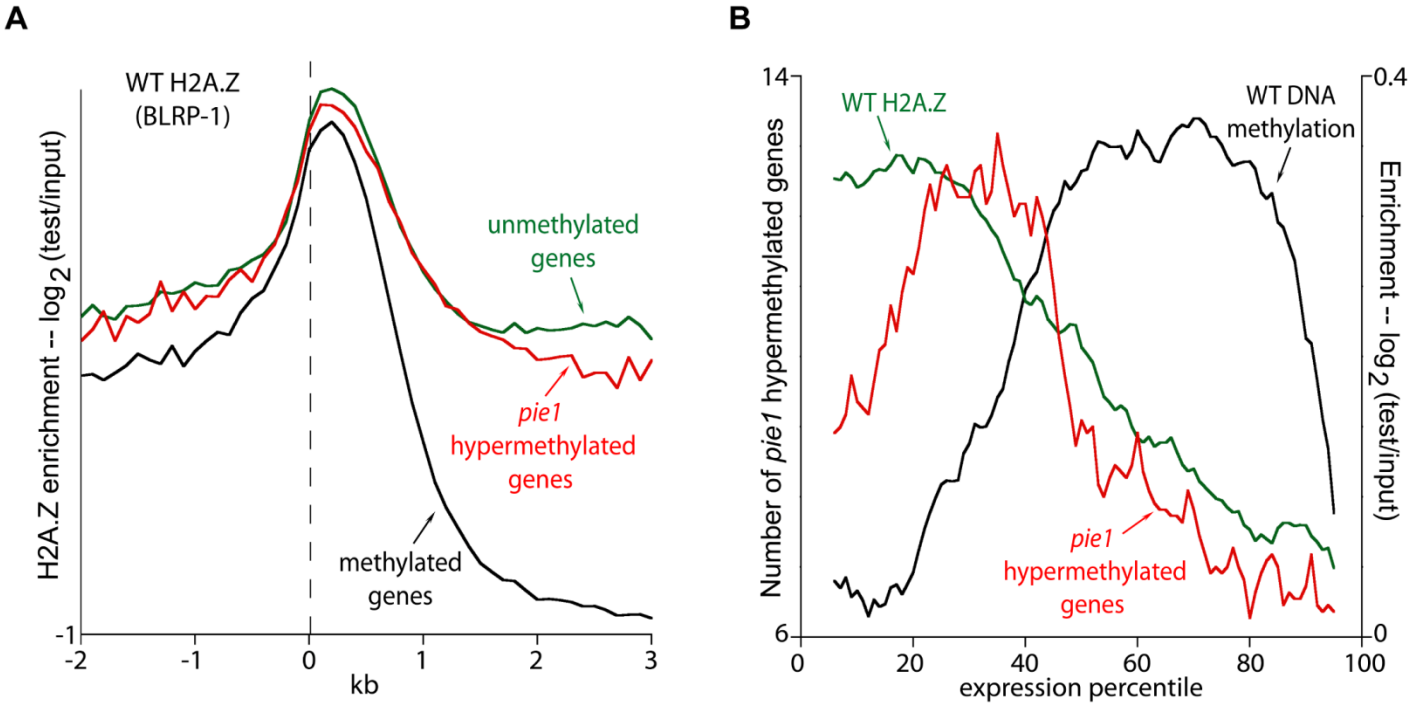

Figure 18. Transcription and methylation at *pie1* hypermethylated genes. a, All genes were aligned at the 5' end and average scores for each 100-bp interval are plotted from 2 kb away from the gene (negative numbers) to 3 kb into the gene (positive numbers). The data were smoothed with a 5-point sliding window. The dashed line represents the point of alignment. **b,** Genes were grouped into percentiles based on transcription levels. The red line traces the number of genes hypermethylated in *pie1* within each percentile (left Y-axis). The black line traces DNA methylation enrichment (all genes) and the green line traces H2A.Z enrichment in unmethylated genes (right Y-axis). The data were smoothed with a 10-point sliding window. The scale of the right Y-axis was set to start at zero to enable comparison between methylation and H2A.Z.

Chapter 3

Deposition of histone variant H2A.Z within gene bodies regulates responsive genes

3.1 Abstract

The regulation of eukaryotic chromatin relies on interactions between many epigenetic factors, including histone modifications, DNA methylation, and the incorporation of histone variants. H2A.Z, one of the most conserved but enigmatic histone variants that is enriched at the transcriptional start sites of genes, has been implicated in a variety of chromosomal processes. Recently, we reported a genome-wide anticorrelation between H2A.Z and DNA methylation, an epigenetic hallmark of heterochromatin that has also been found in the bodies of active genes in plants and animals. Here, we investigate the basis of this anticorrelation using a novel *h2a.z* loss-of-function line in *Arabidopsis thaliana.* Through genome-wide bisulfite sequencing, we demonstrate that a loss of H2A.Z in Arabidopsis does not affect the level or profile of DNA methylation in genes, and we propose that the global anticorrelation between DNA methylation and H2A.Z is caused by the exclusion of H2A.Z from methylated DNA. RNA-seq and genomic mapping of H2A.Z show that H2A.Z enrichment across gene bodies, rather than at the TSS, is correlated with lower transcription levels and higher measures of gene responsiveness. We find that a loss of H2A.Z causes misregulation of many genes that are disproportionately associated with response to both endogenous and exogenous stimuli. We propose that H2A.Z deposition in gene bodies promotes variability in levels and patterns of gene expression, and that a major function of genic DNA methylation is to exclude H2A.Z from constitutively expressed genes.

3.2 Introduction

In addition to packaging the DNA to fit within the cell, histones function to control the structure and accessibility of the chromatin environment by altering the biochemical properties of the nucleosome or through the recruitment of distinct binding partners. These actions promote changes in transcription that regulate the proper timing of developmental decisions and appropriate responses to the external environment. One such method of histone-mediated control comes from the exchange of the canonical histones with non-allelic histone variants, which alter the fundamental structure and stability of the nucleosome (Henikoff and Ahmad 2005; Sarma and Reinberg 2005; Talbert and Henikoff 2010; Barzily-Rokni, Friedman et al. 2011).

H2A.Z is one of the most enigmatic of these histone variants, as well as the most well-conserved, with a single origin at the root of eukaryotic evolution (Talbert and Henikoff 2010). H2A.Z has been implicated in a number of apparently disparate and even contrary chromosomal processes, including heterochromatic silencing, gene activation, transcriptional memory, cell-cycle progression and thermal-sensory response (Dhillon, Oki et al. 2006; Brickner, Cajigas et al. 2007; Zlatanova and Thakar 2008; Kumar and Wigge 2010; Light, Brickner et al. 2010). A common aspect of H2A.Z biology is its enrichment within the few nucleosomes surrounding transcription start sites (TSS), which has been demonstrated by genome-wide localization experiments in protozoa, fungi, animals, and plants (Guillemette, Bataille et al. 2005; Li, Pattenden et al. 2005; Raisner, Hartley et al. 2005; Albert, Mavrich et al. 2007; Barski, Cuddapah et

al. 2007; Creyghton, Markoulaki et al. 2008; Whittle, McClinic et al. 2008; Zilberman, Coleman-Derr et al. 2008; Siegel, Hekstra et al. 2009; Petter, Lee et al. 2011), and which implies a role in the regulation of transcription.

Considerable effort has been made to determine the specific effects of H2A.Z on transcription. In the yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, H2A.Z regulates genes that respond to changes in the environment (Millar, Xu et al. 2006; Wan, Saleem et al. 2009; Sadeghi, Bonilla et al. 2011), and loss-of-function mutants fail to react appropriately to external cues (Jackson and Gorovsky 2000; Santisteban, Kalashnikova et al. 2000). *Arabidopsis thaliana* plants lacking PIE1 (*AT3G12810*), the plant homolog of the SWR1 catalytic subunit of protein complexes responsible for the deposition of H2A.Z in yeast and mammals (Noh and Amasino 2003; Kobor, Venkatasubrahmanyam et al. 2004; Krogan, Baetz et al. 2004; Mizuguchi, Shen et al. 2004; Ruhl, Jin et al. 2006; Choi, Park et al. 2007; Deal, Topp et al. 2007; Wong, Cox et al. 2007), exhibit misregulation of many genes involved in the innate immune response (March-Diaz, Garcia-Dominguez et al. 2008). Recent work has shown that Arabidopsis plants with a mutated *ARP6*, which encodes a component of the PIE1 complex, inappropriately express temperature response genes, leading to the proposal that H2A.Z may act specifically as a thermosensor in plants (Kumar and Wigge 2010).

The genomic distribution and biological functions of DNA methylation, another well-conserved feature of chromatin, are in many aspects strikingly different from those of H2A.Z. DNA methylation in the form of 5-methylcytosine is present in all vertebrates examined to date, as well as in many invertebrates, fungi, and plants (Feng, Cokus et al. 2010; Zemach, McDaniel et al. 2010; Glastad, Hunt et al. 2011). The primary function of eukaryotic DNA methylation has long been considered to be the silencing of the sequences it decorates, particularly transposable elements (Law and Jacobsen 2010), although the recent discovery of gene body methylation in plants and animals, the functional significance of which is still unknown, has complicated this view (Zhang and Jacobsen 2006; Vaughn, Tanurd Ic et al. 2007; Zilberman, Gehring et al. 2007; Cokus, Feng et al. 2008; Lister, O'Malley et al. 2008; Feng, Cokus et al. 2010; Zemach, McDaniel et al. 2010). Whereas H2A.Z is enriched near the TSS of most genes, TSS-proximate DNA methylation is strongly associated with transcriptional repression in plants and vertebrates (He, Chen et al. 2011).

Recently, we reported a strong, genome-wide anticorrelation between H2A.Z and DNA methylation in Arabidopsis, including in bodies of active genes (Zilberman, Coleman-Derr et al. 2008). Results from similar studies in vertebrates suggest that this anticorrelation is a conserved feature of eukaryotes (Conerly, Teves et al. 2010; Edwards, O'Donnell et al. 2010; Zemach, McDaniel et al. 2010). In Arabidopsis, we showed that changes in DNA methylation caused by a mutation in the DNA methyltransferase *MET1* induced reciprocal alterations in H2A.Z deposition, demonstrating that DNA methylation antagonizes H2A.Z recruitment (Zilberman, Coleman-Derr et al. 2008). We also used a null mutation in *PIE1* (*pie1-5*) to examine the effect of disrupted H2A.Z function on DNA methylation. By coupling methylated DNA

immunoprecipitation (MeDIP) to microarray analysis, we found a low magnitude but genome-wide DNA methylation increase in genes that suggested a mutual antagonism between H2A.Z and DNA methylation (Zilberman, Coleman-Derr et al. 2008).

There is now considerable evidence that the PIE1 complex deposits H2A.Z into chromatin in Arabidopsis, though whether it has H2A.Z-independent functions, as has been shown for other eukaryotic SWR1 homologs, remains unclear (Kobor, Venkatasubrahmanyam et al. 2004; Auger, Galarneau et al. 2008; Bowman, Wong et al. 2011). It is also likely that other chromatin remodelers in Arabidopsis can deposit H2A.Z, as the yeast INO80 complex can deposit H2A.Z into chromatin (Papamichos-Chronakis, Watanabe et al. 2011), and H2A.Z is incorporated into nucleosomes at low levels in *pie1* and *swr1* mutants (Kobor, Venkatasubrahmanyam et al. 2004; Wu, Alami et al. 2005; Deal, Topp et al. 2007). Given that in both *S. cerevisiae* and Arabidopsis the sets of genes that are misregulated in H2A.Z and SWR1-related mutants only partially overlap (Kobor, Venkatasubrahmanyam et al. 2004; March-Diaz, Garcia-Dominguez et al. 2008), we sought to use an H2A.Z-deficient plant line, as opposed to SWR1-related mutants, for further analysis of H2A.Z function.

Here, we describe the characterization of an H2A.Z loss-of-function line in *Arabidopsis thaliana*. We find that loss of H2A.Z in Arabidopsis does not significantly affect the level or profile of DNA methylation in genes, and propose that the global anticorrelation between DNA methylation and H2A.Z is caused by the exclusion of H2A.Z from methylated DNA. We show that the level of H2A.Z enrichment in gene bodies is generally correlated with gene responsiveness and that lack of H2A.Z causes misregulation of many types of response genes, including response to endogenous stimuli and exogenous biotic and abiotic stimuli. We propose that H2A.Z deposition in gene bodies promotes gene responsiveness, but may prevent stable and constitutive expression, and that a major function of gene body DNA methylation is to exclude H2A.Z from constitutively expressed genes.

3.3 Results

3.3.1 Construction of a near-null Arabidopsis *h2a.z* mutant line

Three of the thirteen Arabidopsis *H2A* genes, *HTA8 (AT2G38810)*, *HTA9 (AT1G52740)*, and *HTA11 (AT3G54560)*, have been classified as encoding H2A.Z based on phylogenetic analyses (Redon, Pilch et al. 2002; Yi, Sardesai et al. 2006), and distribution patterns and genetic studies suggests that these proteins are largely functionally redundant (Choi, Park et al. 2007; Deal, Topp et al. 2007; Zilberman, Coleman-Derr et al. 2008). Recently published work has demonstrated that a double mutant of *hta9-1* and *hta11-1* produced plants with phenotypes similar to those found in null *pie1-5* mutants (March-Diaz, Garcia-Dominguez et al. 2008). To generate a line devoid of H2A.Z, we crossed *hta9-1* and *hta11-1* plants with a line bearing an insertion in *HTA8*, *hta8-1* (Figure 1A). Contrasting with recent evidence that individual knockouts of the two vertebrate H2A.Z isoforms exhibit different phenotypes (Matsuda, Hori et al. 2010), we did not observe morphological abnormalities in any of the three single mutant

lines. The resulting triple mutant line, which we will refer to as *h2a.z,* is both viable and phenotypically distinguishable from WT (Figure 1B). Transcripts of *HTA8* and *HTA11* were not detectable in the *h2a.z* mutant by RT-PCR, but low levels of *HTA9* RNA were present (~26% of wild-type; Figure 1C and 1D) in *h2a.z* plants but not in *hta9-1* single mutants, suggesting that the intronic T-DNA insertion in *HTA9* is spliced out in a fraction of transcripts, as confirmed by sequencing of the cDNA (Figure 2). To test whether this low level of expression was the result of a genetic rearrangement at the *HTA9* locus that occurred in our crosses, we recreated the *h2a.z* line using *hta9-1* plants lacking *HTA9* transcript (Figure 1D). The *h2a.z* progeny from the independent cross produced similar phenotypes to the original *h2a.z* line and similar RT-PCR results for *HTA9*, suggesting upregulation of *HTA9* in the triple mutant.

A fourth gene, *HTA4 (AT4G13570)*, is the closest H2A family member to the three *H2A.Z* genes and has been categorized as H2A.Z-like (Yi, Sardesai et al. 2006), but all publically available data indicate that *HTA4* is not expressed at significant levels in any WT tissue. To ensure that *HTA4* is not upregulated as a result of the drop in H2A.Z levels in our *h2a.z* line, we tested the expression of *HTA4* by RT-PCR (Figure 1E), and did not detect *HTA4* RNA in *h2a.z* or in WT. Taken together, our data indicate that the *h2a.z* line has less than ten percent of wild-type *H2A.Z* transcript levels. Despite reduced fertility (Figure 1F), *h2a.z* plants are viable and produce offspring, differing markedly from the lethality of strong H2A.Z mutations in other multicellular organisms (van Daal and van der Leij 1992; Faast, Thonglairoam et al. 2001; Rangasamy, Berven et al. 2003; Ridgway, Brown et al. 2004; Whittle, McClinic et al. 2008).

3.3.2 The *h2a.z* mutant phenotype is distinct from that caused by lack of PIE1

We measured the number of leaves present when the plant produced its first flower buds in *h2a.z* and WT (Figure 3A). In short days (SD), the *h2a.z* line flowered significantly earlier than WT, with 23.2 +/- 1.1 leaves vs. 49.7 +/- 1.5 leaves (P-value < 0.0001, two sample T-test). In long days (LD), the difference in flowering time between *h2a.z* and WT was less pronounced, with 8.3 +/- 0.2 leaves and 10.6 +/- 0.2 leaves, respectively (P-value < 0.0001), but the difference in rosette size and plant stature was greater in LD than SD (Figure 4). Of the first ten flowers, 22 +/- 3.1% in LD and 76 +/- 4.6% in SD exhibited extra petals (between 5 and 8) in the *h2a.z* mutant line, compared to 1.5 +/- 0.6% (LD) and 2 +/- 0.8% (SD) in WT (Figure 3B and 3C). The *h2a.z* mutant also exhibited short, thickened siliques, a phenomenon potentially related to decreased fertility. The *h2a.z* siliques averaged 4.8 +/- 0.1 and 5.6 +/- 0.1 mm in length under LD and SD conditions, compared to 10.6 +/- 0.1 and 11 +/- 0.2 mm for WT (Figure 3D-F). The *h2a.z* phenotypes described above, as well as increased leaf serration and petiole length in SD (Figure 3G), are similar to those previously published for *hta9-1*; *hta11-1* and *pie1-5* mutants (Noh and Amasino 2003; Deal, Kandasamy et al. 2005; March-Diaz, Garcia-Dominguez et al. 2008).

The *h2a.z* line exhibited several phenotypes not previously reported for *pie1-5* or *hta9-1;hta11-1.* First, while both *pie1-5* and *h2a.z* have reduced stature, *pie1-5* plants tend to be bushy, while *h2a.z* plants are spindly and have trouble remaining upright (Figure 3H). Second, many of the siliques in the *h2a.z* mutant exhibited a strong asymmetric curvature, most likely due to the improper fusion of its carpels (Figure 3F). Other novel phenotypes occurred only rarely, but taken together are suggestive of an inappropriate regulation of developmental timing (Figure 5). These phenotypes affected multiple aerial plant tissues, including leaf and stem structures, but were most prevalent among floral organs. The most striking examples were the inappropriate emergence of petals and stamens directly from the stem, and flowers with improperly fused carpels, leading to severely compromised reproductive structures.

A cross between yeast *htz1* (*h2a.z*) and *swr1* (*pie1*) mutants ameliorates many of the strong synthetic lethal phenotypes observed with the *htz1* single mutant (Morillo-Huesca, Clemente-Ruiz et al. 2010; Hang and Smith 2011). The cause of the strong *htz1* phenotypes was proposed to be chromatin disruption by the SWR1 complex in the absence of its proper substrate, a hypothesis supported by SWR1-dependent accumulation of DNA damage in the absence of *htz1*. To test whether simultaneous removal of the PIE1 chromatin remodeler and H2A.Z would reduce the severity of phenotypes seen in *h2a.z* plants, we crossed the *h2a.z* mutant line to *pie1-5*. Contrary to the results from yeast, the phenotype of the Arabidopsis double mutant is more severe than that of either parent – progeny exhibit early developmental arrest, dying shortly after germination (Figure 6). Taken together with the phenotypic disparity, our results indicate that H2A.Z and PIE1 have non-redundant functions in Arabidopsis. Because *h2a.z* is not a complete loss-of-function line, the stronger phenotype of *h2a.z*; *pie1-5* plants may be caused by a further reduction of H2A.Z incorporation into chromatin, but nevertheless demonstrates that *pie1-5* does not entirely abolish H2A.Z function.

3.3.3 Lack of H2A.Z does not perturb genic DNA methylation

To test our hypothesis that H2A.Z protects genes from DNA methylation, we generated genome-wide methylation profiles for the *h2a.z* mutant and WT using shotgun bisulfite sequencing. Because plants have DNA methylation in three different sequence contexts, CG, CHG, and CHH (H = A, T or C), which are largely controlled by distinct families of methyltransferases and have different genome-wide distributions (Feng, Cokus et al. 2010; Zemach, McDaniel et al. 2010), it is advantageous to use an assay that has single base-resolution to distinguish between these contexts. Two biological replicates each of *h2a.z* and WT were generated for each of three different tissue types that represent different stages along a developmental continuum: 14 day-old whole seedlings, 6 week-old rosette leaves, and 6 week-old cauline leaves. One biological replicate was taken from the original *h2a.z* mutant line, and the second from the additional *h2a.z* line generated from independent crosses with the same T-DNA insertional alleles. Analysis of the average methylation levels across all genes revealed that a loss of H2A.Z in Arabidopsis does not alter the global patterns of DNA methylation in CG, CHG or CHH contexts (Figure 7A and Figure 8). For comparison, we

generated bisulfite sequencing data for two biological replicates each of *pie1* and sibling WT seedlings, and one replicate of *h2a.z;pie1* seedlings. As with the results for the *h2a.z* mutant, the *pie1* and *h2a.z;pie1* mutants showed no change compared with WT in the global patterns of genic DNA methylation (Figure 9).

Unexpectedly, the *h2a.z* mutant exhibited tissue-specific DNA methylation changes in transposable elements (TEs; Figure 7B and Figure 10). CG methylation was marginally increased over wild-type in four of the six replicates, with the most consistent change in seedlings, whereas CHG methylation decreased more heavily in the older tissues, though there is considerable variation between replicates (Figure 7B and Figure 10). CHH methylation was substantially reduced specifically in cauline leaves (Figure 7B and Figure 10). Kernel density estimations of these changes demonstrate that the majority of transposons show a modest change in methylation, rather than a larger effect in a small subset of TEs (Figure 11). Analyses of the DNA methylation in *pie1* and *h2a.z;pie1* seedlings show that, like *h2a.z* seedlings, these lines exhibit increased CG methylation in TEs (Figure 12 and Figure 13). Curiously, the *h2a.z;pie1* seedlings exhibit decreases in CHG and CHH TE methylation that are not seen in seedlings of *pie1* or *h2a.z*, but which are reminiscent of the decreases in *h2a.z* plants later during development (cauline and rosette leaves; Figure 7B and Figure 12). Our data indicate that whereas a loss of H2A.Z does not change DNA methylation within genes, lack of H2A.Z affects TE methylation in all three sequence contexts in a tissue-specific manner, with different effects in cauline, rosette, and seedling samples.

One potential explanation for the lack of change in genic methylation in the *h2a.z* and *pie1* mutant lines is that perturbation of normal methylation targeting may be required in addition to loss of H2A.Z. To test this hypothesis, we performed crosses of *h2a.z* and *pie1* plants to *ibm1-6* and *met1-6* mutant lines. *IBM1* (*AT3G07610*) encodes a H3 lysine 9 demethylase, *MET1* (*AT5G49160*) encodes the primary CG DNA methyltransferase, and both *ibm1* and *met1* mutations cause increased CHG methylation in gene bodies (Finnegan and Dennis 1993; Kankel, Ramsey et al. 2003; Saze, Mittelsten Scheid et al. 2003; Cokus, Feng et al. 2008; Lister, O'Malley et al. 2008; Saze, Shiraishi et al. 2008; Miura, Nakamura et al. 2009). Single mutant plants are viable and fertile (Figure 14A), but *h2a.z*;*ibm1*, *h2a.z*;*met1*, *pie1;ibm1*, and *pie1;met1* double mutants die shortly after germination and exhibit severe developmental abnormalities, including the production of undifferentiated callus-like material, under-sized root systems, and premature flowering (Figure 14B).

Bisulfite sequencing of *h2a.z;ibm1*, *h2a.z*;*met1,* and *pie1;ibm1* seedlings revealed that a loss of H2A.Z does not obviously alter the genic methylation profile in any context from that seen in the parental backgrounds (Figures 15-17). The *h2a.z;ibm1* and *pie1;ibm1* double mutant lines were generated such that *h2a.z;ibm1* seedlings were newly homozygous for *ibm1* (1st generation), whereas *pie1;ibm1* seedlings came from first generation *ibm1* homozygous parents (2nd generation). The *h2a.z;ibm1* seedlings in their first generation of *ibm1* homozygosity have higher levels of CHG methylation than 1st generation *ibm1* seedlings, and *pie1;ibm1* seedlings in their second generation of *ibm1* homozygosity have lower levels of CHG methylation than second generation *ibm1*

seedlings (Figure 16). Both first generation datasets, *h2a.z;ibm1* and *ibm1,* show similar levels of CHH hypermethylation to one another; likewise, the second generation *pie1;ibm1* and *ibm1* data exhibit similar CHH hypermethylation levels (Figure 17). Importantly, the control data show that genic CHG and CHH methyaltion is unstable in ibm1, increasing greatly in the second generation (Figure 16), making interpretation of changes in *h2a.z;ibm1* and *pie1;ibm1* CHG methylation difficult.

Whereas there is little difference between the double mutant lines and their parental lines in TE CG methylation (Figure 18), we found CHG hypomethylation in the double mutants as compared to their respective parental lines (Figure 19). Additionally, while CHH methylation is unaltered in *h2a.z* and *pie1* seedlings, there is a significant reduction of TE CHH methylation in *h2a.z*;*ibm1*, *h2a.z*;*met1,* and *pie1;ibm1* seedlings compared to the *ibm1* and *met1* single mutants, which is similar to the reduction seen in *h2a.z;pie1* seedlings (Figure 19 and Figure 20). Taken together, our results suggest that while H2A.Z may play a modest role in the regulation of DNA methylation in TEs, the genome-wide anticorrelation between H2A.Z and DNA methylation is due to DNA methylation preventing the incorporation of H2A.Z.

3.3.4 H2A.Z is enriched in responsive genes

Given the published work linking H2A.Z with regulation of several types of genes that respond to the environment, we sought to examine H2A.Z enrichment with respect to gene responsiveness. To do so, we generated a genome-wide map of H2A.Z using our published tagged H2A.Z Arabidopsis line (Zilberman, Coleman-Derr et al. 2008) by coupling affinity purification of H2A.Z-bound DNA with high-throughput sequencing. Consistent with our previous findings, metaanalyses of the new dataset demonstrate a strong peak of H2A.Z at the 5' end and a smaller peak at the 3' end of most genes, with varying levels of H2A.Z distributed within gene bodies (Figure 21). Also consistent with our previous results (Zilberman, Coleman-Derr et al. 2008; Zemach, McDaniel et al. 2010), we found a negative correlation between H2A.Z enrichment in gene bodies and WT transcript levels (Spearman's rho = -0.4039, P-value < 0.0001). Genes with the most gene body H2A.Z (n=4,081) have median WT expression more than six-fold lower than that of genes with the lowest H2A.Z within their bodies (n=3,920) (Figure 22). By comparison, levels of H2A.Z enrichment near the TSS showed a different trend: genes with the most and least H2A.Z at the TSS had lower levels of expression than those with intermediate levels of H2A.Z (Figure 22), as we showed earlier for both Arabidopsis and pufferfish (Zilberman, Coleman-Derr et al. 2008; Zemach, McDaniel et al. 2010).

We discovered a positive correlation between enrichment of H2A.Z across gene bodies and gene responsiveness – the degree to which a gene is differentially expressed among different tissue types or experimental conditions (including hormone, nutrient, and chemical treatments, as well as biotic or abiotic stimulus), with higher response scores associated with greater differential expression (Aceituno, Moseyko et al. 2008). H2A.Z body-enriched genes (n=4,081) have a six-fold higher median gene responsiveness score than that of genes with the lowest H2A.Z levels across their bodies (n = 3,920) (Figure 22). Levels of H2A.Z at the TSS are considerably less

correlated with response score than levels of H2A.Z in the body (Spearman's rho = 0.0748 and 0.3325, P-values < 0.0001, respectively). These results suggest that H2A.Z deposition in the gene body may facilitate rapid activation or inactivation of genes.

3.3.5 H2A.Z regulates responsive genes

To determine which genes are misregulated upon loss of H2A.Z, we profiled the transcriptomes of the *h2a.z* mutant and WT in 4-week old rosette leaves with three replicates each of RNA-seq. 1,800 genes were upregulated and 614 genes were downregulated in *h2a.z* with a P-value cut-off of 0.001. This is consistent with transcriptome analyses of *hta9;hta11* and *pie1,* which showed three-fold and two-fold more genes upregulated than downregulated, respectively (March-Diaz, Garcia-Dominguez et al. 2008). Gene Ontology analysis of the misregulated genes revealed enrichment of categories related to immune response (P-value = $8.6x10^{-9}$) and temperature response (P-value = $4.8x10^{-8}$), consistent with previous studies of *pie1* and *arp4* mutants (March-Diaz, Garcia-Dominguez et al. 2008; Kumar and Wigge 2010) (Table S1). Strikingly, all of the most-enriched categories (P-value < $1x10^{-5}$) are specifically response-related, and include a variety of previously unreported GO-terms involved in the perception of both endogenous and external cues (Figure 23). These results indicate that a loss of H2A.Z in plants leads to the misregulation of inducible genes generally, rather than a specific category of response genes. This hypothesis is consistent with published transcriptional studies from yeast and animals, which have shown that H2A.Z is involved in the transcriptional regulation of many different classes of inducible genes involved in environmental response and development (Adam, Robert et al. 2001; Millar, Xu et al. 2006; Creyghton, Markoulaki et al. 2008; Whittle, McClinic et al. 2008; Amat and Gudas 2011; Petter, Lee et al. 2011; Sadeghi, Bonilla et al. 2011).

Consistent with our Gene Ontology analysis (Tables S1-2), we discovered a relationship between the degree of misregulation in the *h2a.z* mutant and the responsiveness score of a gene (Figure 23B). Genes exhibiting greater than 4-fold upregulation (n=938) had a 2.5-fold higher median responsiveness score than that of the least upregulated genes (less than 1.4-fold up or downregulated, n=9,300). The relationship between downregulation and response score, on the other hand, was roughly parabolic, with the most downregulated and least downregulated genes showing the lowest levels of responsiveness, and genes with intermediate levels of downregulation (2 to 4-fold) showing the greatest responsiveness (Figure 23B). Because H2A.Z is enriched in bodies of response genes, we investigated whether changes in transcriptional regulation in the *h2a.z* mutant correlated with specific H2A.Z enrichment patterns in WT. As expected, we found a positive relationship between misregulation in the *h2a.z* line and H2A.Z gene body enrichment (Figure 23C) (Spearman's rho = 0.2634 for downregulated genes and 0.2540 for upregulated genes, P-value < 0.0001). Genes with the greatest misregulation (greater than four-fold up or downregulated, n=1,258) have more than a 36-fold higher median H2A.Z-body enrichment score than that of genes with the lowest levels of change in transcription between *h2a.z* and WT (less than 1.4-fold up or downregulated, n=9,300). Taken together, our data demonstrate that a loss of H2A.Z leads to a general transcriptional

misregulation of response genes that are enriched for H2A.Z within the gene body in wild type. This suggests that one function of gene body methylation, which is strongly anticorrelated with gene responsiveness in plants and animals (Zhang, Yazaki et al. 2006; Aceituno, Moseyko et al. 2008; Foret, Kucharski et al. 2009), is the exclusion of H2A.Z from the bodies of constitutively expressed genes.

3.4 Discussion

We have generated a viable H2A.Z-deficient mutant line in *Arabidopsis thaliana* that shares many, but not all of the phenotypic characteristics of *pie1* mutants. We show that unlike in yeast, combining Arabidopsis *h2a.z* and *pie1* mutations exacerbates the phenotype. Loss of H2A.Z does not significantly affect the level or profile of DNA methylation in genes, even when combined with mutations that alter the normal genic methylation landscape, whereas DNA methylation in transposons is perturbed in a tissue-dependent manner. We show that differences in gene body H2A.Z levels are correlated with gene expression and gene responsiveness. Finally, we show that a loss of H2A.Z causes misregulation of many genes involved in the response to environmental and developmental cues, and that these genes tend to have high levels of gene-body H2A.Z.

3.4.1 Residual H2A.Z function remains in *pie1* mutant plants

Whereas the fungi *S. pombe* and *S. cerevisiae* can tolerate mutations in H2A.Z (Carr, Dorrington et al. 1994; Jackson and Gorovsky 2000), H2A.Z is essential in many species, including *Tetrahymena thermophila*, *Drosophila melanogastor*, *Xenopus laevis*, *Caenorhabditis elegans* and mice (Liu, Li et al. 1996; Faast, Thonglairoam et al. 2001; Ridgway, Brown et al. 2004; Swaminathan, Baxter et al. 2005; Whittle, McClinic et al. 2008). Consequently, many studies of H2A.Z function outside of yeast have utilized mutants in components of the chromatin remodelers that deposit H2A.Z to emulate H2A.Z loss-of-function (Deal, Topp et al. 2007; March-Diaz, Garcia-Dominguez et al. 2008; Smith, Jain et al. 2009; Kumar and Wigge 2010). The substantial overlap between the phenotypes of Arabidopsis *pie1* and *h2a.z* mutants suggests that PIE1 is the primary remodeler responsible for H2A.Z deposition. However, *h2a.z;pie1* double mutants exhibit early developmental arrest not seen in either of the single mutant lines, indicating that H2A.Z can be deposited in the absence of the PIE1 complex, potentially by the Arabidopsis homolog of INO80 (Fritsch, Benvenuto et al. 2004), which can deposit H2A.Z in yeast (Papamichos-Chronakis, Watanabe et al. 2011). The PIE1 complex might also have H2A.Z-indpendent roles, as has been hypothesized for the PIE1/SWR1 orthologs in animals (Auger, Galarneau et al. 2008; Bowman, Wong et al. 2011). Indeed, a recent study showed that H2A.Z deposition by p400 and SRCAP, the human orthologs of SWR1, could not account for all the regulatory roles of these complexes (Bowman, Wong et al. 2011). These results emphasize that phenotypes caused by mutations in chromatin remodeling complexes must be interpreted with caution.

3.4.2 DNA methylation excludes H2A.Z from chromatin

DNA methylation and H2A.Z are tightly anticorrelated in plants and animals (Zilberman, Coleman-Derr et al. 2008; Conerly, Teves et al. 2010; Edwards, O'Donnell et al. 2010; Zemach, McDaniel et al. 2010), and we have shown that DNA methylation quantitatively excludes H2A.Z from chromatin (Zilberman, Coleman-Derr et al. 2008). Here, we demonstrate that H2A.Z does not have a detectable influence on DNA methylation in genes, even when genic DNA methylation is in flux, indicating that exclusion of H2A.Z from methylated DNA is the cause of the observed anticorrelation (Figure 24). Our earlier experiments examining methylation in *pie1* plants, which showed a modest relative DNA methylation increase in genes compared with transposons, were conducted using MeDIP-chip (Zilberman, Coleman-Derr et al. 2008), a technique that cannot differentiate between methylation in different sequence contexts and relies on the normalization of signal intensities. The bisulfite sequencing data presented here reveal that whereas genes show no change in DNA methylation, there are changes in TE methylation in the *h2a.z* mutant, the largest of which are global decreases in CHG and CHH methylation, which might account for our original findings of relative genic hypermethylation. Changes in TE methylation could be a direct result of H2A.Z loss, or may be caused by a variety of indirect effects. Given the depletion of H2A.Z from methylated transposons and the substantial transcriptional and developmental changes in *h2a.z* plants, we consider indirect explanations to be more probable.

3.4.3 H2A.Z in gene bodies regulates transcription of responsive genes

The significance of H2A.Z enrichment near transcriptional start sites has been a major focus of research (Adam, Robert et al. 2001; Gevry, Chan et al. 2007; Mavrich, Jiang et al. 2008; Hardy, Jacques et al. 2009; Jin, Zang et al. 2009), but a distinct function for H2A.Z in gene bodies has been recently hypothesized (Fujimoto, Seebart et al. 2012). Consistent with this idea, we have previously shown that H2A.Z abundance within gene bodies correlates negatively with transcription in Arabidopsis and the pufferfish *Tetraodon nigroviridis*, whereas H2A.Z near the TSS is most enriched in moderately transcribed genes in both organisms (Zilberman, Coleman-Derr et al. 2008; Zemach, McDaniel et al. 2010). Human studies also show that gene body H2A.Z correlates with silencing (Barski, Cuddapah et al. 2007) and that H2A.Z is depleted from the bodies of actively transcribed genes (Hardy, Jacques et al. 2009). The presence of this relationship in plants and animals implies that it is an ancient property of eukaryotes. Interestingly, recent studies in yeast have shown that mutation of the IN080 complex causes loss of H2A.Z near the TSS and gain of H2A.Z across the coding region (Papamichos-Chronakis, Watanabe et al. 2011), suggesting that competing nucleosome remodelers may shape the genic patterns of H2A.Z.

Here, we show that H2A.Z within gene bodies is correlated with gene responsiveness. A similar conclusion was recently made in yeast, where H2A.Z was shown to be significantly enriched across coding sequences of genes that are

differentially transcribed after environmental stress (Sadeghi, Bonilla et al. 2011). We also demonstrate that loss of H2A.Z leads to misregulation of Arabidopsis genes with high responsiveness scores, which measure differential expression across both tissue types and environmental conditions. Our results are consistent with evidence from many other species, where loss of H2A.Z leads to misregulation of various inducible genes, including environmental response genes in yeast (Adam, Robert et al. 2001; Millar, Xu et al. 2006; Sadeghi, Bonilla et al. 2011) and developmentally regulated and tissue-specific genes in animals (Updike and Mango 2006; Creyghton, Markoulaki et al. 2008; Whittle, McClinic et al. 2008; Amat and Gudas 2011; Petter, Lee et al. 2011). Genes that show little change in transcription in our *h2a.z* mutant plants tend to have H2A.Z depleted from the gene body, whereas those genes with either strong up- or downregulation tend to have much more gene-body H2A.Z. Taken together, these results indicate that H2A.Z within transcribed sequences is necessary for proper regulation of responsive genes but may antagonize constitutive and high-level expression, and that this relationship is both ancient and well-conserved across many eukaryotic lineages.

3.4.4 Gene body methylation may regulate gene expression by preventing H2A.Z incorporation

The presence of DNA methylation within the bodies of animal and plant genes has been known for some time (Jones and Laird 1999; Tran, Henikoff et al. 2005). Recent genome-wide bisulfite sequencing in various eukaryotic species has revealed that gene body methylation is an ancient and widely conserved feature of eukaryotic chromatin predating the divergence of animals and plants (Zhang, Yazaki et al. 2006; Zilberman, Gehring et al. 2007; Cokus, Feng et al. 2008; Lister, O'Malley et al. 2008; Lister, Pelizzola et al. 2009; Feng, Cokus et al. 2010; Xiang, Zhu et al. 2010; Zemach, McDaniel et al. 2010). In both animals and plants, gene body methylation exists almost exclusively within the CG context and follows a remarkably consistent pattern, with precise depletion of DNA methylation from the 5' and 3' ends of genes. Taken together with the finding that many species of invertebrates have DNA methylation primarily or exclusively within gene bodies (Xiang, Zhu et al. 2010; Zemach, McDaniel et al. 2010; Sarda, Zeng et al. 2012), these results strongly suggest that genic methylation plays an important and conserved function in at least some eukaryotic lineages (Suzuki and Bird 2008).

Despite the prevalence of gene body methylation in diverse eukaryotes, its function remains mysterious (Zemach and Zilberman 2010). A potential clue comes from the correlation between genic methylation and transcription. Gene body methylation is highest in moderately transcribed genes in plants and animals, with the lowest levels of genic methylation at either transcriptional extreme (Zhang, Yazaki et al. 2006; Zilberman, Gehring et al. 2007; Zemach, McDaniel et al. 2010). Additionally, there is an unexplained negative linear correlation between genic methylation and gene responsiveness in Arabidopsis and the honeybee *Apis mellifera* (Zhang, Yazaki et al. 2006; Aceituno, Moseyko et al. 2008; Zeng and Yi 2010). High levels of body methylation tend to be found in slowly evolving genes with vital housekeeping functions

in honeybee, silkworm (*Bombyx mori*), sea squirt (*Ciona intestinalis*), sea anemone (*Nematostella vectensis*), poplar (*Populus tricharpa*), and Arabidoopsis (Zhang, Yazaki et al. 2006; Sarda, Zeng et al. 2012; Takuno and Gaut 2012; Vining, Pomraning et al. 2012). These results indicate that DNA methylation of the transcribed region may be important for proper regulation of constitutively expressed genes.

Here, we show that the genome-wide anticorrelation between DNA methylation and H2A.Z is established by the exclusion of H2A.Z from methylated DNA. Because gene body DNA methylation and H2A.Z show opposing correlations with gene responsiveness, and the anticorrelation between DNA methylation and H2A.Z is ancient, we propose that a basal function of genic DNA methylation is the stabilization of constitutive expression patterns within housekeeping genes by antagonizing H2A.Z deposition (Figure 24). As H2A.Z has been linked to the regulation of inducible genes in many organisms, including species such as *S. cerevisae* and *C. elegans* that lack DNA methylation (Adam, Robert et al. 2001; Millar, Xu et al. 2006; Zanton and Pugh 2006; Creyghton, Markoulaki et al. 2008; Whittle, McClinic et al. 2008; Smith, Jain et al. 2009; Kumar and Wigge 2010; Amat and Gudas 2011; Petter, Lee et al. 2011; Sadeghi, Bonilla et al. 2011), and DNA methylation can exclude H2A.Z but not vice versa, we believe that the presence or absence of H2A.Z in the gene body is a better candidate for direct gene regulation than DNA methylation. The functional significance of DNA methylation of constitutive genes may be primarily to prevent incorporation of H2A.Z.

3.5 Materials and Methods

Biological Materials. The Arabidopsis T-DNA lines *hta9-1* (SALK_054814), *hta11-1* (SALK_017235), *ibm1-6* (SALK_006042), and *pie1-5* (SALK_096434) were obtained from the SALK collection (*Col-0* ecotype) (http://signal.salk.edu/). The Arabidopsis T-DNA line *hta8-1* (FLAG_593B04) was obtained from the INRA (http:// www-ijpb.versailles.inra.fr/) collection (*WS* ecotype). Sequencing of the 5' promoter region of *HTA8* confirmed the T-DNA insertion site for *hta8-1* at position 16,220,917 on Chr2 (NC_003071.1), 8 bp downstream of the 5' end of gene model AT2G38810.2. The Arabidopsis EMS mutant *met1-6* is described in (Xiao, Gehring et al. 2003) (*Col-0* ecotype). For bisulfite sequencing of seedling tissues, seeds were planted on 1x Murashige and Skoog Media with micronutrients and 1.5% Sucrose (Caisson Laboratories) and grown under 16h light/ 8h dark for 14 days in a growth chamber. For bisulfite sequencing of rosette and cauline leaf tissue, seeds were planted on soil and grown in greenhouse conditions with LD 16h light / 8h dark. For phenotype analysis of the *h2a.z* mutant, seeds were planted on soil and grown in greenhouse conditions with either 16h light / 8h dark (LD) or 8h light/ 16h dark (SD). Genotyping of SALK and INRA T-DNA lines was carried out by PCR with primers listed in Table S3. Genotyping of the *met1-6* line was carried out by dCAPS-PCR with primers listed in Table S3 and subsequent digestion with *BglII* .

Transcript Analysis of H2A.Z genes. Expression analyses for the *h2a.z* and *hta9-1* mutant lines and for the WT control were performed on total RNA extracted from 4 week post germination rosette leaves grown on soil in LD conditions using the

RNeasy Plant Extraction Kit (Qiagen) with the optional on-column DNAse treatment. RT-PCR reactions were carried out on cDNA generated using 1ug total RNA and the Superscript III Kit (Invitrogen) using gene specific primers listed in Table S3. qPCR was carried out on similarly generated cDNA using EvaGreen Detection chemistry on an ABI 7500 FAST Real-Time PCR System with primers in exons flanking the single intron in *HTA9*. The gene *UBQ5* (AT3G62250) was used as in internal control. Three biological replicates, each with three technical replicates, were averaged.

Bisulfite Sequencing. Approximately 100-500 ng genomic DNA was isolated from either seedling, rosette or cauline leaf tissues. Seedling tissue was obtained from 14 days post germination seedlings grown on Murishige and Skoog media in LD (16h light/ 8h dark). Mature rosette leaves and mature cauline leaves were obtained from 4 week post germination mature plants grown on soil in LD (16h light/ 8h dark). In general, multiple biological replicates were generated for each mutant and WT line; a complete list of all generated libraries is available in Table S4. WT datasets for each mutant were generated from plants derived from recent relatives of the relevant mutant. Bisulfite conversion and Illumina library construction and sequencing were performed as described in (Hsieh, Ibarra et al. 2009). We used single ends (SE) Illumina sequencing for bisulfite sequencing on the GAII and HiSeq platforms and sequence alignments were performed using Bowtie (Langmead, Trapnell et al. 2009) and the TAIR8 Genome Annotation (http://www.arabidopsis.org/) as in (Zemach, McDaniel et al. 2010). The average percent methylation plots were generated as described in (Hsieh, Ibarra et al. 2009) and (Zemach, McDaniel et al. 2010).

RNA Sequencing. Approximately 30 ug total RNA was isolated from 4 week post germination mature rosette leaves using the RNEasy Plant Extraction Kit (Qiagen) with the optional on-column DNAse treatment. mRNA was purified from total RNA by two treatments of poly-A enrichment using the Oligotex kit (Qiagen #72022), followed by a rRNA removal step using the RiboMinus Plant Kit for RNA-sequencing (Invitrogen #A1083702). Illumina library construction and RNA-sequencing were performed as described in (Zemach, McDaniel et al. 2010). We used single ends (SE) Illumina sequencing for RNA-sequencing on the GAII platform and sequence alignments were performed using Bowtie and the TAIR8 Genome Annotation and cDNA Annotation (http://www.arabidopsis.org/) as in (Zemach, McDaniel et al. 2010).

H2A.Z ChAP Sequencing. H2A.Z-containing nucleosomes were chromatin affinity purified (ChAP) from 4 week post germination Arabidopsis roots of our H2A.Z-BLRP transgenic lines grown in LD conditions as in (Zilberman, Coleman-Derr et al. 2008). Illumina libraries were constructed for IP and input DNA samples and sequenced generating SE 50 bp reads. Nucleosomal midpoints were estimated based on an average 150-bp nucleosome length by adding 75bp to the start position of each read. Differences between IP and input over each single-base window were generated to give an overall genome-wide map of H2A.Z-enrichment. Genes were aligned at either their 5' or 3' ends, and average H2A.Z-enrichment values were calculated over each 50bp window between 3kb up and downstream of the point of alignment. Sequencing was

carried out on the HiSeq platform and sequence alignments were performed using Bowtie and the TAIR8 Genome Annotation as in (Zemach, McDaniel et al. 2010).

Differential Expression Sequence Analysis. For differential expression analysis of the RNA-sequencing datasets, a strategy was employed to account for expression differences between WS and Col ecotypes. In brief, we used the recently published list of 144,879 SNPs between the WS and Col ecotypes (Ashelford, Eriksson et al. 2011) to obtain reads per kilobase of exon model per million reads (RPKM) scores for each gene in *h2a.z* and WT from either the WS or Col backgrounds.

First, using Bowtie with no tolerance for mismatches, reads from each of the three *h2a.z* and WT RNA-seq datasets were mapped to small 75bp scaffolds containing either the WS or Col SNP around each SNP locus that mapped within an exon of a gene greater than 200bp in length and with at least 10 mapped reads. We removed all SNPs that were less than one read-length (36bp) from the end of the exon, which left approximately 5,000 SNPs across the genome. The number of reads mapping to the WS and Col scaffolds were compared at each SNP locus and used to determine whether the region was homozygous for WS, Col or heterozygous for the two ecotypes in each dataset. For SNPs at heterozygous loci, a Read Count Contribution from each WS or Col genome was determined by dividing the number of reads mapping to either WS or Col genome by the total reads mapping to the SNP scaffold for each ecotype. As SNPs within a given heterozygous region generally exhibited similar ratios of WS to Col mapped reads, a rolling 20-window (where the windows are the 5,000 SNPs) smoothing function was applied to these read count contribution values.

Next, the six RNA-seq datasets were mapped to the TAIR cDNA scaffold, and each cDNA model was assigned a score equal to the number of mapped RPKM. For both the *h2a.z* and WT datasets, the normalized read counts of the three replicates were partitioned into reads contributed by WS and by Col using the smoothed read count contribution value obtained from the nearest SNP. In this way, approximate WS and Col read count scores were determined for each gene in both *h2a.z* and WT.

To test for statistical significance of the difference between the *h2a.z* and WT, we repeated the above partitioning process using read counts normalized to the size of the smallest library, rather than per million of reads. This alternate normalization less drastically underestimate the number of reads per locus, which is important as the statistical significance is dependent on the number of reads. We calculated the probability that a gene's expression deviates from expectation using a Fisher's two-tailed exact test of *h2a.z* vs. WT scores for each ecotype. Genes were determined to be differentially expressed if for either ecotype they exhibited a two-fold change in expression between *h2a.z* and WT and had a P-value < 0.001, or if for both ecotypes they exhibited a two-fold change in expression and had p-values < 0.005. Gene Ontology analysis was performed on the up- and downregulated gene lists using the GO FAT Ontology on the DAVID web server (http://david.abcc.ncifcrf.gov) (Huang da, Sherman et al. 2009; Huang da, Sherman et al. 2009) and categories with P-values < $1x10^{-5}$ were considered enriched.

Figure 1

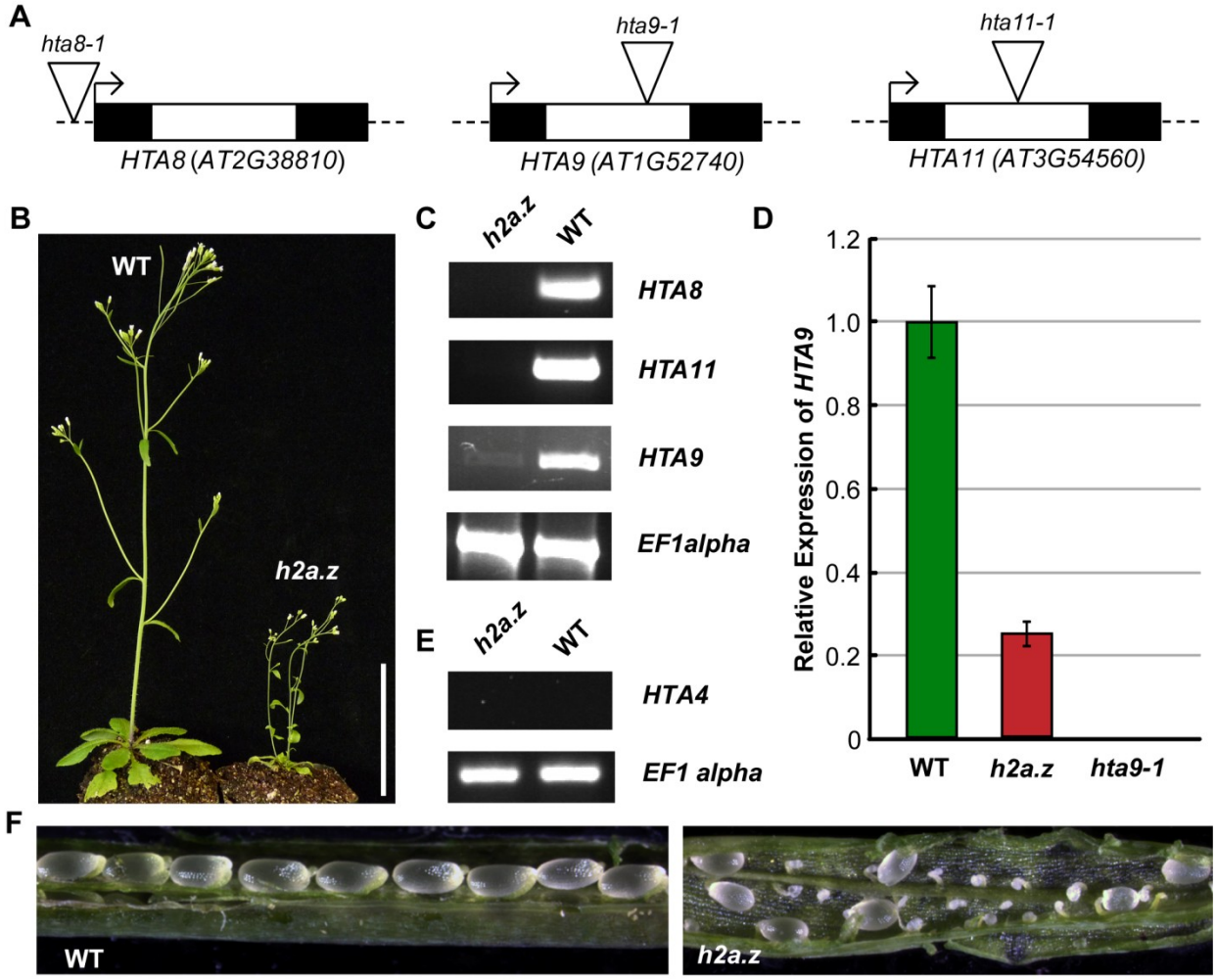

Figure 1. Construction of an *h2a.z* mutant line. (A) The three T-DNA insertions in *HTA8*, *HTA9*, and *HTA11*. Exons are represented by dark boxes, introns by unfilled boxes. Triangles represent the T-DNA insertion points. *hta9-1* and *hta11-1* are intronic insertions, while *hta8-1* is an insertion in the 5' UTR. (B) WT and *h2a.z* plants at 21 days post germination. Scale bar = 3cm. (C) RT-PCR of *HTA8*, *HTA9*, and *HTA11* in the *h2a.z* mutant line and WT control line. RT-PCR of *EF1alpha* was used as a loading control. (D) qPCR results for *HTA9* transcripts in the *h2a.z* and *hta9-1* mutants. Expression is normalized to WT levels; WT is shown in green, and *h2a.z* in red. Note that the level of *HTA9* in the *hta9-1* mutant is 0.040 percent of WT. (E) RT-PCR of *HTA4* in the *h2a.z* mutant line and a WT control. RT-PCR of *EF1alpha* was used as a loading control. (F) Ovule development in *h2a.z* mutant and WT siliques grown in LD conditions.

Figure 2

Figure 2. Residual expression of *HTA9* in the *h2a.z* mutant. Sequencing of residual HTA9 transcripts in the *h2a.z* mutant. Genomic DNA sequence of the 15 bps (5 amino acids) on either side of the exon/inton boundary are shown in black. The WT and *h2a.z* mRNA transcripts from the corresponding region are shown in green and red, respectively.

Figure 3

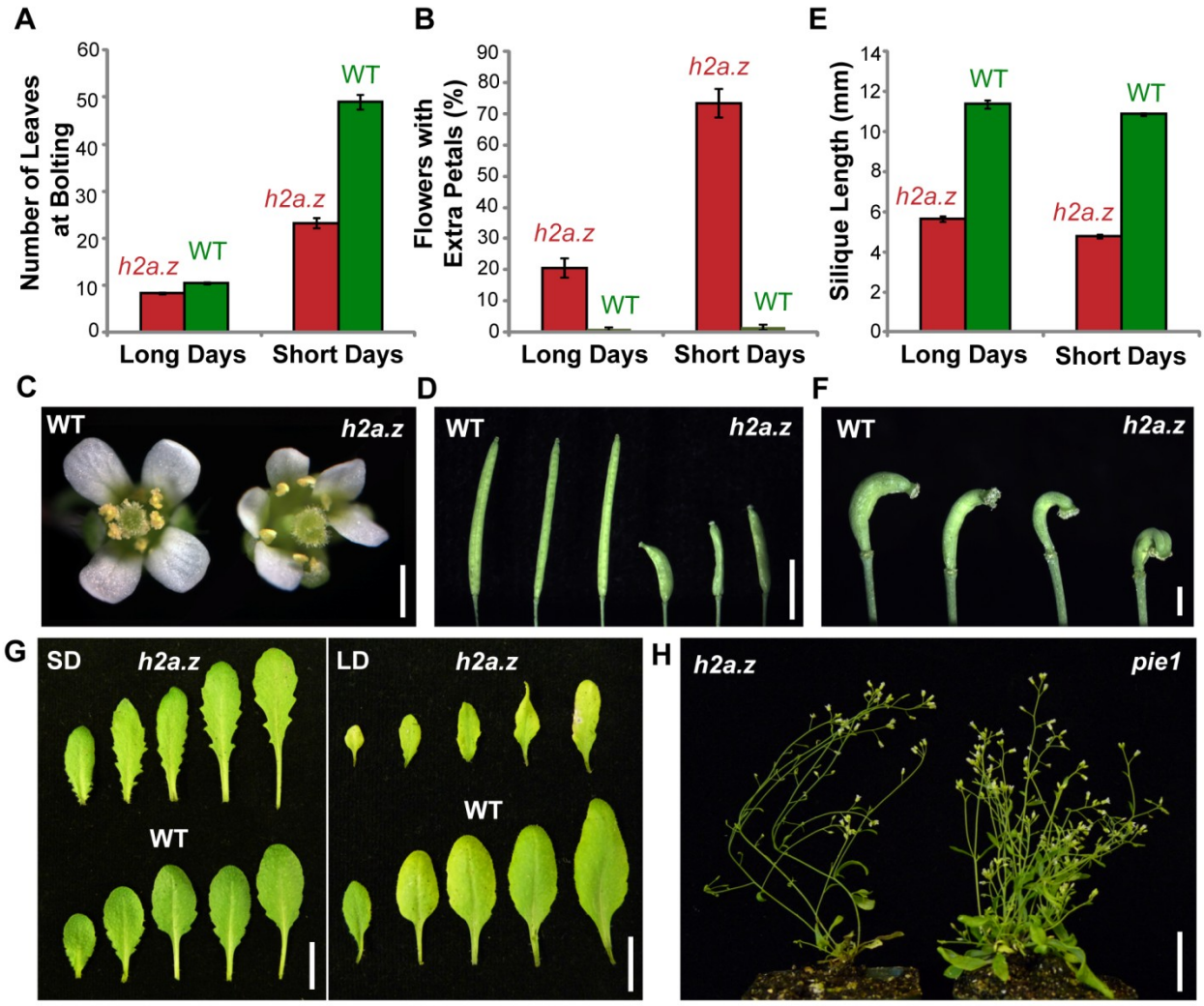

Figure 3. Characterization of the *h2a.z* mutant phenotype. (A) Average flowering time of the *h2a.z* mutant and WT control, as measured by the average number of rosette leaves at the time of bolting (cotyledons not included). Measurements were taken for 50 *h2a.z* and WT plants in both long day (LD, 16h light/ 8h dark) and short day (SD, 8h light/ 16h dark) conditions. (B) Average percentage of early flowers with extra petals, as measured in the first ten flowers of *h2a.z* mutant and WT plants. Measurements were averaged for 50 *h2a.z* and WT plants in both LD (16h light/ 8h dark) and SD (8h light/ 16h dark) conditions. (C) *h2a.z* mutant (right) and WT (left) flowers grown in SD. Scale bar = 2mm. (D) *h2a.z* mutant (right) and WT (left) siliques. Scale bar = 4mm. (E) Average mature silique length in *h2a.z* mutant and WT plants. Measurements were averaged for 10 mature siliques from each of 50 *h2a.z* and WT plants in both LD (16h light/ 8h dark) and SD (8h light/ 16h dark) conditions. (F) Strong curvature of developing *h2a.z* mutant siliques. Scale bar = 1mm. (G) Rosette leaf morphology in *h2a.z* mutant and WT plants grown in SD (40 days; left panel) and in LD (28 days; right panel). Scale bar = 1cm. (H) Phenotypes of *h2a.z* and *pie1-5* mutants grown in LD at 6 weeks post germination. Scale bar = 1cm.

Figure 4

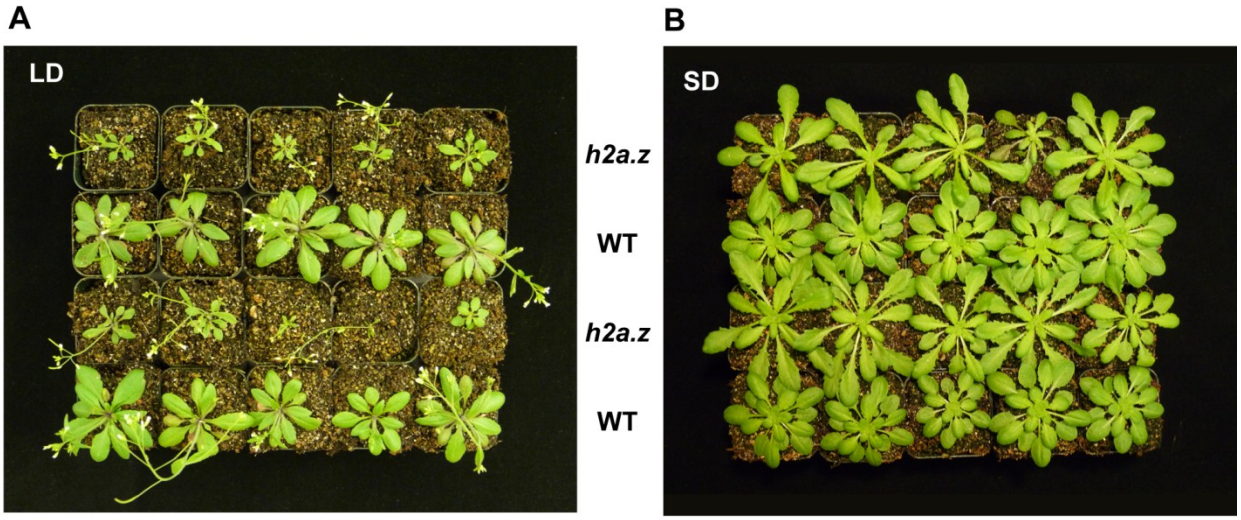

Figure 4. Phenotypes of the *h2a.z* mutant grown in long and short day conditions. Phenotypes of *h2a.z* mutant and WT plants at 28 days post germination grown in LD conditions (A) and 40 days post germination grown in SD conditions (B).

Figure 5

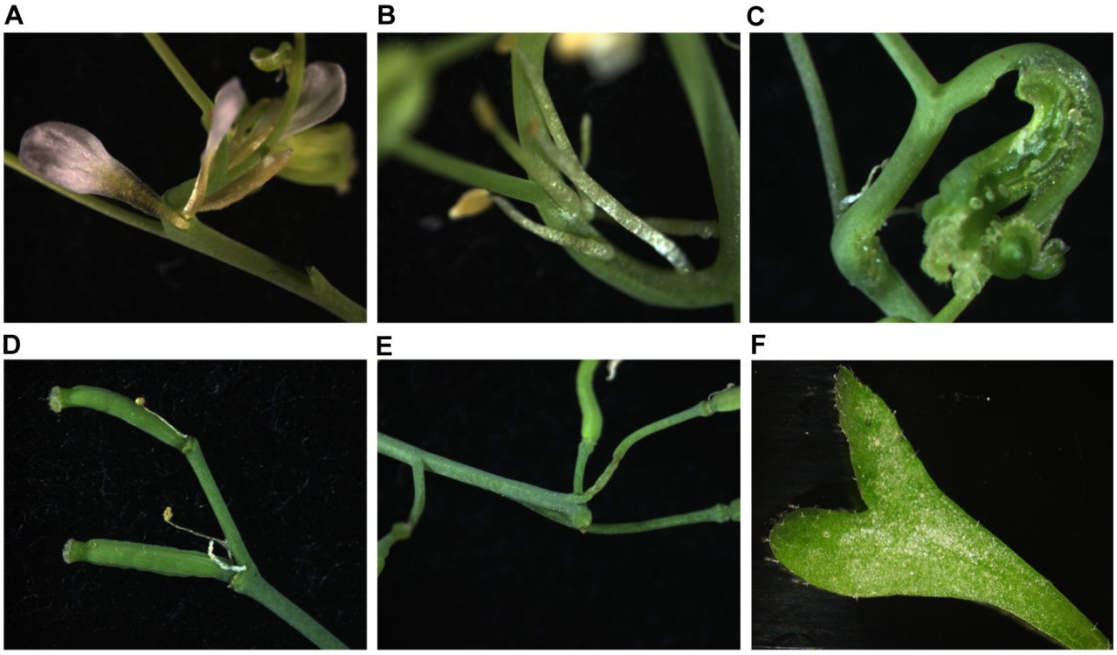

Figure 5. Rare developmental phenotypes of the *h2a.z* mutant. Additional uncommon phenotypes found in the *h2a.z* mutant. (A) and (B) Inappropriate emergence of petals and stamens directly from stem structures. (C) Abnormal floral morphology with ovules presented externally. **(**D) Transformation of outer sepal into floral inflorescence. (E) Phyllotaxy defects. (F) Abnormal leaf morphology.

Figure 6

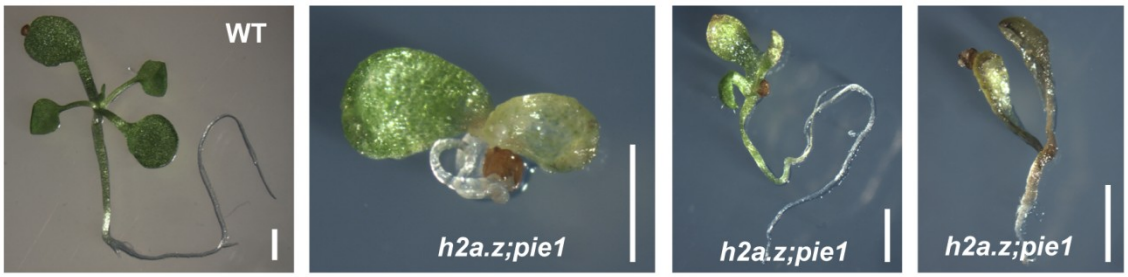

Figure 6. Phenotypes of the *h2a.z;pie1* mutant phenotype. Phenotypes of WT and *h2a.z;pie1* double mutants at 14 days post germination. Scale bar = 1mm.

Figure 7

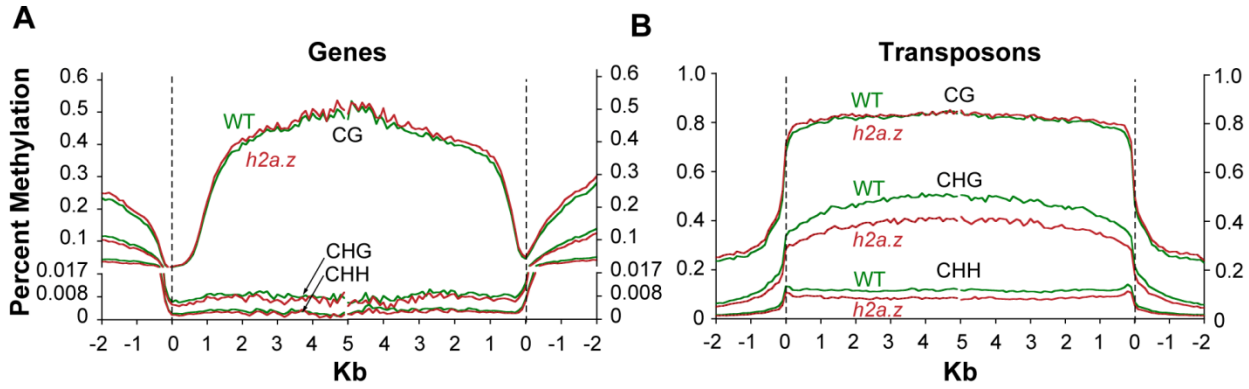

Figure 7. DNA methylation profiles of *h2a.z*-related mutants. (A) Profiles of CG, CHG, and CHH DNA methylation in *h2a.z* and WT. Genes were aligned at the 5′ end (left panel) and the 3′ end (right panel) and average methylation levels for each 100-bp interval are plotted from 2 kb away from the gene (negative numbers) to 5 kb into the gene (positive numbers). WT methylation is represented by the green traces, while *h2a.z* methylation is represented by red traces. The dashed lines at zero represents the point of alignment. The Y-axis was partitioned at 0.017 and the lower portion expanded to aid in the visibility of CHG and CHH traces. (B) Profiles of CG, CHG, and CHH DNA methylation in transposons in *h2a.z* and WT, aligned as genes were in (A). WT methylation is represented by the green traces, while *h2a.z* methylation is represented by red traces.

Figure 8

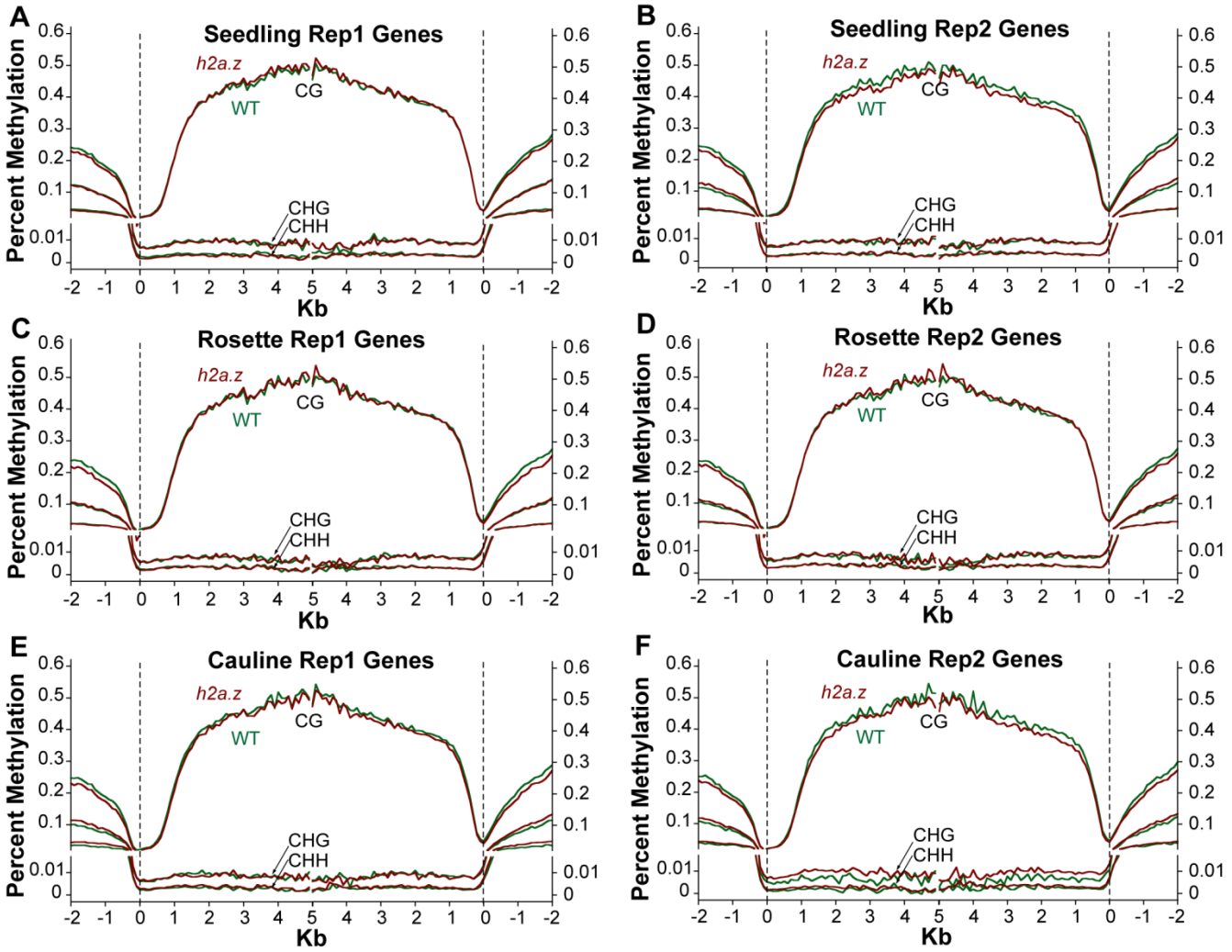

Figure 8. Loss of H2A.Z does not affect genic DNA methylation. Profiles of CG, CHG, and CHH DNA methylation in *h2a.z* and WT, for two independent replicates of seedlings (A-B), rosettes (C-D), and cauline leaves (E-F). Genes were aligned as in (Figure3). WT methylation is represented by the green traces, while *h2a.z* methylation is represented by red traces. The Y-axis was partitioned at 0.017 and the lower portion expanded to aid in the visibility of CHG and CHH traces.

Figure 9

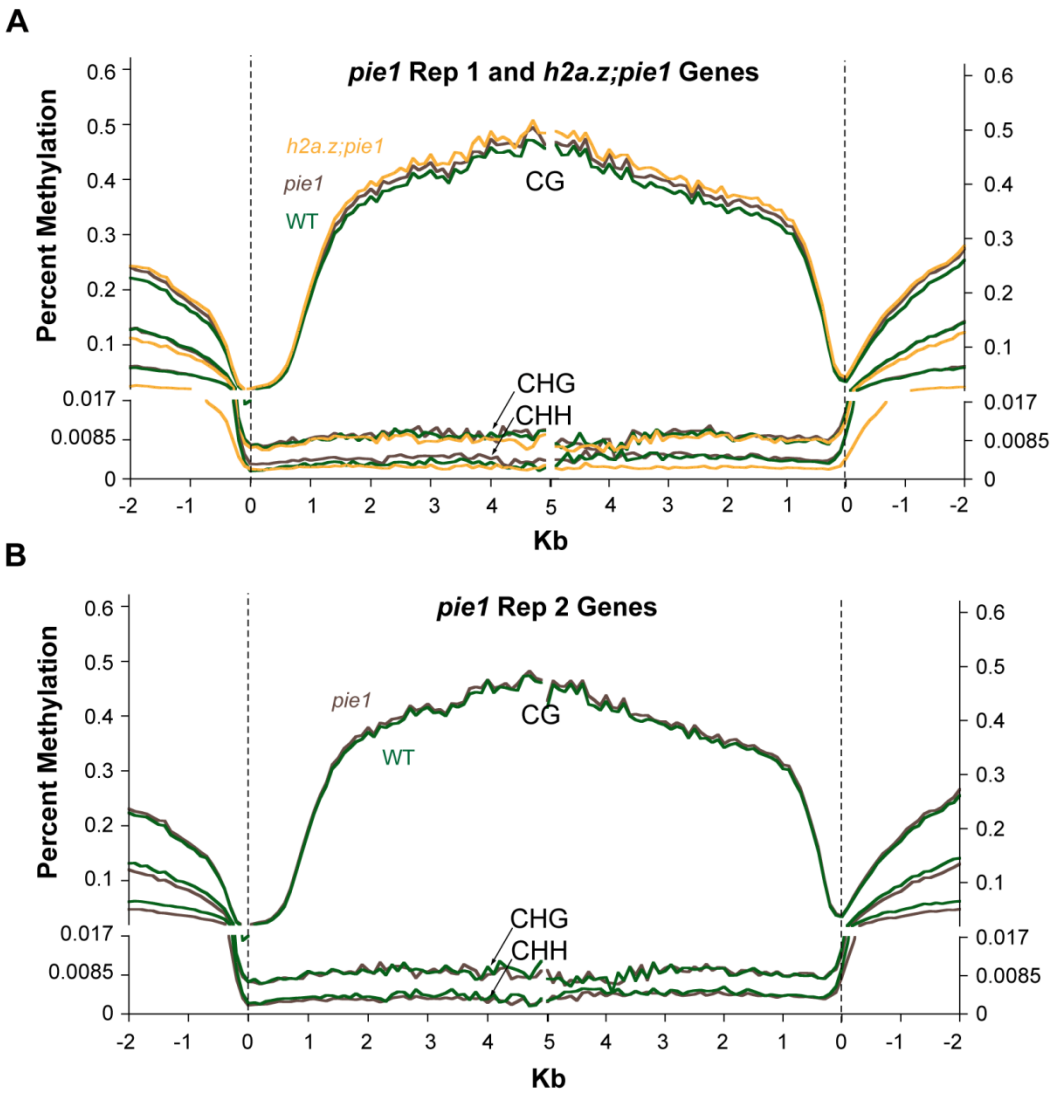

Figure 9. Loss of PIE1 does not affect genic DNA methylation. Profiles of CG, CHG, and CHH DNA methylation in (A) for one replicate of *pie1*, *h2a.z;pie1*, and WT, and an additional replicate each of *pie1* and WT in (B). Genes were aligned as in (Figure 3) and average methylation levels for each 100-bp interval are plotted from 2 kb away from the gene (negative numbers) to 5 kb into the gene (positive numbers). WT methylation is represented by the green traces; *pie1* methylation is represented by brown traces; *h2a.z;pie1* methylation is represented by yellow traces. The Y-axis was partitioned at 0.017 and the lower portion expanded to aid in the visibility of CHG and CHH traces.

Figure 10

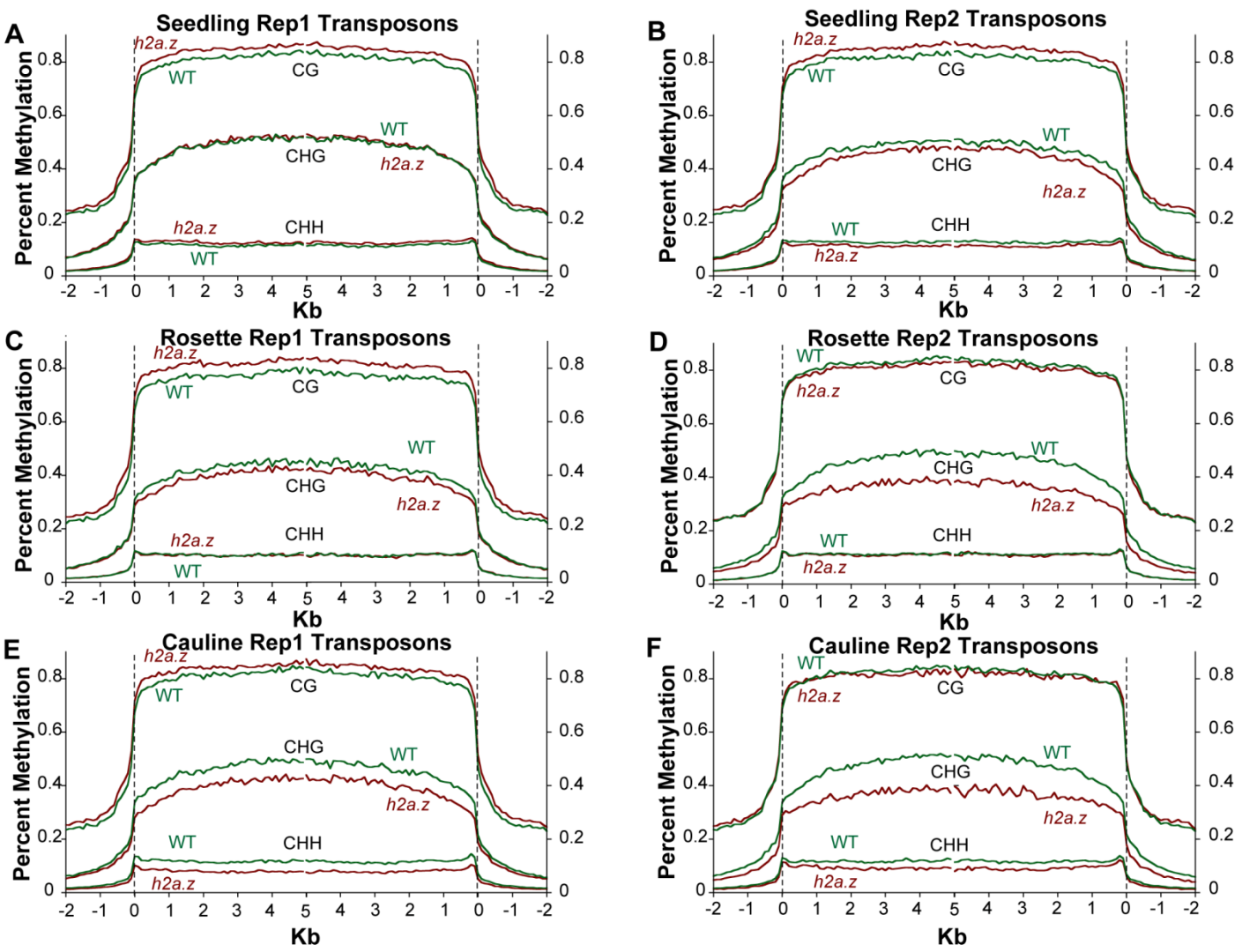

Figure 10. The *h2a.z* mutant exhibits changes in transposon methylation. Profiles of CG, CHG, and CHH DNA methylation in *h2a.z* and WT, for two independent replicates of seedlings (A-B), rosettes (C-D), and cauline leaves (E-F). Transposons were aligned as in (Figure3). WT methylation is represented by the green traces, while *h2a.z* methylation is represented by red traces.

Figure 11

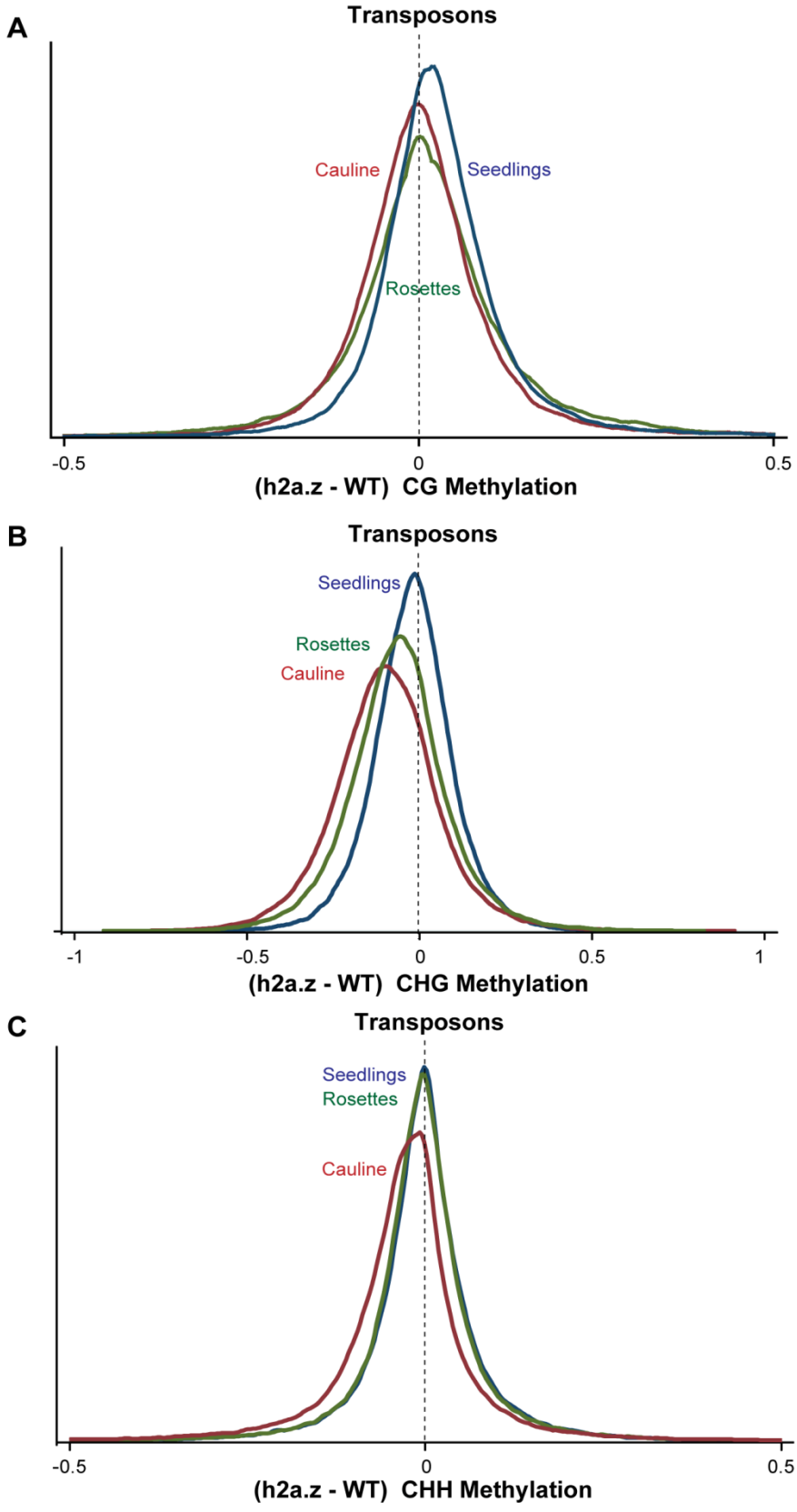

Figure 11. *h2a.z* induces global changes in transposon methylation. Kernel density plots, which have the effect of tracing the frequency distribution, for the differences between *h2a.z* and WT CG (A), CHG (B), and CHH (C) methylation in TEs for seedlings (blue traces), rosettes (green traces) and cauline leaves (red traces). The distributions of methylation differences for each 50bp window located within TEs and having average levels of methylation greater than zero in both *h2a.z* and WT are shown.

Figure 12

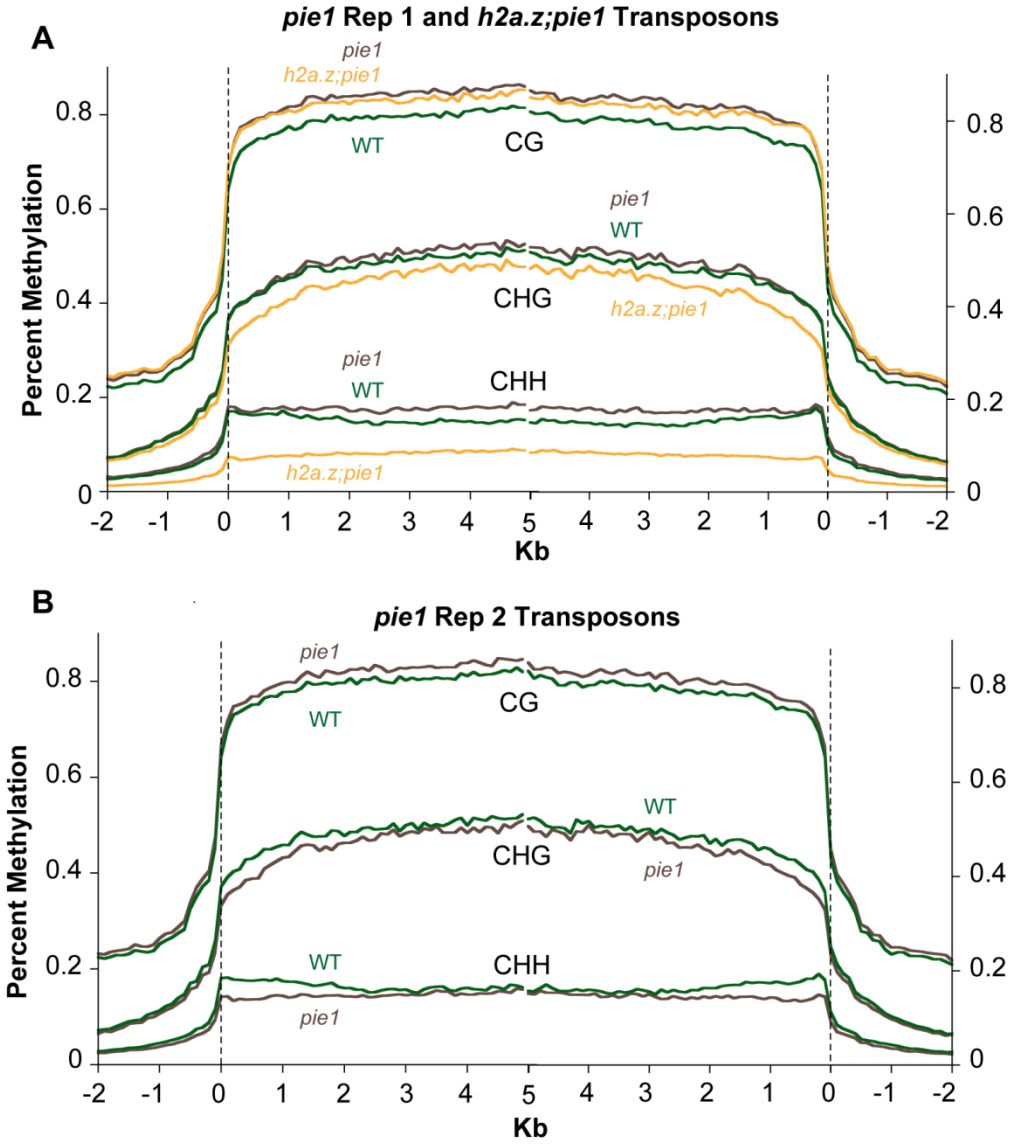

Figure 12. *h2a.z;pie1* causes greater loss of transposon methylation than either *h2a.z* or *pie1*. (A) Profiles of CG, CHG, and CHH DNA methylation in TEs in seedlings for one replicate each of *pie1*, *h2a.z;pie1* and WT. TEs were aligned as in (Figure3) and average methylation levels for each 100-bp interval are plotted. WT methylation is represented by green traces, *pie1* by brown traces, and *h2a.z;pie1* by yellow traces. An additional replicate each of *pie1* and WT are shown in (B).

Figure 13

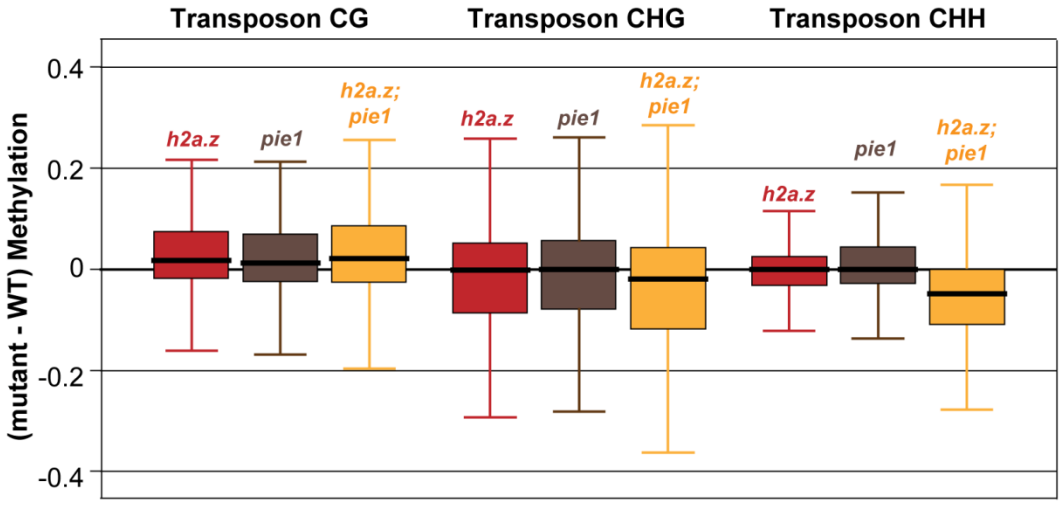

Figure 13. *h2a.z;pie1* causes greater loss of transposon methylation than either *h2a.z* or *pie1*. Box plots of differences in CG, CHG and CHH methylation between WT and either *h2a.z*, *pie1* or *h2a.z;pie1* seedlings for all 50 bp windows within TEs. Each box encloses the middle 50% of the distribution, with the horizontal black line marking the median. The lines extending vertically from each box mark the minimum and maximum values that fall within 1.5 times the height of the box.

Figure 14

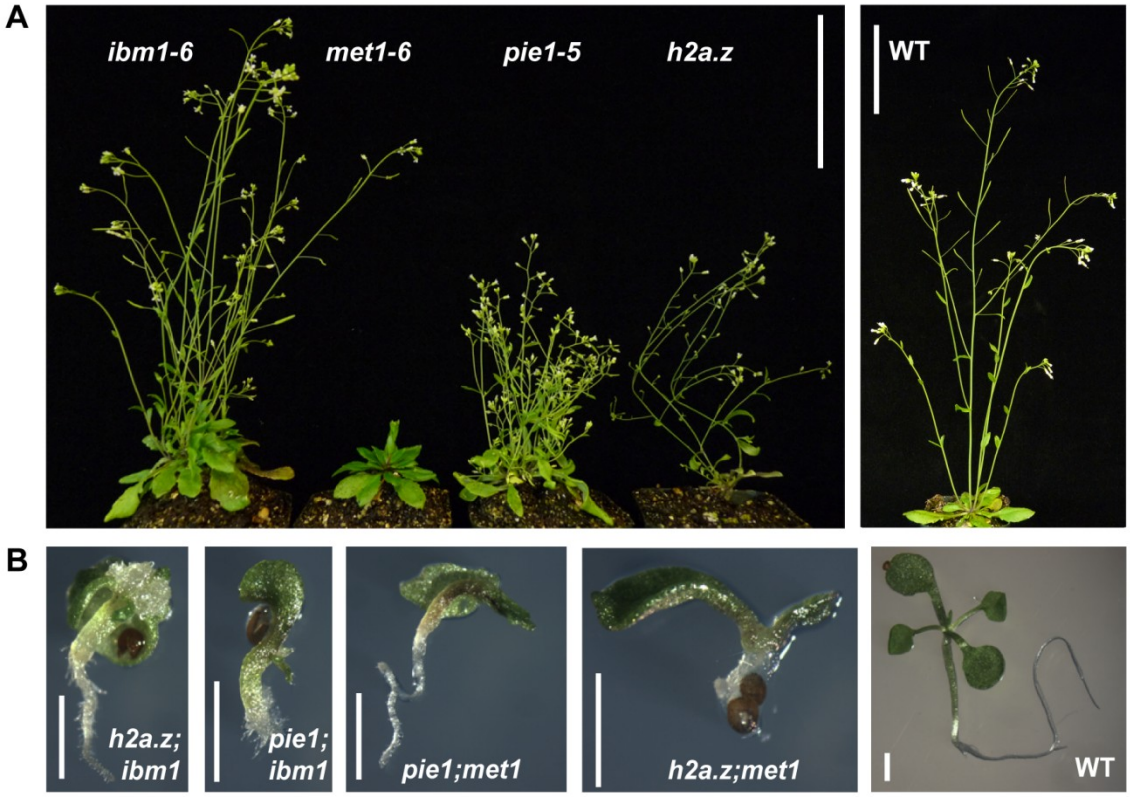

Figure 14. Phenotypes of *h2a.z*-related and methylation double mutants. (A) Phenotypes of *ibm1-6*, *met1-6* , *pie1-5*, *h2a.z* and WT plants grown in LD conditions at 6 weeks post germination. Scale bars = 3cm. (B) Phenotypes of *h2a.z*;*ibm1*, *h2a.z*;*met1*, *pie1*;*ibm1*, *pie1*;*met1* double mutants and WT seedling at 14 days post germination grown in LD conditions. Scale bars = 1mm.

Figure 15

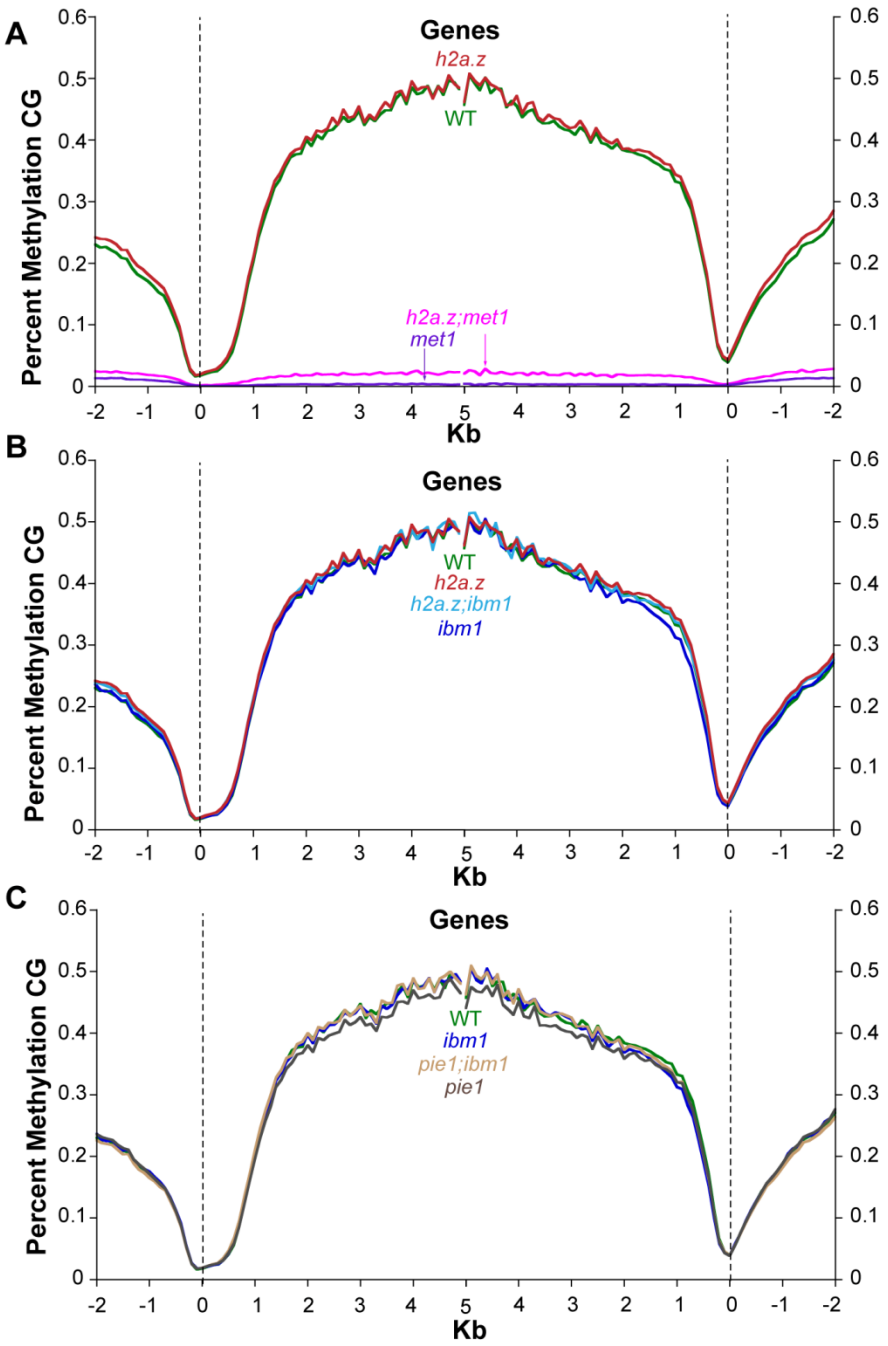

Figure 15. Genic CG DNA methylation profiles of H2A.Z-deficient and DNA methylation-perturbed double mutants. (A) Profiles of CG DNA methylation in *h2a.z*, *h2a.z;met1*, *met1* and WT. Genes were aligned as in (Figure3). WT methylation is represented by the green trace, *h2a.z* methylation by the red trace, *met1* methylation by the purple trace, and *h2a.z;met1* methylation by the pink trace. The dashed line at zero represents the point of alignment. (B) Profiles of CG DNA methylation in genes, as in (A), for *h2a.z* (red trace), *h2a.z;ibm1* (light blue trace), *ibm1* (dark blue trace) and WT (green trace). (C) Profiles of CG DNA methylation in genes, as in (A), for *pie1* (brown trace), *pie1;ibm1* (beige trace), *ibm1* (dark blue trace) and WT (green trace).

Figure 16

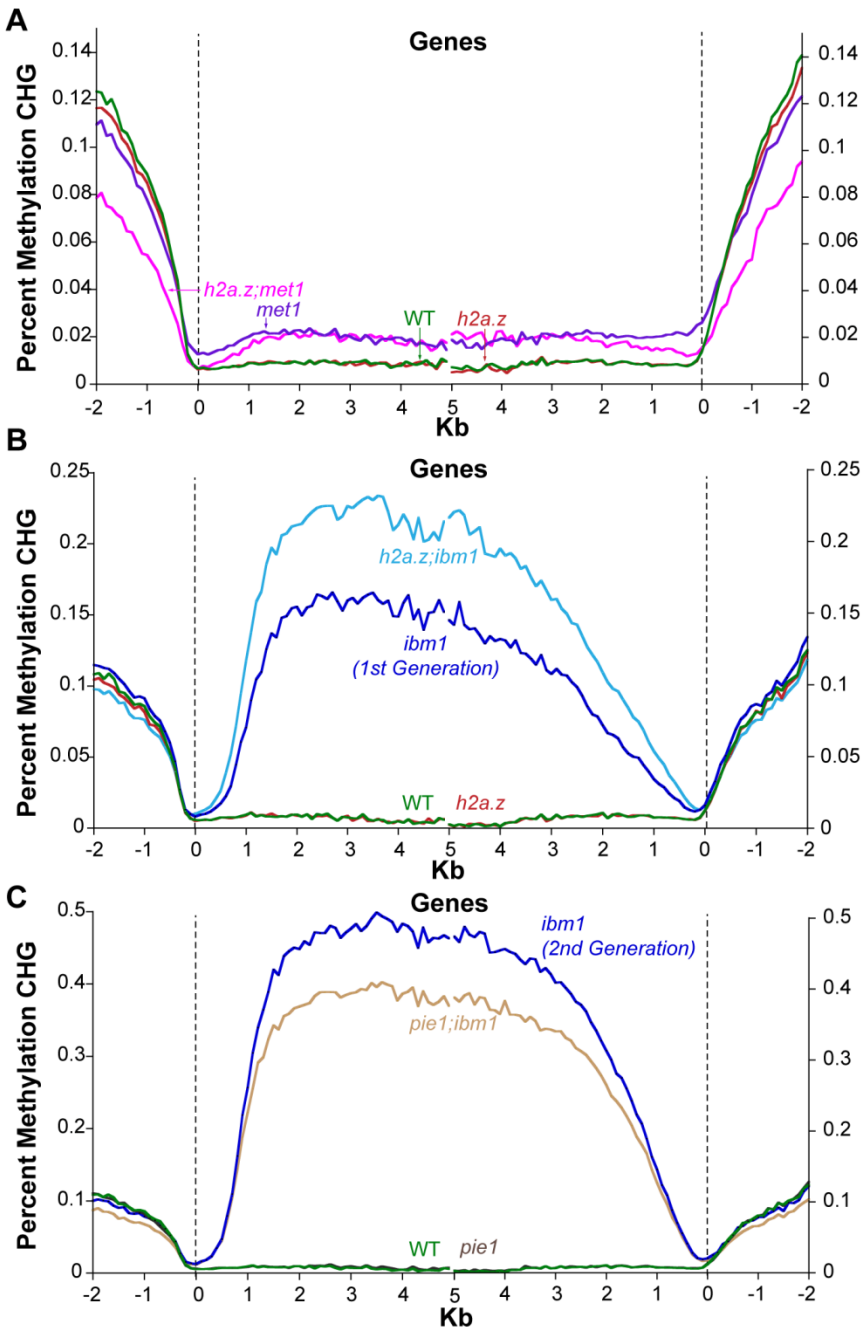

Figure 16. Genic CHG DNA methylation profiles of H2A.Z-deficient and DNA methylation-perturbed double mutants. (A) Profiles of CHG DNA methylation in *h2a.z*, *h2a.z;met1*, *met1* and WT. Genes were aligned as in (Figure 3). WT methylation is represented by the green trace, *h2a.z* methylation by the red trace, *met1* methylation by the purple trace, and *h2a.z;met1* methylation by the pink trace. The dashed line at zero represents the point of alignment. (B) Profiles of CHG DNA methylation in genes, as in (A), for *h2a.z* (red trace), *h2a.z;ibm1* (light blue trace), *ibm1* (dark blue trace) and WT (green trace). (C) Profiles of CHG DNA methylation in genes, as in (A), for *pie1* (brown trace), *pie1;ibm1* (beige trace), *ibm1* (dark blue trace) and WT (green trace).

Figure 17

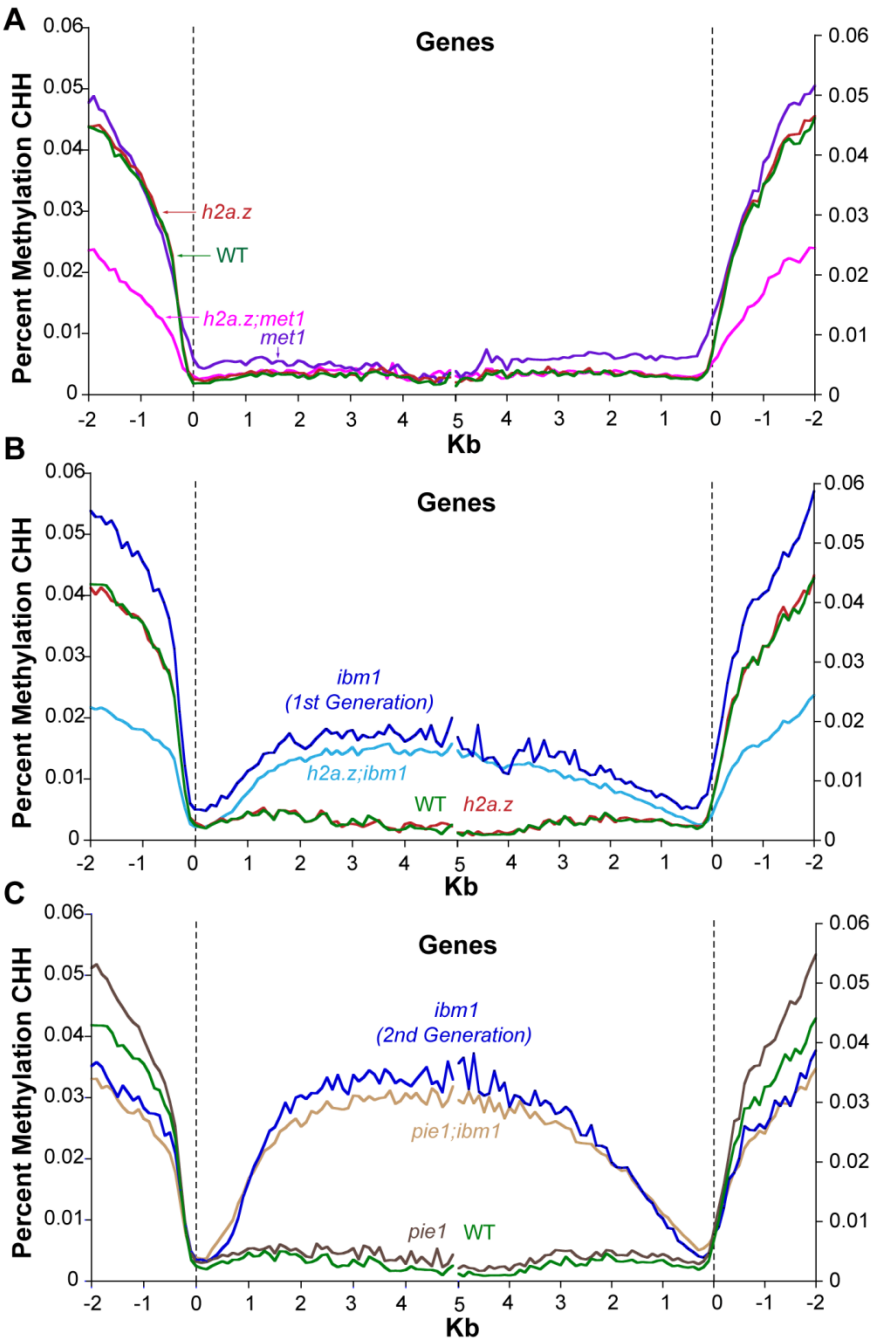

Figure 17. Genic CHH DNA methylation profiles of H2A.Z-deficient and DNA methylation-perturbed double mutants. (A) Profiles of CHH DNA methylation in *h2a.z*, *h2a.z;met1*, *met1* and WT. Genes were aligned as in (Figure3). WT methylation is represented by the green trace, *h2a.z* methylation by the red trace, *met1* methylation by the purple trace, and *h2a.z;met1* methylation by the pink trace. The dashed line at zero represents the point of alignment. (B) Profiles of CHH DNA methylation in genes, as in (A), for *h2a.z* (red trace), *h2a.z;ibm1* (light blue trace), *ibm1* (dark blue trace) and WT (green trace). (C) Profiles of CHH DNA methylation in genes, as in (A), for *pie1* (brown trace), *pie1;ibm1* (beige trace), *ibm1* (dark blue trace) and WT (green trace).

Figure 18

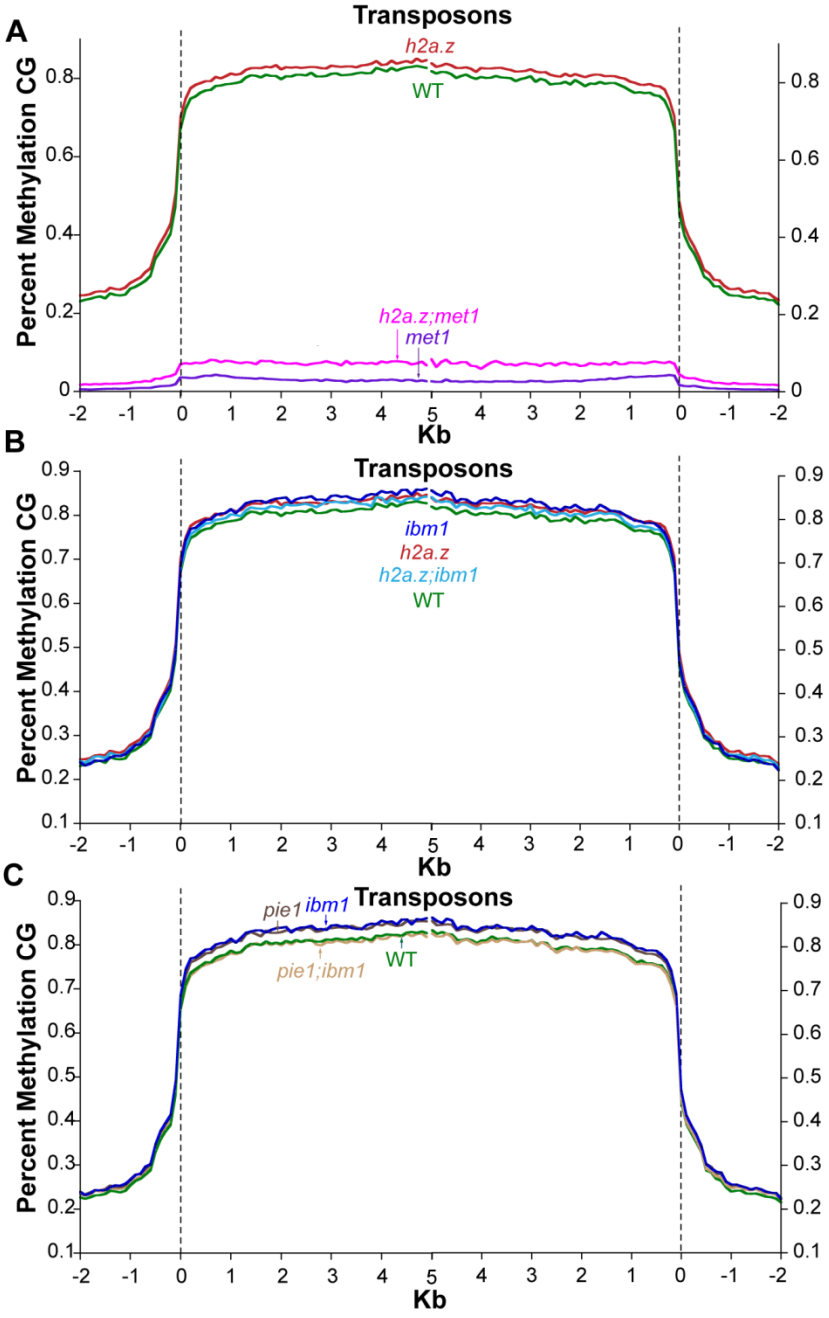

Figure 18. Transposon CG DNA methylation profiles of H2A.Z-deficient and DNA methylation-perturbed double mutants. (A) Profiles of CG DNA methylation in *h2a.z;met1*, *h2a.z*, *met1* and WT. Transposons were aligned as in (Figure3) WT methylation is represented by the green trace, *h2a.z* methylation by the red trace, *met1* methylation by the purple trace, and *h2a.z;met1* methylation by the pink trace. The dashed line at zero represents the point of alignment. (B) Profiles of CG DNA methylation in TEs, as in (A), for *h2a.z* (red trace), *h2a.z;ibm1* (light blue trace), *ibm1* (dark blue trace) and WT (green trace). (C) Profiles of CG DNA methylation in TEs, as in (A), for *pie1* (brown trace), *pie1;ibm1* (beige trace), *ibm1* (dark blue trace) and WT (green trace).

Figure 19

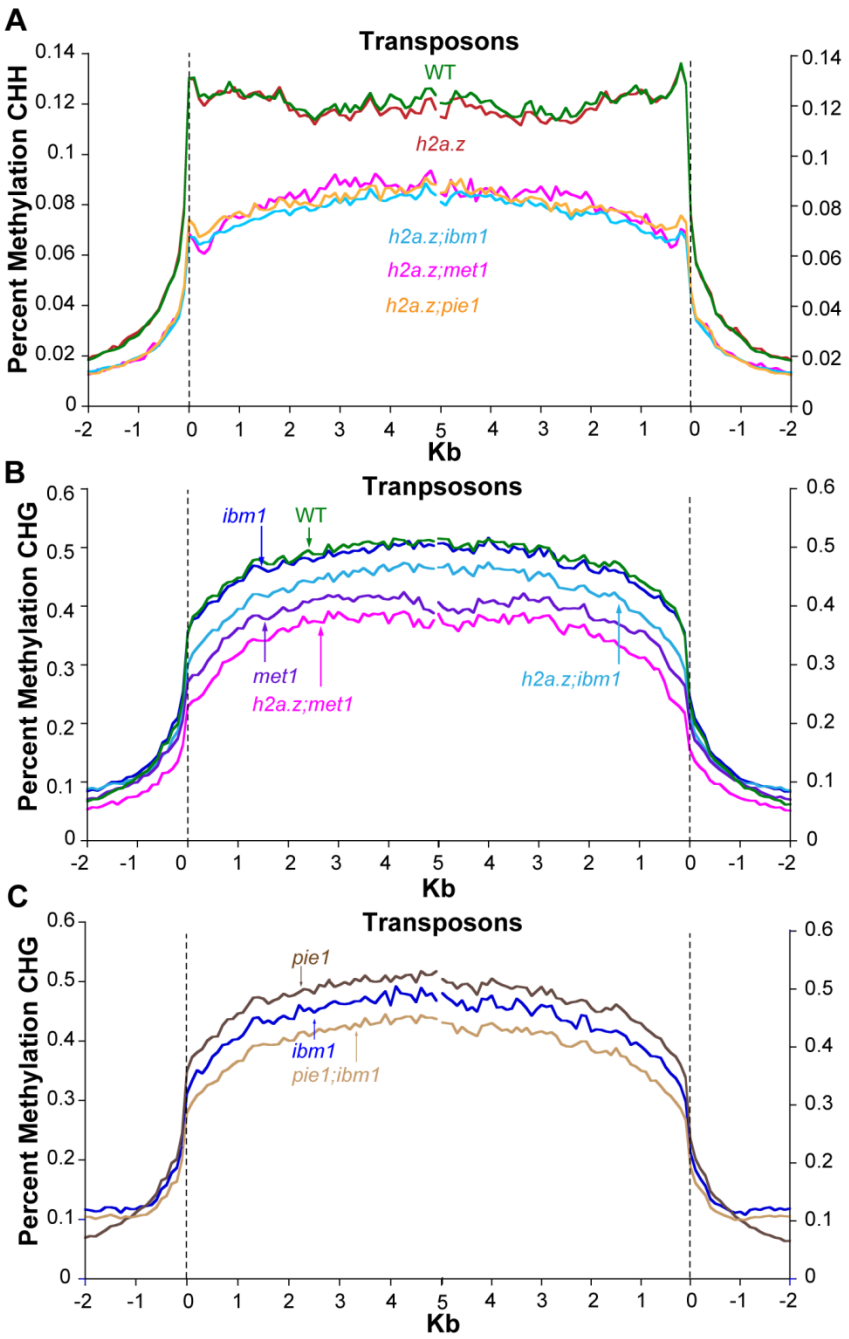

Figure 19. DNA methylation profiles of *h2a.z*-related and methylation double mutants. (A) Profiles of average CHH DNA methylation in TEs. TEs were aligned and average methylation levels determined as in as in (Figure3). WT methylation is represented by the green trace, *h2a.z* by the red trace, *h2a.z;met1* by the pink trace, *h2a.z;ibm1* by the light blue trace, and *h2a.z;pie1* by the orange trace. (B) Profiles of average CHG DNA methylation in TEs, as in (A). WT methylation is represented by the green trace, *met1* by the purple trace, *h2a.z;met1* by the pink trace, and *h2a.z;ibm1* by the light blue trace, and *ibm1* by the dark blue trace. (C) Profiles of CHG DNA methylation in *pie1* (brown trace), *pie1;ibm1* (beige trace), and *ibm1* (dark blue trace).

Figure 20

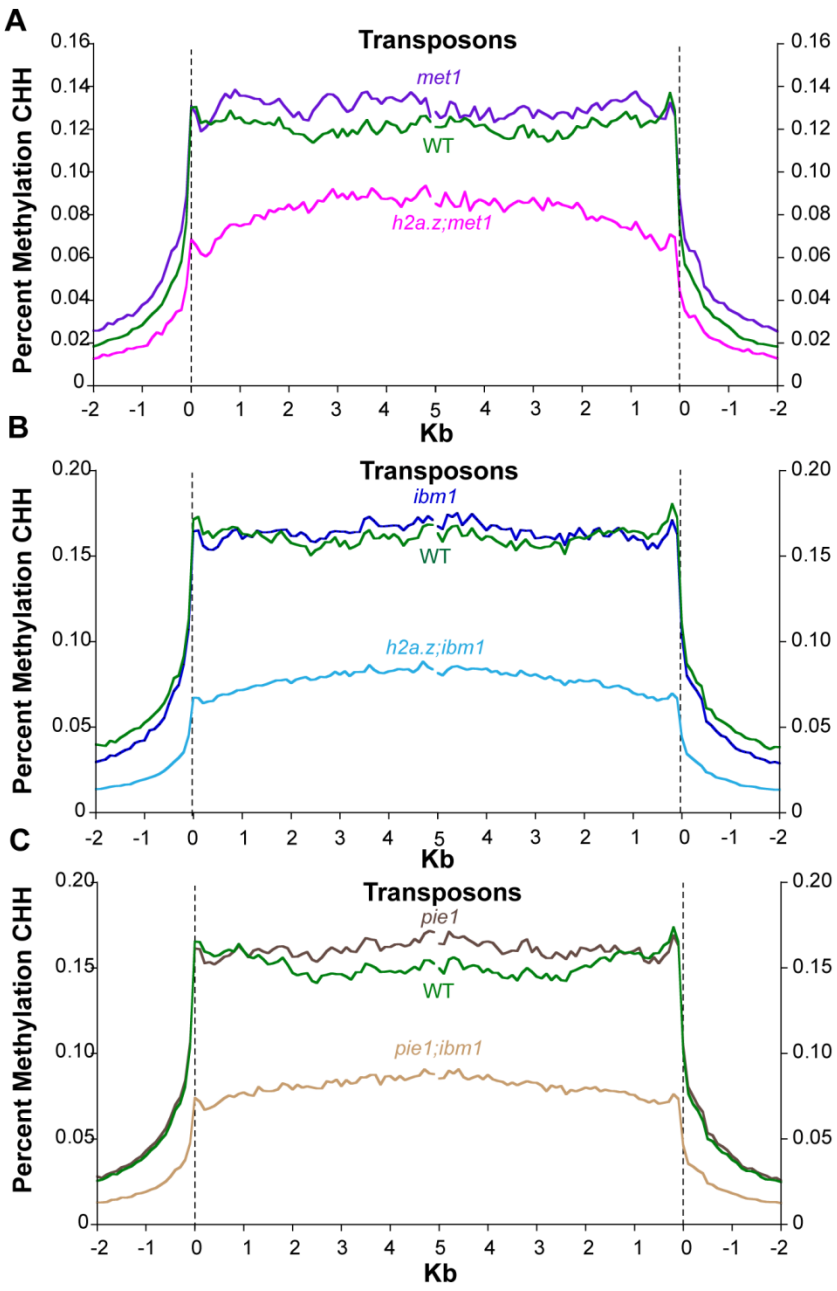

Figure 20. *h2a.z*-related and methylation double mutants show greater loss of TE CHH methylation. (A) Profiles of CHH DNA methylation in *h2a.z;met1*, *met1* and WT. Transposons were aligned as in (Figure3) and average methylation levels for each 100-bp interval are plotted from 2 kb away from the TE to 5 kb into the TE. WT methylation is represented by the green trace, *met1* methylation by the purple trace, and *h2a.z;met1* methylation by the pink trace. The dashed line at zero represents the point of alignment. (B) Profiles of CHH DNA methylation in TEs, as in (A) for *h2a.z;ibm1* (light blue trace), *ibm1* (dark blue trace) and WT (green trace). (C) Profiles of CHH DNA methylation in TEs, as in (A), for *pie1* (brown trace), *pie1;ibm1* (beige trace), and WT (green trace).

Figure 21

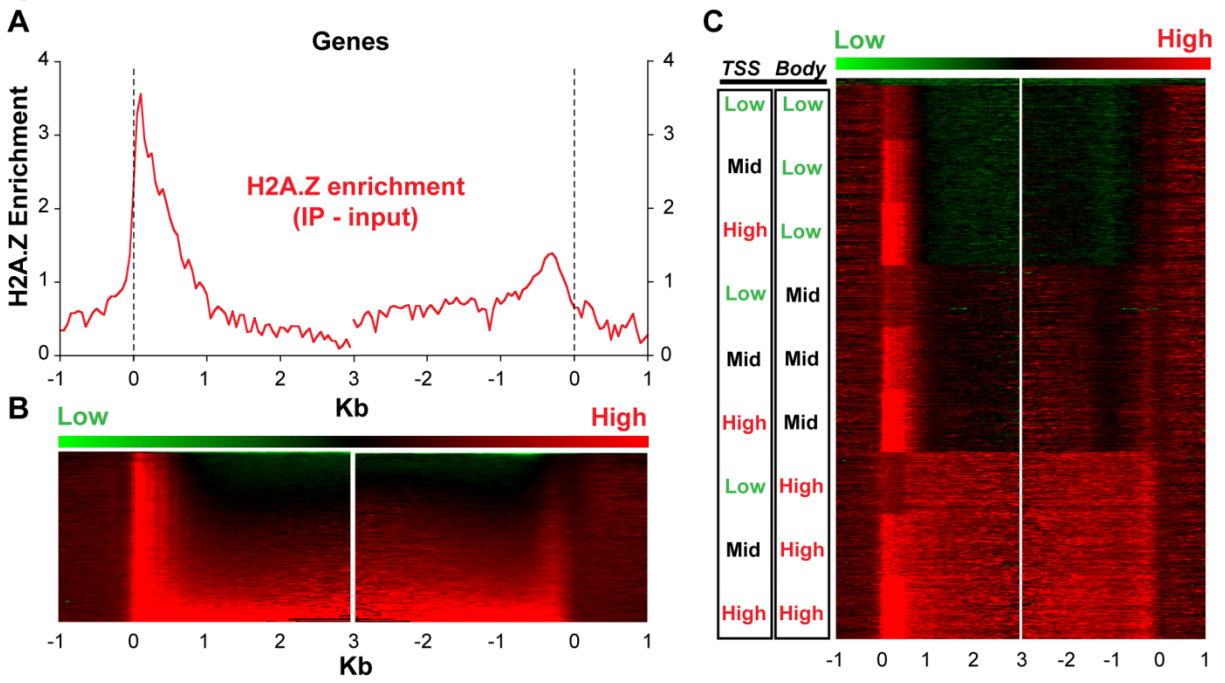

Figure 21. H2A.Z is enriched in gene bodies of some genes. (A) Average profile of H2A.Z enrichment (IP - input) across all genes, shown in red. Genes were aligned at the 5′ end (left panel) and the 3′ end (right panel) and average H2A.Z enrichment levels for each 50-bp interval are plotted from 1 kb away from the gene (negative numbers) to 3 kb into the gene (positive numbers). (B) A heat map of H2A.Z enrichment (IP - input) data described in (A). Genes were first sorted from lowest to highest H2A.Z enrichment score across their gene bodies (from 500bp after the TSS to 500bps before the 3' end), and then stacked vertically, one on top of another. The heat map represents the 5' (left) and 3' (right) ends. (C) The H2A.Z enrichment data were clustered based on levels of H2A.Z enrichment at the TSS (0 to 500bp) and across the gene body (1kb after TSS to 1kb from the 3' end). Genes (n = 12,237) were clustered into 9 approximately equal sized groups based on three tiers (low, mid, high) of H2A.Z enrichment at the TSS and across the body; 158 pseudogenes were removed from the category representing lowest enrichment and clustered together at the top of the heat map.

Figure 22

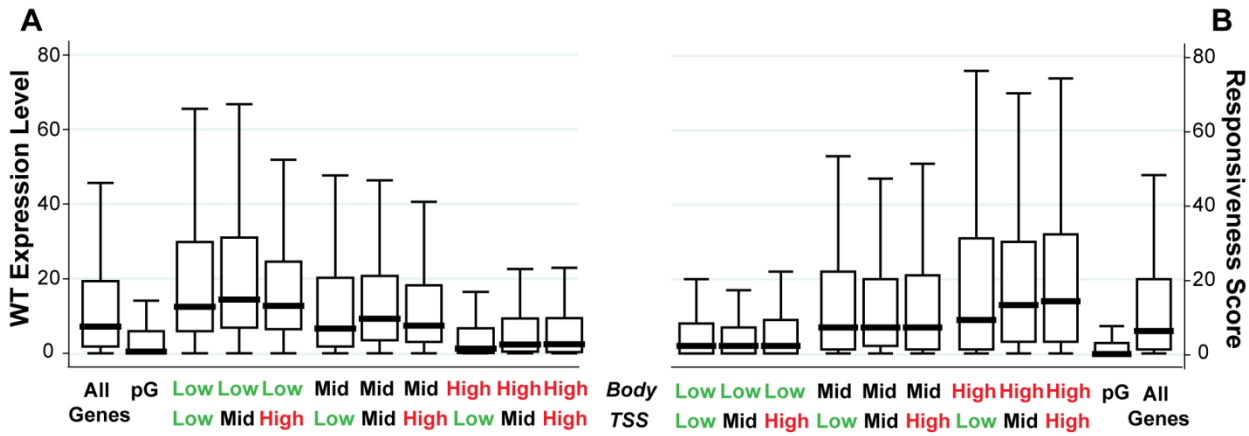

Figure 22. H2A.Z enrichment in gene bodies associated with lower expression and higher responsiveness. (A) and (B) Box plots of WT expression levels (A) and Responsiveness Score (B) for all genes partitioned by the nine H2A.Z-enrichment clusters in (Figure 22C**)**, as well as for all genes and for 158 pseudogenes (pG). Each box represents the middle 50% of the distribution, with the horizontal black lines marking the medians. The lines extending vertically from each box represent the minimum and maximum values that fall within 1.5 times the height of the box.

Figure 23

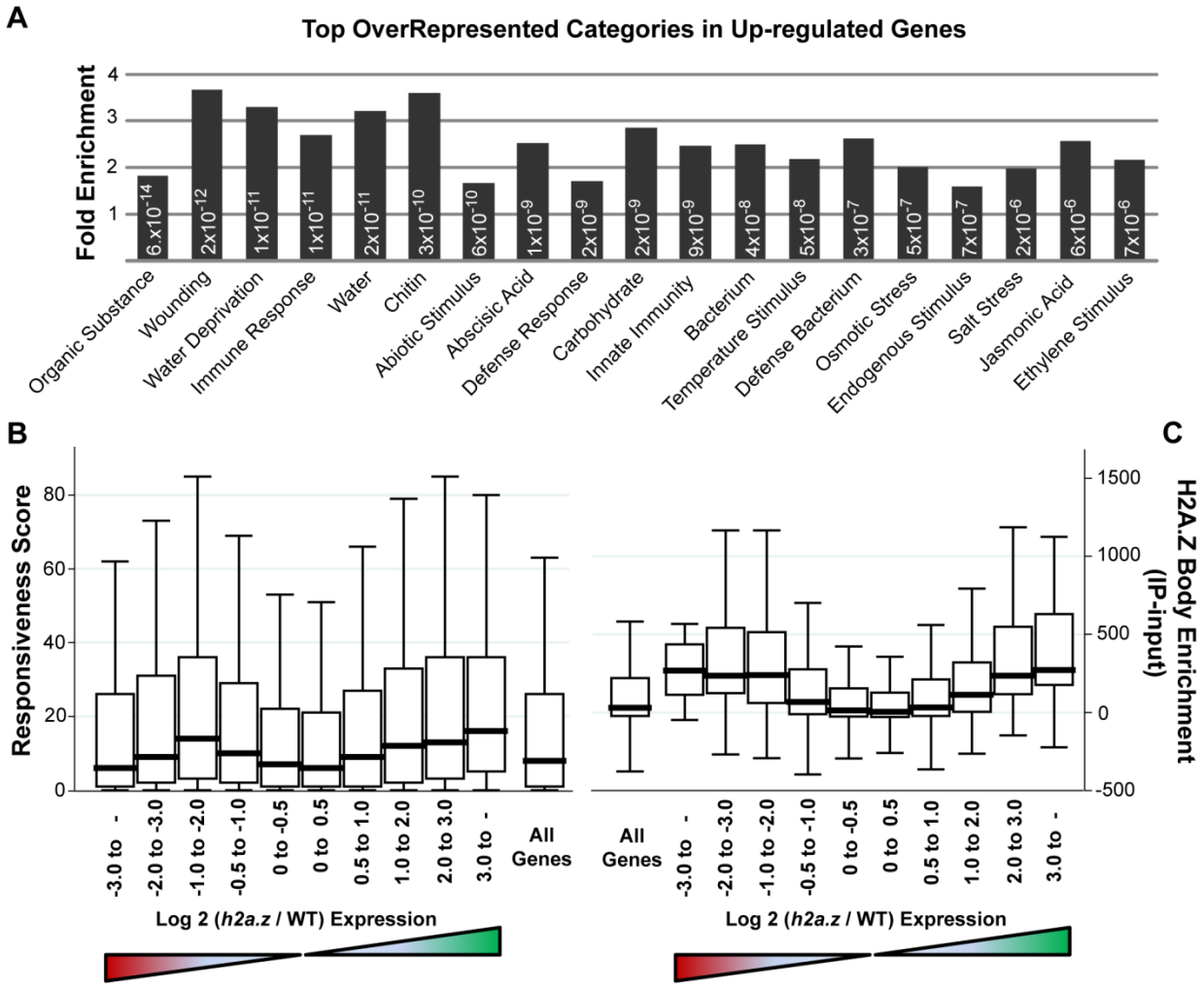

Figure 23. Genes with H2A.Z gene body enrichment are misregulated in *h2a.z*. (A) Fold enrichment of the top 19 over-represented GO Categories in the 1800 upregulated genes in the *h2a.z* mutant, all of which are response-related. For simplicity, all categories have had "Response to" removed from their names, with the exception of "Immune response" and "Defense response". All categories have P-values for over-representation less than $1x10^{-5}$, and each P-value is indicated within the respective bar. (B) Box plots of Responsiveness Score for all genes partitioned by the degree of up and down regulation in the *h2a.z* mutant. Genes are grouped into bins based on increasing $\log_2$ (*h2a.z* / WT) expression level, ranging from -8.6 to 9.8; a separate bin shows the Responsiveness Score for all genes. The red and green triangles below the X-axis respectively represent decreased and increased expression in the mutant over WT. Each box represents the middle 50% of the distribution, with the horizontal black lines marking the medians. The lines extending vertically from each box represent the minimum and maximum values that fall within 1.5 times the height of the box. (C) Box plots of H2A.Z body-enrichment (IP - input) for all genes with H2A.Z body enrichment scores (n = 12,237) partitioned by degree of up and down regulation. Genes are grouped into bins as they are in (B).

Figure 24

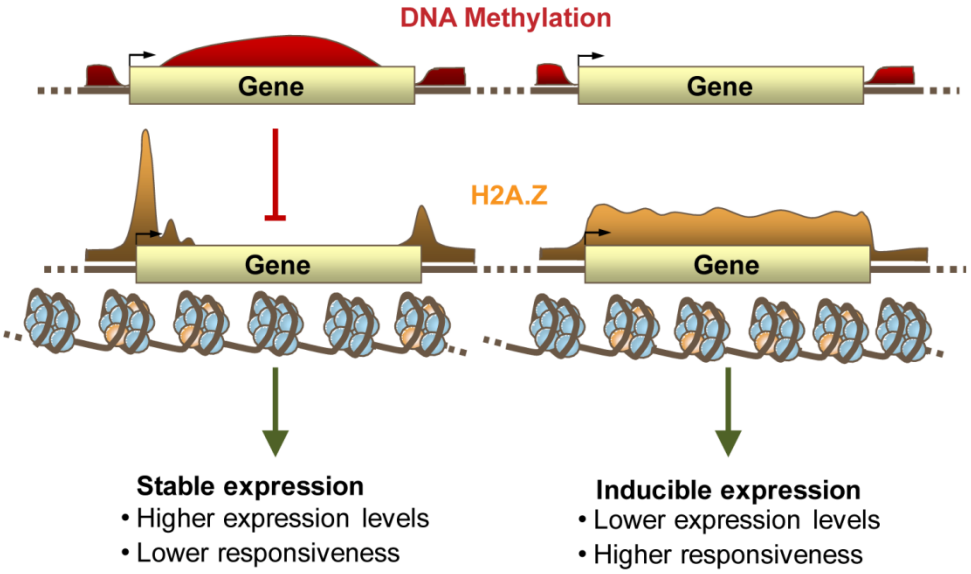

Figure 24. DNA Methylation, H2A.Z, and expression patterning. A proposed model for the relationship between DNA methylation (shown in red) and H2A.Z (shown in yellow) within genes, and their relationship with gene responsiveness and transcript level. Gene body methylation prevents the incorporation of H2A.Z within the bodies of highly and constitutively expressed genes. H2A.Z within unmethylated gene bodies regulates the expression of inducible genes.

Table 1. Overrepresented GO Terms in upregulated Genes in the h2a.z mutant					
GO Term ID	GO Term	Fold Enrichmen	P-value	Gene Count	%
GO:0010033	response to organic substance	1.82	6.92E-14	155	8.61
GO:0009611	response to wounding	3.67	2.36E-12	39	2.17
GO:0009414	response to water deprivation	3.30	1.23E-11	42	2.33
GO:0006955	immune response	2.69	1.27E-11	57	3.17
GO:0009415	response to water	3.22	1.70E-11	43	2.39
GO:0010200	response to chitin	3.59	2.63E-10	33	1.83
GO:0009628	response to abiotic stimulus	1.66	5.69E-10	144	8.00
GO:0009737	response to abscisic acid stimulus	2.52	1.20E-09	52	2.89
GO:0006952	defense response	1.71	1.61E-09	127	7.06
GO:0009743	response to carbohydrate stimulus	2.85	2.44E-09	41	2.28
GO:0045087	innate immune response	2.46	8.55E-09	49	2.72
GO:0009617	response to bacterium	2.49	3.83E-08	44	2.44
GO:0009266	response to temperature stimulus	2.19	4.83E-08	56	3.11
GO:0042742	defense response to bacterium	2.63	3.21E-07	35	1.94
GO:0006970	response to osmotic stress	2.01	5.06E-07	58	3.22
GO:0009719	response to endogenous stimulus	1.59	7.27E-07	112	6.22
GO:0009651	response to salt stress	1.99	2.42E-06	53	2.94
GO:0009753	response to jasmonic acid stimulus	2.57	6.21E-06	29	1.61
GO:0009723	response to ethylene stimulus	2.16	6.98E-06	40	2.22
GO:0016265	death	2.12	1.40E-05	39	2.17
GO:0008219	cell death	2.12	1.40E-05	39	2.17
GO:0009725	response to hormone stimulus	1.52	1.98E-05	100	5.56
GO:0009409	response to cold	2.14	2.76E-05	36	2.00
GO:0009408	response to heat	2.61	3.53E-05	24	1.33
GO:0006468	protein amino acid phosphorylation	1.45	4.13E-05	112	6.22
GO:0006793	phosphorus metabolic process	1.40	5.77E-05	129	7.17
GO:0006796	phosphate metabolic process	1.40	5.83E-05	129	7.17
GO:0009738	abscisic acid mediated signaling	3.18	6.08E-05	17	0.94
GO:0042430	indole and derivative metabolic process	3.95	1.41E-04	12	0.67
GO:0042434	indole derivative metabolic process	3.95	1.41E-04	12	0.67
GO:0012501	programmed cell death	2.05	1.45E-04	33	1.83
GO:0042435	indole derivative biosynthetic process	4.11	2.13E-04	11	0.61
GO:0009873	ethylene mediated signaling pathway	2.20	2.14E-04	27	1.50

Table 1. Overrepresented GO Terms in upregulated Genes in the h2a.z mutant					
GO Term ID	GO Term	Fold Enrichmen	P-value	Gene Count	%
GO:0009755	hormone-mediated signaling	1.66	3.39E-04	52	2.89
GO:0032870	cellular response to hormone stimulus	1.66	3.39E-04	52	2.89
GO:0009751	response to salicylic acid stimulus	2.21	4.66E-04	24	1.33
GO:0009814	defense response, incompatible interaction	2.59	4.70E-04	18	1.00
GO:0016310	phosphorylation	1.35	6.35E-04	115	6.39
GO:0002252	immune effector process	4.81	8.81E-04	8	0.44
GO:0000160	two-component signal transduction system	1.90	0.00102284	30	1.67
GO:0010604	positive regulation of macromolecule metabolism	2.81	0.00117096	14	0.78
GO:0006790	sulfur metabolic process	2.02	0.00128555	25	1.39
GO:0009615	response to virus	3.55	0.00151676	10	0.56
GO:0006979	response to oxidative stress	1.74	0.00158526	36	2.00
GO:0010557	positive regulation of macromolecule biosynthesis	2.85	0.00162144	13	0.72
GO:0051607	defense response to virus	5.10	0.00167453	7	0.39
GO:0019748	secondary metabolic process	1.57	0.00205251	48	2.67
GO:0034976	response to endoplasmic reticulum stress	5.93	0.00224506	6	0.33
GO:0007242	intracellular signaling cascade	1.38	0.00275497	79	4.39
GO:0031328	positive regulation of cellular biosynthetic process	2.64	0.00317411	13	0.72
GO:0009891	positive regulation of biosynthetic process	2.64	0.00317411	13	0.72
GO:0019438	aromatic compound biosynthetic process	1.80	0.00343678	28	1.56
GO:0009626	plant-type hypersensitive response	3.07	0.00429606	10	0.56
GO:0034050	host programmed cell death induced by symbiont	3.01	0.00500638	10	0.56
GO:0006350	transcription	1.26	0.00610641	119	6.61
GO:0006915	apoptosis	1.85	0.00628089	23	1.28
GO:0010193	response to ozone	3.87	0.00753099	7	0.39
GO:0006972	hyperosmotic response	2.82	0.00769156	10	0.56
GO:0042538	hyperosmotic salinity response	2.96	0.00939344	9	0.50
GO:0044272	sulfur compound biosynthetic process	2.07	0.00983504	16	0.89
GO:0042343	indole glucosinolate metabolic process	7.90	0.01054432	4	0.22
GO:0006575	cellular amino acid derivative metabolic process	1.58	0.01149216	32	1.78
GO:0045859	regulation of protein kinase activity	3.46	0.01331705	7	0.39
GO:0043549	regulation of kinase activity	3.46	0.01331705	7	0.39
GO:0051173	positive regulation of nitrogen metabolism	2.42	0.01422949	11	0.61
GO:0045935	positive regulation of nucleoside	2.42	0.01422949	11	0.61

Table 1. Overrepresented GO Terms in upregulated Genes in the h2a.z mutant					
GO Term ID	GO Term	Fold Enrichmen	P-value	Gene Count	%
GO:0043405	regulation of MAP kinase activity	13.83	0.01488339	3	0.17
GO:0046283	anthocyanin metabolic process	4.94	0.01507846	5	0.28
GO:0009625	response to insect	4.94	0.01507846	5	0.28
GO:0002213	defense response to insect	6.92	0.0159735	4	0.22
GO:0052317	camalexin metabolic process	6.92	0.0159735	4	0.22
GO:0010120	camalexin biosynthetic process	6.92	0.0159735	4	0.22
GO:0045941	positive regulation of transcription	2.51	0.01618819	10	0.56
GO:0016311	dephosphorylation	2.14	0.01714487	13	0.72
GO:0010628	positive regulation of gene expression	2.47	0.01809196	10	0.56
GO:0009070	serine family amino acid biosynthetic process	3.23	0.01856596	7	0.39
GO:0009812	flavonoid metabolic process	2.43	0.02015341	10	0.56
GO:0051338	regulation of transferase activity	3.12	0.02165596	7	0.39
GO:0006355	regulation of transcription, DNA-dependent	1.23	0.02263294	97	5.39
GO:0009700	indole phytoalexin biosynthetic process	6.15	0.02269082	4	0.22
GO:0052314	phytoalexin metabolic process	6.15	0.02269082	4	0.22
GO:0046217	indole phytoalexin metabolic process	6.15	0.02269082	4	0.22
GO:0042431	indole metabolic process	6.15	0.02269082	4	0.22
GO:0052315	phytoalexin biosynthetic process	6.15	0.02269082	4	0.22
GO:0042398	cellular amino acid derivative biosynthetic process	1.64	0.02342174	23	1.28
GO:0051252	regulation of RNA metabolic process	1.23	0.02499476	97	5.39
GO:0009627	systemic acquired resistance	3.03	0.02507509	7	0.39
GO:0030001	metal ion transport	1.54	0.02552002	28	1.56
GO:0009759	indole glucosinolate biosynthetic process	10.37	0.0283451	3	0.17
GO:0009069	serine family amino acid metabolic process	2.44	0.02866933	9	0.50
GO:0042325	regulation of phosphorylation	2.93	0.0288358	7	0.39
GO:0010038	response to metal ion	1.40	0.029303	39	2.17
GO:0042440	pigment metabolic process	1.92	0.02961263	14	0.78
GO:0009718	anthocyanin biosynthetic process	5.53	0.03070506	4	0.22
GO:0006816	calcium ion transport	3.32	0.03077317	6	0.33
GO:0046686	response to cadmium ion	1.44	0.03177367	34	1.89
GO:0016138	glycoside biosynthetic process	2.23	0.03304627	10	0.56
GO:0009813	flavonoid biosynthetic process	2.35	0.03519354	9	0.50
GO:0033554	cellular response to stress	1.36	0.03619667	42	2.33

Table 1. Overrepresented GO Terms in upregulated Genes in the h2a.z mutant					
GO Term ID	GO Term	Fold Enrichmen	P-value	Gene Count	%
GO:0007568	aging	2.08	0.03633072	11	0.61
GO:0052543	callose deposition in cell wall	3.84	0.03663507	5	0.28
GO:0019757	glycosinolate metabolic process	2.51	0.03683604	8	0.44
GO:0019760	glucosinolate metabolic process	2.51	0.03683604	8	0.44
GO:0016143	S-glycoside metabolic process	2.51	0.03683604	8	0.44
GO:0016137	glycoside metabolic process	1.98	0.03874171	12	0.67
GO:0009696	salicylic acid metabolic process	5.03	0.04000096	4	0.22
GO:0006457	protein folding	1.49	0.04005913	27	1.50
GO:0019220	regulation of phosphate metabolic process	2.69	0.04226996	7	0.39
GO:0051174	regulation of phosphorus metabolic process	2.69	0.04226996	7	0.39
GO:0006855	multidrug transport	2.03	0.04264778	11	0.61
GO:0052386	cell wall thickening	3.64	0.04383433	5	0.28
GO:0009130	pyrimidine nucleoside monophosphate biosynthesis	8.30	0.04499988	3	0.17
GO:0009314	response to radiation	1.32	0.04528634	46	2.56
GO:0046148	pigment biosynthetic process	1.91	0.04814715	12	0.67
GO:0010224	response to UV-B	2.35	0.05024963	8	0.44
GO:0009631	cold acclimation	3.46	0.05176761	5	0.28
GO:0052545	callose localization	3.46	0.05176761	5	0.28
GO:0019758	glycosinolate biosynthetic process	2.86	0.05431908	6	0.33
GO:0019761	glucosinolate biosynthetic process	2.86	0.05431908	6	0.33
GO:0016144	S-glycoside biosynthetic process	2.86	0.05431908	6	0.33
GO:0009698	phenylpropanoid metabolic process	1.58	0.05441711	19	1.06
GO:0033037	polysaccharide localization	3.29	0.0604285	5	0.28
GO:0009911	positive regulation of flower development	2.77	0.0614697	6	0.33
GO:0015893	drug transport	1.90	0.06159593	11	0.61
GO:0045449	regulation of transcription	1.12	0.06394053	164	9.11
GO:0009129	pyrimidine nucleoside monophosphate metabolism	6.92	0.0643172	3	0.17
GO:0043255	regulation of carbohydrate biosynthetic process	6.92	0.0643172	3	0.17
GO:0032270	positive regulation of cellular protein metabolism	6.92	0.0643172	3	0.17
GO:0010050	vegetative phase change	6.92	0.0643172	3	0.17
GO:0051247	positive regulation of protein metabolic process	6.92	0.0643172	3	0.17
GO:0042493	response to drug	1.88	0.06594773	11	0.61
GO:0048507	meristem development	1.80	0.06705239	12	0.67

Table 1. Overrepresented GO Terms in upregulated Genes in the h2a.z mutant					
GO Term ID	GO Term	Fold Enrichmen	P-value	Gene Count	%
GO:0006470	protein amino acid dephosphorylation	1.95	0.06871375	10	0.56
GO:0042362	fat-soluble vitamin biosynthetic process	3.95	0.0751575	4	0.22
GO:0006775	fat-soluble vitamin metabolic process	3.95	0.0751575	4	0.22
GO:0033674	positive regulation of kinase activity	3.95	0.0751575	4	0.22
GO:0045860	positive regulation of protein kinase activity	3.95	0.0751575	4	0.22
GO:0046395	carboxylic acid catabolic process	1.83	0.07522805	11	0.61
GO:0016054	organic acid catabolic process	1.83	0.07522805	11	0.61
GO:0009699	phenylpropanoid biosynthetic process	1.62	0.07806967	15	0.83
GO:0009416	response to light stimulus	1.28	0.07994998	43	2.39
GO:0006811	ion transport	1.26	0.08456362	45	2.50
GO:0006984	ER-nuclear signaling pathway	5.93	0.08582635	3	0.17
GO:0010189	vitamin E biosynthetic process	5.93	0.08582635	3	0.17
GO:0042360	vitamin E metabolic process	5.93	0.08582635	3	0.17
GO:0051094	positive regulation of developmental process	2.25	0.08688677	7	0.39
GO:0006357	regulation of transcription from RNA polymerase II p	2.88	0.09063042	5	0.28

Table 2. Overrepreseted GO Terms in downregulated Genes in the h2a.z mutant					
GO Term ID	GO Term	Fold Enrichment	P-value	Gene Count	%
GO:0006412	translation	1.98	6.69E-08	68	11.07
GO:0034728	nucleosome organization	4.67	6.40E-04	9	1.47
GO:0006334	nucleosome assembly	4.67	6.40E-04	9	1.47
GO:0031497	chromatin assembly	4.54	7.76E-04	9	1.47
GO:0065004	protein-DNA complex assembly	4.48	8.52E-04	9	1.47
GO:0006323	DNA packaging	4.18	0.001328	9	1.47
GO:0042254	ribosome biogenesis	2.37	0.003255	16	2.61
GO:0009617	response to bacterium	2.35	0.003543	16	2.61
GO:0022613	ribonucleoprotein complex biogenesis	2.30	0.00427	16	2.61
GO:0006333	chromatin assembly or disassembly	3.46	0.00438	9	1.47
GO:0034621	cellular macromolecular complex subunit organization	2.38	0.004451	15	2.44
GO:0042742	defense response to bacterium	2.53	0.005481	13	2.12
GO:0034622	cellular macromolecular complex assembly	2.41	0.005596	14	2.28
GO:0043933	macromolecular complex subunit organization	2.11	0.007197	17	2.77
GO:0009651	response to salt stress	1.94	0.007602	20	3.26
GO:0006970	response to osmotic stress	1.88	0.008326	21	3.42
GO:0065003	macromolecular complex assembly	2.12	0.008827	16	2.61
GO:0009628	response to abiotic stimulus	1.44	0.010134	48	7.82
GO:0000097	sulfur amino acid biosynthetic process	4.30	0.012404	6	0.98
GO:0044275	cellular carbohydrate catabolic process	2.47	0.020265	10	1.63
GO:0006790	sulfur metabolic process	2.30	0.021577	11	1.79
GO:0006820	anion transport	3.17	0.022737	7	1.14
GO:0022900	electron transport chain	2.28	0.023172	11	1.79
GO:0010033	response to organic substance	1.37	0.027188	45	7.33
GO:0016052	carbohydrate catabolic process	2.11	0.027904	12	1.95
GO:0010035	response to inorganic substance	1.59	0.029055	24	3.91
GO:0006091	generation of precursor metabolites and energy	1.68	0.034178	19	3.09
GO:0042545	cell wall modification	2.39	0.035196	9	1.47
GO:0009825	multidimensional cell growth	3.89	0.038649	5	0.81
GO:0010038	response to metal ion	1.68	0.040542	18	2.93
GO:0040007	growth	1.89	0.043215	13	2.12
GO:0006325	chromatin organization	1.96	0.04322	12	1.95
GO:0042592	homeostatic process	1.80	0.048042	14	2.28

Table 2. Overrepreseted GO Terms in downregulated Genes in the h2a.z mutant					
GO Term ID	GO Term	Fold Enrichment	P-value	Gene Count	%
GO:0009719	response to endogenous stimulus	1.36	0.050847	37	6.03
GO:0000096	sulfur amino acid metabolic process	2.90	0.055282	6	0.98
GO:0019725	cellular homeostasis	1.88	0.056165	12	1.95
GO:0051276	chromosome organization	1.80	0.058038	13	2.12
GO:0009414	response to water deprivation	2.03	0.058111	10	1.63
GO:0006535	cysteine biosynthetic process from serine	7.16	0.064013	3	0.49
GO:0032535	regulation of cellular component size	1.81	0.06984	12	1.95
GO:0046686	response to cadmium ion	1.64	0.073831	15	2.44
GO:0009415	response to water	1.94	0.074426	10	1.63
GO:0016049	cell growth	1.84	0.07687	11	1.79
GO:0044272	sulfur compound biosynthetic process	2.34	0.078411	7	1.14
GO:0009650	UV protection	23.87	0.081282	2	0.33
GO:0005976	polysaccharide metabolic process	1.80	0.086464	11	1.79
GO:0007047	cell wall organization	1.68	0.087541	13	2.12

Table 3. Primers				
Gene	TAIR ID	Primer Name	Primer Type	Primer Sequence
PIE1-6	AT3G12810	PIE1-F	PCR	TTGAGGAGCTGCTCAGACGTTATACCTCT
		PIE1-R	PCR	TCTCTGGCATTTGATGAAATGACTGACCT
		SALK-TDNA-LB1	PCR	TGGTTCACGTAGTGGGCCATCG
HTA8	AT2G38810	HTA8-F	PCR	TCCAGAGAGGATGATGATGAAACTGAA
		HTA8-R	PCR	TCCAGATCCGTCGTCGAAATCA
		FLAG-TDNA-LB1	PCR	CGTGTGCCAGGTGCCCACGGAATAGT
HTA11	AT3G54560	HTA11-F	PCR	TTGTTTGACTGTGTACGGTTAGATCTAGA
		HTA11-R	PCR	CAGAACATTGCACATACTAGAACCAAACA
		SALK-TDNA-LB1	PCR	TGGTTCACGTAGTGGGCCATCG
HTA9	AT1G52740	HTA9-F	PCR	TAACACAAAGTACGCTGATCGAACGAACA
		HTA9-R	PCR	GATCGCGTTTAAGTGGTAGATCGATGAGA
		SALK-TDNA-LB1	PCR	TGGTTCACGTAGTGGGCCATCG
HTA4	AT4G13570	HTA4-F	PCR	TTGAAAATGTTCGGATGATTATGGTGGAGG
		HTA4-R	PCR	ATAATTGTTCCTTTGATGAGTGTGTCAAGC
		SALK-TDNA-LB1	PCR	TGGTTCACGTAGTGGGCCATCG
IBM1-6	AT3G07610	IBM1-F	PCR	TAGTTTTGATATCTGCCTTAGCTGTTGCC
		IBM1-R	PCR	CTTTTACCCAATGGTGCTGAAAATGCTTC
		SALK-TDNA-LB1	PCR	TGGTTCACGTAGTGGGCCATCG
MET1-6	AT5G49160	MET1-F	PCR	TGTGACTGAGAACCGCTGTCAGGATCGTTTAAAGAGATC
		MET1-R	PCR	CGTACTATAAGACCTCCGAAG
HTA8	AT2G38810	HTA8-RT-F	RT-PCR	GCTGGTAAAGGTGGGAAAGGGCTTCTAGC
		HTA8-RT-R	RT-PCR	TCAGCAGTCAAGTATTCTAGAATTGATGAAGCGTAAACAG
HTA11	AT3G54560	HTA11-RT-F	RT-PCR	TCGTTCCCTTGATGAGTGTGTCCAGC
		HTA11-RT-R	RT-PCR	GACTCGTAGCTGCGAAGACGATGGCTGC
HTA9	AT1G52740	HTA9-RT-F	RT-PCR	ACCCAGCGGTAGCGACAAGGATAAGGAC
		HTA9-RT-R	RT-PCR	TTGTGCGATCTCCTTTCTCTTTCAATCTC
HTA4	AT4G13570	HTA4-RT-F	RT-PCR	TTGAAAATGTTCGGATGATTATGGTGGAGG
		HTA4-RT-R	RT-PCR	ATAATTGTTCCTTTGATGAGTGTGTCAAGC
EF1alpha	AT5G60390	EF1-RT-F	RT-PCR	TCGAGACCACCAAGTACTACTGC
		EF1-RT-R	RT-PCR	ATCATACCAGTCTCAACACGTCC
HTA9	AT1G52740	HTA9-Q-F	QPCR	AGCGGTAGCGACAAGGATAAG
		HTA9-Q-R	QPCR	TGGTAGGGTGCATCGTCTGT
UBQ5	AT3G62250	UBQ-Q-F	QPCR	TTCTGGTAAACGTAGGTGACTCC
		UBQ-Q-R	QPCR	GAAGACTTACCAAGCCGAAG

Table 4. Illumina sequencing libraries					
Genotype	Replicate	Library name	Library Type	Tissue Type	Age of Tissue
h2a.z	Rep 1	H2AZ_HO_1	Bisulfite-Seq	cauline	6 weeks
h2a.z	Rep 2	H2AZ_HO_2	Bisulfite-Seq	rosettes	6 weeks
h2a.z	Rep 3	H2AZ_HO_3	Bisulfite-Seq	rosettes	6 weeks
h2a.z	Rep 1	H2AZ_HO_4	Bisulfite-Seq	seedlings	2 weeks
h2a.z	Rep 1	H2AZ_HO_5	Bisulfite-Seq	seedlings	2 weeks
h2a.z	Rep 1	H2AZ_HO_6	Bisulfite-Seq	rosettes	6 weeks
h2a.z	Rep 1	H2AZ_HO_7	Bisulfite-Seq	cauline	6 weeks
H2A.Z WT	Rep 1	H2AZ_WT_1	Bisulfite-Seq	cauline	6 weeks
H2A.Z WT	Rep 2	H2AZ_WT_2	Bisulfite-Seq	rosettes	6 weeks
H2A.Z WT	Rep 1	H2AZ_WT_4	Bisulfite-Seq	seedlings	2 weeks
H2A.Z WT	Rep 1	H2AZ_WT_5	Bisulfite-Seq	seedlings	2 weeks
H2A.Z WT	Rep 1	H2AZ_WT_6	Bisulfite-Seq	rosettes	6 weeks
H2A.Z WT	Rep 1	H2AZ_WT_7	Bisulfite-Seq	cauline	6 weeks
pie1	Rep 1	PIE1-5_HO_1	Bisulfite-Seq	seedlings	2 weeks
pie1	Rep 2	PIE1-5_HO_2	Bisulfite-Seq	seedlings	2 weeks
PIE1 WT	Rep 1	PIE1-5_WT_1	Bisulfite-Seq	seedlings	2 weeks
PIE1 WT	Rep 2	PIE1-5_WT_2	Bisulfite-Seq	seedlings	2 weeks
met1	Rep 1	MET HO_1	Bisulfite-Seq	seedlings	2 weeks
met1	Rep 2	MET HO_2	Bisulfite-Seq	seedlings	2 weeks
ibm1	Rep 1	IBM_HO_1	Bisulfite-Seq	seedlings	2 weeks
ibm1	Rep 2	IBM_HO_2	Bisulfite-Seq	seedlings	2 weeks
ibm1	Rep 3	IBM_HO_3	Bisulfite-Seq	seedlings	2 weeks
IBM1 WT	Rep 1	IBM_WT_1	Bisulfite-Seq	seedlings	2 weeks
IBM1 WT	Rep 2	IBM_WT_2	Bisulfite-Seq	seedlings	2 weeks
pie1;ibm1	Rep 1	PIE IBM_HO_1	Bisulfite-Seq	seedlings	2 weeks
pie1;ibm1	Rep 2	PIEIBM_HO_2	Bisulfite-Seq	seedlings	2 weeks
pie1;ibm1	Rep 3	PIEIBM_HO_3	Bisulfite-Seq	seedlings	2 weeks
h2a.z;ibm1	Rep 1	IBMH2AZ_HO_1	Bisulfite-Seq	seedlings	2 weeks
h2a.z;ibm1	Rep 3	IBMH2AZ_HO_3	Bisulfite-Seq	seedlings	2 weeks
h2a.z; met1	Rep 1	METH2AZ HO_1	Bisulfite-Seq	seedlings	2 weeks
h2a.z; pie1	Rep 1	PIEH2AZ_HO_1	Bisulfite-Seq	seedlings	2 weeks

Table 4. Illumina sequencing libraries					
Genotype	Replicate	Library name	Library Type	Tissue Type	Age of Tissue
H2AZ_ChAP_IP_1	Rep 1	H2AZ_ChAP_IP_1	ChIP-Seq	roots	3 weeks
H2AZ_ChAP_in_2	Rep 1	H2AZ_ChAP_in_2	ChIP-Seq	roots	3 weeks
h2a.z	Rep 1	H2AZ_HO_1_RNA	RNA-Seq	rosettes	4 weeks
h2a.z	Rep 2	H2AZ_HO_2_RNA	RNA-Seq	rosettes	4 weeks
h2a.z	Rep 3	H2AZ_HO_3_RNA	RNA-Seq	rosettes	4 weeks
H2AZ WT	Rep 1	H2AZ_WT_1_RNA	RNA-Seq	rosettes	4 weeks
H2AZ WT	Rep 2	H2AZ_WT_2_RNA	RNA-Seq	rosettes	4 weeks
H2AZ WT	Rep 3	H2AZ_WT_3_RNA	RNA-Seq	rosettes	4 weeks

Chapter 4

The *Arabidopsis* Snf2 chromatin remodelers DDM1 and DRD1 regulate DNA methylation at distinct chromatin domains

4.1 Abstract

DNA methylation is usually localized to particular genomic sequences, in particular genes and transposons; however, its targeting mechanisms remain elusive. Several Snf2 chromatin remodeling proteins, which hydrolyze ATP to remodel histone proteins over the DNA they decorate, play an important role in DNA methylation. In *Arabidopsis*, two Snf2 proteins, DDM1 and DRD1, have been shown to be required for the methylation of heterochromatic DNA. Here, we find that DDM1 and DRD1 regulate DNA methylation in genes, as well as in transposons. In genes, a loss of DDM1 causes CHG hypermethylation that is distinct from the patterns and extent of genic hypermethylation reported in two other mutants, *met1* and *ibm1*, suggesting that distinct mechanisms mediate these processes. In transposons, DRD1 and the RNAi pathway target non-CG methylation mostly to short elements and transposon-edges, whereas DDM1 is mainly required for methylation within the bodies of longer transposons. Consequently, DRD1 and DDM1 are preferentially required to silence short and long transposons, respectively, but also collaborate to repress the transcriptional activity of a number of mobile elements. Further analyses revealed that short transposons and transposon edges, which are preferentially targeted by DRD1, are enriched for small RNA molecules and depleted of nucleosomes, whereas the bodies of long transposons methylated by DDM1 are enriched in histone H3 lysine 9-methylated nucleosomes. Our data show that *Arabidopsis* heterochromatin is unevenly structured and is only partially targeted for methylation by RNAi, and that the bulk of DNA methylation in all sequence contexts is mediated by DDM1 and histone methylation.

4.2 Introduction

DNA methylation in eukaryotes is usually not randomly distributed in the genome but restricted to specific regions, in particular genes and transposable elements (TEs). While a few targeting mechanisms from different species have been characterized, the full picture is far from being resolved. Here, we aimed to study the role of chromatin structure and Snf2 chromatin remodelers in targeting DNA methylation in *Arabidopsis thaliana*. DNA methylation is catalyzed by methyltransferses (MTases), which share a conserved catalytic domain and most likely originated from a common ancestral gene family that predates the divergence of eukaryotes and prokaryotes (Goll and Bestor 2005). MTases are distinguished both by their structure and function. *Arabidopsis* has three types of MTases: MET1, the canonical enzyme that maintains CG methylation; CMT3, which catalyzes methylation of CHG sites (H = A,C,T); and DRM2, which has a rearranged catalytic domain and methylates CHH sites (Law and Jacobsen 2010).

In *Arabidopsis*, like in other plants, TEs are frequently methylated at all three nucleotide contexts, CG, CHG and CHH. This DNA methylation combined with additional chromatin marks, such as small RNAs (smRNA) and methylation of histone H3 at lysine 9 (H3K9me2), act to epigenetically silence these sequences (Law and Jacobsen 2010). In contrast, DNA methylation in genes is targeted almost exclusively at CG sites (Zemach, McDaniel et al. 2010), and generally correlates with transcriptional activity (Suzuki and Bird 2008). The *Arabidopsis* histone demethylase *Increase in*

BONSAI Methylation 1 (IBM1) was shown to prevent ectopic non-CG methylation in genes by removing H3K9me2 from within genes (Saze and Kakutani 2011). Similarly to TEs, some genes are methylated in all three contexts. Many of those genes are misannotated or pseudogenes, whereas the rest are functional genes that are demethylated and activated in plant endosperms (Zemach, Kim et al. 2010; Hsieh, Shin et al. 2011).

On naked DNA, MTases are unable to distinguish between genes and transposons; however, *in vivo,* genomic DNA is contextualized through its adornment with histone proteins and other nuclear components, which together form the chromatin fiber. Chromatin structure varies across the genome and is defined by local variation in these epigenetic components, including histone variants, DNA methylation, histone posttranscriptional modifications (e.g. acetylation and methylation), and secondary chromatin factors (e.g. smRNAs). MTases have been found to interact with and depend on many of these chromatin factors; for example, MET1, CMT3 and DRM2 can each be targeted by VIM proteins, H3K9me2 and the RNAi machinery, respectively (Law and Jacobsen 2010).

While serving as a scaffold for the recruitment of certain nuclear components, chromatin structure is simultaneously acting to inhibit the binding and the activities of others (Khorasanizadeh 2004). For example, DNA methylation was found to be dependent on the activity of Snf2 chromatin remodelers (Felle, Joppien et al. 2011). Snf2 remodelers hydrolyze ATP molecules and use the released energy to unleash DNA from nucleosomes (Erdel and Rippe 2011), a process that most likely enables MTases increased access to cytosine residues. From this perspective, Snf2 proteins can be seen as important effectors of DNA methylation targeting.

Numerous Snf2 remodelers are present in most eukaryotic organisms, for instance, the *Arabidopsis* genome has 42 (www.chromdb.org) (Huettel, Kanno et al. 2007). This variety is useful for regulating the many DNA activities that occur within the nucleus, as well as at diverse chromatin environments across the genome (Langst and Becker 2004).Two Snf2 remodelers, DDM1 (*Decrease in DNA methylation 1*) and DRD1 (*Defective in RNA-directed DNA methylation 1*), were previously found to regulate DNA methylation in *Arabidopsis* (Jeddeloh, Stokes et al. 1999; Kanno, Mette et al. 2004). The first of these, DDM1 is an evolutionary conserved Snf2 remodeler, with homologues in animals. Repetitive sequences and TEs are reported to be the major genomic sequence classes affected by loss of DDM1 and its rodent homologue LSH (Lippman, Gendrel et al. 2004; Tao, Xi et al. 2011). Both DDM1 and LSH colocalize with pericentromeric heterochromatin (Yan, Cho et al. 2003; Zemach, Li et al. 2005). Additionally, LSH was found to associate with Dnmt3a or Dnmt3b and DDM1 was shown to bind methyl-CpG binding domain (MBD) proteins (Zemach, Li et al. 2005; Zhu, Geiman et al. 2006), suggesting a direct link to the DNA methylation process. The loss of either of these remodelers results in a strong decrease in cytosine methylation and subsequent transcriptional reactivation of endogenous TEs (Huang, Fan et al. 2004; Lippman, Gendrel et al. 2004); *ddm1* mutants have also been shown to have increased rates of TE transposition (Tsukahara, Kobayashi et al. 2009) These TE targets of LSH

and DDM1 are found to correspond with the presence of small interfering RNAs (siRNAs), suggesting a potential role for RNAi in guiding DDM1 activity (Lippman, Gendrel et al. 2004). In addition, both proteins have been shown to affect histone methylation patterns at certain genomic loci (Gendrel, Lippman et al. 2002; Tao, Xi et al. 2011). Despite all of this information, the mechanisms by which DDM1 and LSH regulate DNA methylation are still poorly understood. Biochemical analyses have demonstrated that DDM1 is capable of remodeling nucleosomes *in vitro* in an ATP-dependent manner (Brzeski and Jerzmanowski 2003). In addition to their effects on TEs, *ddm1* and *lsh* mutants also induce ectopic non-CG methylation in genes, as well (Jacobsen and Meyerowitz 1997; Cokus, Feng et al. 2008; Tao, Xi et al. 2011). *ddm1* genic hypermethylation was recently found to be dependent on the presence of H3K9me2, but not on the RNAi pathway (Sasaki, Kobayashi et al. 2012). Interestingly, non-CG methylation was also found to accumulate in genes in two other mutants, *met1* (Cokus, Feng et al. 2008; Lister, O'Malley et al. 2008) and *ibm1* (Miura, Nakamura et al. 2009; Inagaki and Kakutani 2010); whether these hypermethylations phenotypes are caused by similar phenomena remains unknown.

DRD1, a member of the only plant-specific Snf2 subfamily, was found in a genetic screen looking for positive regulators of RNA directed DNA methylation (RdDM) (Kanno, Mette et al. 2004). RdDM is a transcriptional silencing mechanism discovered in plants that targets DNA methylation through homology to smRNAs (Kanno, Mette et al. 2004). DRD1 was found to interact with many RdDM components, including *Defective In Meristem Silencing 3* (DMS3), *RNA-Directed DNA Methylation 1* (RDM1) and *RNA polymerase V* (Pol V), and is required for the production of the RdDM-generated Pol V-dependent transcripts (Law and Jacobsen 2010). Several presumed targets of the DDM1 remodeler, which have been shown to be reactivated and/or lose CG methylation in the *ddm1* mutant, were shown to not be affected by a loss of DRD1 (Lippman, May et al. 2003; Kanno, Mette et al. 2004; Kanno, Aufsatz et al. 2005). This suggested that the mechanisms by which DDM1 and DRD1 act may be distinct. Further examinations of specific loci found that *ddm1* induced hypomethylation primarily in CG sites, while *drd1* had an effect on methylation in all three contexts (Kanno, Mette et al. 2004; Huettel, Kanno et al. 2006). The targets of DRD1 appeared to be primarily transposons and repeats, located in and around euchromatin. Unlike DDM1, DRD1 has not been shown to possess nucleosome remodeling activity. Based on relatedness to another family of SNF2 remodeler that is responsible for homologous strand pairing in the recombination repair pathway, it has been suggested that DRD1 may provide a sequence homology search functionality to the RdDM pathway (Huettel, Kanno et al. 2007). It has also been posited that DRD1 may help open chromatin at smRNA target sites, thereby exposing the DNA to the rest of the RdDM machinery (Kanno, Huettel et al. 2005).

Although both DDM1 and DRD1 have been shown to have an effect on DNA methylation, their precise targets and their functional mechanisms remain unresolved. To better understand the role of these two remodelers in regulating DNA methylation, we profiled the complete methylomes and transcriptomes of *ddm1*, *drd1* and a *ddm1drd1* double mutant, and compared them to genome-wide datasets for other

chromatin features, including smRNAs, nucleosome occupancy, and H3K9me2. Our data reveal that in addition to their previously characterized role in targeting DNA methylation to TEs, DDM1 and DRD1 also play a role in regulating DNA methylation within genes. In TEs, DRD1 and the RNAi pathway target non-CG methylation mostly to short transposons and transposon edges, whereas DDM1 is mainly required for methylation within long transposons. Additionally, DRD1 and DDM1 collaborate to repress transposition of a number of mobile elements. Furthermore, the short transposon and transposon edge targets of DRD1 are enriched for small RNA molecules and depleted of nucleosomes, whereas the bodies of long, DDM1-dependent transposons are enriched in histone H3 lysine 9-methylated nucleosomes. Our data show that Arabidopsis heterochromatin is unevenly structured and that, contrary to conventional dogma, the methylation of TEs is only partially targeted by RNAi, with the bulk of all CG, CHG, and CHH methylation mediated by a DDM1-mediated process that is likely independent of RdDM.

4.3 Results

4.3.1 *De novo* CHH methylation is regulated by both DRD1 (RNAi) and DDM1

To study the precise role of DDM1 and DRD1 in regulating DNA methylation, we mapped methylation in the full genomes of *ddm1* and *drd1* mutant plants, as well as in a double mutant of *ddm1drd1* (Figure 1). Analyses of total genomic methylation frequencies revealed that each of the mutants have a distinctive and aberrant methylation phenotype (Figure 2). While *ddm1* has a strong reduction in all three methylation contexts CG, CHG and CHH (59%, 57% and 32%, respectively), the *drd1* hypomethylation phenotype was much weaker and present mostly at CHH sites (33%) (Figures 1 and 2). Previously, CHH methylation was reported to be primarily targeted by RNAi (Law and Jacobsen 2010), however the substantial CHH methylation that remains in *drd1* suggests that additional pathways may well have a role in targeting CHH methylation. To check if the *drd1* methylation phenotype is the result of an incomplete disruption of the RNAi pathway, we mapped methylation in the mutant of another RNAi component, *RNA Dependent RNA Polymerase 2* (*RDR2*), that causes a loss of all endogenous short interfering RNA (siRNA) (Xie, Allen et al. 2005). Our data show that the *rdr2* mutant has nearly identical hypomethylation levels as *drd1* (Figure2). This suggests that DRD1 is a crucial component in the RdDM pathway and that plants can target CHH methylation independently of RNAi. Double mutants of *ddm1drd1* and *ddm1rdr2* exhibit additive and synergistic hypomethylation effects at CG and non-CG sites, respectively (Figure 2), which suggests that the DNA methylation activities of DDM1 and DRD1 (RNAi) might overlap with each other and that DDM1 significantly contributes to the methylation of CHH sites. We also mapped methylation in the *met1* mutant and found that its methylation phenotype is quite different from that of *ddm1* at both CG sites and non-CG sites. For example, CHG methylation in *ddm1* is reduced by 57%, which is more than three times higher than the level of CHG hypomethylation in *met1* (Figure 2). These results suggest that DDM1 and MET1 activities are not completely functionally redundant.

Previous analyses , performed over limited genomic regions, have shown DDM1 and DRD1 to regulate DNA methylation mostly at TEs or DNA repeats (Lippman, Gendrel et al. 2004; Huettel, Kanno et al. 2006). To better understand the role of DDM1 and DRD1 in regulating the genomic distribution of DNA methylation, we examined the methylation phenotypes of their mutants across both genes and transposable elements (TEs). In TEs, the *ddm1* and *drd1* hypomethylation phenotypes were found to be different with respect to their methylation contexts, levels and patterns. For instance, *ddm1* showed a strong hypomethylation at all three methylation contexts, whereas *drd1* showed a weaker reduction that was localized mostly at non-CG sites (Figure 3). Additionally, the *drd1* hypomethylation phenotypes were stronger near the edges of TEs, whereas *ddm1* had a larger affect towards the center of the elements (Figure 3). In *ddm1drd1* and *ddm1rdr2,* most TE methylation was completely abolished (Figures 2 and 3), which suggests that together DDM1 and DRD1 (RNAi) are responsible for targeting the majority of the methylation in TEs. It also implies that the residual methylation found in the *drd1* single mutants is mediated mostly by DDM1, and similarly that DRD1 is responsible for targeting the methylation remaining in the *ddm1* mutant.

4.3.2 DDM1 and DRD1 target methylation at distinct TE sizes and domains

So far, our results suggest that DDM1 and DRD1 target DNA methylation differently in TEs. In *Arabidopsis,* most TEs are shorter than 1000 bp, suggesting that one possible explanation for the differences between *ddm1* and *drd1* TE hypomethylation seen at the points of alignment in Figure 3 could be caused by differential targeting of short TEs. To check whether DDM1 and DRD1 methylation activities are associated with different size of TEs, we plotted the methylation levels of wild-type and mutant plants over TE size (Figure 4). This analysis shows that the *ddm1* hypomethylation effect is positively correlated with the size of TEs (Figure 4); by contrast, the *drd1* hypomethylation effect is negatively correlated with TE size (Figure 4). The *ddm1drd1* double mutant showed almost a complete reduction in all TE methylation independent of their size (Figure 4). As chromocenters and chromosome arms are enriched for long and short TEs, respectively, the preferential targeting of *ddm1* and *drd1* methylation toward different size classes of TEs has broader implications for the chromosomal distribution of these two remodelers. DDM1 is found to target DNA methylation mostly in the heterochromatin rich chromocenters, whereas DRD1 predominantly enhances methylation along the chromosome arms (Figure 5).

Next, we examined if, in addition to their preferences with respect to TE size, DDM1 and DRD1 also have preferred target domains within the bodies of TEs. To that end, we averaged the methylation across long TEs (TEs>4 kb), so that short TEs would not influence the methylation level at the edges of the aligned TEs. We found the edges of long transposons in wild-type to be hypermethylated at CHH sites in comparison to their internal domains (Figure 6). Furthermore, the hypomethylation induced by *drd1* was found to be strongest at TE-edges, whereas *ddm1* hypomethylation was found to be higher within TE bodies (Figure 6). The *ddm1drd1* double mutant showed an additive reduction of methylation in all three contexts with respect to its parental lines, with the

remaining methylation evenly dispersed across the entire TE sequence (Figures 2 and 6). Taken together, these results suggest that DDM1 targets DNA methylation preferentially within long TEs, whereas DRD1 and the RNAi pathway are targeting DNA methylation mostly at short TEs and at TE-edges. Additionally, they demonstrate that CG methylation in TEs is regulated mainly by DDM1, regardless of TE size. (Figures 4 and 6).

4.3.3 DDM1 and DRD1 target methylation to distinct heterochromatin subtypes

Next, we examined the mechanisms responsible for the distinctive patterns of TE methylation conferred by DDM1 and DRD1. As TE-edges are known to be abundant in DNA repeats, which can serve as templates for smRNA production (Martienssen 2003), we profiled the distribution of smRNAs (previously mapped by Lister et al., 2008) in TEs. We found an enrichment of smRNAs at TE-edges as compare to TE-'bodies' (Figure 6). Interestingly, smRNAs were found to be only slightly enriched at short TEs over long ones (Figure 4), which suggests that another factor in addition to smRNAs might be responsible for targeting DRD1-mediated methylation of TEs.

While TE-edges frequently contain repeat sequences, TE bodies often code for genic elements that are necessary for TE transposition. The repeats found in TE edges are usually A/T rich elements, whereas open-reading-frames are often associated with high level of G/C nucleotides. Because long TEs are more likely to encode these TE-specific genes, we expected the GC content to increase with the size of TEs. To test this, we measured GC content across the *Arabidopsis* genome and found that, indeed, both regular genes and the bodies of long TEs are globally enriched for G/C nucleotides (Figures 4 and 6); by comparison, short TEs and TE-‘edges’ were found to be mostly G/C depleted (Figures 4 and 6). These results suggest that the nucleotide composition biases of the DNA sequence itself may play a role in targeting DDM1 and DRD1 in TEs.

The nucleotide content of a sequence is the strongest single factor in determining nucleosome occupancy, and A/T-rich sequences are known to be nucleosome-depleted (Segal and Widom 2009). To check whether GC content affects nucleosome occupancy in TEs, we mapped the genome-wide distribution of nucleosomes in the *Arabidopsis* genome. As predicted, nucleosomes were found to be enriched within the 'bodies' of long TEs, and depleted from TE edges and from short TEs (Figures 4 and 6). Because nucleosomes were mapped using MNase digestion without a prior crosslinking step, it is also possible that short TEs and TE-edges are actually not depleted of nucleosomes, but rather accommodate unstable nucleosomes that dissociate during the nuclei preparation process. However, we think that this is unlikely since these DNA sites are enriched for A/T nucleotides (Figures 4 and 6), which are known to strongly resist the incorporation of nucleosomes not just in *Arabidopsis,* but in eukaryotes generally (Segal and Widom 2009).

As DDM1 was previously shown to reposition nucleosomes *in vitro* (Brzeski and Jerzmanowski 2003), it makes sense that we have found it to function mainly within long

TEs, regions where nucleosomes are particularly enriched (Figures 4 and 6). However, the presence of nucleosomes alone is not necessarily sufficient for targeting DDM1, as the nucleosome occupancy in genes was found to be as high as that in long TEs (Figure 4), however DDM1 methylation activity has been reported to be associated with TEs. Hence, it is likely that DDM1 was designed to remodel only certain nucleosomes, specifically those that are found within long TEs, but are absent from genes and short TEs. One method of differentiating nucleosomes is through the presence of post-translational histone modifications. H3K9 methylation is a well-conserved nucleosome modification, and is located almost exclusively in heterochromatin and absent from active genes (Bernatavichute, Zhang et al. 2008) (Figure 4). To determine how H3K9me2 is distributed across the different domains and size classes of TEs, we analyzed a previously published H3K9me2 dataset (Bernatavichute, Zhang et al. 2008). Our analysis show that H3K9me2 is not uniformly distributed in TEs, but it is highly enriched within long TEs and depleted from short TEs and TE-edges (Figure 4 and 6). Furthermore, kernel density plots of DDM1 methylation activity, specifically at short TEs (<1 kb), demonstrate that that DDM1 induced methylation is significantly higher at H3K9-methylated enriched domains (Figure 7A). Overall theses results suggest that DDM1 is targeting DNA methylation mainly within H3K9-methylated nucleosomes, which coincidently are most enriched within the bodies of long TEs.

The *Arabidopsis* genome contains various classes of TEs which are distinguishable by their mechanism of transposition, internal structure, as well as their chromosomal localization (Initiative 2000). Next, we examined if any of these specific types of TE correlate with DDM1 and DRD1 activities. To do so, we compared the level and distribution of DNA methylation, smRNAs, nucleosome enrichment, nucleotide content and H3K9me2 in four classes of TE: LINEs (a non-LTR retrotransposon), MuDRs (a DNA transposon), and Gypsies and Copias (two types of LTR-retrotransposon).

Our data show that chromatin structure is varied among these four classes of TE. Gypsy elements, which are located mostly at the core of the chromocenters, (Figure 8) have high level of smRNAs across their entire sequence; Copia and MuDR elements accumulate siRNAs mostly at their 5' and 3' edges; LINEs, which are missing DNA repeats at their 5’ edges, are enriched by smRNAs mostly at their 3’ edges, where they have a poly-A sequence (Figure 8). In general, nucleosomes are enriched in TE-'bodies' (Figure 8) and depleted from the TE-edges though the different TE types were found to vary in their nucleosome distribution patterns: nucleosomes are sharply depleted from both ends of Copias and from the 3' ends of LINEs, whereas in MuDRs the depletion at edges was found to be wider and more gradual (Figure 8). H3K9me2 patterns are positively correlated with nucleosome occupancy in all four TE types (Figure 8). DDM1 methylation activity (i.e. the residual methylation that is observed in the *drd1* mutant) at CHH sites is positively associated with nucleosomes, GC content, and in particular, with H3K9me2. This is clearly shown in MuDR elements, in which the parabolic distributions of nucleosomes and H3K9me2 strongly resemble the CHH methylation pattern in *drd1* (Figure 8). These associations are also supported by statistical analyses that show a positive correlation between DDM1 methylation activity and TE size, nucleosome

enrichment, and above all with H3K9me2 (Spearman's R = 0.59, 0.51, and 0.71). The CHH methylation conferred by DRD1 (i.e. the residual methylation that is left in the *ddm1* mutant) is positively correlated with the level and distribution pattern of smRNA in Copia, MuDR and LINE elements. Interestingly, though Gypsies are enriched by smRNAs throughout their 'bodies', DRD1 activity was still localized mostly at their edges (Figure 8), which suggests that the presence of smRNA are not sufficient for targeting DRD1. Indeed, when we measured DRD1 methylation activity in TEs with similar levels of smRNAs, we found it was reduced at nucleosome enriched sites (Figure 7B). Based on this, we conclude that DRD1 methylation activity is restricted to chromatin domains that are enriched by smRNAs and also depleted of nucleosomes, which in gypsies are more frequently found at their edges.

The low levels of DDM1 targeted DNA methylation activity at sequences enriched by smRNAs, such as short TEs and TE-edges, might suggest that DDM1 is inhibited by smRNAs. However, within the bodies of Gypsy elements, which are highly enriched in smRNAs, the methylation levels in the *drd1* mutant are nearly identical to wild-type (Figure 8), which suggests that DDM1 is capable of targeting methylation at these sites and is not substantially affected by the presence by smRNAs. Altogether, these results show that TE chromatin is unevenly structured and consequently requires the action of two different pathways, governed by DDM1 and DRD1, respectively, for targeting DNA methylation.

4.3.4 DDM1 and DRD1 are essential for gene silencing and TE transposition

Both *ddm1* and *drd1* mutants have been previously reported to reactivate TEs and repeats (Lippman, Gendrel et al. 2004; Huettel, Kanno et al. 2006; Tsukahara, Kobayashi et al. 2009). In *ddm1*, this effect was quite strong and localized mainly within the chromocenters, whereas in *drd1* the reactivation was weak and found mainly within the chromosome arms (Lippman, Gendrel et al. 2004; Huettel, Kanno et al. 2006). To study the respective roles of DDM1 and DRD1 in the silencing of TEs, we sequenced and examined the global mRNA expression profiles of *ddm1*, *drd1*, and *ddm1drd1* mutants. Based on our data, we found 1777 reactivated TEs in *ddm1* and 51 in *drd1*, only 14 of which were reactivated in both mutants; this suggests that DDM1 and DRD1 are required to silence different groups of TEs (Figure 9A). In both mutants, reactivated TEs were associated with DNA hypomethylation (Figure 9B). Consistent with our conclusions about DDM1s role in the targeting of DNA methylation within the bodies of long TEs, where the TE-encoded genes required for transposition are located, many TEs were reactivated in *ddm1*. By contrast, the DRD1-specific methylation at short TEs and TE-edges appears to play a relatively weak role in the transcriptional repression of TEs on its own. Interestingly, the *ddm1drd1* double mutant had a much stronger transcriptional reactivation of TEs than either of the single mutants, both in terms of the number of reactivated TEs (2240) as well as in their transcriptional level (Figure 9C). These results suggest that, similar to our results for DNA methylation, DDM1 and DRD1 can partially compensate for each other's loss with respect to TE silencing.

Furthermore, DDM1 and DRD1 appear to act redundantly in the transcriptional silencing of a large number of TEs.

We have also found that in addition to repressing TEs, DDM1 and DRD1 are also essential for proper activity of protein-coding genes. According to our data, 8 and 170 genes were upregulated in *drd1* and *ddm1*, respectively, with no overlap between the two groups of genes (Figure 10). The *ddm1drd1* double mutant had the largest group of upregulated genes (322), which include most of the upregulated genes in the *ddm1* and *drd1* single mutants(Figure 10). Very few genes were silenced in *drd1*, *ddm1* and *ddmdrd1*, i.e. 10, 16 and 25, respectively. By contrast, 185 genes where significantly silenced in the *met1* mutant.

4.3.5 DDM1 and DRD1 regulate DNA methylation in genes

Previously, DDM1 was found to affect DNA methylation mostly in heterochromatic regions. Here we show that DDM1 is also required for methylating genes, as well. On average CG methylation in gene bodies in the *ddm1* mutant was reduced approximately 25% (Figure 11). Our analyses show that compared with *drd1*, *ddm1* exhibits hypomethylation of a large number of genes (Figure 12). These results imply that DDM1 is an essential regulator for maintaining CG methylation in genes, which to our knowledge makes it the first chromatin factor, besides MET1, that regulates CG methylation in the gene bodies of plants.

Normally, DNA methylation in gene bodies is restricted to CG sites (Cokus, Feng et al. 2008; Lister, O'Malley et al. 2008), and frequently occurs in semi-discrete, highly-methylated regions or "islands" within genes (Figure 1). Hypermethylation of non-CG sites in genes was previously reported in mutants of IBM1, MET1 and DDM. IBM1 is a histone demethylase that was found to be responsible for specifically removing the H3K9me2 modification (which serves as a docking site for the CHG methylase CMT3) from the bodies of some genes (Saze and Kakutani 2011). The mechanism/s for the genic hypermethylation in *met1* and *ddm1* remain unclear, though it has been suggested that reactivation of TEs in these mutants might recruit the IBM1 protein away from genes, thus leading to an accumulation of H3K9me2 and CHG methylation (Inagaki and Kakutani 2010). Here we found that in contrast to *met1*, which shows global non-CG hypermethylation in many genes throughout the genome, *ddm1* non-CG hypermethylation was found mostly in gene bodies that normally contain CG methylation in WT (Figure 13). Whereas the level of CHG hypermethylation in *ddm1* rises proportionately to the level of WT CG methylation, *met1* CHG hypermethylation showed the opposite trend (Figure 14). Additionally, *ddm1* hypermethylation was found to be higher at the 3' ends of genes, whereas *met1* hypermethylation level was enriched more evenly across the gene body (Figures 11). Taken together, these results suggest that the non-CG hypermethylation of genes in *ddm1* and *met1* are not necessarily the same phenomena, and therefore challenges the previously postulated hypermethylation mechanism. If titration of IBM1 away from genes were the only cause for non-CG hypermethylation in *ddm1* and *met1,* we would expect the patterns and distribution of hypermethylation of genes to be similar in both mutants.

In order to better understand the mechanism controlling non-CG methylation in *Arabidopsis* genes we mapped methylation in an *ibm1* mutant and compared its genic methylation phenotype with that of *ddm1* and *met1* mutants. Similarly to *ddm1*, *ibm1* hypermethylation was found at both CHG and CHH sites (Figure 11). The *ibm1* induced CHH hypermethylation correlated well with the presence of CHG hypermethylation (Pearson's R = 0.7715), suggesting that the two phenomena are mechanistically linked (Figure 15). Both CHH and CHG hypermethylation in *ibm1* were strongly correlated with the presence of genic CG methylation in WT (Pearson's R= 0.7671 and 0.5918, respectively (Figure 16A). Furthermore, unlike *met1*, both *ibm1* and *ddm1* exhibited hypermethylation that corresponded to the boundaries of WT CG methylation (Figure 16B).

Unlike the hypermethylation in *ddm1*, however, *ibm1* induced genic CHG and CHH methylation had a strong 5' bias, decaying towards the 3' ends of genes (Figure 11). These different biases in *ibm1* and *ddm1* suggest that the respective mechanisms responsible for hypermethylation are distinct from one another, and yet both potentially linked to the process of transcription. When compared to transcription rates, both *ibm1* and *ddm1* hypermethylation in genes show roughly similar trends to that of WT CG methylation, with roughly parabolic distributions (Figure 17). On the other hand, hypermethylation in *met1* shows a nearly linear negative correlation with increasing transcription (Figure 17). Taken together, our analysis of genic hypermethylation in *ibm1*, *ddm1*, and *met1* suggests different mechanisms are responsible for the increases in non-CG methylation in each mutant.

Interestingly, we found that the genic hypermethylation in *ibm1* takes at least two generations to accumulate, and differs substantially between mutants in their first and second generations of *ibm1* homozygosity (Figure 18). On the other hand, hypermethylation in *ibm1* mutants of the same generation were well correlated (Pearson's R = 0.941, Figure 18), as has been shown previously (Miura, Nakamura et al. 2009; Inagaki and Kakutani 2010). Additionally, we find that transposons show decreased CHG and CHH methylation in the second generation of *ibm1* homozygosity, potentially the result of reduced occupancy of CMT3 at heterochromatin after recruitment to genic sequences (Figure 19). Previous reports have found that DNA methylation in transposons was unaffected by a loss of *ibm1* (Saze, Shiraishi et al. 2008; Miura, Nakamura et al. 2009); as our data show similar profiles between first generation *ibm1* mutants and WT for both CHG and CHH in TEs, it is possible that this discrepancy can be explained by generational differences of the *ibm1* mutants used in these experiments.

Both *ibm1* and *ddm1* genic non-CG hypermethylation were reported to be targeted by H3K9me2 and not by the RNAi pathway (Saze, Shiraishi et al. 2008; Sasaki, Kobayashi et al. 2012). We found that *ddm1* genic hypermethylation is mediated by DRD1. In the *ddm1drd1* double mutant, the non-CG hypermethylation level was as low as in wild-type (Figure 11). These results suggest that RNAi is involved in targeting non-CG methylation in genes in the *ddm1* mutant. Consistent with results published in Sasaki et al. (Sasaki, Kobayashi et al. 2012), we found that some genes in

ddm1, including *BONSAI*, were hypermethylated independently of RNAi (Figure 20). Hence, it is likely that genes in *ddm1* are hypermethylated by at least two different mechanisms, including both RNAi-dependent and RNAi-independent pathways. To check if DRD1 targets non-CG methylation in an *ibm1* mutant, we crossed *drd1* with *ibm1* and discovered that genes in the *ibm1drd1* double mutant are still strongly hypermethylated at non-CG sites (Figure 21), which verifies that the RNAi pathway is not involved in targeting non-CG methylation in genes in plants deficient for *ibm1*.

4.4 Discussion

In plants, DNA methylation targeting at heterochromatin has been suggested to rely on both smRNAs and histone H3K9me2 (Lippman, Gendrel et al. 2004; Law and Jacobsen 2010), however the precise relationship between these three components has remained unclear. Here we found that DNA methylation in *Arabidopsis* is targeted to heterochromatin regions by two distinct pathways: the first pathway is controlled by DRD1 and the RNAi machinery, and the second pathway is regulated by DDM1 and H3K9me2. Along the chromosome arms, DRD1 and the RNAi machinery are responsible for methylating non-CG sites at short TEs, TE-edges and other DNA repeats, whereas DDM1 is required for methylating DNA in all three contexts within the bodies of TEs. At the chromocenters, which are rich in both smRNAs and H3K9me2, DDM1 and DRD1 methylation activities were found to significantly overlap. In conclusion, we found that the while the various sizes and types of TEs are generally all targeted for methylation, a modification which is crucial for keeping them silenced, the chromatin structure of different TEs can be quite different, and these differences are associated with distinct targeting mechanisms.

Thus far, the only mechanism shown to be instrumental in the targeting of CHH methylation in plants is the RNAi-directed DNA methylation pathway (Law and Jacobsen 2010). By examining the transgenerational stability of *ddm1*-induced hypomethylation in 56 TEs and repeats, Teixeira et al., (Teixeira, Heredia et al. 2009) have shown that after several backcrosses of *ddm1* to wild-type, remethylation occurred in only a subset of heavily methylated repeats targeted by the RNAi machinery. This suggests that the RNAi machinery is responsible for restoring only a portion of the methylation. Our analyses of the genome-wide methylomes of *drd1* and *rdr2* support this finding and identify the targets of RdDM as short TEs and TE-edges, sites that generally are depleted of nucleosomes. By contrast, most of the remaining CHH sites, found in TE bodies, are targeted for methylation by DDM1 in a RNAi-independent way.

A genome-wide nucleosome positioning analysis in *Arabidopsis* found nucleosomal DNA to be more highly methylated than flanking DNA (Chodavarapu, Feng et al. 2010). These results indicated that nucleosome positioning may influence DNA methylation patterns throughout the genome and that DNA methyltransferases preferentially target nucleosome-bound DNA. However, our data show that DNA methylation is also very efficiently targeted by DRD1 and smRNAs in nucleosome-depleted DNA. Furthermore, we found RdDM methylation activity to negatively correlate with nucleosome enrichment. In conclusion, we find that in plants, nucleosomes can

have both positive as well as negative effects on targeting DNA methylation to TEs, depending on the targeting system.

A previous genome-wide analysis of DNA methylation in Arabidopsis using a tiling microarray suggested that genes are insulated from DDM1-mediated heterochromatic DNA methylation (Lippman, Gendrel et al. 2004). This early study included only 1% of the Arabidopsis genome and had a low resolution of 1 kb per probe. In our study, using single base resolution methylation data for the entire genome, we found that DDM1 is required for methylating genes, as well as TEs. In *ddm1*, the presence of CG gene body methylation was found to be reduced globally at all typically methylated genes, and the ectopic appearance of non-CG methylation was found in the bodies of some genes. Interestingly, both *met1* and *ibm1* mutants have also been shown to exhibit an increase in non-CG methylation in the bodies of some genes (Jacobsen and Meyerowitz 1997; Cokus, Feng et al. 2008; Lister, O'Malley et al. 2008; Miura, Nakamura et al. 2009; Inagaki and Kakutani 2010). It has been hypothesized that the hypermethylations found in these three mutants might be caused by the same mechanism; specifically, the decreases in DNA methylation in and depression of TEs in both *met1* and *ddm1* might lead to a titration of IBM1 protein from genic to TE sequences, similar to though less severe than the loss of IBM1 caused by the *ibm1* mutant itself, leaving genes without a means of adequately pruning H3K9me2e sequences (Miura, Nakamura et al. 2009; Inagaki and Kakutani 2010). Here we present several lines of evidence that suggests that this is unlikely. While the CHG hypermethylation of genes in both *ibm1* and *ddm1* are associated with the presence of CG methylation in WT, the hypermethylation in the *met1* mutant does not. Additionally, the non-CG hypermethylation in *ibm1* and *ddm1* mutants exhibit distinct distribution patterns across the bodies of genes, with hypermethylation showing preferences for the beginning and ends of genes, respectively. Taken together with the facts that double mutants of *ddm1* and various RNAi components abolish this non-CG hypermethylation in genes, while *ibm1drd1* mutants do not, these results suggest that the hypermethylation phenomena in *met1*, *ibm1* and *ddm1* are the result of distinct mechanisms.

4.5 Material and Methods

Biological materials. Genomic DNA was purified from mature rosette and cauline leaves of 6 week-old plants grown on soil under 16h light/ 8h dark in greenhouse conditions. Total RNA was purified from Arabidopsis roots (28 days post germination) grown in MS media under 16h light/ 8h dark at 25 degrees Centigrade (RNA). The *ddm1-2*, *met1-6*, *rdr2-1*, and *ago4-1* mutant lines were described previously (Jeddeloh, Stokes et al. 1999; Xiao, Gehring et al. 2003; Zilberman, Cao et al. 2003; Xie, Allen et al. 2005). The *drd1* mutant (GABI_503F06) was obtained from the GABI-KAT collection (Kleinboelting, Huep et al. 2012). GABI_503F06 has a T-DNA insertion within predicted exon 3 (after nucleotide 1,318 from ATG of the genomic DNA) of DRD1 (At2g16390). The *ibm1* (*ibm1*-6) line was obtained from the SALK collection (SALK_006042).

Bisulfite sequencing. Bisulfite conversion, Illumina library construction and sequencing were performed exactly as described in (Zemach, Kim et al. 2010).

RNA-Seq library preparation. RNA-Seq library construction and Illumina sequencing were performed exactly as described in (Zemach, McDaniel et al. 2010).

MNase-seq library preparation. Arabidopsis roots (1 g) were ground in liquid nitrogen, resuspended in 20 ml of HBM buffer (25mM Tris, pH 7.6, 0.44M Sucrose, 10 mM $MgCl_2$, 2 mM spermine and 0.1% Triton X-100), homogenized using a polytron, filtered through a miracloth, transferred to a 30 ml round bottom glass tube, centrifuged at 2000g (4^0C) for 10 min and resuspended in 1 ml HBB buffer (25mM Tris, pH 7.6, 0.44M Sucrose, 10 mM $MgCl_2$ and 0.1% Triton X-100). Nuclei were further spun down at 200g, 4^0C for 2 min and resuspended in 1ml TNE buffer (10 mM Tris, pH 8.0, 100 mM NaCl and 1 mM EDTA). MNase digestion was done with 4 ul of 1 M $CaCl_2$ and 1ul of diluted (1/20) MNases (200ul/ml; Sigma #N-3755) per 100 ul of pellet nuclei. Nuclei were then divided into several tubes and the digestion reaction was stopped every 45 sec with 10mM EDTA. Digested nuclei were spin down at maximum speed at 4^0C, for 5 min and soluble released nucleosomes were collected from the supernatent. Following RNaseA and proteinaseK digestion, DNA was purified using phenol/chloroform methodology. The purified DNA sample was run on a 2% agarose gel, and the digested samples with the most enriched intact mononucleosomes were chosen; bands corresponding to ~150 bp were cut and purified with a Gel Purification kit (Qiagen). Illumina libraries were constructed and sequenced at the UC Berkeley Genomic Sequencing Laboratory, generating paired ends (SE) 36 base reads.

Sequence analysis. We used a mixture of paired ends (PE) and single ends (SE) Illumina sequencing datasets for bisulfite sequencing exactly as described in (Hsieh, Ibarra et al. 2009) and (Zemach, Kim et al. 2010), respectively.

Genome annotation. For TEs, we used TAIR8 annotations with the following modifications: we removed TEs with less than 5% of CG methylation in wild-type cells and smaller than 30 bases. Also, we merged overlapped and sequential TEs

Figure 1

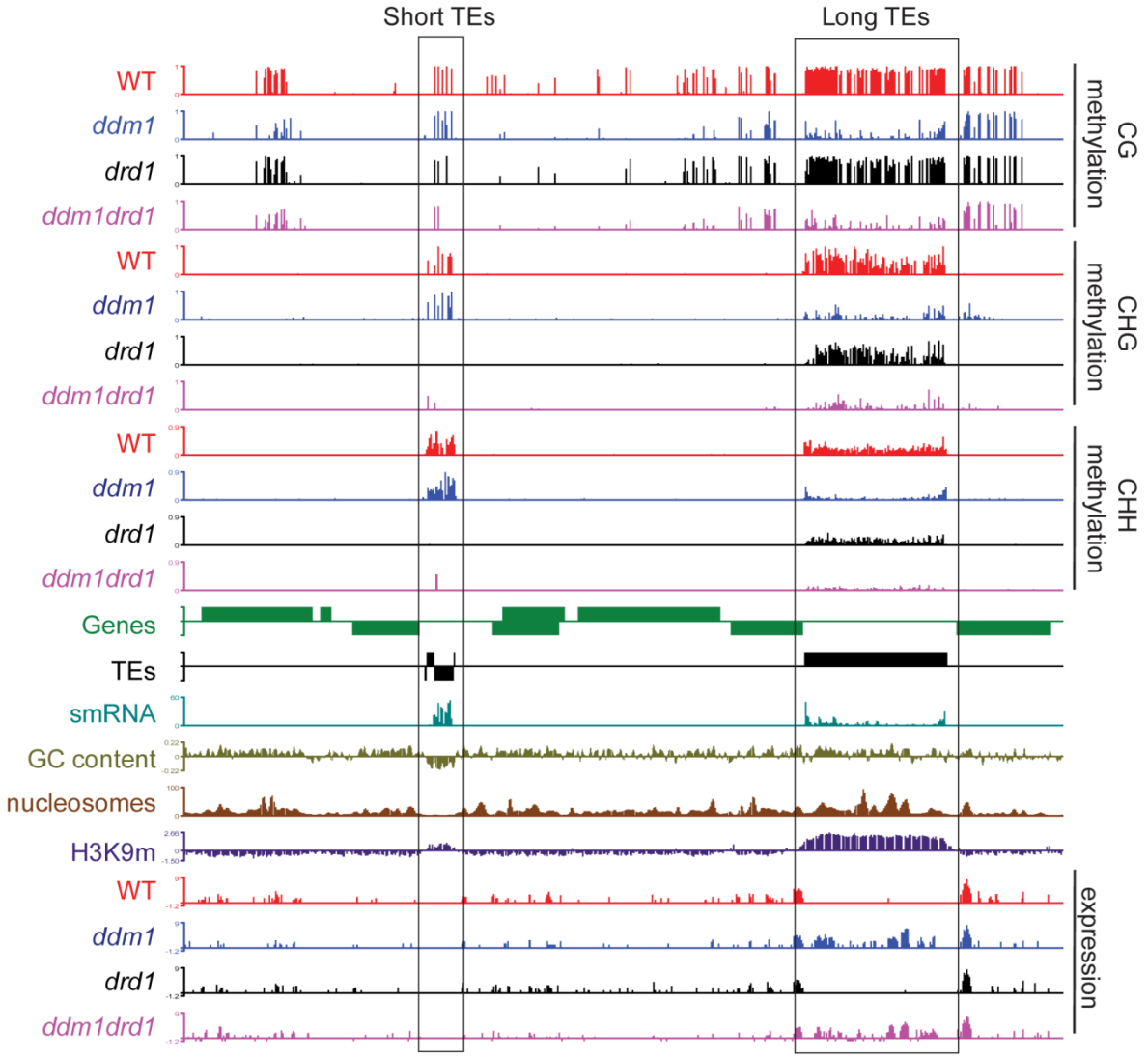

Figure 1. Mapping DNA methylation in *ddm1*, *drd1* and *ddm1drd1* mutants. DNA methylation (CG, CHG and CHH), smRNAs, GC content, nucleosomes, H3K9m and transcription profiles for positions 5,891,000 – 5,920,000 on chromosome 1, including both genes (green boxes) and TEs (black boxes). Genes and TEs, oriented 5′ to 3′ and 3′ to 5′, are shown above and below the line, respectively.

Figure 2

	CG	CHG	CHH
met1	96 %	16 %	41 %
ddm1	59 %	57 %	32 %
drd1	1 %	9 %	33 %
rdr2	~1 %	~7 %	~31 %
ddm1drd1	60 %	81 %	83 %
ddm1rdr2	~69 %	~84 %	~78 %

Figure 2. Percentage of global DNA hypomethylation in the *ddm1* and *drd1* mutants. The level (percent) of hypomethylation across the genome was calculated for *ddm1*, *drd1*, *met1*, *rdrd2*, *ddm1drd1* and *ddm1rdr2* mutants. *ddm1* shows a strong reduction at all three methylation contexts, whereas *drd1* hypomethylation is localized mostly at non-CG sites. Also, note the synergistic hypomethylation effect in CHH sites in *ddm1drd1* versus the *ddm1*

Figure 3

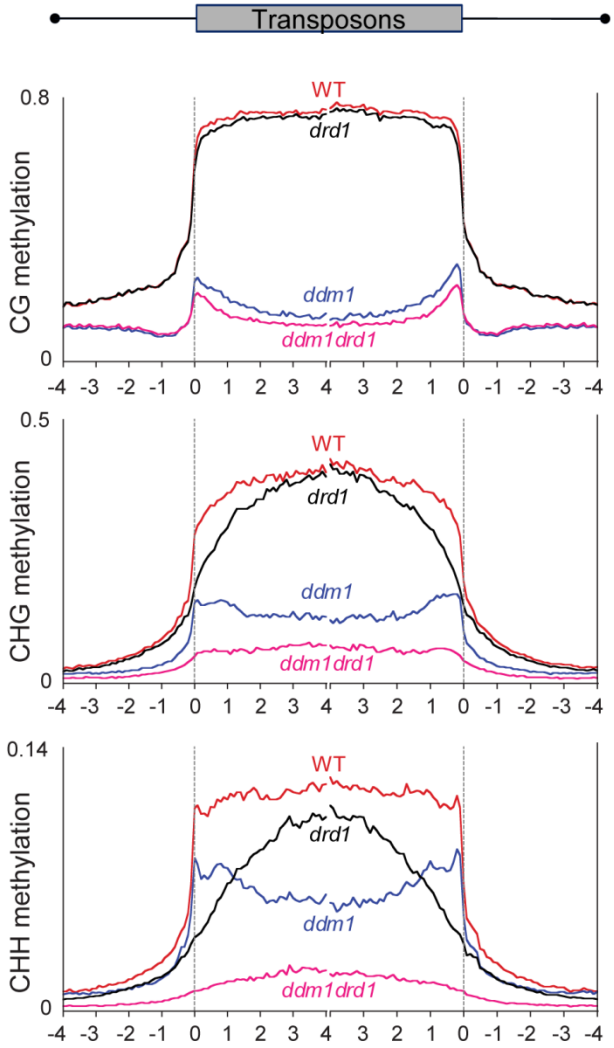

Figure 3. Patterns of DNA methylation in *ddm1*, *drd1*, *ddm1drd1*, and *ibm1* mutants. Arabidopsis TEs were aligned at the 5′ end or the 3′ end, and average methylation levels for each 100-bp interval are plotted. The dashed lines represent the points of alignment. Note that *ddm1* and *drd1* show an opposite hypomethylation patterns, which suggest that they regulate methylation at different TE sites. The *ddm1drd1* double mutant shows an almost complete loss of DNA methylation, which suggests that DDM1 and DRD1 together regulate most of the DNA methylation in heterochromatin.

Figure 4

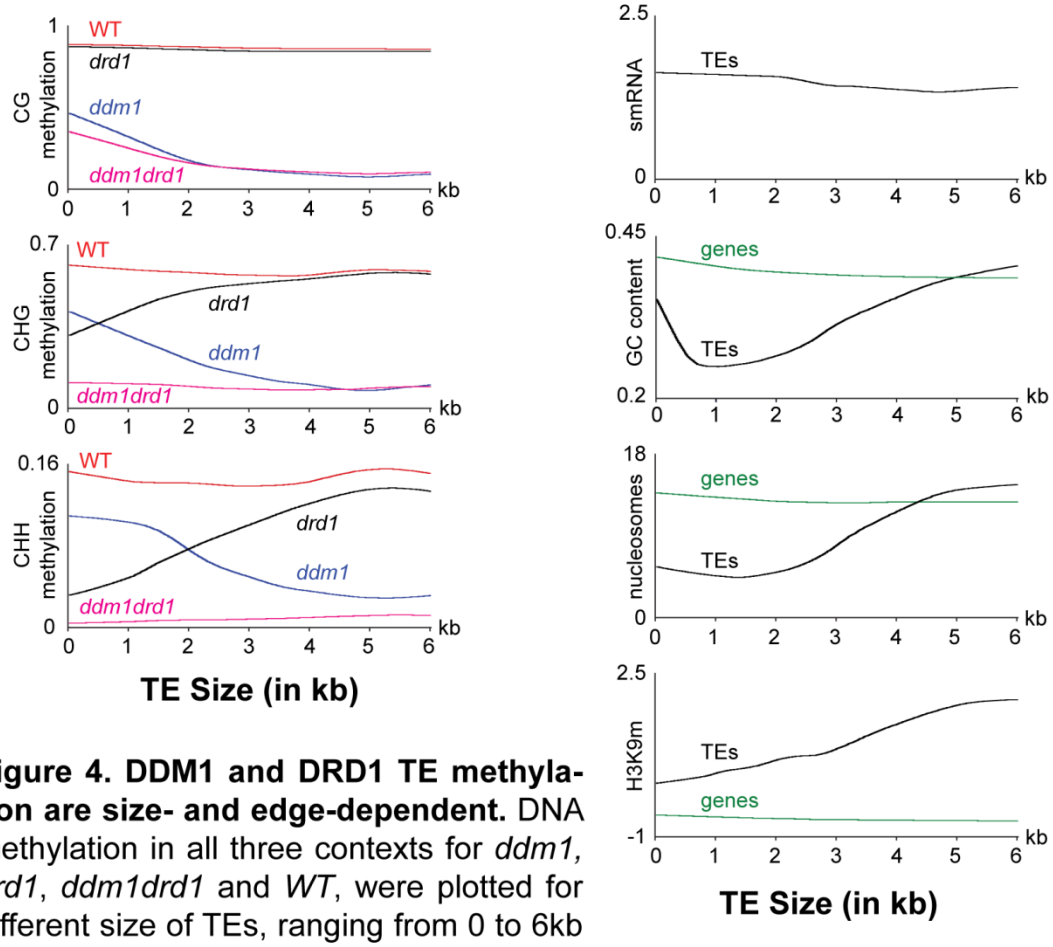

Figure 4. DDM1 and DRD1 TE methylation are size- and edge-dependent. DNA methylation in all three contexts for *ddm1, drd1*, *ddm1drd1* and *WT*, were plotted for different size of TEs, ranging from 0 to 6kb (X-axis). The level of smRNAs (Lister *et al*, *Cell*, 2008), GC content, nucleosomes and H3K9m levels are plotted below.

Figure 5

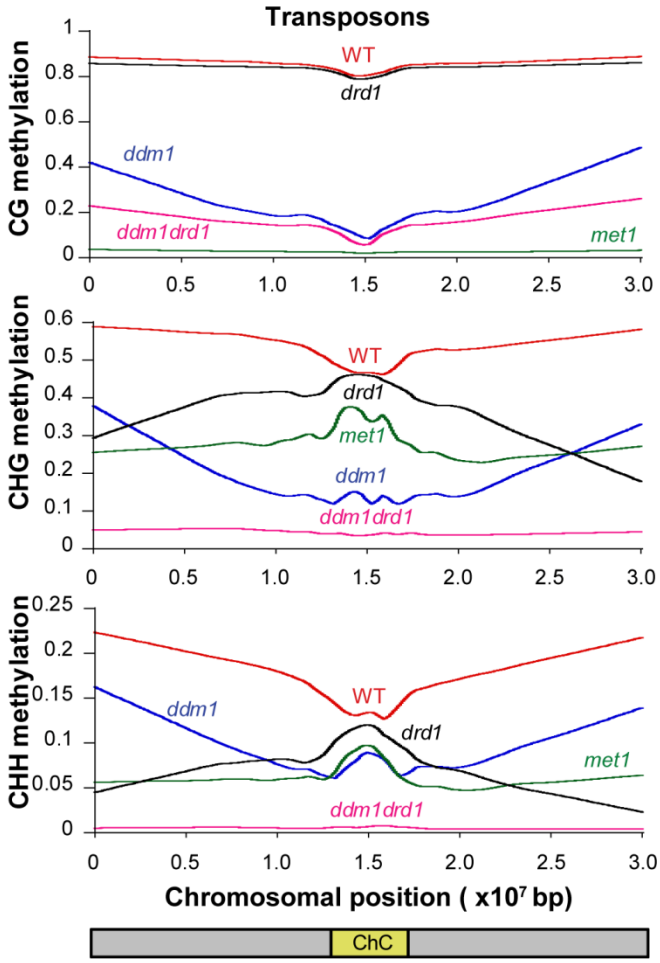

Figure 5. DDM1 and DRD1 act at distinct chromatin domains. Distribution of the DNA methylation in TEs along Chr1 in wild-type and the *ddm1*, *drd1*, and *ddm1drd1* mutants. Fractional methylation was calculated for each context within 50 base pair windows that have an absolute fractional methylation of at least 0.5 in the CG context, 0.3 in the CHG context or at least 0.1 in the CHH context. Plots represent a weighted (30) curve of the data, fitted by KalidaGraph. Note the synergistic hypomethylation in *ddm1drd1* at CHG and CHH sites in the chromocenters (ChC).

Figure 6

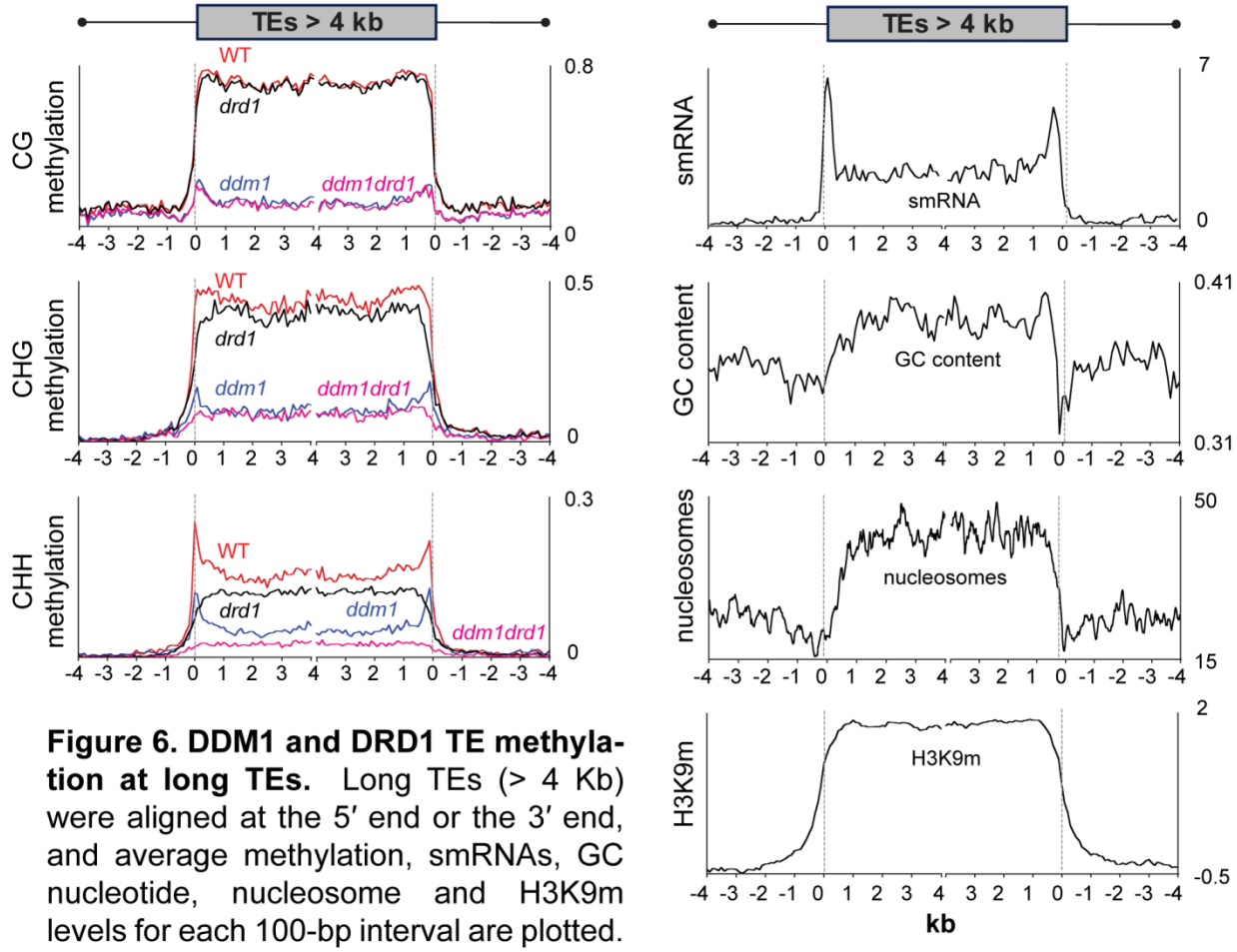

Figure 6. DDM1 and DRD1 TE methylation at long TEs. Long TEs (> 4 Kb) were aligned at the 5′ end or the 3′ end, and average methylation, smRNAs, GC nucleotide, nucleosome and H3K9m levels for each 100-bp interval are plotted. The dashed line represents the point of alignment. DDM1 regulates DNA methylation activity at H3K9 methylated nucleosomes, which are mostly found within long TEs. In contrast, DRD1 function is associated with smRNA-enriched and nucleosome-depleted sites, which are mainly found at short TEs and TE-edges.

Figure 7

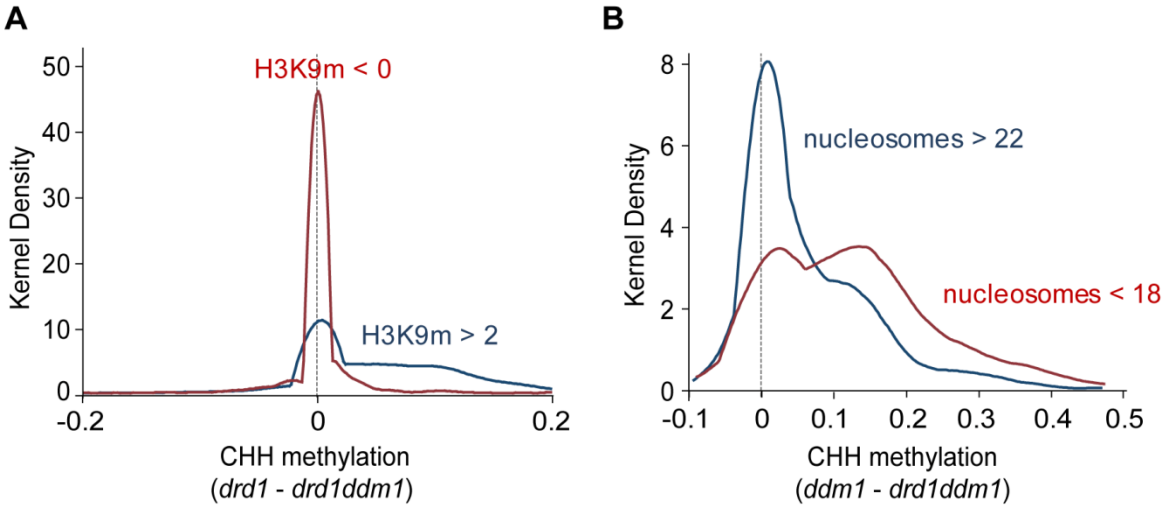

Figure 7. DDM1 DNA methylation activity is correlated with H3K9m. (A) Kernel density plots of CHH methylation differences within 50-bp windows between *drd1* and the *ddm1drd1* double mutant at H3K9 depleted sites (red trace), and H3K9m enriched (blue trace) sites. (B) Kernel density plot of CHH methylation differences within 50-bp windows between *ddm1* and the *ddm1drd1* double mutants at nucleosome depleted sites (red trace), and nucleosome enriched (blue trace) sites.

Figure 8

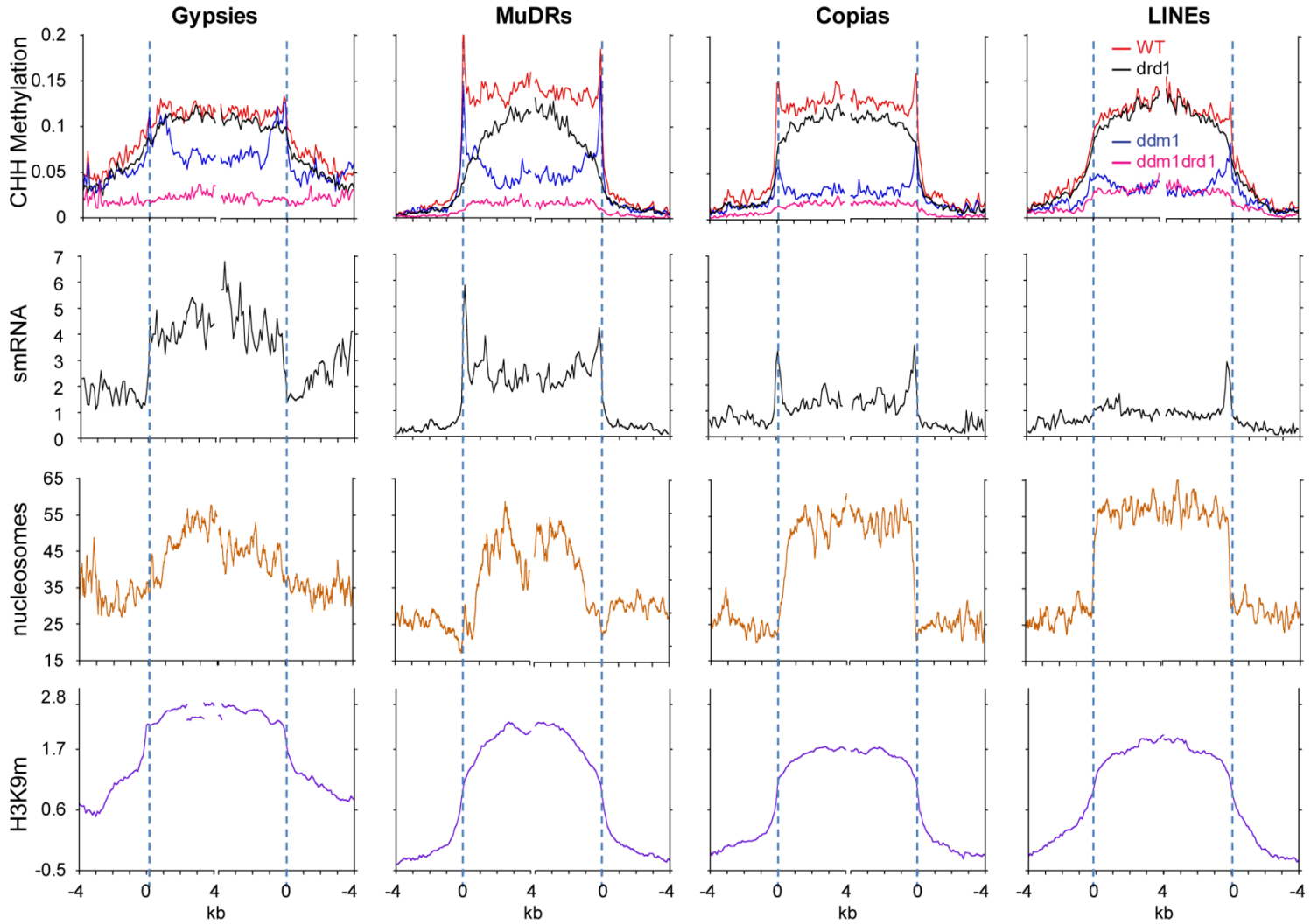

Figure 8. Chromatin features at different classes of TE. (A) TEs were grouped by their TE type (Gypsy, MuDR, Copia, and LINEs) and aligned at the 5′ end or the 3′ end, and average methylation, smRNAs (Lister *et al*, *Cell*, 2008), nucleosome and H3K9m (Bernatavichute YV *et al., PLoS One,* 2008) levels for each 100-bp interval are plotted. The dashed line represents the point of alignment. Note the low DRD1 methylation activity (as represented by the residual methylation in *ddm1*) at the smRNA-rich TEs (Gypsy class).

Figure 9

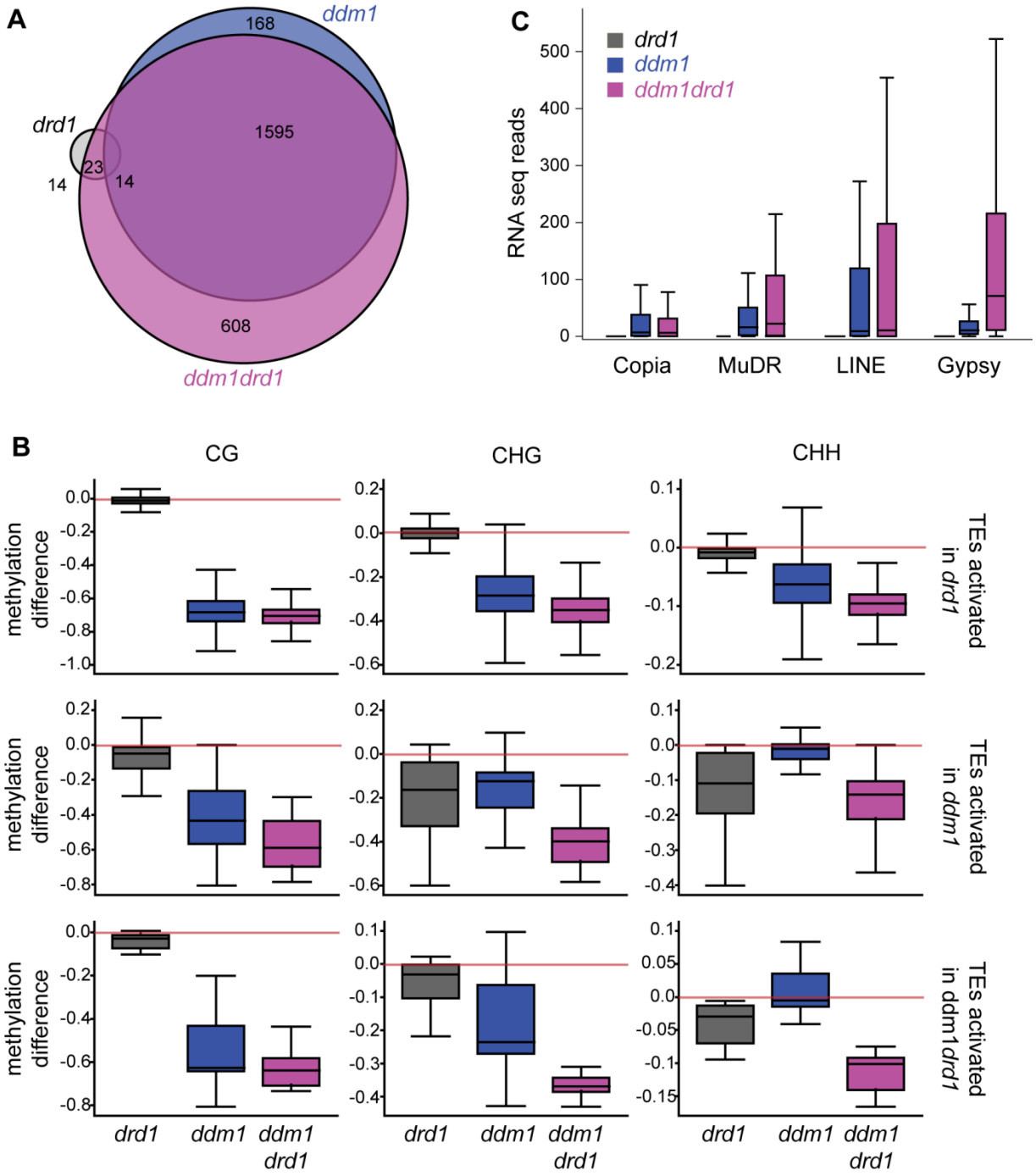

Figure 9. DDM1 and DRD1 regulate TE transcriptional activity. (A) Venn diagram of the TEs upregulated in *ddm1*, *drd1* and the *ddm1drd1* double mutant. (B) Box plots of methylation differences for CG, CHG, and CHH between each mutant and wild-type for TEs upregulated in each of the three mutant background. (C) Box plots of RNA expression read counts at TEs divided by TE type in the three mutants.

Figure 10

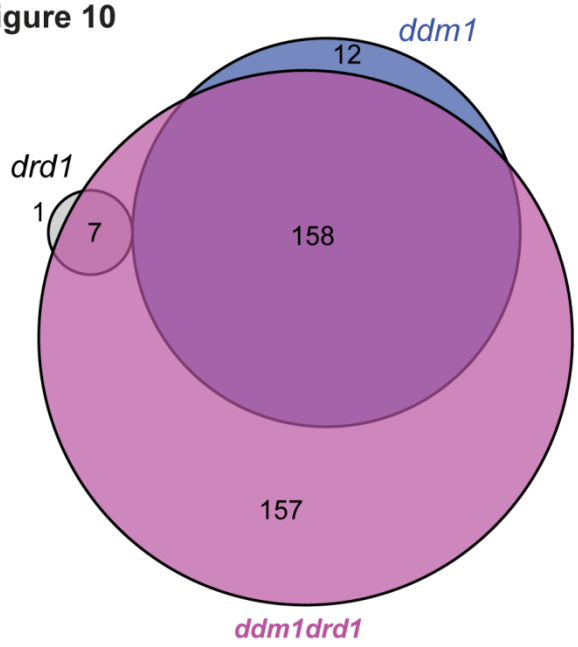

Figure 10. DDM1 and DRD1 regulation of gene expression. A Venn diagram showing the upregulated genes in the three mutant backgrounds, *ddm1*, *drd1*, and the double mutant *ddm1drd1*.

Figure 11

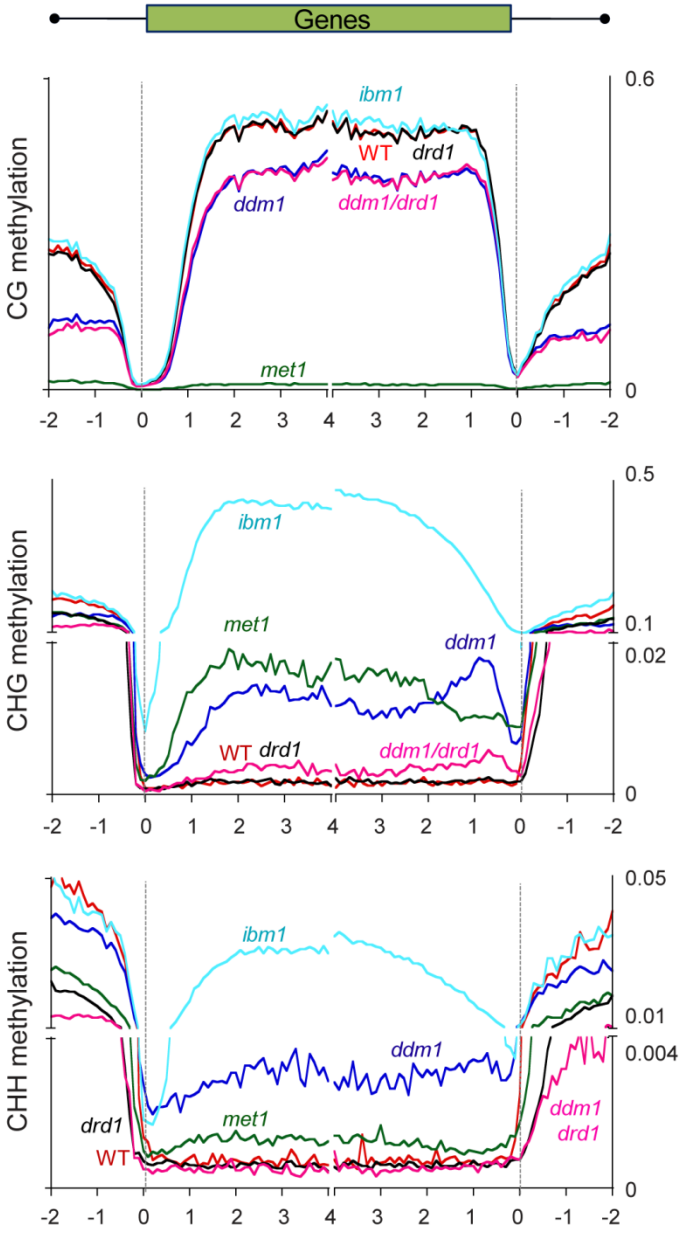

Figure 11. DDM1 affects genic methylation profiles. Patterns of DNA methylation for CG, CHG, and CHH in *ddm1*, *drd1*, *ddm1drd1*, *met1* and *ibm1* mutants. Arabidopsis TEs genes were aligned at the 5′ end or the 3′ end, and average methylation levels for each 100-bp interval are plotted. The dashed lines represent the points of alignment.

Figure 12

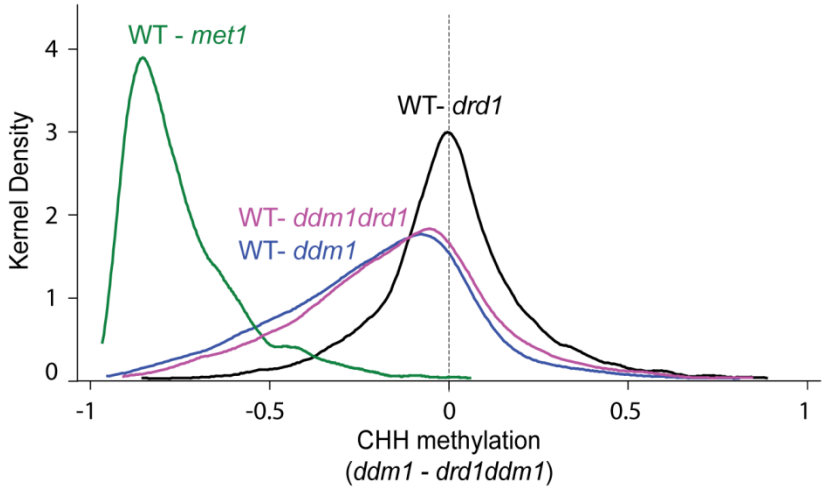

Figure 12. Chromatin remodeling regulates global gene-body methylation. Kernel density plots of the fractional difference in CHH methylation of mutant versus wild-type in genes. Note the global decrease in CHH-methylation in *ddm1* and *met1*, while *drd1* exhibits no change.

Figure 13

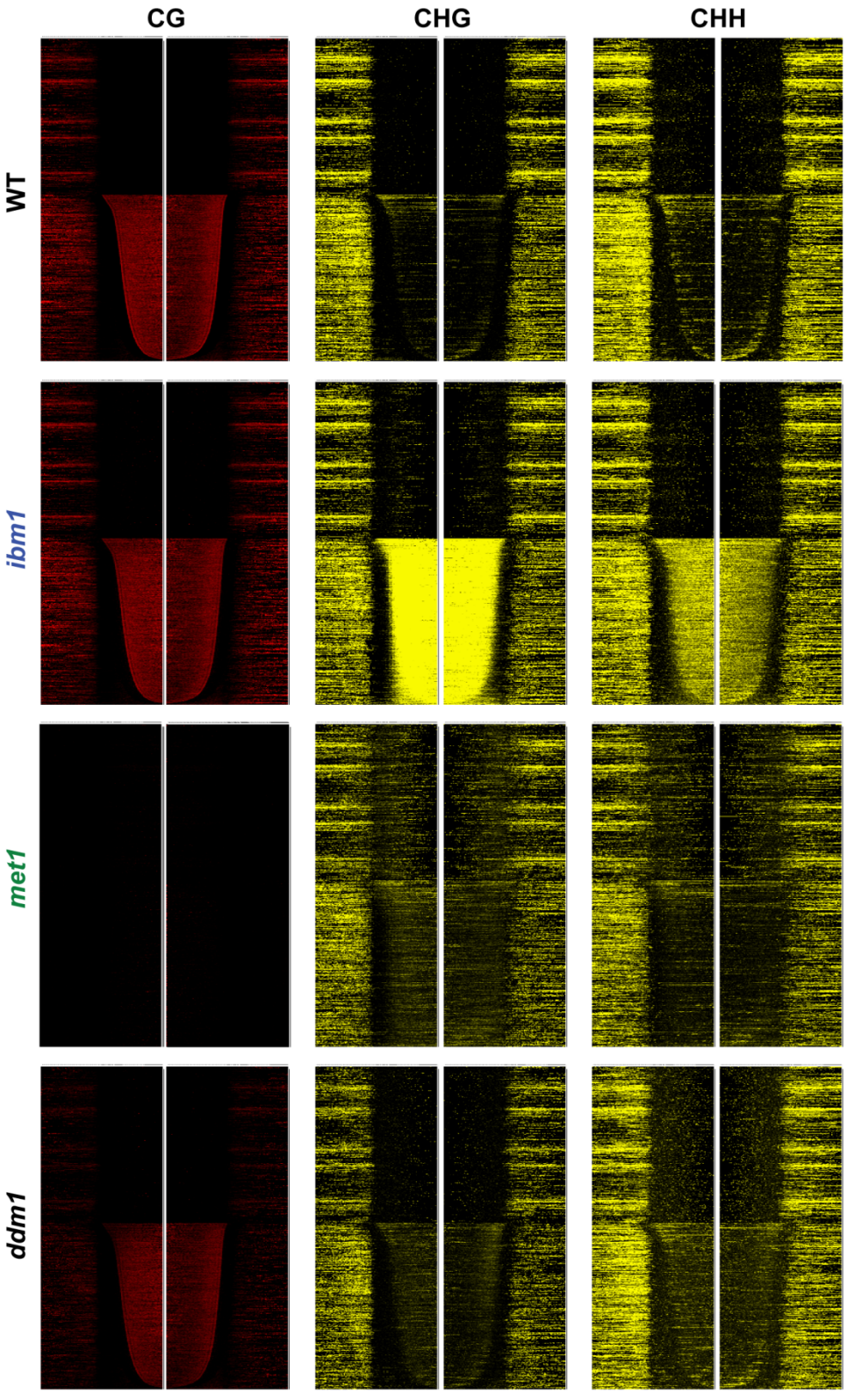

Figure 13. Hyper-methylation and WT CG. TAIR8 genes were filtered according to presence of WT CG methylation. Genes without WT CG methylation (shown in the top half of each panel) were stacked from the top of chromosome 1 to the bottom of chromosome 5; genes containing CG methylation islands (shown at the bottom of each panel) were sorted based on the starting position (for 5' panels) or ending position (for 3' panels) of the WT CG methylation island. Panels represent the 5' and 3' alignments, for CG (red), CHG and CHH (yellow). Different intensity scales were adopted for the three contexts of methylation (CG = 1.0, CHG = 0.1, CHH = 0.02), in order to visualize the distribution patterns.

Figure 14

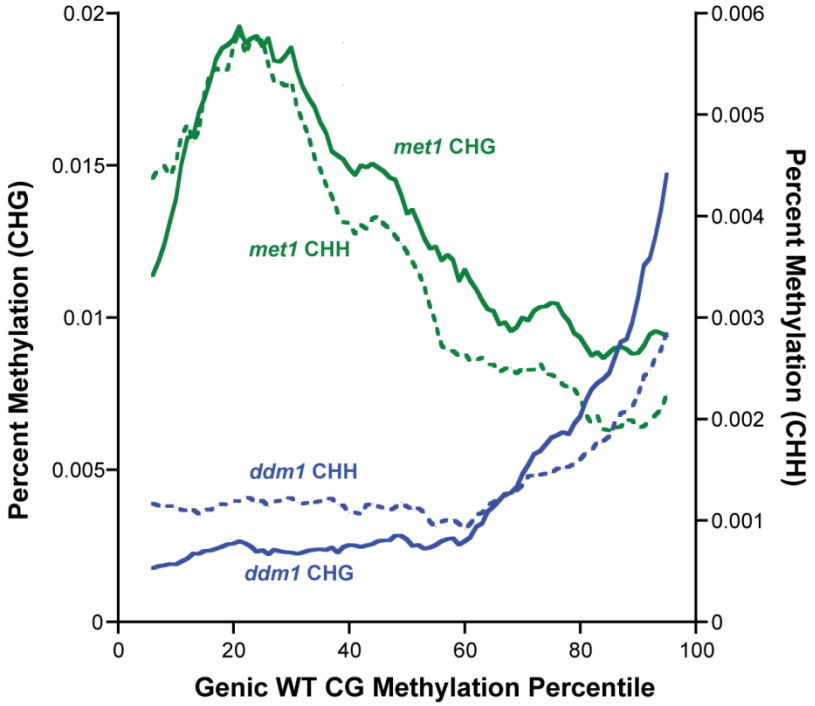

Figure 14. *ddm1* hypermethylation correlates with WT genic methylation. Genes were grouped into percentiles on the basis of WT CG methylation levels. Genes with no WT CG methylation were removed from this analysis. The blue and dotted blue lines represent *ddm1* CHG and CHH, respectively. The green and dotted green lines represent *met1* CHG and CHH, respectively. The scales on the left and right axis represent percent methylation in CHG and CHH respectively.

Figure 15

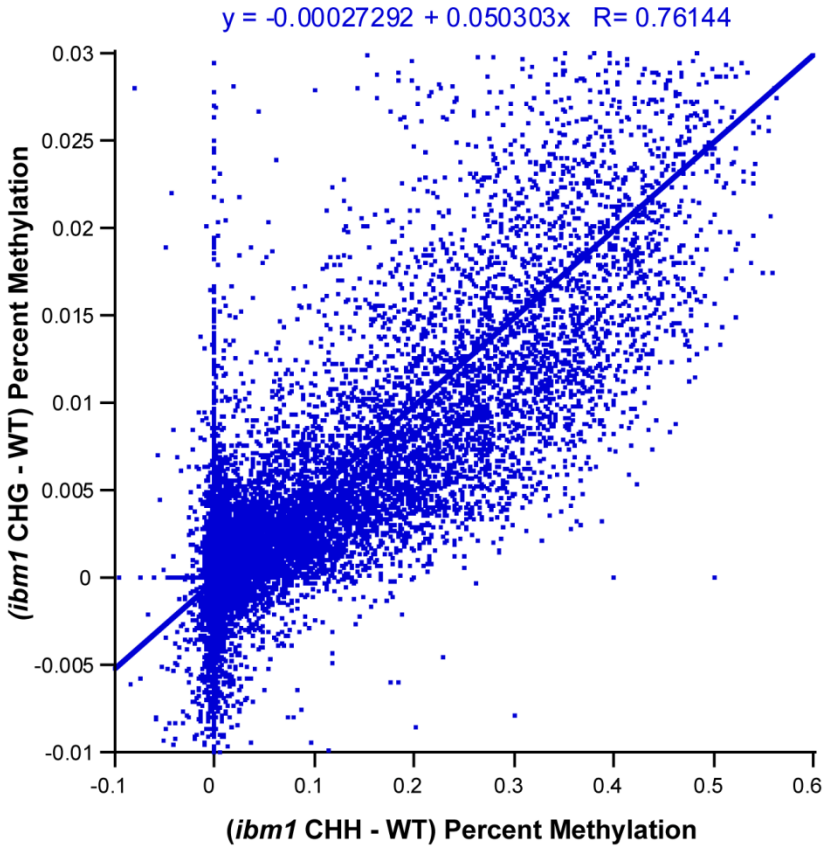

Figure 15. *ibm1* CHG and CHH genic hypermethylation are correlated. The *ibm1* CHG hypermethylation (*ibm1* -WT) in each gene (Y-axis) are plotted against the *ibm1* CHH hypermethylation (*ibm1* -WT) in each gene (X-axis) from the same dataset.

Figure 16

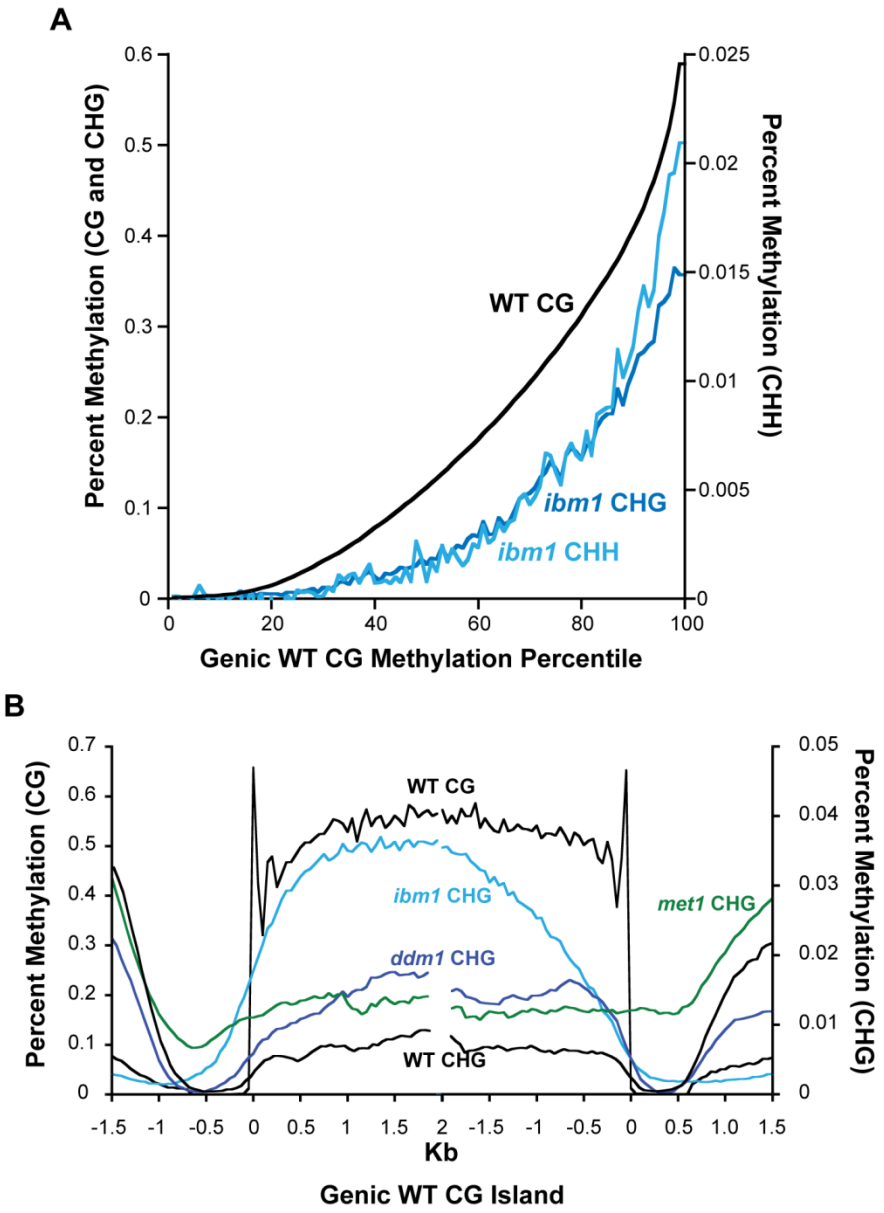

Figure 16. *ibm1* hypermethylation correlates with WT genic methylation. (A) Genes were grouped into percentiles on the basis of WT CG methylation levels. The black line represents WT CG methylation; the blue and green lines represent *ibm1* CHG and CHH, respectively. The left and right axis represent percent methylation in CG/CHG and CHH, respectively. (B) Profiles of WT CG and CHG methylation, compared to *ibm1* , *ddm1*, and *met* 1 CHG methylation, in genic WT CG islands. TAIR8-annotated genes were screened for the presence of WT CG islands, as described in the methods section. These islands were aligned at the 5′ end (left panel) and the 3′ end (right panel) and average methylation levels for each 100-bp interval are plotted from 1.5 kb away from the gene to 2 kb into the gene. WT CG and CHG methylation are represented by the black traces, *ibm1* CHG methylation by the light blue trace, *ddm1* by the blue trace, *met1* by the green trace.

Figure 17

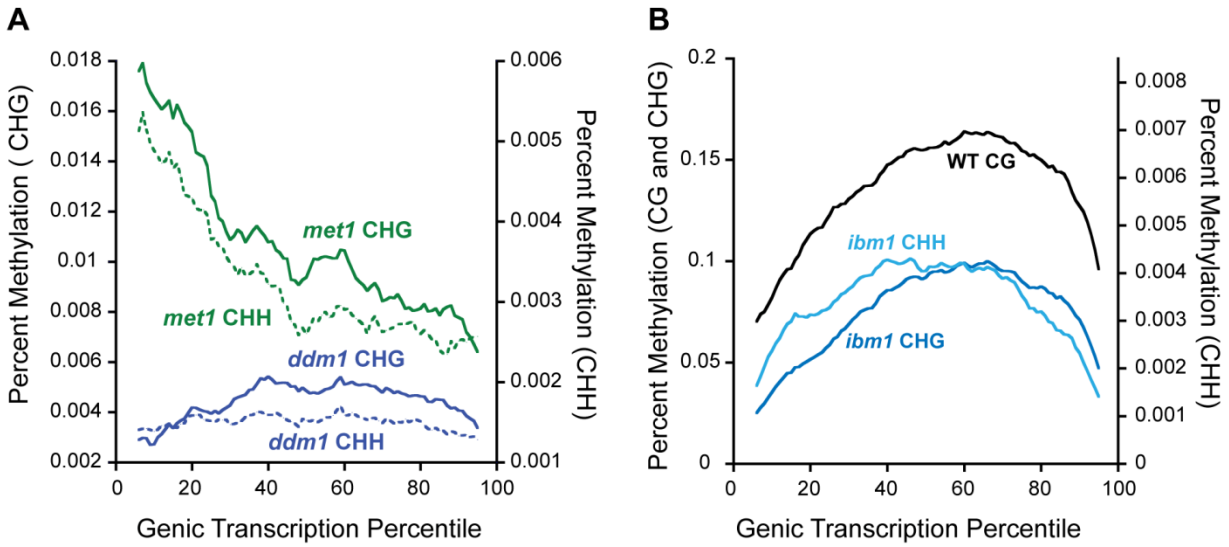

Figure 17. ***ibm1*, *ddm1*, and *met1* hypermethylation by transcription percentile.** (A) Genes were grouped into percentiles on the basis of WT transcription levels. The green and dotted green lines represent *met1* CHG and CHH, respectively. The blue and dashed blue lines represent *ddm1* CHG and CHH, respectively. The scales on the left and right axis represent percent methylation in CHG and CHH respectively. (B) Genes were grouped into percentiles on the basis of WT transcription levels. The black line represents average WT CG methylation for genes within each percentile, and the blue and light blue lines represent *ibm1* CHG and CHH, respectively. The scales on the left and right axis represent percent methylation in CG/CHG and CHH respectively.

Figure 18

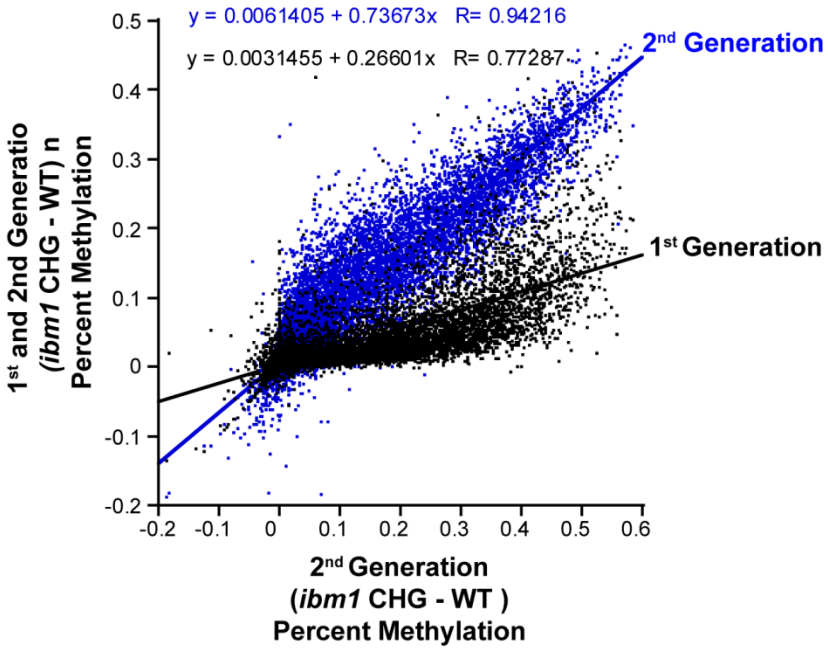

Figure 18. *ibm1* hypermethylation of genes increases across generations. The *ibm1* CHG hypermethylation (*ibm1* - WT) in each gene (Y-axis) for a second generation *ibm1* mutant are plotted against *ibm1* CHG hypermethylation (*ibm1* -WT) in each gene (X-axis) for an independent second generation *ibm1* mutant (shown in blue) and for a first generation *ibm1* mutant (shown in black).

Figure 19

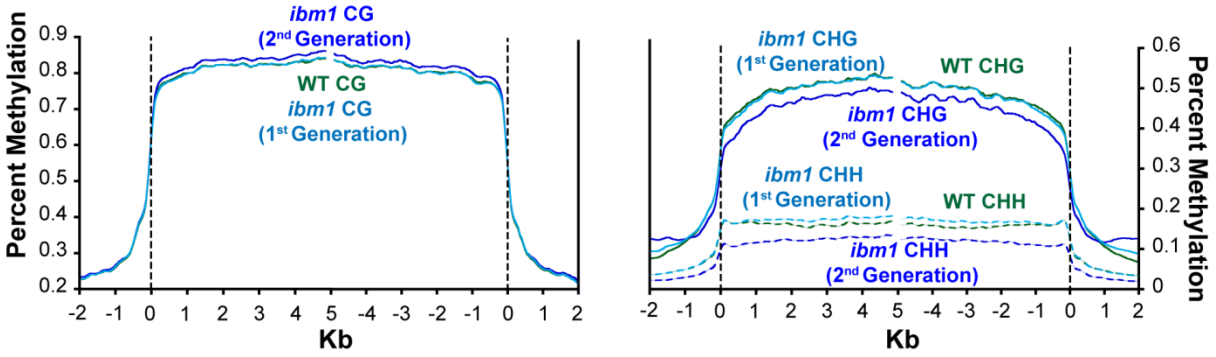

Figure 19. *ibm1* hypomethylation of TEs appears in the second generation. (A) Profiles of WT and *ibm1* CG methylation for both first and second generation *ibm1* mutants. (B) Profiles of WT and *ibm1* CHG and CHH methylation for both first and second generation *ibm1* mutants. WT methylation is represented by the green traces. *ibm1* CG and CHG methylation from the first and second generations are represented by light blue and dark blue traces, respectively. *ibm1* CHH methylation from the first and second generations are represented by dashed light blue and dashed dark blue traces, respectively. The vertical dashed lines at zero represents the point of alignment.

Figure 20

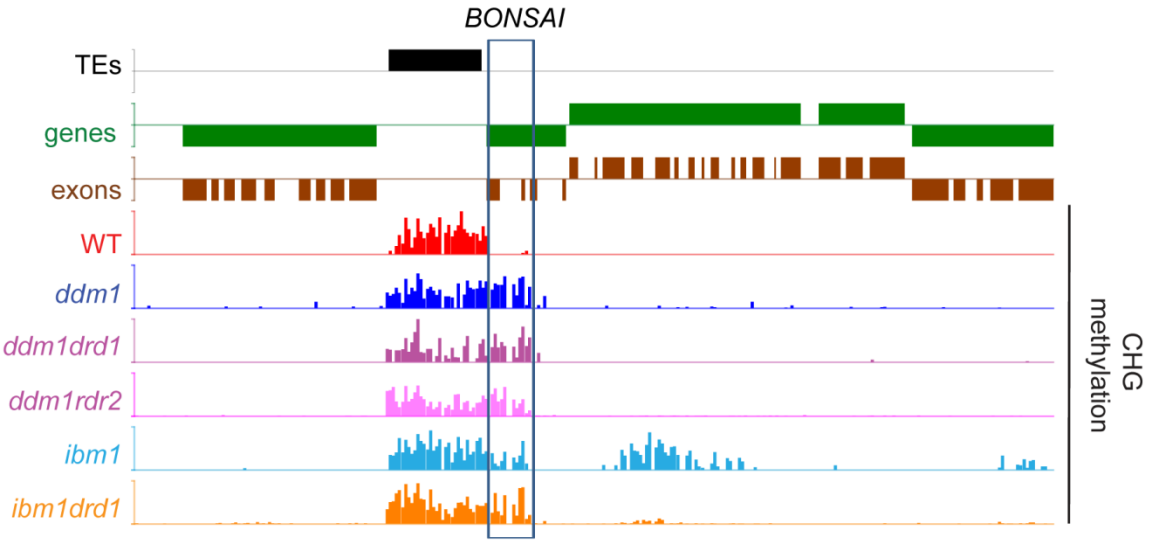

Figure 20. Hypermethylation of the *BONSAI* locus. CHG Methylation profiles for the BONSAI locus are shown for *ddm1*, *ddm1drd1*, *ddm1rdr2*, *ibm1* and *ibm1drd1*.

Figure 21

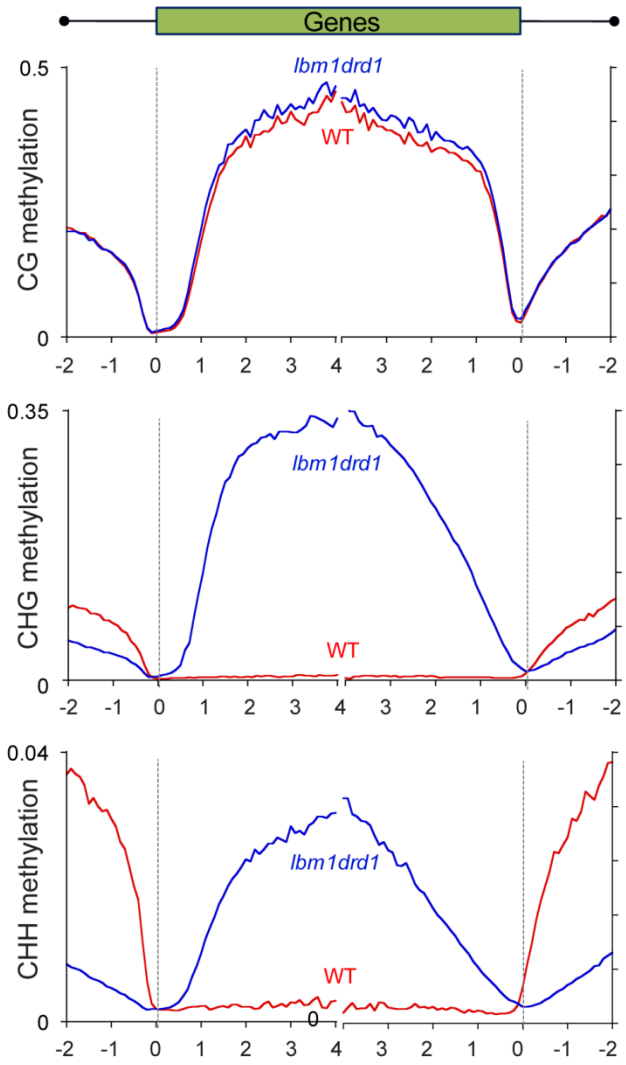

Figure 21. Patterns of DNA methylation the *ibm1drd1* mutant. Arabidopsis genes were aligned at the 5′ end or the 3′ end, and average methylation levels for each 100-bp interval are plotted. The dashed lines represent the points of alignment. Similar to results from *ibm1*, the *ibm1drd1* double mutant shows non-CG hypermethyion of genes compared with WT.

6.1 References

Aceituno, F. F., N. Moseyko, et al. (2008). "The rules of gene expression in plants: organ identity and gene body methylation are key factors for regulation of gene expression in Arabidopsis thaliana." BMC Genomics **9**: 438.

Adam, M., F. Robert, et al. (2001). "H2A.Z is required for global chromatin integrity and for recruitment of RNA polymerase II under specific conditions." Mol Cell Biol **21**(18): 6270-9.

Albert, I., T. N. Mavrich, et al. (2007). "Translational and rotational settings of H2A.Z nucleosomes across the Saccharomyces cerevisiae genome." Nature **446**(7135): 572-6.

Amat, R. and L. Gudas (2011). "RAR-gamma is Required for Correct Deposition and Removal of Suz12 and H2A.Z in Embryonic Stem Cells." Journal of Cellular Physiology **226**: 293-298.

Ashelford, K., M. E. Eriksson, et al. (2011). "Full genome re-sequencing reveals a novel circadian clock mutation in Arabidopsis." Genome Biol **12**(3): R28.

Auger, A., L. Galarneau, et al. (2008). "Eaf1 is the platform for NuA4 molecular assembly that evolutionarily links chromatin acetylation to ATP-dependent exchange of histone H2A variants." Mol Cell Biol **28**(7): 2257-70.

Barski, A., S. Cuddapah, et al. (2007). "High-resolution profiling of histone methylations in the human genome." Cell **129**(4): 823-37.

Barzily-Rokni, M., N. Friedman, et al. (2011). "Synergism between DNA methylation and macroH2A1 occupancy in epigenetic silencing of the tumor suppressor gene p16(CDKN2A)." Nucleic Acids Res **39**(4): 1326-35.

Bernatavichute, Y. V., X. Zhang, et al. (2008). "Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in Arabidopsis thaliana." PLoS One **3**(9): e3156.

Bernstein, E. and S. B. Hake (2006). "The nucleosome: a little variation goes a long way." Biochem Cell Biol **84**(4): 505-17.

Bhaumik, S. R., E. Smith, et al. (2007). "Covalent modifications of histones during development and disease pathogenesis." Nat Struct Mol Biol **14**(11): 1008-16.

Billon, P. and J. Cote (2011). "Precise deposition of histone H2A.Z in chromatin for genome expression and maintenance." Biochim Biophys Acta **1819**(3-4): 290-302.

Bowman, T. A., M. M. Wong, et al. (2011). "Loss of H2A.Z Is Not Sufficient to Determine Transcriptional Activity of Snf2-Related CBP Activator Protein or p400 Complexes." Int J Cell Biol **2011**: 715642.

Brickner, D. G., I. Cajigas, et al. (2007). "H2A.Z-mediated localization of genes at the nuclear periphery confers epigenetic memory of previous transcriptional state." PLoS Biol **5**(4): e81.

Brzeski, J. and A. Jerzmanowski (2003). "Deficient in DNA methylation 1 (DDM1) defines a novel family of chromatin-remodeling factors." J Biol Chem **278**(2): 823-8.

Buck, M. J., A. B. Nobel, et al. (2005). "ChIPOTle: a user-friendly tool for the analysis of ChIP-chip data." Genome Biol **6**(11): R97.

Carr, A. M., S. M. Dorrington, et al. (1994). "Analysis of a histone H2A variant from fission yeast: evidence for a role in chromosome stability." Mol Gen Genet **245**(5): 628-35.

Chan, S. W., D. Zilberman, et al. (2004). "RNA silencing genes control de novo DNA methylation." Science **303**(5662): 1336.

Chodavarapu, R. K., S. Feng, et al. (2010). "Relationship between nucleosome positioning and DNA methylation." Nature **466**(7304): 388-92.

Choi, K., C. Park, et al. (2007). "Arabidopsis homologs of components of the SWR1 complex regulate flowering and plant development." Development **134**(10): 1931-41.

Clarkson, M. J., J. R. Wells, et al. (1999). "Regions of variant histone His2AvD required for Drosophila development." Nature **399**(6737): 694-7.
Cokus, S. J., S. Feng, et al. (2008). "Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning." Nature **452**(7184): 215-9.
Coleman-Derr, D., Zilberman, D (2012). "Deposition of histone variant H2A.Z within gene bodies regulates responsive genes." PlOS Genetics **in review**.
Conerly, M. L., S. S. Teves, et al. (2010). "Changes in H2A.Z occupancy and DNA methylation during B-cell lymphomagenesis." Genome Res **20**(10): 1383-90.
Creyghton, M. P., S. Markoulaki, et al. (2008). "H2AZ is enriched at polycomb complex target genes in ES cells and is necessary for lineage commitment." Cell **135**(4): 649-61.
Cuadrado, A., N. Corrado, et al. (2011). "Essential role of p18Hamlet/SRCAP-mediated histone H2A.Z chromatin incorporation in muscle differentiation." EMBO J **29**(12): 2014-25.
Deal, R. B., M. K. Kandasamy, et al. (2005). "The nuclear actin-related protein ARP6 is a pleiotropic developmental regulator required for the maintenance of FLOWERING LOCUS C expression and repression of flowering in Arabidopsis." Plant Cell **17**(10): 2633-46.
Deal, R. B., C. N. Topp, et al. (2007). "Repression of flowering in Arabidopsis requires activation of FLOWERING LOCUS C expression by the histone variant H2A.Z." Plant Cell **19**(1): 74-83.
Dhillon, N., M. Oki, et al. (2006). "H2A.Z functions to regulate progression through the cell cycle." Mol Cell Biol **26**(2): 489-501.
Edwards, J. R., A. H. O'Donnell, et al. (2010). "Chromatin and sequence features that define the fine and gross structure of genomic methylation patterns." Genome Res **20**(7): 972-80.
Erdel, F. and K. Rippe (2011). "Chromatin remodelling in mammalian cells by ISWI-type complexes--where, when and why?" FEBS J **278**(19): 3608-18.
Faast, R., V. Thonglairoam, et al. (2001). "Histone variant H2A.Z is required for early mammalian development." Curr Biol **11**(15): 1183-7.
Fan, J. Y., F. Gordon, et al. (2002). "The essential histone variant H2A.Z regulates the equilibrium between different chromatin conformational states." Nat Struct Biol **9**(3): 172-6.
Feinberg, A. P., R. Ohlsson, et al. (2006). "The epigenetic progenitor origin of human cancer." Nat Rev Genet **7**(1): 21-33.
Felle, M., S. Joppien, et al. (2011). "The USP7/Dnmt1 complex stimulates the DNA methylation activity of Dnmt1 and regulates the stability of UHRF1." Nucleic Acids Res **39**(19): 8355-65.
Feng, S., S. J. Cokus, et al. (2010). "Conservation and divergence of methylation patterning in plants and animals." Proc Natl Acad Sci U S A **107**(19): 8689-94.
Finnegan, E. J. and E. S. Dennis (1993). "Isolation and identification by sequence homology of a putative cytosine methyltransferase from Arabidopsis thaliana." Nucleic Acids Res **21**(10): 2383-8.
Foret, S., R. Kucharski, et al. (2009). "Epigenetic regulation of the honey bee transcriptome: unravelling the nature of methylated genes." BMC Genomics **10**: 472.
Fritsch, O., G. Benvenuto, et al. (2004). "The INO80 protein controls homologous recombination in Arabidopsis thaliana." Mol Cell **16**(3): 479-85.
Fujimoto, S., C. Seebart, et al. (2012). "Proteome analysis of protein partners to nucleosomes containing canonical H2A or the variant histones H2A.Z or H2A.X." Biol Chem.
Furner, I. J. and M. Matzke (2011). "Methylation and demethylation of the Arabidopsis genome." Curr Opin Plant Biol **14**(2): 137-41.
Gehring, M. and S. Henikoff (2007). "DNA methylation dynamics in plant genomes." Biochim Biophys Acta.
Gendrel, A. V., Z. Lippman, et al. (2002). "Dependence of heterochromatic histone H3 methylation patterns on the Arabidopsis gene DDM1." Science **297**(5588): 1871-3.

Gevry, N., H. M. Chan, et al. (2007). "p21 transcription is regulated by differential localization of histone H2A.Z." Genes Dev **21**(15): 1869-81.

Glastad, K. M., B. G. Hunt, et al. (2011). "DNA methylation in insects: on the brink of the epigenomic era." Insect Mol Biol **20**(5): 553-65.

Goll, M. G. and T. H. Bestor (2005). "Eukaryotic cytosine methyltransferases." Annu Rev Biochem **74**: 481-514.

Guillemette, B., A. R. Bataille, et al. (2005). "Variant histone H2A.Z is globally localized to the promoters of inactive yeast genes and regulates nucleosome positioning." PLoS Biol **3**(12): e384.

Guillemette, B. and L. Gaudreau (2006). "Reuniting the contrasting functions of H2A.Z." Biochem Cell Biol **84**(4): 528-35.

Guy, J., B. Hendrich, et al. (2001). "A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome." Nat Genet **27**(3): 322-6.

Hang, M. and M. M. Smith (2011). "Genetic analysis implicates the Set3/Hos2 histone deacetylase in the deposition and remodeling of nucleosomes containing H2A.Z." Genetics **187**(4): 1053-66.

Hardy, S., P. E. Jacques, et al. (2009). "The euchromatic and heterochromatic landscapes are shaped by antagonizing effects of transcription on H2A.Z deposition." PLoS Genet **5**(10): e1000687.

He, X. J., T. Chen, et al. (2011). "Regulation and function of DNA methylation in plants and animals." Cell Res **21**(3): 442-65.

Hendrich, B., J. Guy, et al. (2001). "Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development." Genes Dev **15**(6): 710-23.

Henikoff, S. and K. Ahmad (2005). "Assembly of variant histones into chromatin." Annu Rev Cell Dev Biol **21**: 133-53.

Hsieh, T. F., C. A. Ibarra, et al. (2009). "Genome-wide demethylation of Arabidopsis endosperm." Science.

Hsieh, T. F., J. Shin, et al. (2011). "Regulation of imprinted gene expression in Arabidopsis endosperm." Proc Natl Acad Sci U S A **108**(5): 1755-62.

Huang da, W., B. T. Sherman, et al. (2009). "Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists." Nucleic Acids Res **37**(1): 1-13.

Huang da, W., B. T. Sherman, et al. (2009). "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources." Nat Protoc **4**(1): 44-57.

Huang, J., T. Fan, et al. (2004). "Lsh, an epigenetic guardian of repetitive elements." Nucleic Acids Res **32**(17): 5019-28.

Huettel, B., T. Kanno, et al. (2006). "Endogenous targets of RNA-directed DNA methylation and Pol IV in Arabidopsis." EMBO J **25**(12): 2828-36.

Huettel, B., T. Kanno, et al. (2007). "RNA-directed DNA methylation mediated by DRD1 and Pol IVb: a versatile pathway for transcriptional gene silencing in plants." Biochim Biophys Acta **1769**(5-6): 358-74.

Inagaki, S. and T. Kakutani (2010). "Control of genic DNA methylation in Arabidopsis." J Plant Res **123**(3): 299-302.

Initiative, T. A. G. (2000). "Analysis of the genome sequence of the flowering plant Arabidopsis thaliana." Nature **408**(6814): 796-815.

Ishibashi, T., D. Dryhurst, et al. (2009). "Acetylation of vertebrate H2A.Z and its effect on the structure of the nucleosome." Biochemistry **48**(22): 5007-17.

Jackson, J. D. and M. A. Gorovsky (2000). "Histone H2A.Z has a conserved function that is distinct from that of the major H2A sequence variants." Nucleic Acids Res **28**(19): 3811-6.

Jacobsen, S. E. and E. M. Meyerowitz (1997). "Hypermethylated SUPERMAN epigenetic alleles in arabidopsis." Science **277**(5329): 1100-3.

Jeddeloh, J. A., T. L. Stokes, et al. (1999). "Maintenance of genomic methylation requires a SWI2/SNF2-like protein [see comments]." Nat Genet **22**(1): 94-7.

Jensen, K., M. S. Santisteban, et al. (2011). "Histone H2A.Z acid patch residues required for deposition and function." Mol Genet Genomics **285**(4): 287-96.

Jin, C., C. Zang, et al. (2009). "H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions." Nat Genet **41**(8): 941-5.

Johnson, L. M., M. Bostick, et al. (2007). "The SRA methyl-cytosine-binding domain links DNA and histone methylation." Curr Biol **17**(4): 379-84.

Jones, P. A. and P. W. Laird (1999). "Cancer epigenetics comes of age." Nat Genet **21**(2): 163-7.

Kankel, M. W., D. E. Ramsey, et al. (2003). "Arabidopsis MET1 cytosine methyltransferase mutants." Genetics **163**(3): 1109-22.

Kanno, T., W. Aufsatz, et al. (2005). "A SNF2-like protein facilitates dynamic control of DNA methylation." EMBO Rep **6**(7): 649-55.

Kanno, T., B. Huettel, et al. (2005). "Atypical RNA polymerase subunits required for RNA-directed DNA methylation." Nat Genet **37**(7): 761-5.

Kanno, T., M. F. Mette, et al. (2004). "Involvement of putative SNF2 chromatin remodeling protein DRD1 in RNA-directed DNA methylation." Curr Biol **14**(9): 801-5.

Khorasanizadeh, S. (2004). "The nucleosome: from genomic organization to genomic regulation." Cell **116**(2): 259-72.

Kinoshita, T., A. Miura, et al. (2004). "One-way control of FWA imprinting in Arabidopsis endosperm by DNA methylation." Science **303**(5657): 521-3.

Kleinboelting, N., G. Huep, et al. (2012). "GABI-Kat SimpleSearch: new features of the Arabidopsis thaliana T-DNA mutant database." Nucleic Acids Res **40**(Database issue): D1211-5.

Klose, R. J. and A. P. Bird (2006). "Genomic DNA methylation: the mark and its mediators." Trends Biochem Sci **31**(2): 89-97.

Kobor, M. S., S. Venkatasubrahmanyam, et al. (2004). "A protein complex containing the conserved Swi2/Snf2-related ATPase Swr1p deposits histone variant H2A.Z into euchromatin." PLoS Biol **2**(5): E131.

Krogan, N. J., K. Baetz, et al. (2004). "Regulation of chromosome stability by the histone H2A variant Htz1, the Swr1 chromatin remodeling complex, and the histone acetyltransferase NuA4." Proc Natl Acad Sci U S A **101**(37): 13513-8.

Kumar, S. V. and P. A. Wigge (2010). "H2A.Z-containing nucleosomes mediate the thermosensory response in Arabidopsis." Cell **140**(1): 136-47.

Kusch, T., L. Florens, et al. (2004). "Acetylation by Tip60 is required for selective histone variant exchange at DNA lesions." Science **306**(5704): 2084-7.

Langmead, B., C. Trapnell, et al. (2009). "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome." Genome Biol **10**(3): R25.

Langst, G. and P. B. Becker (2004). "Nucleosome remodeling: one mechanism, many phenomena?" Biochim Biophys Acta **1677**(1-3): 58-63.

Larochelle, M. and L. Gaudreau (2003). "H2A.Z has a function reminiscent of an activator required for preferential binding to intergenic DNA." EMBO J **22**(17): 4512-22.

Law, J. A. and S. E. Jacobsen (2010). "Establishing, maintaining and modifying DNA methylation patterns in plants and animals." Nat Rev Genet **11**(3): 204-20.

Li, B., S. G. Pattenden, et al. (2005). "Preferential occupancy of histone variant H2AZ at inactive promoters influences local histone modifications and chromatin remodeling." Proc Natl Acad Sci U S A **102**(51): 18385-90.

Light, W. H., D. G. Brickner, et al. (2010). "Interaction of a DNA zip code with the nuclear pore complex promotes H2A.Z incorporation and INO1 transcriptional memory." Mol Cell **40**(1): 112-25.
Lindroth, A. M., X. Cao, et al. (2001). "Requirement of CHROMOMETHYLASE3 for Maintenance of CpXpG Methylation." Science **10**: 10.
Lindroth, A. M., D. Shultis, et al. (2004). "Dual histone H3 methylation marks at lysines 9 and 27 required for interaction with CHROMOMETHYLASE3." EMBO J **23**(21): 4286-96.
Lippman, Z., A. V. Gendrel, et al. (2004). "Role of transposable elements in heterochromatin and epigenetic control." Nature **430**(6998): 471-6.
Lippman, Z., B. May, et al. (2003). "Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification." PLoS Biol **1**(3): E67.
Lister, R., R. C. O'Malley, et al. (2008). "Highly Integrated Single-Base Resolution Maps of the Epigenome in Arabidopsis." Cell.
Lister, R., R. C. O'Malley, et al. (2008). "Highly Integrated Single-Base Resolution Maps of the Epigenome in Arabidopsis." Cell **133**(3): 523-36.
Lister, R., M. Pelizzola, et al. (2009). "Human DNA methylomes at base resolution show widespread epigenomic differences." Nature **462**(7271): 315-22.
Liu, X., B. Li, et al. (1996). "Essential and nonessential histone H2A variants in Tetrahymena thermophila." Mol Cell Biol **16**(8): 4305-11.
Lorincz, M. C., D. R. Dickerson, et al. (2004). "Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells." Nat Struct Mol Biol **11**(11): 1068-75.
Luco, R. F., Q. Pan, et al. (2010). "Regulation of alternative splicing by histone modifications." Science **327**(5968): 996-1000.
Malik, H. S. and S. Henikoff (2003). "Phylogenomics of the nucleosome." Nat Struct Biol **10**(11): 882-91.
March-Diaz, R., M. Garcia-Dominguez, et al. (2007). "SEF, a new protein required for flowering repression in Arabidopsis, interacts with PIE1 and ARP6." Plant Physiol **143**(2): 893-901.
March-Diaz, R., M. Garcia-Dominguez, et al. (2008). "Histone H2A.Z and homologues of components of the SWR1 complex are required to control immunity in Arabidopsis." Plant J **53**(3): 475-87.
Mathieu, O., J. Reinders, et al. (2007). "Transgenerational stability of the Arabidopsis epigenome is coordinated by CG methylation." Cell **130**(5): 851-62.
Matsuda, R., T. Hori, et al. (2010). "Identification and characterization of the two isoforms of the vertebrate H2A.Z histone variant." Nucleic Acids Res **38**(13): 4263-73.
Mavrich, T. N., C. Jiang, et al. (2008). "Nucleosome organization in the Drosophila genome." Nature **453**(7193): 358-62.
Meneghini, M. D., M. Wu, et al. (2003). "Conserved histone variant H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin." Cell **112**(5): 725-36.
Meyer, P. (2010). "DNA methylation systems and targets in plants." FEBS Letters **585**(13): 2008-20015.
Millar, C. B., F. Xu, et al. (2006). "Acetylation of H2AZ Lys 14 is associated with genome-wide gene activity in yeast." Genes Dev **20**(6): 711-22.
Mito, Y., J. G. Henikoff, et al. (2005). "Genome-scale profiling of histone H3.3 replacement patterns." Nat Genet **37**(10): 1090-7.
Miura, A., M. Nakamura, et al. (2009). "An Arabidopsis jmjC domain protein protects transcribed genes from DNA methylation at CHG sites." EMBO J **28**(8): 1078-86.
Mizuguchi, G., X. Shen, et al. (2004). "ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex." Science **303**(5656): 343-8.
Morillo-Huesca, M., M. Clemente-Ruiz, et al. (2010). "The SWR1 histone replacement complex causes genetic instability and genome-wide transcription misregulation in the absence of H2A.Z." PLoS One **5**(8): e12143.

Morrison, A. J. and X. Shen (2009). "Chromatin remodelling beyond transcription: the INO80 and SWR1 complexes." Nat Rev Mol Cell Biol **10**(6): 373-84.

Noh, Y. S. and R. M. Amasino (2003). "PIE1, an ISWI family gene, is required for FLC activation and floral repression in Arabidopsis." Plant Cell **15**(7): 1671-82.

Papamichos-Chronakis, M., S. Watanabe, et al. (2011). "Global regulation of H2A.Z localization by the INO80 chromatin-remodeling enzyme is essential for genome integrity." Cell **144**(2): 200-13.

Park, Y. J., P. N. Dyer, et al. (2004). "A new fluorescence resonance energy transfer approach demonstrates that the histone variant H2AZ stabilizes the histone octamer within the nucleosome." J Biol Chem **279**(23): 24274-82.

Petter, M., C. C. Lee, et al. (2011). "Expression of P. falciparum var genes involves exchange of the histone variant H2A.Z at the promoter." PLoS Pathog **7**(2): e1001292.

Placek, B. J., L. N. Harrison, et al. (2005). "The H2A.Z/H2B dimer is unstable compared to the dimer containing the major H2A isoform." Protein Sci **14**(2): 514-22.

Raisner, R. M., P. D. Hartley, et al. (2005). "Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin." Cell **123**(2): 233-48.

Rangasamy, D., L. Berven, et al. (2003). "Pericentric heterochromatin becomes enriched with H2A.Z during early mammalian development." EMBO J **22**(7): 1599-607.

Redon, C., D. Pilch, et al. (2002). "Histone H2A variants H2AX and H2AZ." Curr Opin Genet Dev **12**(2): 162-9.

Ridgway, P., K. D. Brown, et al. (2004). "Unique residues on the H2A.Z containing nucleosome surface are important for Xenopus laevis development." J Biol Chem **279**(42): 43815-20.

Ruhl, D. D., J. Jin, et al. (2006). "Purification of a human SRCAP complex that remodels chromatin by incorporating the histone variant H2A.Z into nucleosomes." Biochemistry **45**(17): 5671-7.

Sadeghi, L., C. Bonilla, et al. (2011). "Podbat: a novel genomic tool reveals Swr1-independent H2A.Z incorporation at gene coding sequences through epigenetic meta-analysis." PLoS Comput Biol **7**(8): e1002163.

Santisteban, M. S., T. Kalashnikova, et al. (2000). "Histone H2A.Z regulats transcription and is partially redundant with nucleosome remodeling complexes." Cell **103**(3): 411-22.

Sarda, S., J. Zeng, et al. (2012). "The Evolution of Invertebrate Gene Body Methylation." Mol Biol Evol.

Sarma, K. and D. Reinberg (2005). "Histone variants meet their match." Nat Rev Mol Cell Biol **6**(2): 139-49.

Sasaki, T., A. Kobayashi, et al. (2012). "RNAi-independent de novo DNA methylation revealed in Arabidopsis mutants of chromatin remodeling gene DDM1." Plant J.

Saze, H. and T. Kakutani (2007). "Heritable epigenetic mutation of a transposon-flanked Arabidopsis gene due to lack of the chromatin-remodeling factor DDM1." EMBO J **26**(15): 3641-52.

Saze, H. and T. Kakutani (2011). "Differentiation of epigenetic modifications between transposons and genes." Curr Opin Plant Biol **14**(1): 81-7.

Saze, H., O. Mittelsten Scheid, et al. (2003). "Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis." Nat Genet **34**(1): 65-9.

Saze, H., A. Shiraishi, et al. (2008). "Control of genic DNA methylation by a jmjC domain-containing protein in Arabidopsis thaliana." Science **319**(5862): 462-5.

Schmid, M., T. S. Davison, et al. (2005). "A gene expression map of Arabidopsis thaliana development." Nat Genet **37**(5): 501-6.

Segal, E. and J. Widom (2009). "What controls nucleosome positions?" Trends Genet **25**(8): 335-43.

Siegel, T. N., D. R. Hekstra, et al. (2009). "Four histone variants mark the boundaries of polycistronic transcription units in Trypanosoma brucei." Genes Dev **23**(9): 1063-76.

Smith, A. P., A. Jain, et al. (2009). "Histone H2A.Z regulates the expression of several classes of phosphate starvation response genes but not as a transcriptional activator." Plant Physiol **152**(1): 217-25.

Suto, R. K., M. J. Clarkson, et al. (2000). "Crystal structure of a nucleosome core particle containing the variant histone H2A.Z." Nat Struct Biol **7**(12): 1121-4.

Suzuki, M. M. and A. Bird (2008). "DNA methylation landscapes: provocative insights from epigenomics." Nat Rev Genet **9**(6): 465-76.

Suzuki, M. M., A. R. Kerr, et al. (2007). "CpG methylation is targeted to transcription units in an invertebrate genome." Genome Res **17**(5): 625-31.

Swaminathan, J., E. M. Baxter, et al. (2005). "The role of histone H2Av variant replacement and histone H4 acetylation in the establishment of Drosophila heterochromatin." Genes Dev **19**(1): 65-76.

Takuno, S. and B. S. Gaut (2012). "Body-methylated genes in Arabidopsis thaliana are functionally important and evolve slowly." Mol Biol Evol **29**(1): 219-27.

Talbert, P. B. and S. Henikoff (2010). "Histone variants--ancient wrap artists of the epigenome." Nat Rev Mol Cell Biol **11**(4): 264-75.

Tao, Y., S. Xi, et al. (2011). "Lsh, chromatin remodeling family member, modulates genome-wide cytosine methylation patterns at nonrepeat sequences." Proc Natl Acad Sci U S A **108**(14): 5626-31.

Teixeira, F. K., F. Heredia, et al. (2009). "A role for RNAi in the selective correction of DNA methylation defects." Science **323**(5921): 1600-4.

Thambirajah, A. A., D. Dryhurst, et al. (2006). "H2A.Z stabilizes chromatin in a way that is dependent on core histone acetylation." J Biol Chem **281**(29): 20036-44.

Tolstorukov, M. Y., P. V. Kharchenko, et al. (2009). "Comparative analysis of H2A.Z nucleosome organization in the human and yeast genomes." Genome Res **19**(6): 967-77.

Tran, R. K., J. G. Henikoff, et al. (2005). "DNA methylation profiling identifies CG methylation clusters in Arabidopsis genes." Curr Biol **15**(2): 154-9.

Tsukahara, S., A. Kobayashi, et al. (2009). "Bursts of retrotransposition reproduced in Arabidopsis." Nature **461**(7262): 423-6.

Turck, F., F. Roudier, et al. (2007). "Arabidopsis TFL2/LHP1 specifically associates with genes marked by trimethylation of histone H3 lysine 27." PLoS Genet **3**(6): e86.

Updike, D. L. and S. E. Mango (2006). "Temporal regulation of foregut development by HTZ-1/H2A.Z and PHA-4/FoxA." PLoS Genet **2**(9): e161.

van Daal, V. H. and A. van der Leij (1992). "Computer-based reading and spelling practice for children with learning disabilities." J Learn Disabil **25**(3): 186-95.

Vaughn, M. W., M. Tanurd Ic, et al. (2007). "Epigenetic Natural Variation in Arabidopsis thaliana." PLoS Biol **5**(7): e174.

Venkatasubrahmanyam, S., W. W. Hwang, et al. (2007). "Genome-wide, as opposed to local, antisilencing is mediated redundantly by the euchromatic factors Set1 and H2A.Z." Proc Natl Acad Sci U S A **104**(42): 16609-14.

Vining, K. J., K. R. Pomraning, et al. (2012). "Dynamic DNA cytosine methylation in the Populus trichocarpa genome: tissue-level variation and relationship to gene expression." BMC Genomics **13**: 27.

Wan, Y., R. A. Saleem, et al. (2009). "Role of the histone variant H2A.Z/Htz1p in TBP recruitment, chromatin dynamics, and regulated expression of oleate-responsive genes." Mol Cell Biol **29**(9): 2346-58.

Wang, A. Y., M. J. Aristizabal, et al. (2011). "Key functional regions in the histone variant H2A.Z C-terminal docking domain." Mol Cell Biol **31**(18): 3871-84.

Whittle, C. M., K. N. McClinic, et al. (2008). "The genomic distribution and function of histone variant HTZ-1 during C. elegans embryogenesis." PLoS Genet **4**(9): e1000187.
Wong, M. M., L. K. Cox, et al. (2007). "The chromatin remodeling protein, SRCAP, is critical for deposition of the histone variant H2A.Z at promoters." J Biol Chem **282**(36): 26132-9.
Wu, W. H., S. Alami, et al. (2005). "Swc2 is a widely conserved H2AZ-binding module essential for ATP-dependent histone exchange." Nat Struct Mol Biol **12**(12): 1064-71.
Xiang, H., J. Zhu, et al. (2010). "Single base-resolution methylome of the silkworm reveals a sparse epigenomic map." Nat Biotechnol **28**(5): 516-20.
Xiao, W., M. Gehring, et al. (2003). "Imprinting of the MEA Polycomb gene is controlled by antagonism between MET1 methyltransferase and DME glycosylase." Dev Cell **5**(6): 891-901.
Xie, Z., E. Allen, et al. (2005). "DICER-LIKE 4 functions in trans-acting small interfering RNA biogenesis and vegetative phase change in Arabidopsis thaliana." Proc Natl Acad Sci U S A **102**(36): 12984-9.
Yan, Q., E. Cho, et al. (2003). "Association of Lsh, a regulator of DNA methylation, with pericentromeric heterochromatin is dependent on intact heterochromatin." Mol Cell Biol **23**(23): 8416-28.
Yi, H., N. Sardesai, et al. (2006). "Constitutive expression exposes functional redundancy between the Arabidopsis histone H2A gene HTA1 and other H2A gene family members." Plant Cell **18**(7): 1575-89.
Zanton, S. J. and B. F. Pugh (2006). "Full and partial genome-wide assembly and disassembly of the yeast transcription machinery in response to heat shock." Genes Dev **20**(16): 2250-65.
Zemach, A., M. Y. Kim, et al. (2010). "Local DNA hypomethylation activates genes in rice endosperm." Proc Natl Acad Sci U S A **107**(43): 18729-34.
Zemach, A., Y. Li, et al. (2005). "DDM1 binds Arabidopsis methyl-CpG binding domain proteins and affects their subnuclear localization." Plant Cell **17**(5): 1549-58.
Zemach, A., I. E. McDaniel, et al. (2010). "Genome-wide evolutionary analysis of eukaryotic DNA methylation." Science **328**(5980): 916-9.
Zemach, A. and D. Zilberman (2010). "Evolution of eukaryotic DNA methylation and the pursuit of safer sex." Curr Biol **20**(17): R780-5.
Zeng, J. and S. V. Yi (2010). "DNA methylation and genome evolution in honeybee: gene length, expression, functional enrichment covary with the evolutionary signature of DNA methylation." Genome Biol Evol **2**: 770-80.
Zhang, H., D. N. Roberts, et al. (2005). "Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss." Cell **123**(2): 219-31.
Zhang, X. and S. E. Jacobsen (2006). "Genetic Analyses of DNA Methyltransferases in Arabidopsis thaliana." Cold Spring Harb Symp Quant Biol **71**: 439-47.
Zhang, X., J. Yazaki, et al. (2006). "Genome-wide High-Resolution Mapping and Functional Analysis of DNA Methylation in Arabidopsis." Cell **126**(6): 1189-201.
Zhu, H., T. M. Geiman, et al. (2006). "Lsh is involved in de novo methylation of DNA." EMBO J **25**(2): 335-45.
Zilberman, D., X. Cao, et al. (2003). "ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation." Science **299**(5607): 716-9.
Zilberman, D., D. Coleman-Derr, et al. (2008). "Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks." Nature **456**(7218): 125-9.
Zilberman, D., M. Gehring, et al. (2007). "Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription." Nat Genet **39**(1): 61-9.
Zlatanova, J. and A. Thakar (2008). "H2A.Z: view from the top." Structure **16**(2): 166-79.