UCLA
UCLA Electronic Theses and Dissertations

Title

Epigenetic regulation of gene silencing and DNA replication in Arabidopsis thaliana

Permalink

https://escholarship.org/uc/item/4dt855fr

Author

Akahori Stroud-Drinkwater, Hume

Publication Date

2013

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Epigenetic regulation of gene silencing and DNA replication in *Arabidopsis thaliana*

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy in

Molecular, Cell, and Developmental Biology

by

Hume Akahori Stroud-Drinkwater

2013

ABSTRACT OF THE DISSERTATION

Epigenetic regulation of gene silencing and DNA replication in *Arabidopsis thaliana*

by

Hume Akahori Stroud-Drinkwater

Doctor of Philosophy in Molecular, Cell, and Developmental Biology

University of California, Los Angeles, 2013

Professor Steven E. Jacobsen, Chair

Chemical modifications to histones and DNA regulate various biological processes including transcriptional silencing and DNA replication. One of the most well studied chromatin modifications in *Arabidopsis thaliana* is DNA methylation and its role in silencing transposable elements (TEs) and genes. In *Arabidopsis*, certain DNA methylation pathways are controlled by histone H3 lysine 9 methylation, a histone modification associated with heterochromatin. A much less characterized heterochromatic mark is histone H3 lysine 27 monomethylation (H3K27me1). Two SET domain proteins, ARABIDOPSIS TRITHORAX-RELATED PROTEIN5 (ATXR5) and ATXR6, were found to catalyze H3K27me1. In *atxr5 atxr6* double mutants, transcriptional reactivation of TEs was observed without global defects in DNA methylation levels. Thus, unlike H3K9 methylation, H3K27me1 appeared to be involved in a silencing system independent of DNA methylation.

In addition to transcriptional reactivation of TEs, we found that *atxr5 atxr6* mutants show increased copies of heterochromatic DNA, suggesting that ATXR5 and ATXR6 are involved in a pathway that prevents heterochromatin from over-replicating.

To gain better understanding about DNA methylation and H3K27me1, we performed genome-wide mapping of these chromatin marks. We profiled DNA methylation in a comprehensive list of mutants and characterized locations different DNA methylation pathways act in the genome. We also profiled H3K27me1 and found that it is enriched at sites heavily DNA methylated. Sites of over-replication in *atxr5 atxr6* mutants correlated with sites normally enriched with H3K27me1, consistent with the fact that ATXR5 and ATXR6 catalyze this mark.

Given the overlap between H3K27me1 and DNA methylation across the genome, we explored the relationships between ATXR5 ATXR6 and DNA methylation in regulating gene silencing and DNA replication. We found that ATXR5 ATXR6 and DNA methylation cooperatively silence TEs through independent pathways, indicating that multiple pathways redundantly silence many TEs across the *Arabidopsis* genome. In contrast, we found that ATXR5 ATXR6 and DNA methylation antagonistically regulate heterochromatic DNA replication, suggesting a complex relationship between these chromatin marks in regulating transcriptional silencing TEs and heterochromatic DNA replication.

Taken together, our results provide insight into the roles of epigenetic marks in regulating gene silencing and DNA replication.

The dissertation of Hume Akahori Stroud-Drinkwater is approved.

Chentao Lin

Matteo Pellegrini

James A. Wohlschlegel

Steven E. Jacobsen, Committee Chair

University of California, Los Angeles

2013

TABLE OF CONTENTS

FIGURES AND TABLES

Chapter 1

Chapter 2

Chapter 4

ACKNOWLEDGMENTS

Chapter2:

Chapter Two is the version of the research article published in *Nature* 2010 466, 987-91 entitled "Regulation of heterochromatic DNA replication by histone H3 lysine 27 methyltransferases".

Chapter3:

Chapter Three is the version of the research article published in *Nature Structural and Molecular Biology* 2011 18(3): 395-400 entitled "Genome-wide mapping of Arabidopsis origins of DNA replication and their associated epigenetic marks." This paper is authored by Celina Costas, Maria de la Paz Sanchez, Hume Stroud, Yanchun Yu, Juan Carlos Oliveros, Suhua Feng, Alberto Benguria, Irene López-Vidriero, Xiaoyu Zhang, Roberto Solano, Steven E. Jacobsen and Crisanto Gutierrez. Crisanto Gutierrez wrote most of the paper with input from

Chapter4:

Chapter Four is the version of the research article published in *PLoS Genetics* 2012 8(7): e1002808 entitled "DNA methyltransferases are required to induce heterochromatic re-replication in Arabidopsis". This paper is authored by Hume Stroud, Christopher J. Hale, Suhua Feng, Elena Caro, Yannick Jacob, Scott D. Michaels, Steven E. Jacobsen. Hume Stroud wrote the manuscript. Flow cytometry was performed in the UCLA Jonsson Comprehensive Cancer Center Flow Cytometry Core Facility. HS was supported by a Fred Eiserling and Judith Lengyel Graduate Doctorate Fellowship. EC was supported by a Marie Curie International Outgoing Fellowship for Career Development within the 7th European Community Framework Program.

Vita

Education:

2008-present   Graduate Student (Advisor: Steven E. Jacobsen)

                University of California, Los Angeles, CA

2007   Bachelor of Arts, College of Arts and Sciences, Cornell University, Ithaca, NY

Previous Employment:

2007   Clinical Research Associate, Otsuka Pharmaceutical Co., Ltd., Tokyo, Japan

Published manuscripts:

Denotes equal contribution.

16. **Stroud, H.** et al. (2013) *eLife*, 2: e00354.

15. Xu, Y. et al. (2013) *Current Biology*, 23(4): 345-50.

14. **Stroud, H.** et al. (2013) *Cell*, 152(1-2): 352-64.

13. Deleris, A.*, **Stroud H.*** et al. (2012) *PLOS Genetics*, 8(11): e1003062.

12. Caro, E. et al. (2012) *PLOS Genetics*, 8(10): e1002995.

11. Du, J.*, Zhong, X.* et al. (2012) *Cell*, 151(1): 167-80.

10. **Stroud, H.** et al. (2012) *PLOS Genetics*, 8(7): e1002808.

9. Moissiard, G. et al. (2012) *Science*, 336(6087): 1448-51.

8. Pontvianne, F. et al. (2012) *Genes and Development*, 26(9): 945-57.

7. **Stroud, H.***, Otero, S.* et al. (2012) *Proc. Nat. Acad. Sci. U. S. A.,* 109(14): 5370-5.

6. **Stroud, H.** et al. (2011) *Genome Biology*, 12: R54.

5. Kinney, S.M.*, Chin, H.G.* et al. (2011) *Journal of Biological Chemistry*, 286(28): 24685-93.

4. Costas, C.*, Sanchez, M.P.*, **Stroud, H.*** et al. (2011) *Nature Structural and Molecular Biology*, 18(3): 395-400.

3. Jacob, Y.*, **Stroud, H.*** et al. (2010) *Nature*, 466(7309): 987-91.

2. Chodavarapu, R.K.*, Feng, S.* et al. (2010) *Nature*, 466(7304): 388-92.

1. Jacob, Y. et al. (2009) *Nature Structural and Molecular Biology*, 16(7): 763-8.

Conference talks:

Epigenetic regulation of DNA replication in *Arabidopsis thaliana*. American Society of Plant Biologist Annual Meeting Mini-symposium. August 2011, Minneapolis, MN, USA. (Oral presentation). *presenter.

Selected Honors:

2012-2013   Dissertation Year Fellowship, University of California, Los Angeles

2010-2011   Judith Lengyel and Fred Eiserling Graduate Doctoral Fellowship in Life Sciences

2007   Phi Beta Kappa, Cornell University

INTRODUCTION

In eukaryotic cells DNA is packaged into nucleosomes, where approximately 147 base pairs of DNA is wrapped around eight core histone proteins consisting of H2A, H2B, H3 and H4 (Kornberg and Lorch, 1999; Luger et al., 1997). Nucleosomes are basic units of chromatin. The histones are subject to post-translational modifications such as acetylation, methylation, phosphorylation and ubiquitination (Strahl and Allis, 2000). The fifth carbon of cytosine in DNA is also subject to methylation. While genetic information is coded in the DNA, these chromatin modifications provide an extra layer of information. There is growing evidence indicating that modifications on chromatin regulate various biological processes including gene expression and DNA replication. Histone modifications have different functions depending on the type of modification and location the modification occurs on the histone tail (Strahl and Allis, 2000). DNA methylation can also occur in different cytosine sequence contexts that are mediated by different pathways (Law and Jacobsen, 2010), and may potentially have different functions. Chromatin modifications are heritable information that regulate processes such as gene expression without changes in underlying DNA sequences, and therefore are "epigenetic" (Ptashne and Gann, 2002).

In eukaryotes there are generally two distinct types of chromatin states: euchromatin and heterochromatin (Grewal and Elgin, 2007). Euchromatin and heterochromatin are associated with distinct chromatin modifications, nucleosome packing and higher-order chromatin structures (Grewal and Elgin, 2002; Lieberman-Aiden et al., 2009). Euchromatin is a lightly packed form of chromatin that is enriched in protein-coding genes and more accessible to transcription. Actively transcribed genes are associated with certain chromatin modifications such as histone acetylation and histone H3 lysine 4 methylation (Kouzarides, 2007; Li et al.,

2007). In contrast, heterochromatin is a densely packed form of chromatin that is gene-poor and highly enriched in repetitive DNA including transposable elements (TEs). A high proportion of eukaryotic genomes are made up with TEs (Lander et al., 2001; SanMiguel et al., 1996; Waterston et al., 2002). In *Arabidopsis thaliana*, heterochromatin is highly enriched near centromeres (TheArabidopsisGenomeInitiative, 2000). Heterochromatin is generally in a transcriptionally silent state, and certain chromatin modifications such as histone H3 lysine 9 and 27 methylation are required to keep the chromatin in this state (Kouzarides, 2007). Removal of certain modifications associated with heterochromatin may result in ectopic expression of TEs and repeats which potentially may be harmful to the host genome.

Euchromatic and heterochromatic regions of the genome not only are associated with distinct transcriptional states, but also appear to have different properties in timing of DNA replication in S phase of the cell cycle. In eukaryotes including plants, euchromatin is generally early replicating, whereas heterochromatin is generally late replicating (Birney et al., 2007; Hiratani et al., 2008; Karnani et al., 2007a; Lee et al., 2010), although in yeast, certain types of heterochromatin have been reported to be early replicating (Kim et al., 2003; Raghuraman et al., 2001). While it remains unclear how DNA replication timing is regulated, it is tempting to speculate that certain chromatin modifications may play a role. Indeed, building evidence suggest that specific chromatin modifications may regulate origins of replication, which are sites of the genome where DNA replication is initiated.

An important question that is poorly understood is the relationship between transcription and DNA replication. Recent studies in several eukaryotes indicate an association of active

transcription to active origins of replication. However, whether there is a mechanistic link between gene regulation and DNA replication remains poorly understood. The roles of chromatin modifications in regulating these processes are of interest.

Details regarding DNA methylation, histone modifications and DNA replication are discussed in subsequent sections with emphasis on *Arabidopsis thaliana*.

DNA methylation

DNA methylation is an evolutionarily ancient modification involved in various biological processes such as silencing of transposable elements (TEs), imprinting and X chromosome inactivation (Law and Jacobsen, 2010). Loss of DNA methylation results in embryonic lethality in mouse, suggesting that DNA methylation is a critical mark for viability (Meissner, 2010). It is interesting to note that some model organisms such as *Drosophila melanogaster* and *Caenorhadbitis elegans* have lost DNA methylation, and a possibility is that certain histone modifications may have replaced the roles of DNA methylation (Nanty et al., 2011).

Arabidopsis thaliana is an attractive model organism to study DNA methylation since they are able to tolerate mutations that essentially eliminate DNA methylation. In *Arabidopsis thaliana* DNA is methylated in three cytosine contexts, CG, CHG and CHH (where H=A, T or C) (Law and Jacobsen, 2010). This is in contrast to mammalian DNA, which is predominantly methylated in CG contexts by DNMT1, except for certain cell types such as embryonic stem cells and brain cells where non-CG methylation have recently been reported (Chen et al., 2011; Laurent et al.,

2010; Lister et al., 2009; Xie et al., 2012). In *Arabidopsis thaliana,* CG sites are methylated by METHYLTRANSFERASE1 (MET1), the homolog of mammalian DNMT1. CHG sites are methylated by a plant-specific CHROMOMETHYLASE3 (CMT3) and is controlled by binding to histone H3 lysine 9 (H3K9) methylation, indicating cross-talks between modifications on histones and DNA (Du et al., 2012; Ebbs and Bender, 2006; Jackson et al., 2002; Lindroth et al., 2001). Certain CHH sites are methylated by DOMAINS REARRANGED METHYLTRANSFERASE2 (DRM2), a homolog of mammalian DNMT3, in a RNA-directed DNA methylation pathway that is targeted by 24nt small interfering RNAs (siRNAs) (Law and Jacobsen, 2010). In addition, chromatin remodeler DECREASE IN DNA METHYLATION 1 (DDM1) regulates DNA methylation in all sequence contexts possibly by allowing DNA methyltransferases to access chromatin (Jeddeloh et al., 1999; Vongs et al., 1993; Zemach et al., 2013).

Whole-genome profiling methods for mapping DNA methylation have revealed genome-wide DNA methylation patterns in *Arabidopsis thaliana* (Cokus et al., 2008; Lister et al., 2008; Zhang et al., 2006; Zilberman et al., 2007). While the mammalian genome is heavily methylated in CG contexts (Cokus et al., 2008; Feng et al., 2010; Lister et al., 2009; Meissner et al., 2008; Zemach et al., 2010), the *Arabidopsis thaliana* genome contained generally lower levels of DNA methylation, where global CG methylation levels were 24%, CHG methylation levels were 6.7% and CHH methylation levels were 1.7% (Cokus et al., 2008). Genome-wide correlations between nucleosome positioning and DNA methylation levels have indicated that DNA methylation is preferentially targeted to nucleosomal DNA in both *Arabidopsis thaliana* and humans

(Chodavarapu et al., 2010), suggesting that nucleosome positioning in part determines DNA methylation patterning in eukaryotes.

Unlike in mammals, most DNA methylation in plants is concentrated in heterochromatic regions of the genome (Cokus et al., 2008; Feng et al., 2010; Lister et al., 2008; Zemach et al., 2010). Thus TEs in plants are highly methylated in all cytosine contexts. This indicates a distinct difference in DNA methylation patterning control between plants and mammals. Loss of DNA methylation induces transcriptional reactivation of certain TEs across the genome, suggesting that DNA methylation is required for silencing TEs (Lister et al., 2008; Stroud et al., 2012a; Zhang et al., 2006). Furthermore, loss of DNA methylation causes other defects such as loss of H3K9 methylation (Johnson et al., 2002; Johnson et al., 2007; Mathieu et al., 2005; Soppe et al., 2002) and decondensation of normally compact chromocenters (Mittelsten Scheid et al., 2002; Soppe et al., 2002).

DNA methylation at promoters of protein-coding genes is implicated in gene silencing (Bird, 2002). Indeed, this appears to be the case in plants, where loss of DNA methylation at promoters can cause upregulation of expression levels of genes (Becker et al., 2011; Schmitz et al., 2011; Stroud et al., 2013a; Zhang et al., 2006; Zhong et al., 2012). Transcribed regions of genes can also be DNA methylated. Intragenic methylation (gene-body methylation) is a highly conserved feature of eukaryotic genomes (Feng et al., 2010; Zemach et al., 2010), and in *Arabidopsis thaliana* this occurs exclusively in CG contexts and is present on more than a third of expressed genes in the genome (Cokus et al., 2008; Lister et al., 2008; Zhang et al., 2006; Zilberman et al., 2007). Nevertheless, the function of gene-body methylation remains poorly understood. Loss of

6

CG methylation in *Arabidopsis thaliana* causes developmental morphological defects that stochastically aggravates across inbred generations (Finnegan et al., 1996), which at least in part is likely due to ectopic H3K9 hypermethylation (and thus hyper non-CG methylation) and redistribution of H3K27 trimethylation across the genome (Deleris et al., 2012; Mathieu et al., 2007).

Most studies on DNA methylation pathways had been restricted to selected loci. However the locations in the genome that DNA methylation pathways are targeted was poorly characterized. Furthermore, some interplay between DNA methylation pathways have been reported (Cao and Jacobsen, 2002a), but the extent was largely unknown. To gain better understanding of DNA methylation targeting in *Arabidopsis thaliana*, we generated high-resolution genome-wide DNA methylation maps in a comprehensive list of mutants (Chapter 1). By examining DNA methylation patterns in different mutants, we were able to deduce the sites and extent the gene mutated normally controls DNA methylation. In contrast to the current view, we found that DNA methylation is regulated site-specifically and not just based on cytosine sequence contexts. For example, small TEs proximal to genes as well as regulatory regions of TEs were methylated by DRM2, and large TEs distal to genes were methylated by CMT3. We also revealed cross talks between different methylation pathways. For example, asymmetric CHH methylation was highly dependent on symmetric CG and CHG methylation. Finally, we found several new regulators of DNA methylation. For example, we identified a novel methyltransferase, SUVR2, which regulates RNA-directed DNA methylation through DRM2 by acting downstream of small RNA biogenesis. Our results provide a comprehensive view of DNA methylation patterning control (Chapter 1).

Histone H3 lysine 27 monomethylation

While DNA methylation clearly regulates gene and TE silencing, gene-silencing pathways

seemingly acting independent of DNA methylation also exist. For example, mutation in

MORPHEUS' MOLECULE1 (MOM1) results in TE transcriptional derepression without

changes in DNA methylation (Amedeo et al., 2000; Stroud et al., 2013b). In collaboration with

Scott Michaels's laboratory at Indiana University, we found that mutations in two histone H3

lysine 27 (H3K27) monomethyltransferase genes, ARABIDOPSIS TRITHORAX-RELATED

PROTEIN5 (ATXR5) and ATXR6, caused transcriptional derepression of TEs without global

disruption of DNA methylation (Jacob et al., 2009). This suggested that ATXR5 and ATXR6

silence TEs independently of DNA methylation. ATXR5 and ATXR6 were initially identified as

potential interactors of proliferating cell nuclear antigen (PCNA), and were reported to be

preferentially expressed during S phase of the cell cycle (Raynaud et al., 2006). Interestingly,

atxr5 atxr6 double mutants had few other notable phenotypes such as chromocenter

decondensation and morphological defects such that leaves in mutants tended to be smaller than

those in wild type (Jacob et al., 2009). While it is well established that H3K27 trimethylation

(H3K27me3) is a repressive mark mediated by Polycomb group proteins in both plants and

animals (Feng and Jacobsen, 2011), the roles of H3K27 monomethylation (H3K27me1) was

poorly understood. Previous immunostaining studies in *Arabidopsis thaliana* have suggested that

H3K27me1 is enriched in chromocenters (Jacob et al., 2009; Lindroth et al., 2004; Mathieu et

al., 2005). In contrast, in humans, H3K27me1 was reported to be enriched at active promoters of

genes (Barski et al., 2007), indicating a distinct difference in the distribution, and thus possibly

8

function of this mark in plants and animals. To further understand the function of H3K27me1 in *Arabidopsis thaliana*, we mapped and characterized H3K27me1 in the genome (Chapter 2). We found that H3K27me1 is predominantly enriched in pericentromeric heterochromatin as well as isolated heterochromatic patches in the euchromatic arms, correlated with DNA methylation and H3K9 methylation, and its distribution was anticorrelated with transcription levels of protein-coding genes. The distribution of H3K27me1 is significantly different to that of polycomb-mediated H3K27me3 patterns, where H3K27me3 is enriched in euchromatin, depleted in DNA methylation and H3K9 methylation, and largely restricted to protein-coding genes that are over-represented with developmentally important transcription factors (Turck et al., 2007; Zhang et al., 2007b). Our results suggest that distinct H3K27 methylation-mediated silencing systems exist in *Arabidopsis thaliana*.

Regulation of DNA replication

During cell division it is critical that each daughter cell inherits one copy of DNA. DNA replication is initiated at origins of DNA replication that are distributed throughout the genome. While DNA replication occurs in the S phase of the cell cycle, replication origins are determined in the G1 phase. This initiation process is referred to as DNA replication "licensing" (Blow and Laskey, 1988). Replication origins are determined by the formation of the pre-replication complex (pre-RC) that involves Origin recognition complex (ORC) binding to origins which recruit CDT1 and CDC6, which then loads the minichromosome maintenance (MCM) complex (Bell and Dutta, 2002; Blow and Dutta, 2005; Thommes and Blow, 1997). Once cells enter S phase and DNA replication has initiated, pre-RC components are disassembled and re-initiation

9

of origin firing is prevented. Multiple mechanisms exist to regulate origins of replication to make

sure that origins of replication are fired only once per cell cycle (Arias and Walter, 2007; Blow

and Dutta, 2005). Misregulation of licensing factors, such as overexpression of CDT1, can result

in re-initiation origin firing (known as "re-replication") which would threaten genome stability

(Blow and Gillespie, 2008). Re-replication has experimentally been shown to be a potential

inducer of gene duplications in *Saccharomyces cerevisiae* (Green et al., 2010). Studies in

Xenopus laevis egg extracts have shown that re-replication produces small fragments of double-

stranded DNA (Davidson et al., 2006b). In addition, re-replication has been shown to be

associated with double-stranded breaks in several organisms and triggers DNA damage response

pathways (Archambault et al., 2005; Green and Li, 2005; Jin et al., 2006; Lovejoy et al., 2006;

Melixetian et al., 2004; Vaziri et al., 2003; Zhu et al., 2004; Zhu and Dutta, 2006).

Regulation of DNA replication in plants is especially poorly understood. While many replication

factors are conserved between plants and animals, there are clear differences (Shultz et al., 2007).

For example, plants do not have Geminin, a factor that regulates DNA replication through

inhibitory activity against CDT1 (Tada et al., 2001; Wohlschlegel et al., 2000), and interestingly,

plant genomes contain multiple copies of replication factor genes (Shultz et al., 2007). Hence it

is possible that plants may have different mechanisms of regulating origins of replication

compared to animals.

In *Saccharomyces cerevisiae,* replication origins are defined by an 11 base-pair DNA sequence

referred to as the ARS (autonomous replication sequence) core consensus sequence (Bell and

Stillman, 1992b; Mechali, 2010; Wyrick, 2001). However, in other eukaryotes including

Schizosaccharomyces pombe, origins appear not to have consensus DNA sequences (Antequera, 2004a; Remus et al., 2004; Vashee et al., 2003). Because of the lack of consensus DNA sequences at origins, it is tempting to speculate that chromatin modifications may play a role in defining origins. Indeed, building evidence suggest that several chromatin modifications are implicated in regulating origins. Previous studies have suggested that histone acetylation, a mark coupled with open chromatin and active transcription, is associated with active origins in mammals and yeast (Aggarwal and Calvi, 2004; Danis et al., 2004; Goren et al., 2008; Kaplan et al., 2008). On the other hand, ORC has been reported to preferentially localize to heterochromatin in human and *Drosophila melanogaster* (Pak et al., 1997; Prasanth et al., 2010), and furthermore, a proteomics study showed that human ORC recognizes repressive histone modifications, H3K9me3 and H3K27me3, and this binding preference is positively influenced by the presence of DNA methylation (Bartke et al., 2010). There is also some evidence that replication origins maybe regulated by DNA methylation in mammals (Rein et al., 1999) and *Xenopus laevis* egg extracts (Harvey and Newport, 2003). In higher eukaryotes H4K20 methylation has been implicated in regulating origin firing (Abbas et al., 2010; Bartke et al., 2010; Beck et al., 2012; Jorgensen et al., 2007; Kuo et al., 2012; Tardat et al., 2010; Tardat et al., 2007; Vermeulen et al., 2010) where the bromo adjacent homology (BAH) domain of ORC1 recognizes H4K20me2 (Kuo et al., 2012). However, mass spectrometry studies in plants have indicated that plants lack H4K20 methylation, and instead, H4K20 can be acetylated (Zhang et al., 2007a). This suggests that plants lack this system for regulating origins or possibly utilize different histone marks. In fact, in contrast to findings in mammals, one of the *Arabidopsis thaliana* ORC1 proteins, ORC1b, has been suggested to bind the active histone modification H3K4me3 (de la Paz Sanchez and Gutierrez, 2009).

The expression levels of H3K27 methyltransferases, ATXR5 and ATXR6, are cell cycle dependent (Raynaud et al., 2006), and ATXR6 is co-expressed with DNA replication factors such as CDT1 and ORC2. Furthermore, ATXR5 and ATXR6 potentially interact with PCNA (Raynaud et al., 2006) which functions as sliding clamps for DNA polymerases (Moldovan et al., 2007). These facts, as well as preliminary FACS analyses suggesting altered DNA contents in *atxr5 atxr6* double mutant nuclei from Scott Michaels' laboratory, led us to investigate the effect *atxr5 atxr6* double mutants have on DNA replication (Chapter 2). We found increased copies of heterochromatic DNA in *atxr5 atxr6* double mutants, suggesting that ATXR5 and ATXR6 are required to prevent over-replication of heterochromatin. Interestingly, the over-replication defect in *atxr5 atxr6* double mutants was restricted to cells that were endoreduplicating (i.e. cells undergoing DNA replication in the absence of cell division). Distributions of genomic DNA at over-replicating sites in *atxr5 atxr6* double mutants were reminiscent of re-replication. Consistent with the fact that ATXR5 and ATXR6 are H3K27 monomethyltransferases, sites of over-replication in *atxr5 atxr6* double mutants correlated with sites normally enriched in H3K27me1. Our results uncovered a new pathway that regulates DNA replication in heterochromatin in plants (Chapter 2).

Bacterial chromosomes typically have a single sequence-specific origin of replication. Eukaryotes have much larger genomes and it has been suggested the speed of the replication forks are slower than that of bacteria (Mechali, 2010). If the human genome had only one origin of replication per chromosome, it would take more than 20 days to replicate the genome (Mechali, 2010). Therefore, it is necessary that eukaryotic genomes contain multiple origins of

replication. In recent years, with the development of high-throughput sequencing technologies, origins of DNA replication are starting to be mapped and characterized in various organisms including yeast (Eaton et al., 2010; Hayashi et al., 2007; Xu et al., 2012), human (Cadoret et al., 2008; Karnani et al., 2010; Martin et al., 2011; Mesner et al., 2011), mouse (Cayrou et al., 2011; Sequeira-Mendes et al., 2009) and *Drosophila melanogaster* (Cayrou et al., 2011; MacAlpine et al., 2010). Origins of replication in plants had not been defined and were poorly understood. Crisanto Gutierrez's laboratory developed a method to purify newly synthesized DNA using short-pulse-5-bromo-2'-deoxyuridine (BrdU) labeling. We collaborated with Crisanto Gutierrez's laboratory to map replication origins in *Arabidopsis thaliana* by massively parallel sequencing the newly synthesized DNA (Chapter 3). By analyzing whole-genome maps of *Arabidopsis thaliana* origins, we found that plant origins have several notable genomic and epigenomic characteristics, some of which shared similarities with other organisms, such as association with histone acetylation. However, there were certain differences, most notably being the enrichment of nucleosomes at plants origins (Stroud et al., 2012b), which was the opposite of what had been reported in yeast and human (Berbenetz et al., 2010; Eaton et al., 2010; Xu et al., 2012; Yin et al., 2009). Hence, in plants replication-licensing factors may be targeted to nucleosomes. Notably, sites of over-replication in *atxr5 atxr6* double mutants tended to overlap with normal origins of replication. This suggested that ATXR5 and ATXR6 at least in part regulate normal replication origins, opposed to the possibility that ATXR5 ATXR6 system may specially regulate ectopic origins. The enrichment of nucleosomes at origins as well as the possibility that ATXR5 ATXR6 regulate origins are concordant with the roles of ATXR5 and ATXR6 in regulating DNA replication through nucleosome post-translational modification.

Relationship between chromatin marks in regulating gene silencing and DNA replication.

Our results have uncovered that ATXR5 and ATXR6 are involved in regulating transcriptional silencing of TEs as well as regulating DNA replication in heterochromatin. H3K27me1 was enriched in heterochromatic regions of the genome, which are also sites in the genome that have high levels of DNA methylation in plants (Bernatavichute et al., 2008; Cokus et al., 2008; Jacob et al., 2010). In fact, over-replication in *atxr5 atxr6* double mutants was closely confined to heterochromatin, appearing as if the over-replication was unable to spread into flanking euchromatin (Stroud et al., 2012a). The relationship between ATXR5 ATXR6 and DNA methylation in regulating transcriptional silencing and DNA replication had not been explored. DNA methylation was not disturbed in *atxr5 atxr6* double mutants (Jacob et al., 2009; Stroud et al., 2013b), suggesting that H3K27me1 does not regulate DNA methylation. To understand the relationship between these chromatin marks in silencing TEs and regulating DNA replication, we generated plants containing various DNA methylation mutations in an *atxr5 atxr6* double mutant background, and examined genome-wide silencing and DNA replication defects (Chapter 4). We found that ATXR5 ATXR6 and DNA methylation transcriptionally silence a largely overlapping subset of TEs. In addition, we found that ATXR5 ATXR6 and DNA methylation cooperatively transcriptionally silence TEs through independent pathways such that many TEs become deregulated only when losing both ATXR5 ATXR6 and DNA methylation systems. This suggested that multiple pathways redundantly silence many TEs across the genome. In contrast, we observed an antagonistic relationship between ATXR5 ATXR6 and DNA methylation in regulating DNA replication, such that the over-replication defect in *atxr5 atxr6* double mutants is suppressed upon loss of DNA methylation (Chapter 4). Hence our results suggest that there is a

complex relationship between chromatin marks in regulating transcriptional silencing and DNA replication.

CHAPTER 1

Comprehensive analysis of silencing mutants reveals complex regulation of the

Arabidopsis methylome

ABSTRACT

Cytosine methylation is involved in various biological processes such as silencing of transposable elements (TEs) and imprinting. Multiple pathways regulate DNA methylation in different sequence contexts, but the factors that regulate DNA methylation at a given site in the genome largely remain unknown. Here we have surveyed the methylomes of a comprehensive list of 86 Arabidopsis gene silencing mutants by generating single nucleotide resolution maps of DNA methylation. We find that DNA methylation is site-specifically regulated by different factors. Furthermore, we have identified novel regulators of DNA methylation. These data and analyses will serve as a comprehensive community resource for further understanding the control of DNA methylation patterning.

INTRODUCTION

The Arabidopsis genome is methylated in CG, CHG, and CHH (where H=A, T or C) sequence contexts (Law and Jacobsen, 2010). It is understood that distinct pathways regulate methylation in each of the three sequence contexts. CG methylation is maintained by METHYLTRANSFERASE 1 (MET1), the plant homolog of mammalian DNA (cytosine-5)-methyltransferase 1 (DNMT1), and CHG methylation is maintained by CHROMOMETHYLASE 3 (CMT3). KRYPTONITE (KYP/SUVH4), SUVH5, and SUVH6 are the primary H3K9 methyltransferases, and are required for CMT3 activity (Ebbs and Bender, 2006; Jackson et al., 2002; Lindroth et al., 2001). DOMAINS REARRANGED METHYLTRANSFERASES 1 (DRM1) and (DRM2), plant homologs of mammalian DNMT3, are responsible for CHH methylation through the RNA-directed DNA methylation (RdDM) pathway, which involves two plant specific RNA polymerases, RNA Pol IV and Pol V, as well as 24 nucleotide (24-nt) small interfering RNAs (siRNAs). While these three main DNA methylation pathways exist, it is notable that the chromatin remodeler DECREASE IN DNA METHYLATION 1 (DDM1) is required for the maintenance of CG and non-CG methylation (Jeddeloh et al., 1999; Vongs et al., 1993). Some interplay between DNA methylation pathways have been reported, however the extent is largely unknown due to the lack of genome-wide analyses. Studies thus far on DNA methylation have usually been restricted to a few selected loci or insensitive methods such as immunostaining.

Whole genome bisulfite sequencing (BS-seq) enables determination of methylation levels at single nucleotide resolution (Cokus et al., 2008; Lister et al., 2008). Here we have generated high coverage genome-wide maps of the Arabidopsis methylome in 86 mutants in the same genetic background and tissue type. Along with the current view that distinct pathways control

CG, CHG, and CHH methylation, we also found that DNA methylation is regulated in a site-specific manner involving interplays between different pathways. Our results provide a comprehensive view of the regulation of DNA methylation patterning in the Arabidopsis genome. In addition, our results revealed several unexpected features. Close examination of RNAi mutants suggested that specific sites in the genome might be regulated by RNAi factors not involved in the DRM1/2 pathway. Mutation in the chromatin assembly factor 1 (CAF-1) complex, which was thought not to regulate DNA methylation, induced CHG hypermethylation. We also found that RNA Pol II is required for DNA methylation largely independent of Pol IV and Pol V, suggesting a novel pathway for heterochromatin formation in plants. Finally, we found that one of the Su(var)3-9 related genes, SUVR2, is involved in RdDM, revealing a new component in the pathway. Our results open new areas of future research, and our dataset allows one to determine the factor(s) involved in controlling DNA methylation at a given cytosine in the genome, and thus will serve as a platform for further studies.

RESULTS AND DISCUSSION

Single nucleotide resolution maps of DNA methylation

We performed whole genome BS-seq on 86 mutants utilizing tissue of the same developmental stage (3-week-old leaves) and in a single ecotype (Columbia) so that we could carefully detect methylation differences due to genotype. By deeply sequencing each mutant, we obtained an average coverage of 43-fold (Table 1-1). The methylation data are displayed in a modified UCSC genome browser (http://genomes.mcdb.ucla.edu/AthBSseq/). Differentially methylated regions (DMRs) were determined by comparing methylation levels in each mutant to three

independent wild-type replicates in 100 base-pair tiles throughout the genome (see Experimental Procedures).

Regulation of CG methylation

CG methylation is the most abundant type of DNA methylation. CG methylation is present over heterochromatic regions enriched with TEs and repeats, as well as genic regions (Cokus et al., 2008; Lister et al., 2008). This is in contrast to CHG and CHH methylation, which are almost exclusively present in heterochromatin (Cokus et al., 2008; Lister et al., 2008). Mutation of the CG methyltransferase MET1 results in elimination of CG methylation throughout the genome (Figures 1-1A, B, Figures 1-2A, B) (Cokus et al., 2008; Lister et al., 2008). VARIANT IN METHYLATION 1 (VIM1), VIM2, and VIM3 are orthologous to mammalian UBIQUITIN-LIKE, CONTAINING PHD AND RING FINGER DOMAINS 1 (UHRF1) and have been shown to regulate CG methylation (Feng et al., 2010; Woo et al., 2008). In *vim1 vim2 vim3* (*vim1/2/3*), CG methylation was strongly reduced resembling *met1* (Figures 1-1A, B, Figures 1-2A, B). Notably, *vim1*, *vim2*, and *vim3* individually did not affect CG methylation, indicating complete functional redundancy in regulating CG methylation (Figure 1-2C). Either *met1* +/+ or +/- progeny of *met1* +/- heterozygous plants have morphological defects, which led us to investigate their methylomes. We found that while TEs largely had wild-type methylation levels, genic methylation was severely impaired (Figures 1-1A-C). Hence our results suggest that genic methylation cannot be restored once lost, and is consistent with previous studies suggesting that siRNAs (which are exclusively associated with heterochromatin) are required for restoration of DNA methylation in mutants of chromatin remodeler DDM1 (Teixeira et al., 2009). In *ddm1*, some heterochromatic DNA methylation has been shown to be reduced (Lippman et al., 2004),

20

and DNA methylation is lost progressively upon inbreeding (Kakutani et al., 1996). We tested 7th generation homozygous *ddm1*, and found that heterochromatic DNA methylation is severely lost in *ddm1*, however genic methylation remained largely intact (Figures 1-1A-C, Figure 1-2A). Hence DDM1 controls DNA methylation specifically at heterochromatin (Lippman et al., 2004).

Regulation of CHG methylation

CMT3 is the main CHG methyltransferase in Arabidopsis (Law and Jacobsen, 2010). Indeed we observed a strong depletion of CHG methylation in *cmt3* (Figure 1-3A). H3K9 methyltransferases KYP, SUVH5, and SUVH6 have been shown to be required for CMT3 dependent CHG methylation (Ebbs and Bender, 2006). Loss of CHG methylation in *kyp suvh5 suvh6* (*kyp suvh5/6*) closely mimicked the loss of CHG methylation in *cmt3* (Figure 1-3B, Figure 1-4A). Hence KYP SUVH5/6 regulates CHG methylation through CMT3 genome-wide. We further tested redundancies between KYP, SUVH5, and SUVH6. We found that KYP was solely responsible for certain CHG methylation sites in the genome, whereas mutations in SUVH5 or SUVH6 alone did not show any detectable alterations in CHG methylation (Figure 1-3B).

While a strong depletion of CHG methylation was observed in *cmt3*, there were patches of DNA methylation in the genome that were not affected (Figure 1-3B). We found that CHG methylation at these sites often depended on DRM1/2 (Figure 1-3B). While DRM1/2 are suggested to be CHH methyltransferases, our results confirm that they also regulate CHG methylation (Cao and Jacobsen, 2002a). DRM1/2 regulate CHG methylation at specific sites in the genome, 60.4% of which sites are non-overlapping with sites regulated by CMT3 (Figure 1-4A). To test the degree of redundancy between CMT3 and DRM1/2, we profiled *drm1/2 cmt3* triple mutants. We did not observe many additional losses of CHG methylation in *drm1/2 cmt3*

(Figure 1-4B), suggesting that CMT3 and DRM1/2 regulate CHG methylation in a mostly non-redundant fashion.

Regulation of CHH methylation

CHG methylation and CHH methylation highly co-localize in the wild-type genome (Cokus et al., 2008; Lister et al., 2008). Hence we tested whether loss of CHG methylation is associated with loss of CHH methylation. In *cmt3*, loss of CHG methylation is only partially associated with loss of CHH methylation (Figure 1-3B, Figure 1-4M). Hence while CMT3 is required for the majority of CHG methylation in the genome, it is required for a relatively small proportion of CHH methylation. In contrast, in *drm1/2*, loss of CHG methylation was always associated with loss of CHH methylation (Figure 1-4N). Loss of CHH methylation, however, was not always coupled with loss of CHG methylation (Figure 1-4N). Thus CHH methylation maintenance appears more reliant on CHG methylation than CHG methylation is on CHH methylation.

Interestingly, whereas *kyp* and *cmt3* showed very similar losses in CHH methylation, *kyp suvh5/6* showed much stronger losses of CHH methylation compared to *cmt3* (Figures 1-3A-C, Figure 1-4C). KYP SUVH5/6 dependent clusters of CHH methylation were generally non-overlapping with DRM1/2 dependent CHH methylation (Figures 1-3B-D, Figure 1-4C). Hence while KYP SUVH5/6 controls CHG methylation through CMT3, our results suggest that KYP SUVH5/6 strongly regulate CHH methylation through a different pathway. Notably, mutations in factors responsible for siRNA biogenesis such RNA-Dependent RNA Polymerase 2 (RDR2) and DICER-LIKE 2, 3 and 4 (DCL2/3/4) (discussed more in detail below) did not disrupt CHH methylation at most KYP SUVH5/6 regulated sites (Figure 1-4O). Thus KYP SUVH5/6 regulate CHH methylation in a siRNA independent manner.

Different methylation pathways appeared to target different classes of TEs (Figure 1-4R). One insight was that both *cmt3* and *kyp* CHH DMRs were over-represented by LTR/Copia type TEs. The overlap between *cmt3* and *kyp* was high, with 71.0% of *cmt3* CHH TE DMRs overlapping with *kyp* CHH TE DMRs (Figure 1-4S).

Interdependence of CG and non-CG methylation

Previous studies based on immunostaining and single loci ChIP analyses have suggested that mutation in MET1 causes loss of H3K9m2 at certain sites (Soppe et al., 2002; Tariq et al., 2003). Consistent with these findings, we found loss of CHG methylation at certain sites in *met1* (Figure 1-3B). Comparing *met1* and *vim1/2/3*, we found 85.1% overlap between sites that lose CHG methylation, suggesting that CG methylation is required for proper CHG methylation at those sites (Figure 1-2B). Loss of CHG methylation was observed at a subset of sites in *met1* +/+ and *met1* +/- progenies of *met1* +/- (Figure 1-4P). These sites corresponded to the subset of heterochromatic sites that did not restore CG methylation (Figure 1-4P), further supporting the notion that CG methylation is required for maintaining CHG methylation. Loss of CHG methylation in *met1* largely occurred at certain KYP SUVH5/6 and CMT3 dependent CHG sites (Figure 1-3B, Figures 1-4E, G). However loss of CHH methylation in *met1* largely did not overlap with KYP SUVH5/6 and CMT3 dependent CHH sites (Figures 1-4F, H). Hence while MET1 regulates CHG methylation through KYP SUVH5/6 and CMT3, it regulates CHH methylation mostly through a different pathway.

While *met1* CHH DMRs were much more abundant compared to *drm1/2* CHH DMRs, 63.0% of *drm1/2* CHH DMRs overlapped with *met1* CHH DMRs (Figure 1-3B, Figures 1-4I, J). This overlap was significantly higher than observed for CMT3 and KYP SUVH5/6 dependent

sites (11.4% and 26.7%, respectively). This together with the fact that *drm1/2* has minimal disruption of CG methylation (Figure 1-4N) suggest a strong tendency for DRM1/2 targeted methylation to depend on CG methylation. Wild-type CG methylation levels at CG methylation dependent and independent DRM1/2 target sites were similar (Figure 1-4Q). Therefore, the features that determine whether a DRM1/2 site is dependent on CG methylation or not is unclear.

An additional insight was that *met1 cmt3* caused strong reduction in both CHG and CHH methylation. In fact, *met1 cmt3* most severely affected CHH methylation of all the mutants we tested (Figure 1-3A). *met1 cmt3* reduced CHH methylation at many additional sites compared to *met1* or *cmt3* alone (Figures 1-4K, L), suggesting that MET1 and CMT3 cooperatively regulate the bulk of CHH methylation in the genome. Our results indicate a strong genome-wide dependence of asymmetric CHH methylation on symmetrical CG and CHG methylation.

Mutation in DDM1 also disrupted CHG and CHH methylation, where loss of DNA methylation generally occurred at sites regulated by KYP SUVH5/6 rather than sites regulated by DRM1/2 (Figure 1-3B). Only 27.3% of *drm1/2* CHG DMRs and 23.1% of *drm1/2* CHH DMRs overlapped with corresponding *ddm1* DMRs. Hence unlike MET1, DDM1 is largely not required for DRM1/2 dependent methylation.

We also found that CG methylation is dependent on non-CG methylation at certain sites. Loss of CG methylation was associated with loss of non-CG methylation at a subset of sites in *kyp suvh5/6*, *cmt3*, and *drm1/2* (Figures 1-4M, N). For example, while loss of methylation in *drm1/2* mostly occurred in CHH and CHG contexts, 18.5% of *drm1/2* CHH DMRs were associated with loss in CG methylation. Among sites where CHH methylation was lost in both *drm1/2* and *kyp suvh5/6*, sites that lost CG methylation in *drm1/2* and *kyp suvh5/6* were largely overlapping (77.8% of those in *drm1/2* overlapped with *kyp suvh5/6*). This confirms that it is

likely that the loss of non-CG methylation causes the loss of CG methylation. 71.8% of DRM1/2 CHH DMRs that also lost CG methylation were sites where CG methylation was required for CHH methylation. These sites are interesting as CG methylation and non-CG methylation become interdependent.

Comparison of regions methylated by KYP SUVH5/6, CMT3, and DRM1/2

We further examined the characteristics of sites affected in *kyp suvh5/6*, *cmt3*, and *drm1/2*. DRM1/2 target sites were associated with relatively lower G+C sequence composition (Figure 1-5A, Figure 1-6A), and showed a tendency of GC skewing (Figure 1-5B, Figure 1-6B). Association with inverted repeats was also a unique feature of *drm1/2* DMRs (Figure 1-5C, Figure 1-6C). While siRNAs are produced throughout most heterochromatic sites in the genome, it is unclear how DRM1/2 are specifically targeted to a subset of these sites. Our results suggest that sequence composition may be one of the factors that determines which methylation pathway is targeted at a given site in the genome.

We found that DRM1/2 targets small TEs, whereas KYP SUVH5/6 and CMT3 target large TEs (Figure 1-5D, Figure 1-6D) (Tran et al., 2005). TEs regulated by DRM1/2 were proximal to promoters of genes, where around ~70% of *drm1/2* DMRs fell within 2kb of transcription start sites of protein-coding genes (Figures 1-5E, F, Figures 1-6E, F) (Zhong et al., 2012). These proximal genes were not significantly associated with particular biological processes (not shown). Another distinction was that DRM1/2 specifically methylated the boundaries of TEs (Figures 1-5G, H, Figures 1-6G, H). We next examined the levels of enrichment of transcription factor binding sites (TFBS) at DMRs. We found enrichment of TFBS over *drm1/2* DMRs but not over *kyp suvh5/6* and *cmt3* DMRs (Figure 1-5I, Figures 1-6I, L).

Hence DRM1/2 target sites tend to be regulatory sites. We then sought to measure the expression

levels of TEs in wild-type by performing RNA sequencing. We found that DRM1/2 targeted

TEs are more silent compared to CMT3 targeted TEs (Figure 1-5J, Figure 1-6J), presumably

because they are closer to genes and thus potentially harmful if expressed.

Finally, we examined the distribution of nucleosomes and known histone modifications

over DMRs. Consistent with the notion that CMT3 is dependent on H3K9 methylation, *kyp*

suvh5/6 and *cmt3* DMRs were associated with higher levels of H3K9me2 compared to *drm1/2*

DMRs (Figure 1-5K, Figure 1-6K). *kyp suvh5/6* and *cmt3* DMRs were also associated with

higher levels nucleosome occupancy compared to *drm1/2* DMRs (Figure 1-5K, Figure 1-6K).

Hence different methylation pathways regulate sites with distinct genomic and epigenomic

characteristics.

RNA-directed DNA methylation

The RdDM pathway involves many accessory factors which guide DNA methylation by DRM2

(Law and Jacobsen, 2010). We sought to examine whether disruption of components of the

pathway result in similar DNA methylation defects. We tested 29 mutants previously suggested

to affect RdDM (Gu et al., 2011; He et al., 2009; Henderson et al., 2006; Law and Jacobsen,

2010; Zheng et al., 2009). By examining methylation levels at DRM1/2 dependent CHH sites,

we found that there are differential effects when disrupting components of the RdDM pathway

(Figure 1-7A). Broadly, there are four classes of RdDM components: those where mutation in

the gene eliminates DRM1/2 dependent methylation, those where mutation reduces methylation,

those where mutation weakly reduces methylation, and those that only affect a very small

proportion of sites (some of which we describe below) (Figure 1-7A). Importantly, AGO4 and

AGO6 were suggested to be partially redundant (Zheng et al., 2007). However our results suggest that mutation in AGO4 alone is sufficient to eliminate DRM1/2 dependent methylation.

The flowering-time regulators FCA and FPA were suggested to be responsible for DNA methylation at certain RdDM sites (Baurle et al., 2007). We did not observe global reduction of DNA methylation at RdDM sites in *fca fpa*, but did find minor alterations in methylation (Figure 1-8A), where 69 out of 86 (80.3%) defined *fca fpa* CHH DMRs overlapped with *drm1/2* DMRs. Consistent with the overlap, *fca fpa* DMRs were associated with promoters of genes (Figure 1-8B). Analyses of *fca* and *fpa* single mutants revealed partial redundancies (Figure 1-8C).

While DICER-LIKE 3 (DCL3) cleaves double stranded RNA (dsRNA) into 24-nt siRNA and thus functions in RdDM, DCL2 and DCL4 cleaves dsRNA into 22-nt and 21-nt siRNA, respectively, and function in other biological processes (Voinnet, 2008). However, functional redundancies between DCL3 and DCL2/4 have been suggested at some loci (Henderson et al., 2006). Indeed, while *dcl3* was categorized as a "weakly reduced" mutant, *dcl2/3/4* was categorized as a "reduced" mutant (Figure 1-7A). Hence in the absence of DCL3, DCL2 and DCL4 can mediate DNA methylation at most RdDM sites.

RNAi factors are involved in DNA methylation

We further examined whether mutants of known RNAi components not implicated in the canonical RdDM pathway, including AGOs, DCLs, HEN1, RDRs, SDEs and SGS3, affected DNA methylation. Of all mutants tested, RDR1 and RDR6, which are involved in pathways that yield 21- and 22-nt siRNAs, showed the strongest loss of DNA methylation (Figure 1-7B). 38.4% of *rdr6* DMRs overlapped with *rdr1* DMRs (Figure 1-8D). Only 60 out of 215 sites (27.9%) were also DMRs in *drm1/2*. Furthermore, unlike mutations in RdDM components,

where DNA methylation was largely affected in non-CG contexts (Figure 1-4N), DNA methylation was lost in all three cytosine contexts in *rdr1* and *rdr6* (Figures 1-8E, F). One characteristic of these sites was that they were more likely to be associated with genes compared to DRM1/2 sites (Figure 1-7C). Another characteristic of these sites was that 21- and 22-nt siRNA levels in wild-type, measured by small RNA sequencing (Lee et al., 2012), were somewhat enriched compared to 24-nt siRNAs (Figure 1-8G). Hence RDR1 and RDR6 may be involved in DNA methylation independent of the DRM1/2 pathway.

Ectopic gain of CHG methylation

Gain of DNA methylation has been reported in several mutants, including DNA and histone demethylases as well as DNA methylation mutants. In the DNA demethylase mutant, *ros1 dml2 dml3* (*rdd*), we found that hypermethylation occurred in all three cytosine contexts, however *ros3*, which has been suggested to act in the same genetic pathway, showed only very limited hypermethylation (Figure 1-10A). In contrast, other tested mutants exhibited hypermethylation mostly in CHG contexts and to a lesser extent in the CHH contexts (Figures 1-9A-C, Figure 1-10B). Hence ectopic hypermethylation in these mutants occurs through a different mechanism (Figure 1-10C). Importantly, CHG hypermethylated sites in *met1* were not necessarily hypermethylated in *vim1/2/3* (Figure 1-10D) and *vice versa*, suggesting that the hypermethylation phenotype may not be explained directly by loss of CG methylation. This is in contrast to CHG hypomethylation, where *met1* and *vim1/2/3* caused loss of CHG methylation at very similar sites (Figure 1-2B). One interpretation for this result may be that CHG hypermethylation occurs stochastically (Mathieu et al., 2007), whereas CHG hypomethylation is a direct consequence of loss of CG methylation. It is worth noting that *met1* +/+ and *met1* +/-

progeny from *met1 +/-* showed limited CHG hypermethylation and this did not occur at sites that lost CG methylation (Figure 1-10E). Collectively, our results suggest that CHG hypermethylation in *met1* is not likely a phenomenon that compensates for the loss of CG methylation as previously proposed (Cokus et al., 2008; Lister et al., 2008). In *met1 cmt3*, the CHG hypermethylation is eliminated, indicating that the CHG hypermethylation phenotype in *met1* is dependent on CMT3 (Figure 1-9A).

CHG hypermethylation in *ddm1* occurred at distinct sites compared to *met1* (Figure 1-10F). CHG hypermethylation in *met1* and *vim1/2/3* predominantly occurred in normally unmethylated regions (Figure 1-9A, Figure 1-10B), however CHG hypermethylation in *ddm1* occurred predominantly at regions only CG methylated (Figure 1-9B). This suggests that CHG hypermethylation in *ddm1* is not likely due to loss of CG methylation and likely occurs through a different mechanism. Because *ddm1* loses a significant amount of both CG and non-CG methylation, a speculation is that loss of non-CG methylation induces CHG hypermethylation at distinct loci in the genome. Histone H3K9 demethylase IBM1 has been suggested to protect CG methylated genes from becoming CHG methylated (Miura et al., 2009). We have confirmed these findings with BS-seq (Figure 1-9C). Notably, 49.1% of *ddm1* CHG hypermethylation DMRs overlapped with those of *ibm1* (Figure 1-10G). Because IBM1 transcripts are largely unaffected in *ddm1* (Figure 1-10H), CHG hypermethylation in *ddm1* is not likely due to impaired IBM1. These results suggest a relationship between the hypermethylation occurring in *ddm1* and *ibm1*.

We next examined whether RdDM mutants show any CHG hypermethylation. We defined 79 CHG hypermethylated sites in *drm1/2*, and found that they tend to occur at sites immediately flanking TEs (Figure 1-10I). Because of the relatively small number of DMRs, we

cannot rule out stochastic variations; however we observed CHG hypermethylation across RdDM mutants at these same sites (with some variation) (Figure 1-10J). In *drm1/2 cmt3*, the CHG hypermethylation was suppressed (Figure 1-10J), suggesting that CMT3 is again required for CHG hypermethylation. Consistent with the fact that DNA methylation by DRM1/2 is largely regulated by MET1 (Figure 1-4I, J), 51.9% of *drm1/2* CHG hypermethylation DMRs overlapped with those of *met1*, compared to 20.2% of those in *ibm1*. In summary our results suggest that loss of DNA methylation induces CHG hypermethylation through CMT3.

Ectopic gain of CHH methylation

We found that CHG hypermethylated sites were generally associated with CHH hypermethylation in the mutants tested (Figures 1-9A-C, Figure 1-10K). In *met1 cmt3*, the CHH hypermethylation associated with CHG hypermethylation was suppressed, suggesting that CMT3 is responsible for CHH methylation at these sites (Figure 1-10K). However, *met1 cmt3* exhibited comparable genome-wide CHH hypermethylation levels as *met1* (Figure 1-9D). Interestingly, while 65.1% of CHH hypermethylation DMRs were associated with CHG hypermethylation in *ibm1*, this was the case for only 20.6% and 19.4% in *met1* and *ddm1*, respectively (Figure 1-9E). Hence in *met1* and *ddm1*, the bulk of CHH hypermethylation occurs at distinct sites compared to CHG hypermethylation. CHH hypermethylation decoupled from CHG hypermethylation might be explained by transcriptional reactivation of TEs in these mutants, where TE transcripts become processed into siRNA which then directs CHH methylation. Consistent with this idea, CHH hypermethylation DMRs not overlapping with CHG hypermethylation DMRs corresponded predominantly to TEs in *met1* and *ddm1* (Figure 1-9F). TEs that get CHH hypermethylated were over-represented with LTR/Gypsy type TEs (Figure 1-10L). In summary,

loss of global DNA methylation induces CHH hypermethylation that is largely distinct from the CHG hypermethylation phenomenon.

Mutation in the CAF-1 complex induces CHG hypermethylation

The CAF-1 complex is required for proper heterochromatin formation. FASCIATA 1 (FAS1) and FAS2 are subunits of the CAF-1 complex, and their disruption results in reduced heterochromatin without disturbing DNA methylation at certain repeats (Schonrock et al., 2006). We studied the CAF-1 complex by testing *fas2*. We found that *fas2* tended to have local alterations of DNA methylation (1,572 defined sites), especially exhibiting hypermethylation in CHG contexts (Figures 1-9G, H). Whole chromosomal views and average plots over TEs suggested modest genome-wide elevation of DNA methylation (Figures 1-9I, J). There was relatively little overlap between *fas2* CHG hypermethylated DMRs and those of *rdd*, *met1*, *ddm1*, and *ibm1* (4.5%, 18.4%, 3.9%, and 0.9%, respectively) (Figure 1-9K). CHG hypermethylation DMRs tended to overlap with TEs (60.2%) but not genes (14.0%). TEs that get CHG hypermethylated were somewhat over-represented with LTR/Gypsy type TEs (Figure 1-10M). Hence FAS2 is likely involved in an independent pathway to prevent hypermethylation of TEs.

RNA Pol II is involved in DNA methylation independent of Pol IV and Pol V

Pol IV and Pol V have likely evolved from Pol II, and specifically function in RdDM (Haag and Pikaard, 2011). Pol II was suggested to be involved in regulating DNA methylation at certain intergenic sites by recruiting Pol IV and Pol V (Zheng et al., 2009). Using a weak Pol II mutant allele, *nrpb2-3*, we confirmed that *nrpb2* has reduced DNA methylation at certain sites (Figure 1-

11A). We found a tendency of *nrpb2* DMRs to overlap with genic regions compared to *drm1/2* DMRs (Figure 1-11B). Intriguingly, we found that 64.4% and 66.6% of *nrpb2* DMRs did not overlap with *nrpd1* and *nrpe1* DMRs, respectively (Figure 1-11C), suggesting that for the most part Pol II regulates DNA methylation independently of Pol IV and Pol V. Furthermore, unlike *nrpd1* and *nrpe1*, loss of DNA methylation in *nrpb2* occurred in all three cytosine contexts (Figure 1-11D). Because we are limited to analyzing a weak Pol II allele since null mutations in Pol II are lethal (Onodera et al., 2008), it is possible that Pol II regulates a larger proportion of DNA methylation in the genome. Hence we provide evidence that Pol II itself is involved in a novel pathway that regulates DNA methylation.

Relationship between histone modifications and DNA methylation

There is growing evidence regarding interplays between histone modifications and DNA methylation (Cedar and Bergman, 2009). In Arabidopsis, H3K9 methylation is required for CHG methylation (Law and Jacobsen, 2010). Loss of DNA methylation is accompanied by ectopic gain in H3K27me3 and H3K4me3 in plants and animals (Hon et al., 2012; Weinhofer et al., 2010; Zhang et al., 2009). We utilized several mutants known to regulate particular histone modifications, and examined the impact on DNA methylation. We tested mutations that alter histone modifications normally localized at DNA hypomethylated sites (H3K4me3 and H3K27me3) as well as those that are normally localized at DNA methylated sites (H3K36me3 and H3K27me1). We tested *sdg8*, which has reduced H3K36me3 (Xu et al., 2008), *atxr5/6*, which has reduced H3K27me1 (Jacob et al., 2009), *sdg2*, which has reduced H3K4me3 (Berr et al., 2010; Guo et al., 2010), and *ref6*, which has gain of H3K27me3 (Lu et al., 2011). We did not observe notable changes in DNA methylation in these mutants (Figures 1-12A-D). Hence while

H3K36me3 and H3K27me1 co-localize with DNA methylation, loss of these marks does not affect DNA methylation. And while loss of DNA methylation causes gain of H3K4me3 and H3K27me3, alteration of these marks does not affect DNA methylation. These results suggest that H3K9 methylation is the main histone modification that regulates DNA methylation in Arabidopsis.

Histone deacetylase 6 regulates DNA methylation at promoters

Previous studies have suggested that the histone deacetylase 6 (HDA6) plays a role in RdDM (Aufsatz et al., 2002; He et al., 2009). Studies of rDNA suggested that loss of HDA6 causes reduction in CG/CHG methylation as well as CHH hypermethylation (Earley et al., 2010). In addition, other studies have reported that HDA6 physically interacts with MET1 (Liu et al., 2012) and FVE (Gu et al., 2011), and transcriptionally regulates a similar subset of genes as MET1 (To et al., 2011). We analyzed *hda6* and found that, unlike *met1*, the methylome was largely unaltered, except at particular sites of the genome (Figure 1-13A). These hypomethylated sites tended to be at promoters of genes, similar to the extent seen in *drm1/2* (Figure 1-13B). However, only 27.5% of *hda6* hypomethylation DMRs overlapped with those of *drm1/2*. Rather, 91.7% of *hda6* DMRs corresponded with those methylated by CMT3 (Figure 1-13C). However, unlike *drm1/2* and *cmt3,* where loss of methylation was largely restricted to non-CG contexts (Figures 1-4M, N), DNA methylation was lost in all three cytosine contexts (Figure 1-13D). This was similar to what was seen in *nrpb2* (Figure 1-11D). Furthermore, 27.5% of *hda6* DMRs overlapped with *nrpb2* DMRs despite the fact that *nrpb2* affects many fewer sites compared to *drm1/2*. Hence, while HDA6 has been suggested to interact with MET1 and FVE, our results suggest that HDA6 regulates DNA methylation at subsets of promoters through an independent

mechanism. Further, at least in part, HDA6 appears to be associated with Pol II directed DNA methylation.

SUVR2 is involved in the DRM1/2 pathway

Su(var)3-9 is a conserved factor required for gene silencing through H3K9 methylation (Schotta et al., 2003). Arabidopsis has 15 Su(var)3-9 homologs: 10 SUVH genes and 5 SUVR genes (Baumbusch et al., 2001; Pontvianne et al., 2010). We performed BS-seq on each mutant except for *suvr4* (due to lack of a knockout allele in a Col background). Interestingly, we found large losses of methylation in the *suvr2* mutant, especially at CHH sites. We found 1,113 sites CHH hypomethylated in *suvr2*, which, as in *drm1/2* (Figure 1-4N), were often associated with loss of CHG and to a small extent loss in CG (Figure 1-13E). 1,041 (93.5%) of *suvr2* CHH DMRs overlapped with those in *drm1/2* (Figure 1-13F). Comparison of methylation levels between *suvr2* and *drm1/2* suggested that *suvr2* is a weak RdDM mutant (Figure 1-13G), with *suvr2* falling into the "weakly reduced" (Figure 1-7A) class of RdDM mutants (not shown). We also tested methylation levels in a *suvr1/2/3/4/5* quintuple mutant (into which a Nossen ecotype allele of *suvr4* had been introgressed) and did not observe additional methylation loss compared with *suvr2* alone (Figure 1-13H), ruling out functional redundancies with other SUVR genes. Methylation analysis by Southern blot at the known RdDM target, *MEDEA-INTERGENIC SUBTELOMERIC REPEATS* (*MEA-ISR*), supported observations seen at the genome-wide level (Figure 1-12E).

In addition to their role in DNA methylation maintenance, RdDM pathway components also carry out DNA methylation establishment—or *de novo* methylation (Cao and Jacobsen, 2002b; Chan et al., 2004; Greenberg et al., 2011). Given its new potential role as an effector of

RdDM, we wanted to test whether SUVR2 also is required for *de novo* methylation. In order to do so, we utilized the *FLOWERING WAGINENGEN* (*FWA*) transgenic system. In the vegetative tissue of wild-type plants, *FWA* expression is repressed in a DNA methylation-dependent manner at tandem repeats in its 5' UTR (Soppe et al., 2002). When *FWA* transgenes are introduced into wild-type plants, the repeats are targeted for *de novo* DNA methylation and silenced. However, in RdDM mutants, the transgene fails to be methylated, causing ectopic expression that leads to a late-flowering phenotype (Ausin et al., 2009; Chan et al., 2004; Greenberg et al., 2011). When we transformed *suvr2* with the *FWA* transgene, the mutant plants flowered significantly later than wild-type controls (Figure 1-13I). Consistently, bisulfite analysis of the *FWA* transgene showed that DNA methylation was virtually absent in all three cytosine contexts (Figure 1-13J). Taken together, these results strongly indicate that SUVR2 is a canonical RdDM factor that is required for both DRM2 establishment and maintenance methylation.

In order to further place SUVR2 in the RdDM pathway, we performed small RNA northern blots (Figure 1-12F). RdDM proteins that act downstream of 24-nt siRNA biogenesis—such as NRPE1—only affect siRNA accumulation at a subset of targets, known as Type I loci, but not Type II loci (Zheng et al., 2009). We found that *suvr2* behaved similarly to *nrpe1*, indicating that SUVR2 is not required for generation of siRNAs. Consistent with the methylation analysis, higher order *suvr* mutants did not impact siRNA levels any more than *suvr2* alone. In summary, our results indicate that SUVR2 is a new regulator of the DRM2 pathway that acts downstream of siRNA biogenesis.

Conclusion

In summary, by generating single nucleotide resolution maps of the Arabidopsis methylome for a comprehensive list of mutants, we found interplays between different pathways and found new regulators of DNA methylation. All DNA methylation data generated in this study can be viewed at our genome browser along with various epigenomic data. These genome-wide datasets and tools should serve as a community resource for further understanding DNA methylation patterning in Arabidopsis.

MATERIALS AND METHODS

Plant material

All mutant lines used in this study were in the Columbia background. Exceptions are the *ros1 dml2 dml3* line where each allele was introgressed into Col (Penterman et al., 2007), and *suvr1/2/3/4/5*, where *suvr4* allele was in Nossen. First generation homozygous plants of *met1*, 2nd generation plants of *ibm1* and *vim1/2/3*, and 7th generation *ddm1* plants were used. Plants were grown under continuous light, and three-week-old leaves were used for all experiments.

Genome annotations

TAIR10 gene and TE models were obtained from The Arabidopsis Information Resource (www.arabidopsis.org), and transcription factor binding sites were obtained from AGRIS (http://arabidopsis.med.ohio-state.edu).

Whole genome bisulfite sequencing

0.5~1ug of genomic DNA was used to generate BS-seq libraries. Libraries were generated as previously described using premethylated adapters (Feng et al., 2011). Libraries were single-end sequenced on a HiSeq 2000 generating 50mer reads. Sequenced reads were base-called using the standard Illumina software. BS-seq reads were mapped to the TAIR10 genome using BS-seeker (Chen et al., 2010) allowing 2 mismatches. Identical reads were collapsed into one read. Methylation levels were calculated by the ratio of #C/(#C+#T). Differentially methylated regions (DMRs) were defined by tiling the genome into 100 base-pair bins and comparing the number of called Cs and Ts in mutant and wild-type. Bins with absolute methylation difference of 0.4, 0.2, 0.1 for CG, CHG, CHH, respectively, and *Benjamini*-Hochberg corrected FDR<0.01 (Fisher's

exact test) were selected. To avoid 100bp bins with few cytosines, we selected for bins with at least 4 cytosines that are each covered by at least 4 reads in the wild-type replicate. This whole process was performed against three wild-type replicates, and only regions called significant in all 3 comparisons were defined to be DMRs. Finally, because loss and gain of methylation occurred in clusters, DMRs within 200bp of each other were merged. All heat maps in this study were generated by complete linkage and using Euclidean distance as a distance measure. Rows with missing values were omitted for presentation purposes but did not affect the conclusions in the paper. Venn diagrams of DMRs were generated by calculating the overlap between 100 base-pair tiles of mutant DMRs (i.e. without merging).

RNA sequencing

0.1g of tissue was ground in Trizol (Invitrogen). Total RNA was treated with DNaseI (Roche), and cleaned up with phenol-chlorophorm and precipitated with ethanol. Poly(A) purifications were performed using mRNA purification dynabeads (Invitrogen) following manufacture instructions. RNA-seq libraries were generated following manufacturer instructions (Illumina) and sequenced on a Genome Analyzer. RNA-seq reads were uniquely mapped to the TAIR10 version of the Arabidopsis genome with Bowtie (Langmead et al., 2009) allowing 2 mismatches. Expression levels were measured by calculating reads per kilobase per million mapped reads (RPKM).

Histone modification data

Previously published histone ChIP data (Bernatavichute et al., 2008; Roudier et al., 2011; Zhang et al., 2009) were converted into TAIR10 coordinates and used for analyses.

Micrococcal nuclease (MNase) sequencing

One gram of tissue was ground in liquid nitrogen. 30 ml of Extraction Buffer 1 (0.4M sucrose, 10mM Tris-HCl pH 8, 10mM MgCl, 5mM BME, 0,1 mM PMSF, Complete Protease Inhibitor Cocktail Tablets (Roche)) was added, and filtered through Miracloth twice. Solution was spun at 4000 rpm for 20 minutes at 4C. The pellet was resuspended in 1ml of Extraction buffer 2 (0.25M sucrose, 10mM Tris-HCl pH 8, 10mM MgCl2, 1% Triton X-100, 5mM BME, 0.1mM PMSF, one Complete mini tablet (Roche)) and spun at 12,000g for 10 minutes at 4C. The pellet was resuspended in 300ul of Extraction Buffer 3 (1.7M sucrose, 10mM Tris-HCl pH 8, 0.15% Triton X-100, 2mM MgCl2, 5mM BME, 0.1mM PMSF, one Complete mini tablet (Roche)), and this was added to 300ul of Extraction buffer, and spun at 13.2 krpm for 1 hour at 4C. The pellet was washed once with 1ml digestion buffer (0.32M sucrose, 50mM Tris pH 8, 4mM MgCl2, 1mM CaCl2, 0.1mM PMSF), and resuspended in 300ul of digestion buffer. 1ul of MNase (Takara) was added and incubated at 37C. 300ul of Lysis buffer (0.1M Tris pH 8.5, 0.1M NaCl, 50mM EDTA, 1% SDS) was added and incubated at 65C for 45 minutes. DNA was recovered by phenol chlorophorm extraction followed by ethanol precipitation. Libraries were paired-end sequenced on a HiSeq 2000. Sequencing reads were mapped to the TAIR10 genome with Bowtie (Langmead et al., 2009) allowing 2 mismatches. For analyses, we selected for uniquely mapping pairs of reads that were oriented in the expected orientation and that were separated by 131-170bp. Randomly sheared DNA was used as a background control.

Southern Blot

MEA-ISR Southern blot was performed as previously described in Greenberg et al. 2011.

FWA Transgene Assays

The first generation of plants transformed and selected for the *FWA* transgene were assayed for total number of primary rosette and cauline leaves at time of flowering in long day conditions. Detailed methods for generation of transgenic plants that contain *FWA*, as well as subsequent flowering-time and bisulfite analysis can be found in Greenberg et al. 2011.

Small RNA Northern Blot

Northern blots were performed as described (Law et al., 2011).

ACCESSION NUMBER

All sequencing data have been deposited in GEO with accession GSE39901 and GSE38286.

FIGURE LEGENDS

Table 1-1. List of mutants tested in this study.

List of genes, mutant alleles, genome coverage (Arabidopsis genome size = 119 Mb) and error

rates are listed. Error rates were determined based on detected methylation levels of the

chloroplast genome (which is unmethylated).

Figure 1-1. CG methylation.

(A) Average distribution of CG methylation over protein-coding genes (left panel) and TEs

(right panel). Flanking regions are the same length as the gene or TE body (middle region). TSS=

transcription start site. TTS= transcription termination site.

(B) Heat map of CG methylation levels (black, 1; white 0) within all genes and TEs in

chromosome 1. Columns represent data for each indicated genotype, and rows represent the

genes/TEs. The rows were sorted by complete linkage hierarchical clustering with Euclidean

distance as a distance measure.

(C) Genome browser views of CG methylation in chromosome 1. Genes (black bars) and TEs

(grey bars) are shown below.

Figure 1-2. Further analysis of CG methylation.

(A) Chromosomal views of CG methylation levels. TE distribution (TE per bp) is shown to

indicate pericentromeric regions.

(B) Heat map of methylation levels within *met1* CG hypomethylation DMRs.

(C) Heat map of methylation levels within *vim1/2/3* CG hypomethylation DMRs.

Figure 1-3. Non-CG methylation.

(A) Genome coverage of defined CHG and CHH hypomethylation DMRs.

(B) Heat map of methylation levels within 17,437 *met1 cmt3* CHG (top panel) and 13,776 CHH (bottom panel) hypomethylation DMRs.

(C) Overlap between *kyp suvh5/6* and *drm1/2 cmt3* hypomethylation DMRs.

(D) Genome browser views of CHH methylation in chromosome 1. Genes (black bars) are shown below.

Figure 1-4. Further analysis of non-CG methylation.

(A-L) Overlap of CHG (left panels) and CHH (right panels) hypomethylation DMRs between indicated genotypes.

(M) Heat map of methylation levels within *cmt3* CHG and CHH hypomethylation DMRs.

(N) Heat map of methylation levels within *drm1/2* CHG and CHH hypomethylation DMRs.

(O) Heat map of methylation levels within *kyp suvh5/6* CHH hypomethylation DMRs.

(P) Heat map of methylation levels within *met1* CHG and CHH hypomethylation DMRs.

(Q) Average distribution of wild-type methylation levels over *drm1/2* CHH hypomethylation DMRs that overlap (solid lines) and do not overlap (faded lines) with *met1* CHH hypomethylation DMRs.

(R) Distribution of TE families overlapping with CHG (top panel) and CHH (bottom panel) hypomethylation DMRs.

(S) Overlap between TEs associated with *kyp* and *cmt3* CHH hypomethylation DMRs.

Figure 1-5. Characteristics of CHH sites regulated by KYP SUVH5/6, CMT3, and DRM1/2.

(A) G+C content ((G+C)/(G+C+A+T)) in CHH hypomethylation DMRs. Red lines, median; edges of boxes, 25th (bottom) and 75th (top) percentiles; error bars, minimum and maximum points within 1.5xIQR (Interquartile range); red dots, outliers.

(B) Base composition over *drm1/2* CHH hypomethylation DMRs.

(C) Average distribution of CHH hypomethylation DMRs (DMR per bp) over different repeats. Flanking regions are the same length as the repeat (middle region).

(D) Boxplots of sizes of TEs that overlap with CHH hypomethylation DMRs.

(E) Boxplots of distances between CHH hypomethylation DMRs and the closest gene TSS.

(F) Fraction of CHH hypomethylation DMRs that are within 1kb or 2kb from TSS.

(G) Average distribution of DMRs over TEs of indicated sizes. Negative x-axis scale is outside of TEs, and positive x-axis scale is towards the body of TEs.

(H) Genome browser views showing loss of methylation spikes at boundaries of TEs in *drm1/2* in chromosome 1. Genes (black bars) and TEs (grey bars) are shown below.

(I) TFBS (TFBS per bp) over CHH hypomethylation DMRs.

(J) Wild-type expression levels of TEs overlapping with CHH hypomethylation DMRs.

(K) Average histone modification and nucleosome distributions over CHH hypomethylation DMRs.

Figure 1-6. Characteristics of CHG sites regulated by KYP SUVH5/6, CMT3, and DRM1/2.

(A) G+C content ((G+C)/(G+C+A+T)) in CHG hypomethylation DMRs.

(B) Base composition over *drm1/2* CHG hypomethylation DMRs.

(C) Average distribution of CHG hypomethylation DMRs (DMR per bp) over different repeats. Flanking regions are the same length as the repeat (middle region).

(D) Boxplots of sizes of TEs that overlap with CHG hypomethylation DMRs.

(E) Boxplots of distances between CHG hypomethylation DMRs and the closest gene TSS.

(F) Fraction of CHG hypomethylation DMRs that are within 1kb or 2kb from TSS.

(G) Average distribution of DMRs over TEs of indicated sizes. Negative x-axis scale is outside of TEs, and positive x-axis scale is towards the body of TEs.

(H) Genome browser views showing loss of methylation spikes at boundaries of TEs in *drm1/2* in chromosome 1. Genes (black bars) and TEs (grey bars) are shown below.

(I) TFBS (TFBS per bp) over CHG hypomethylation DMRs.

(J) Wild-type expression levels of TEs overlapping with CHG hypomethylation DMRs.

(K) Average histone modification and nucleosome distributions over CHG hypomethylation DMRs.

(L) Average distribution of indicated TFBS over CHG (top panel) and CHH (bottom panel) hypomethylation DMRs.

Figure 1-7. RNA directed DNA methylation.

(A) Heat map of methylation levels within 4,949 *drm1/2* CHH hypomethylation DMRs. Genotypes (columns) have also been clustered.

(B) Genome browser views of DNA methylation in chromosome 1. Genes (black bars) are shown below.

(C) Overlap of *rdr1* and *rdr6* CHG hypomethylation DMRs with TEs and genes.

Figure 1-8. Further analysis of RNA directed DNA methylation.

(A) Genome browser views of sites that lose DNA methylation in *fca fpa* in chromosome 1. Genes (black bars) are shown below.

(B) Fraction of *fca fpa* CHH hypomethylation DMR overlapping with gene promoters, gene bodies, and TEs.

(C) Heat map of methylation levels within *fca fpa* CHG (top panel) and CHH (bottom panel) hypomethylation DMRs.

(D) Overlap between *rdr1* and *rdr6* CHG hypomethylation DMRs.

(E) Heat map of methylation levels within *rdr6* CHG hypomethylation DMRs.

(F) Heat map of methylation levels within *rdr1* CHG hypomethylation DMRs.

(G) Average distribution of small RNA-seq reads over CHG hypomethylation DMRs of *drm1/2* and *rdr1+rdr6*.

Figure 1-9. Ectopic hypermethylation.

(A) Heat map of methylation levels within 4,773 *met1* CHG hypermethylation DMRs.

(B) Heat map of methylation levels within 2,695 *ddm1* CHG hypermethylation DMRs.

(C) Heat map of methylation levels within 13,588 *ibm1* CHG hypermethylation DMRs.

(D) Genome coverage of defined CHG and CHH hypermethylation DMRs.

(E) Fraction of CHH hypermethylation DMRs non-overlapping with CHG hypermethylation DMRs.

(F) Fraction of CHH hypermethylation DMRs non-overlapping with CHG hypermethylation DMRs that overlap with genes and TEs.

(G) Genome browser views of DNA methylation in wild-type and *fas2* in chromosome 1. Genes (black bars) and TEs (grey bars) are shown below.

(H) Heat map of methylation levels within 1,572 *fas2* CHG hypermethylation DMRs.

(I) Chromosomal views of methylation in wild-type (faded lines) and *fas2* (solid lines). Regions of pericentromeric heterochromatin are indicated by back bars below the graphs.

(J) Average distribution of methylation levels over genes and TEs in wild-type (faded lines) and *fas2* (solid lines). Upstream and downstream regions are the same length as the gene/TE (middle region).

(K) Heat map of methylation levels within *fas2* CHG hypermethylation DMRs.

Figure 1-10. Further analysis of hypermethylation.

(A) Heat map of methylation levels within *rdd* CHG hypermethylation DMRs.

(B) Heat map of methylation levels within *vim1/2/3* CHG hypermethylation DMRs.

(C) Heat map of methylation levels within *rdd* CHG hypermethylation DMRs.

(D and E) Heat map of methylation levels within *met1* CHG and CHH hypermethylation DMRs.

(F) Overlap between *met1, ddm1*, and *ibm1* hypermethylation DMRs.

(G) Overlap between *ddm1* and *ibm1* CHG hypermethylation DMRs.

(H) Relative expression levels of IBM1 gene in wild-type and *ddm1*. The average expression levels between 2 biological replicates (Stroud et al., 2012a) was calculated. Data are represented as mean +/- SD.

(I) Average distribution of *drm1/2* CHG hypermethylation DMRs over edges of TEs. DMRs (DMR per bp x10^{-4}) were plotted over boundaries of all TEs in the genome.

(J) Boxplots of CHG methylation differences of indicated mutants relative to wild-type in *drm1/2* CHG hypermethylation DMRs.

(K) Average distribution of CHH methylation levels over *met1*, *ddm1*, and *ibm1* CHG hypermethylation DMRs. Flanking regions are the same length as the DMR (middle region).

(L) Distribution of TE families overlapping with *met1* and *ddm1* CHH hypermethylation DMRs.

(M) Distribution of TE families overlapping with *fas2* CHG hypermethylation DMRs.

Figure 1-11. RNA Pol II directed DNA methylation.

(A) Genome browser views of DNA methylation in wild-type and *nrpb2* in chromosome 1. Genes (black bars) and TEs (grey bars) are shown below.

(B) Fraction of CHG hypomethylation DMRs overlapping with TEs and genes.

(C and D) Heat map of methylation levels within 413 *nrpb2* CHG hypomethylation DMRs.

Figure 1-12. Relationship between histone modifications and DNA methylation.

(A) Top panel: Average distribution of DNA methylation levels over defined H3K27me1 regions in wild-type. Flanking regions are the same length as the H3K27me1 region (middle region). Bottom panel: Heat map of methylation levels within H3K27me1 regions.

(B) Top panel: Average distribution of DNA methylation levels over defined H3K36me3 regions in wild-type. Flanking regions are the same length as the H3K36me3 region (middle region). Bottom panel: Heat map of methylation levels within H3K36me3 regions.

(C) Top panel: Average distribution of DNA methylation levels over defined regions that gain H3K27me3 in *ref6*. Flanking regions are the same length as the regions that gain H3K27me3 in

ref6 (middle region). Bottom panel: Heat map of methylation levels within regions that gain H3K27me3 in *ref6*.

(D) Top panel: Average distribution of DNA methylation levels over defined H3K4me3 regions in wild-type. Flanking regions are the same length as the H3K4me3 region (middle region). Bottom panel: Heat map of methylation levels within H3K4me3 regions.

(E) *MEA-ISR* Southern blot. Genomic DNA was digested with the *MspI* restriction endonuclease, which is sensitive to non-CG methylation. Hybridization with a probe specific to *MEA-ISR* indicates increased digestion in *suvr2* compared to wild-type, thus a decrease in DNA methylation.

(F) Small RNA northern blots. Small RNAs were probed with both Type I and Type II sequences. *miR159* was used as a loading control.

Figure 1-13. Chromatin modifiers involved in DNA methylation.

(A) Genome browser views of DNA methylation in wild-type and *hda6* in chromosome 1. Genes (black bars) are shown below.

(B) Fraction of CHG hypomethylation DMRs that are within 1kb or 2kb from TSS.

(C and D) Heat map of methylation levels within 120 *hda6* CHG hypomethylation DMRs.

(E) Heat map of methylation levels within 1,113 *suvr2* CHH hypomethylation DMRs.

(F) Overlap between *suvr2* and *drm1/2* CHH hypomethylation DMRs.

(G) Average distribution of CHH methylation over *drm1/2* CHH hypomethylation DMRs.

(H) Heat map of methylation levels within 4,949 *drm1/2* CHH hypomethylation DMRs.

(I) *FWA* flowering-time assay. ~20 plants were measured in each population. Data are represented as mean +/- SEM.

(J) *FWA* transgene bisulfite analysis. DNA methylation of the transgenic copy of FWA in FWA transformed plants was analyzed.

Table 1-1

	Gene	Mutant Allele	Uniquely Mapping Reads	Coverage (X)	CG Error Rate	CHG Error Rate	CHH Error Rate
1	Col replicate 1	-	76813710	32.3	1.20%	1.16%	0.72%
2	Col replicate 2	-	66831128	28.1	1.70%	1.75%	0.99%
3	Col replicate 3	-	44697982	18.8	1.55%	1.50%	1.25%
4	AGO1	ago1-27	112546984	47.3	0.79%	0.74%	0.52%
5	AGO2	ago2-1	113380729	47.6	0.86%	0.84%	0.58%
6	AGO3	ago3-1	115140994	48.4	1.55%	1.57%	0.90%
7	AGO4	ago4-5	113021824	47.5	1.02%	1.02%	0.64%
8	AGO5	ago5-2	86162158	36.2	1.15%	1.16%	0.73%
9	AGO6	ago6-2	122321945	51.4	0.97%	0.96%	0.58%
10	AGO7	zip-1	103368628	43.4	1.32%	1.33%	0.77%
11	AGO8	ago8-2	109967887	46.2	1.00%	0.97%	0.64%
12	AGO9	ago9-2	79577300	33.4	2.22%	2.32%	1.23%
13	AGO10	ago10-3	113035513	47.5	1.10%	1.06%	0.74%
14	ATXR5/6	SALK_130607; SAIL_240_H01	101211138	42.5	0.86%	0.82%	0.54%
15	BRU1	bru1-4	104005341	43.7	0.85%	0.83%	0.57%
16	CLSY1	clsy1-7	118106905	49.6	0.79%	0.76%	0.49%
17	CMT1	SALK_030404	128641965	54.1	0.95%	0.92%	0.64%
18	CMT2	WISCDSLOX7E02	116673442	49.0	1.24%	1.23%	0.92%
19	CMT3	cmt3-11	86439602	36.3	0.82%	0.77%	0.50%
20	DCL2	dcl2-1	112385119	47.2	3.40%	3.55%	1.84%
21	DCL3	dcl3-1	106112741	44.6	1.01%	1.01%	0.60%
22	DCL4	dcl4-2	85344423	35.9	1.13%	1.13%	0.62%
23	DCL2/4	dcl2-1; dcl4-2	104554046	43.9	1.33%	1.38%	0.93%
24	DCL2/3/4	dcl2-1; dcl3-1; dcl4-2	89872096	37.8	0.99%	0.94%	0.62%
25	DDM1	ddm1-2	61138399	25.7	1.35%	1.37%	0.83%
26	DMS3	dms3-4	56586550	23.8	0.97%	0.93%	0.60%
27	DMS4	dms4-3	93753769	39.4	2.27%	2.36%	1.24%
28	DNMT2 CMT3	SALK_1366; cmt3-11	113857619	47.8	1.16%	1.13%	0.78%
29	DNMT2 DRM1/2	SALK_1366; drm1-2; drm2-2;	128948098	54.2	1.20%	1.16%	0.80%
30	DRD1	drd1-6	107976314	45.4	1.36%	1.37%	0.84%
31	DRM1/2	drm1-2; drm2-2	56195598	23.6	0.80%	0.74%	0.56%
32	DRM1/2 CMT3	drm1-2; drm2-2; cmt3-11	116152677	48.8	0.77%	0.77%	0.51%
33	DRM3	drm3-1	91597689	38.5	1.02%	1.04%	0.59%
34	FAS2	SALK_033228	111567207	46.9	0.84%	0.81%	0.50%
35	FCA	fca-9	116059441	48.8	0.95%	0.92%	0.61%
36	FCA FPA	fca-9; fpa-7	92928024	39.0	1.89%	1.94%	1.23%
37	FLD	fld-4	122281351	51.4	0.93%	0.92%	0.57%
38	FPA	fpa-7	119272902	50.1	1.16%	1.18%	0.75%
39	FVE	fve-3	94745237	39.8	0.66%	0.61%	0.43%
40	HDA6	axe1-5	118583550	49.8	0.71%	0.69%	0.52%
41	HEN1	hen1-6	96439193	40.5	2.46%	2.56%	1.34%
42	IBM1	SALK_006042	98203032	41.3	1.70%	1.76%	0.96%
43	IDN2	idn2-1	104290706	43.8	1.21%	1.17%	0.73%
44	IDN2 IDNL1/2	idn2-1; idl1-1; idl2-1	130587808	54.9	0.90%	0.87%	0.53%
45	IDNL1 IDNL2	idl1-1; idl2-1	25289459	10.6	0.69%	0.62%	0.44%
46	KTF1	ktf1-1/rdm3-3	95432884	40.1	1.02%	1.00%	0.63%

47	MET1	*met1-3*	111861104	47.0	1.14%	1.10%	0.77%
48	MET1 CMT3	*met1-3; cmt3-11*	98425158	41.4	1.07%	1.04%	0.64%
49	MET1 HET	*met1-3*	118714517	49.9	1.39%	1.37%	0.89%
50	MET1 WT	*met1-3*	106080395	44.6	1.67%	1.68%	1.02%
51	MET2	SAIL_815_G06	121096301	50.9	1.00%	0.96%	0.68%
52	MOM1	SAIL_610_G01	99721205	41.9	1.01%	0.98%	0.66%
53	MSI2	*msi2-1*	126887575	53.3	1.23%	1.20%	0.85%
54	NRPB2	*nrpb2-3*	99304633	41.7	0.60%	0.58%	0.40%
55	NRPD1	*nrpd1a-4*	119641641	50.3	1.25%	1.27%	0.95%
56	NRPE1	*nrpd1b-11*	110711289	46.5	1.20%	1.21%	0.71%
57	ROS1 DML2/3	*ros1-3; dml2-1;dml3-1*	103951405	43.7	0.92%	0.91%	0.61%
58	RDM1	*rdm1-4*	120118111	50.5	0.83%	0.81%	0.61%
59	RDR1	*rdr1-1*	50918266	21.4	1.07%	1.06%	0.66%
60	RDR2	*rdr2-2*	103496056	43.5	1.09%	1.07%	0.73%
61	RDR6	*rdr6-15*	106527813	44.8	1.30%	1.30%	0.78%
62	REF6	SAIL_747_A07	131871583	55.4	1.03%	0.98%	0.68%
63	ROS3	SALK_022363	113826390	47.8	0.81%	0.78%	0.53%
64	RPA2	*rpa2-4*	104308250	43.8	1.22%	1.16%	0.79%
65	SDE3	*sde3-5*	89911818	37.8	1.83%	1.83%	1.02%
66	SDE5	*sde5-2*	106191880	44.6	0.89%	0.89%	0.62%
67	SDG2	SALK_021008	31345638	13.2	1.49%	1.52%	0.93%
68	SDG8	*sdg8-2*	112060908	47.1	0.84%	0.83%	0.52%
69	SGS3	*sgs3-14*	119435088	50.2	0.84%	0.81%	0.49%
70	SUVH1	SALK_003675	109474100	46.0	1.26%	1.25%	0.76%
71	SUVH2	SALK_079574	102275261	43.0	0.99%	0.98%	0.64%
72	SUVH3	Garlic_401b_D01_.b.1a	82520351	34.7	1.51%	1.51%	0.95%
73	SUVH4/KYP	SALK_41474	110536025	46.4	0.99%	0.99%	0.63%
74	SUVH5	GABI-Kat line 263C05	119106559	50.0	1.34%	1.33%	0.81%
75	SUVH6	Garlic_1244_F04.b.1a	115158171	48.4	1.25%	1.27%	0.75%
76	SUVH4/5/6	*kyp suvh5 suvh6*	104134066	43.8	0.97%	0.97%	0.62%
77	SUVH7	Gabi-kat_037C06	45892440	19.3	1.08%	1.07%	0.63%
78	SUVH8	SALK_123140	100941655	42.4	1.01%	1.00%	0.61%
79	SUVH9	SALK_048033	112515012	47.3	1.19%	1.19%	0.76%
80	SUVH10	SALK_152977	92030177	38.7	1.08%	1.07%	0.67%
81	SUVR1	*suvr1-1*/ SALK_012786	111208489	46.7	0.94%	0.92%	0.64%
82	SUVR2	*suvr2-1*/ SAIL_832_E07	91706842	38.5	0.76%	0.71%	0.48%
83	SUVR3	*suvr3-1*/ SALK_063174	103495616	43.5	1.67%	1.71%	0.96%
84	SUVR5	*suvr5-1*/ SALK_026224	90272885	37.9	1.69%	1.74%	1.00%
85	SUVR1/2/3/4/5	*suvr1-1;suvr2-1; suvr3-1;* RATM12-090501; *suvr5-1*	114299373	48.0	1.74%	1.77%	0.99%
86	VIM1	*vim1-2*	110688864	46.5	0.98%	0.96%	0.61%
87	VIM2	*vim2-1*	85407593	35.9	1.20%	1.23%	0.76%
88	VIM3	*vim3-1*	105271053	44.2	0.75%	0.70%	0.56%
89	VIM1 VIM2 VIM3	*vim1-2; vim2-1;vim3-1*	102634982	43.1	1.15%	1.17%	0.69%

Figure 1-1
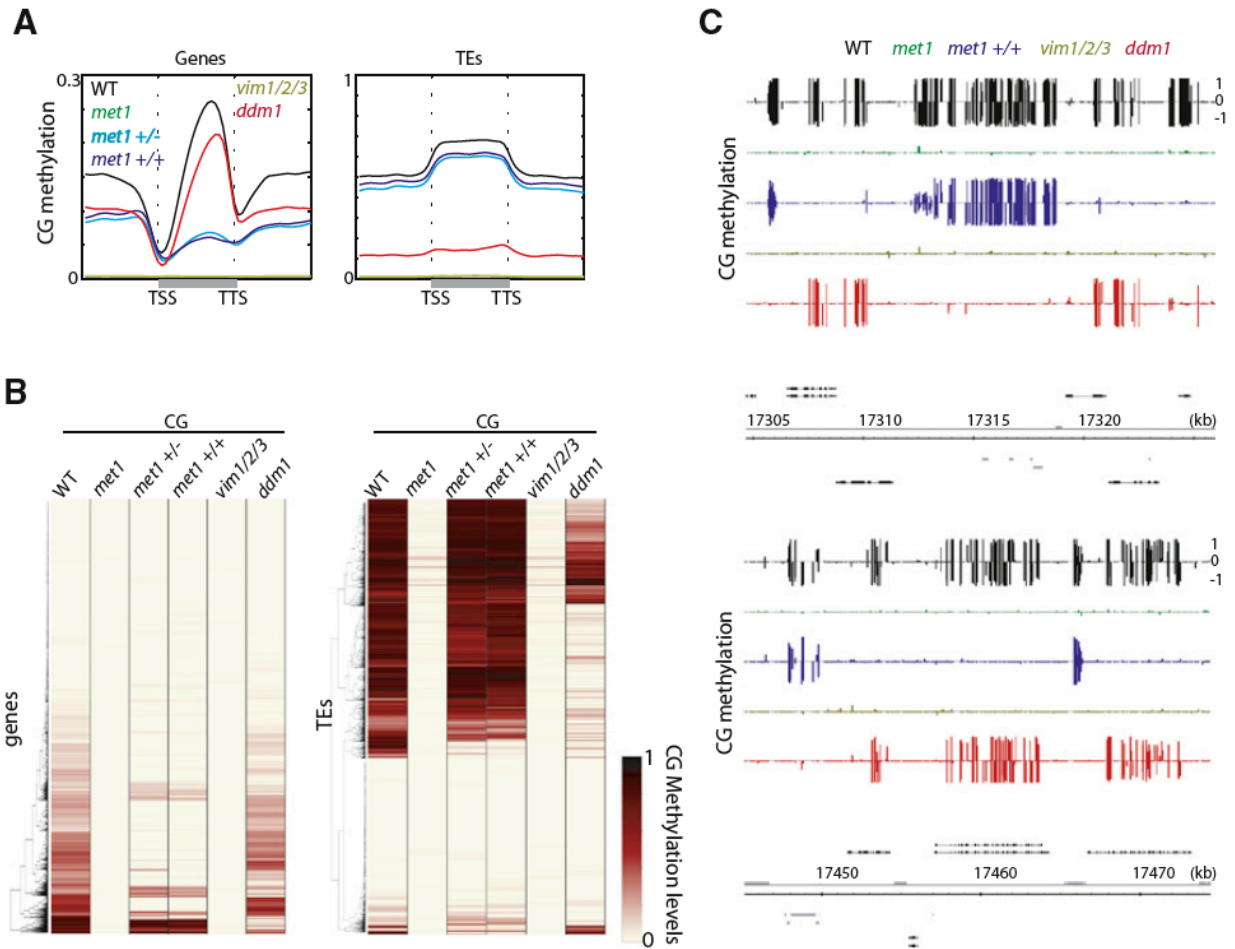
Figure 1-2
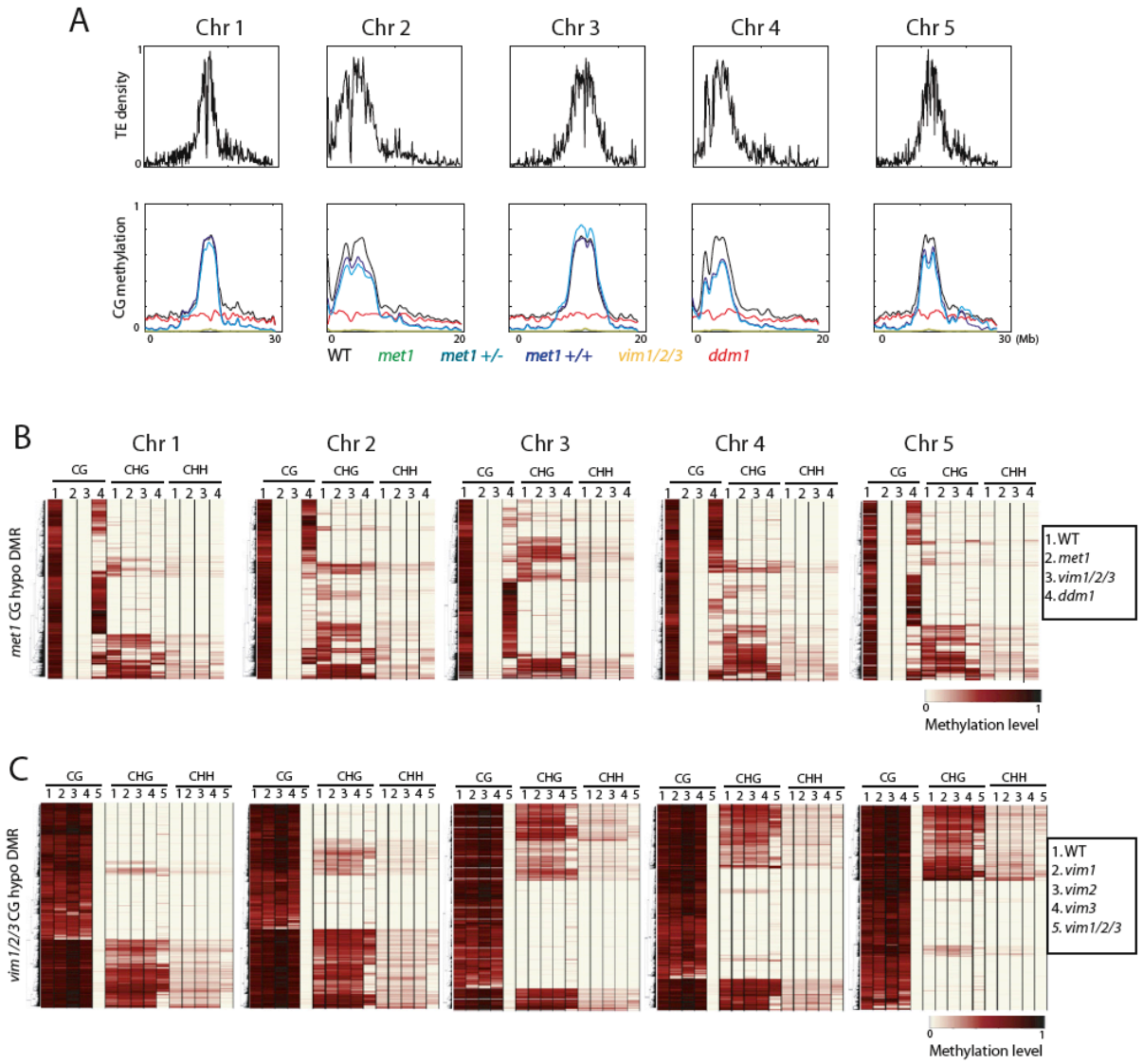
Figure 1-3
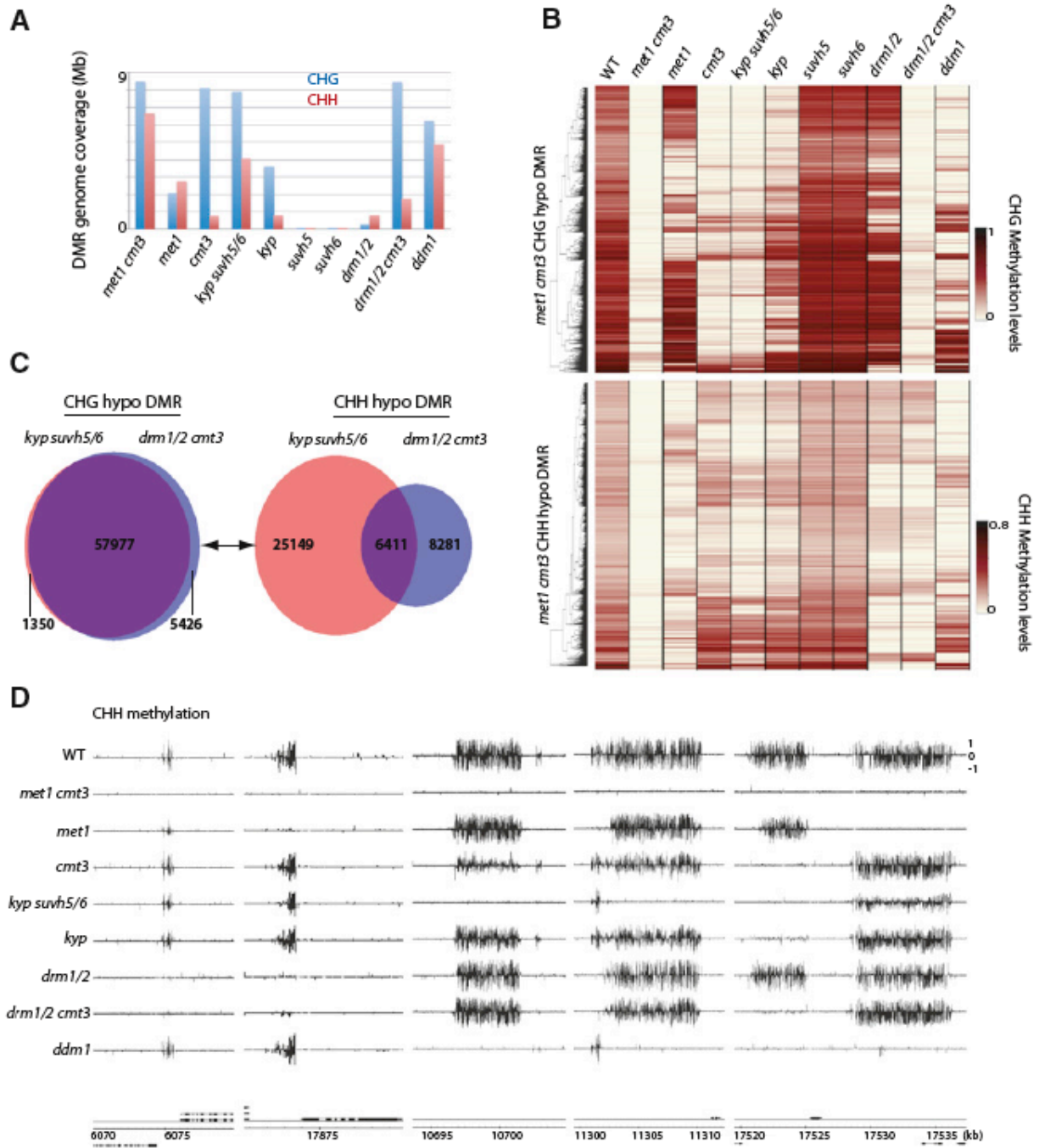
Figure 1-4
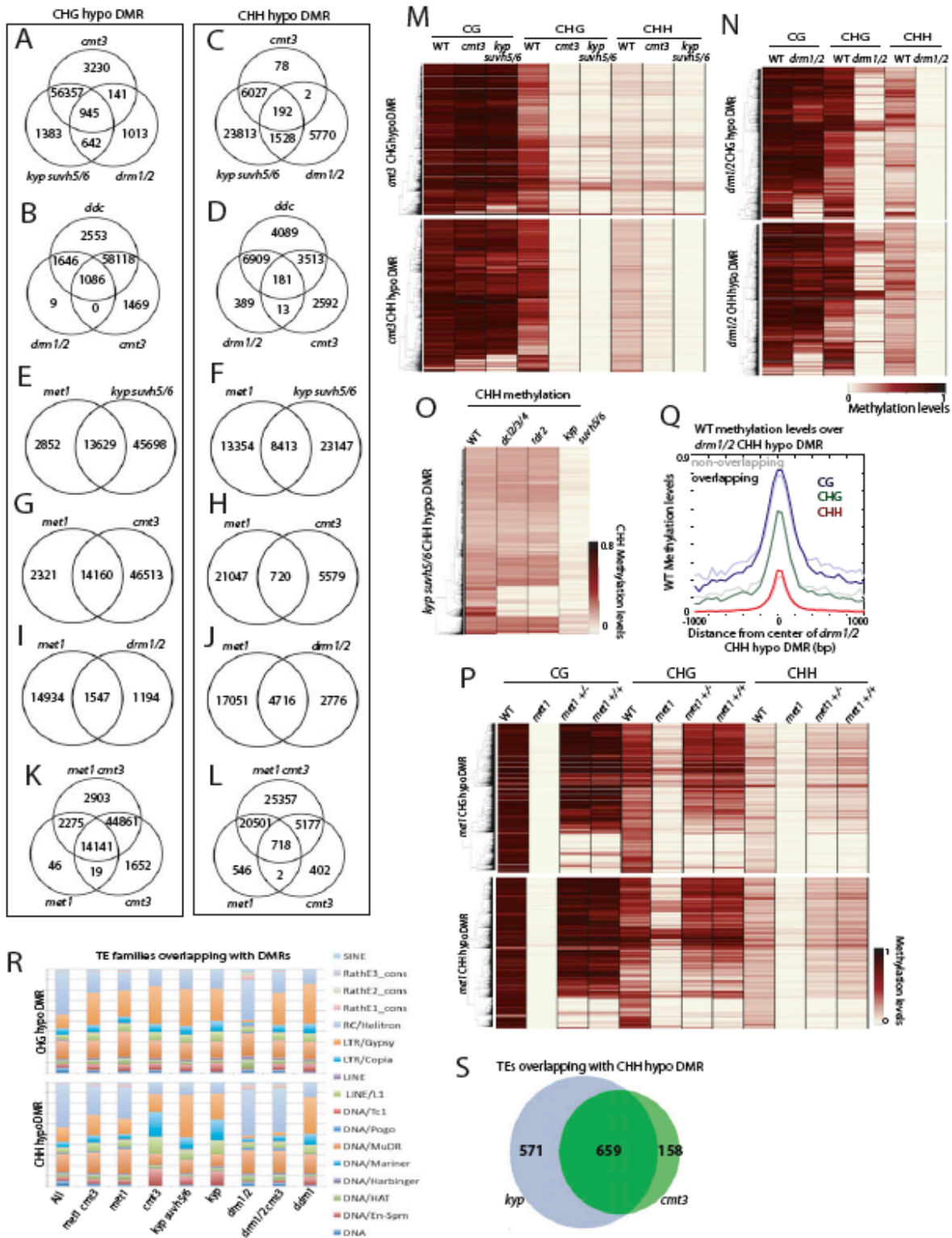
Figure 1-5
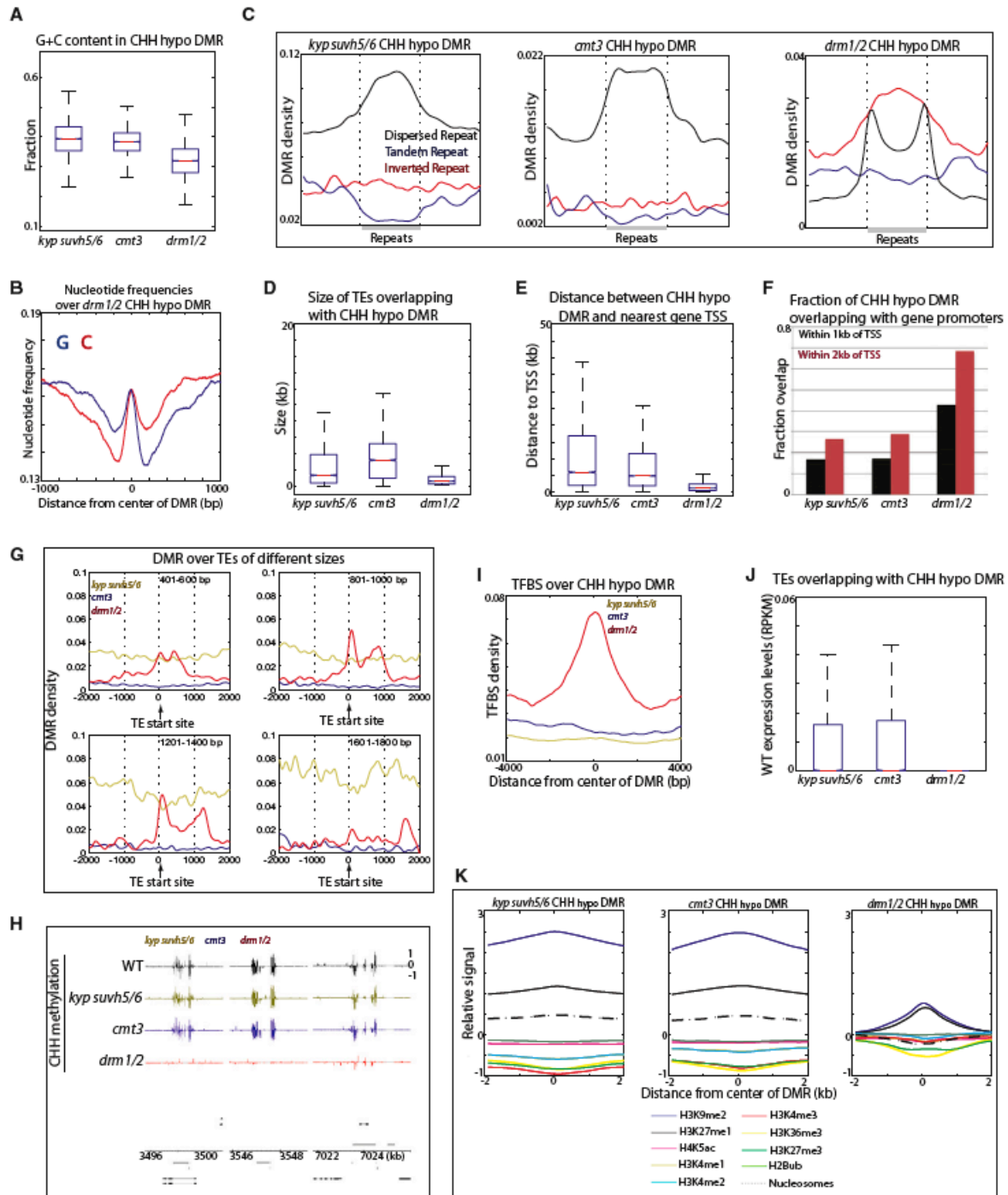
Figure 1-6
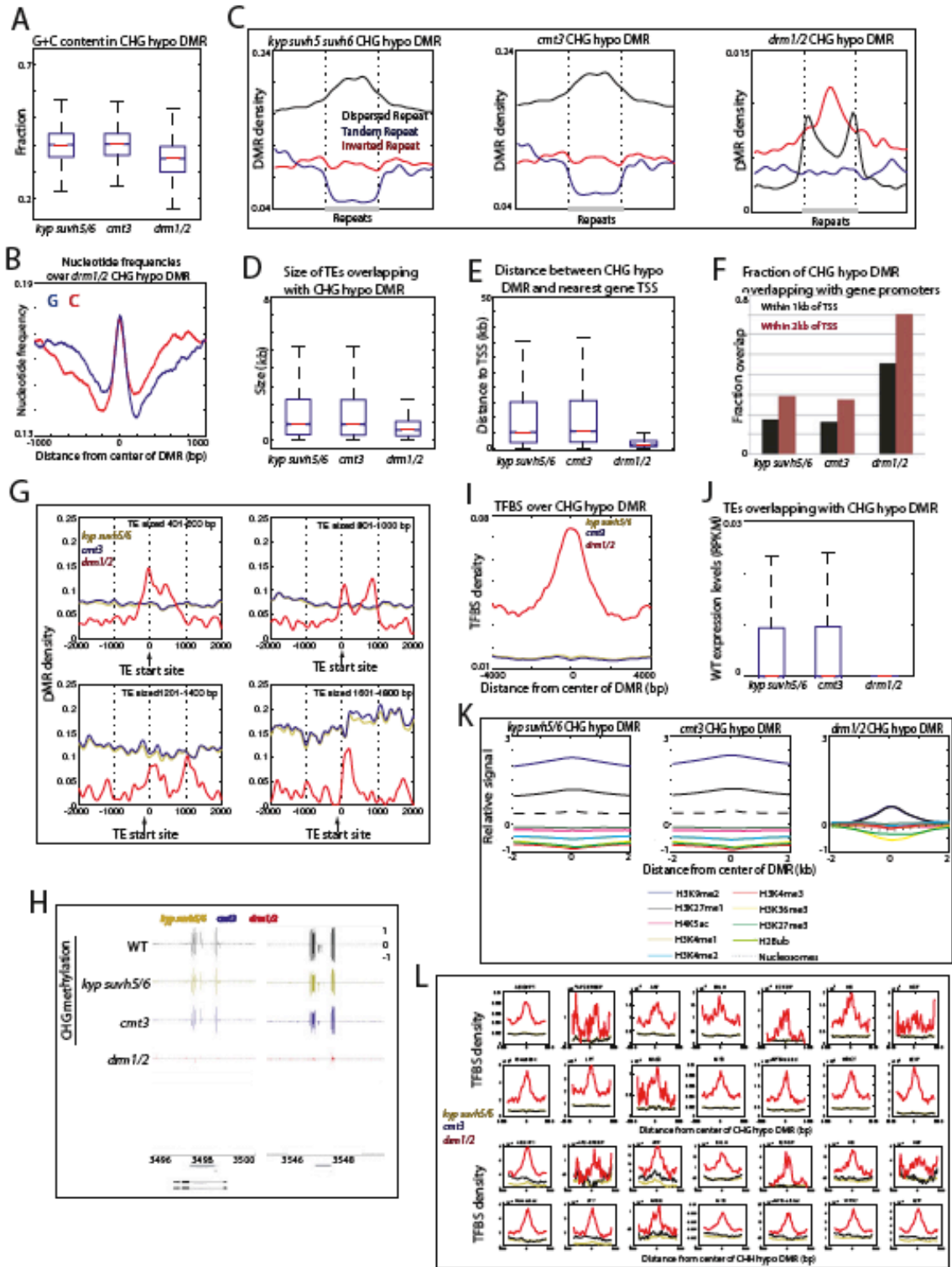
Figure 1-7
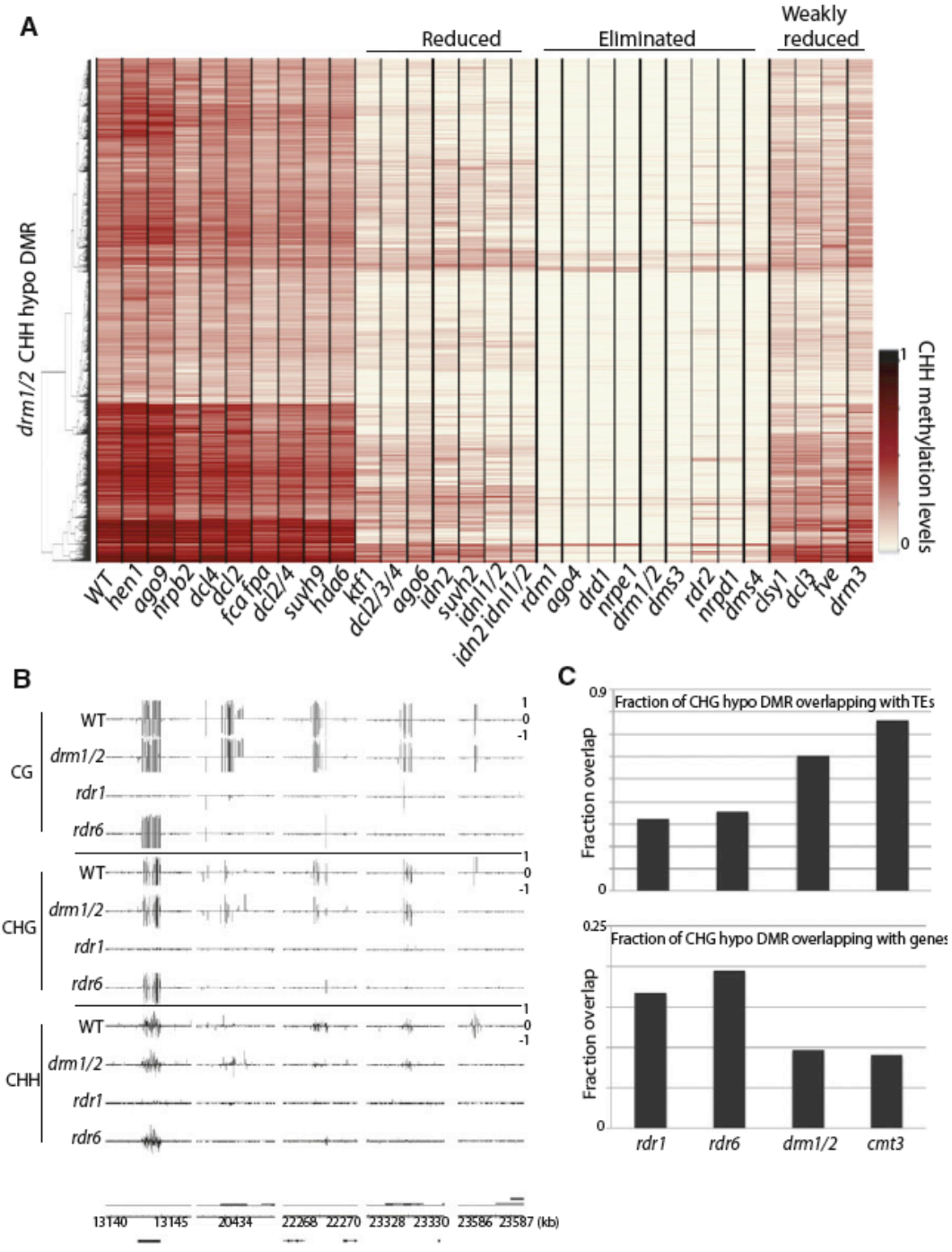
Figure 1-8
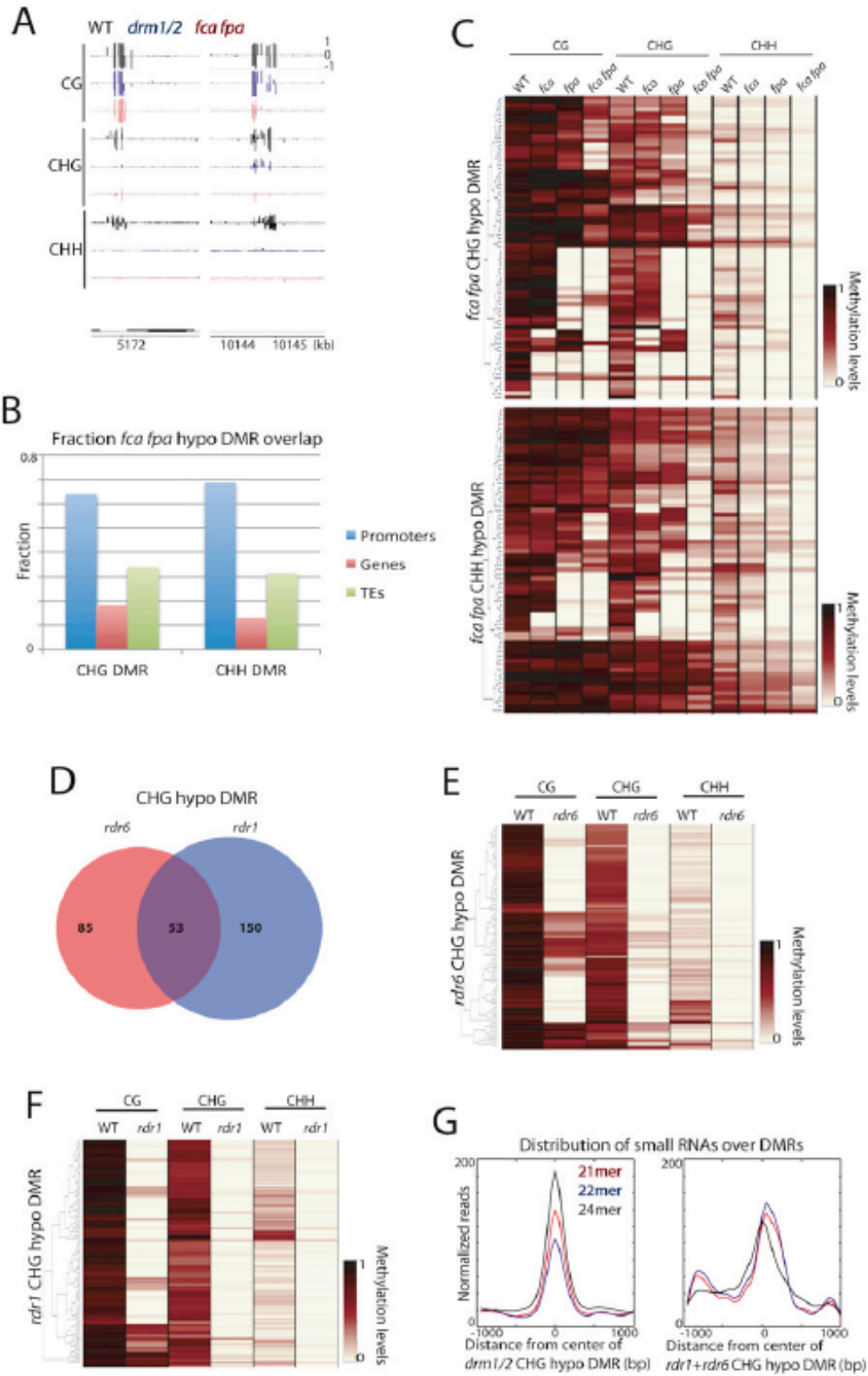
Figure 1-9
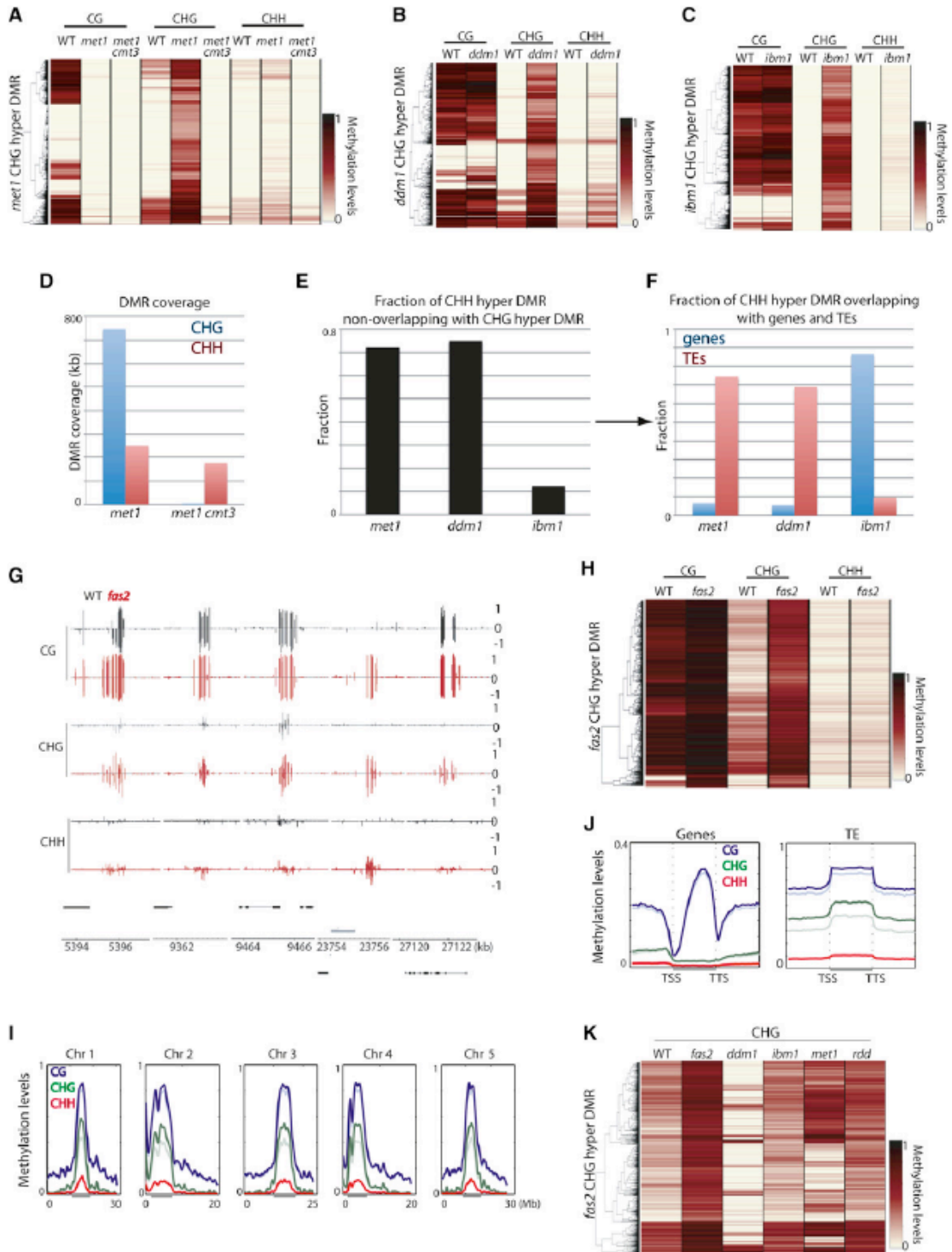
Figure 1-10

Figure 1-11

Figure 1-12

A H3K27me1 enriched regions in WT

B H3K36me3 enriched regions in WT

C H3K27me3 increased regions in *ref6*

D H3K4me3 enriched regions in WT

E

F

63
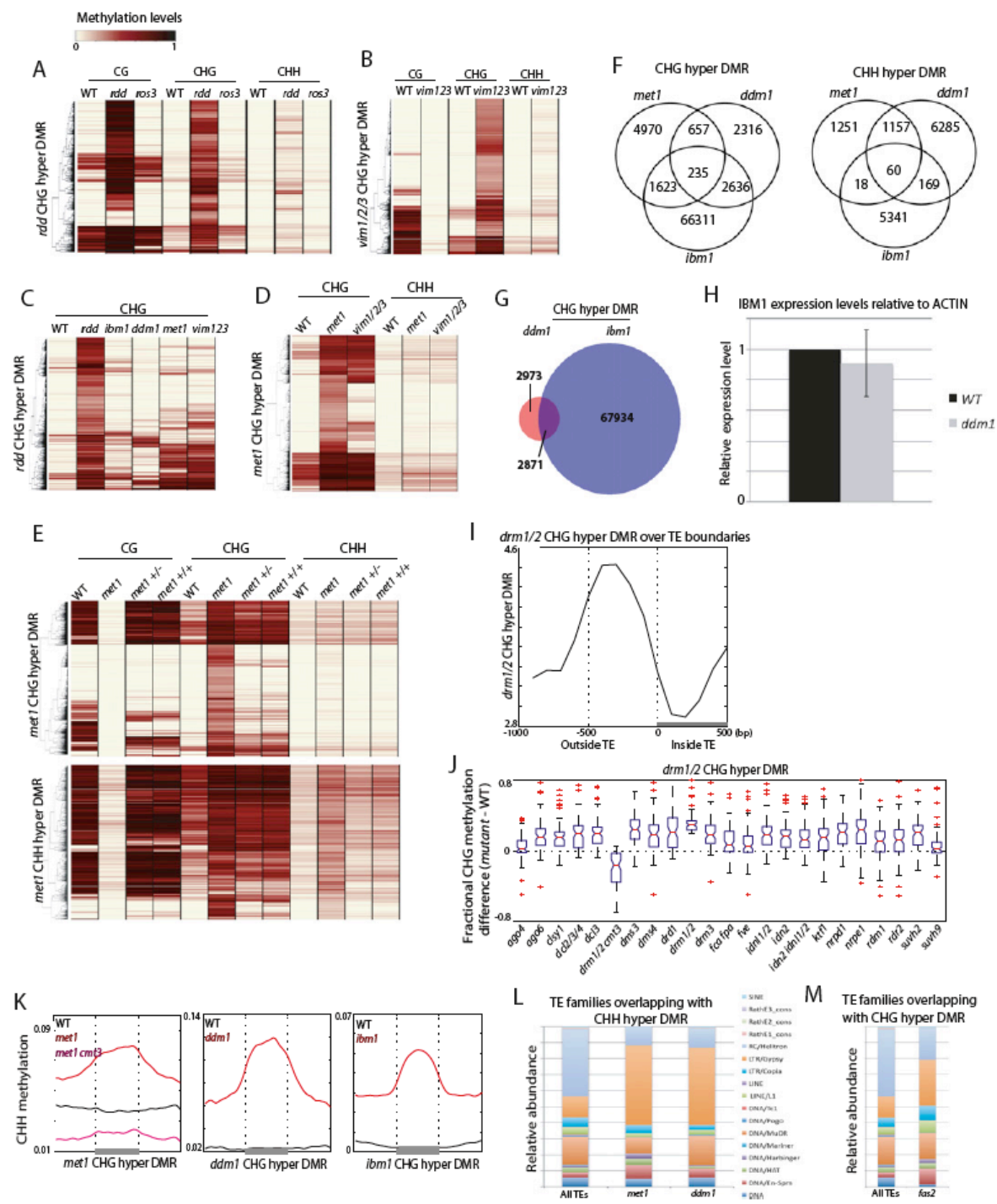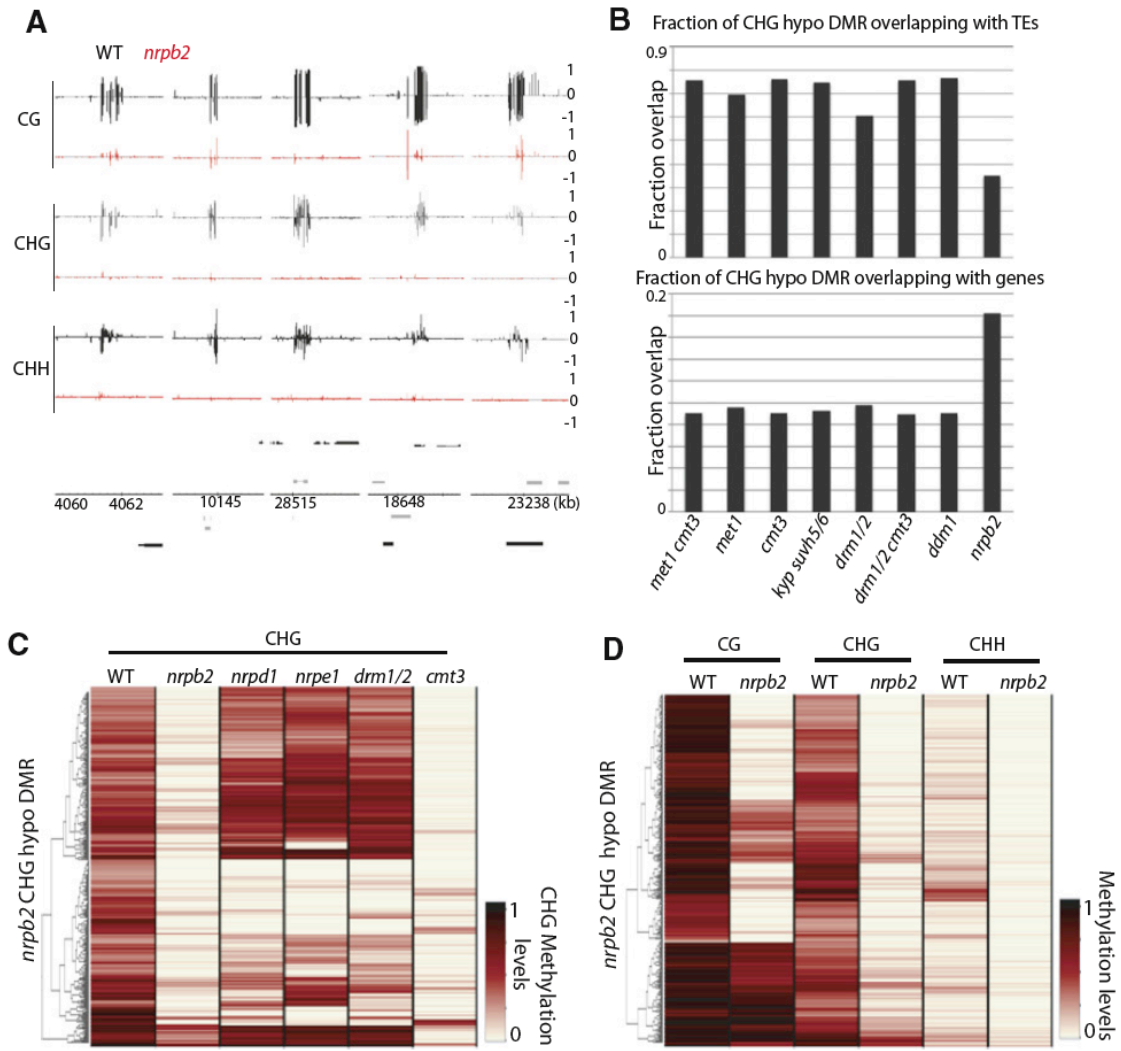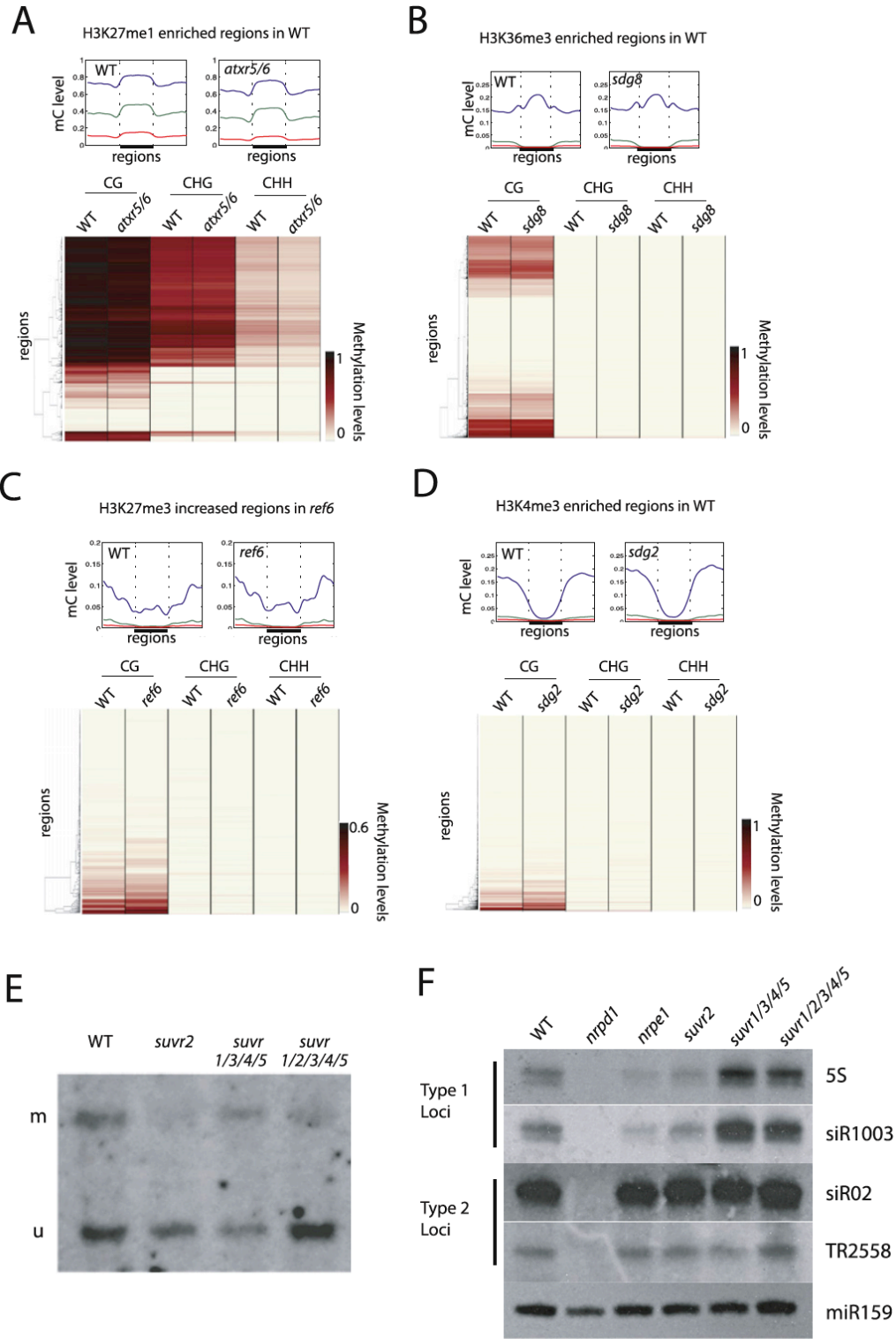
Figure 1-13
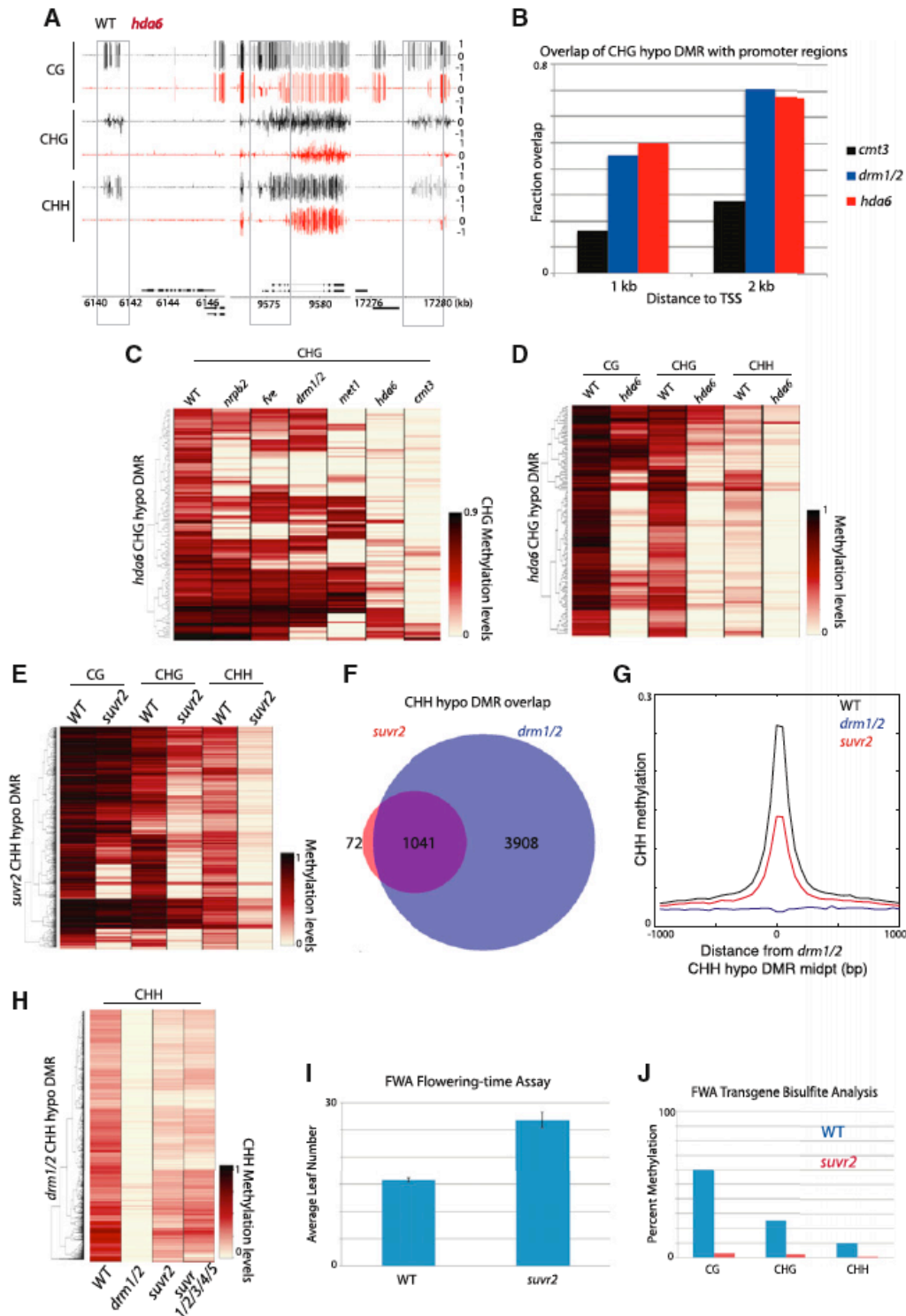
CHAPTER 2

Regulation of heterochromatic DNA replication by histone H3 lysine 27 methyltransferases

ABSTRACT

Multiple pathways prevent DNA replication from occurring more than once per cell cycle(Arias and Walter, 2007). These pathways block re-replication by strictly controlling the activity of pre-replication complexes, which assemble at specific sites in the genome called origins. Here we show that mutations in the homologous histone 3 lysine 27 (H3K27) monomethyltransferases, *ARABIDOPSIS TRITHORAX-RELATED PROTEIN5* (*ATXR5*) and *ATXR6*, lead to re-replication of specific genomic locations. The vast majority of these locations correspond to transposons and other repetitive and silent elements of the Arabidopsis genome. These sites also correspond to high levels of H3K27 monomethylation, and mutation of the catalytic SET domain is sufficient to cause the re-replication defect. Mutation of *ATXR5* and *ATXR6* also causes upregulation of transposon expression and has pleiotropic effects on plant development. These results uncover a novel pathway that prevents over-replication of heterochromatin in Arabidopsis.

We previously characterized two redundant histone methyltransferase genes, *ATXR5* and *ATXR6*, and demonstrated that the ATXR5 and ATXR6 proteins show H3K27 monomethylation (H3K27m1) activity *in vitro*, and that an *atxr5 atxr6* double mutant shows a reduction of H3K27me1 *in vivo(Jacob et al., 2009)*. The *atxr5 atxr6* double mutant exhibits pleiotropic defects in plant development, including smaller misshapen leaves(Jacob et al., 2009). Overexpression of ATXR5 or ATXR6 also causes morphological defects and male sterility(Raynaud et al., 2006). Furthermore, the double mutant displays reactivation of the expression of a variety of both DNA transposons and retrotransposons(Jacob et al., 2009). Interestingly, *atxr5 atxr6* did not disturb DNA methylation or histone H3K9 dimethylation (H3K9me2, a key repressive histone modification correlated with DNA methylation(Bernatavichute et al., 2008 ; Jackson et al., 2002; Malagnac et al., 2002)), suggesting that ATXR5 and ATXR6 act via a novel pathway to maintain gene silencing. Previous work also provided hints that ATXR5 and ATXR6 show links with DNA replication. *ATXR5* and *ATXR6* expression is regulated by the cell cycle, with expression peaking just before DNA replication(Raynaud et al., 2006) and *ATXR6* expression is strongly coregulated with CDT1, ORC2 and other DNA replication proteins(Obayashi et al., 2009). In addition, ATXR5 and ATXR6 contain PCNA-interacting-protein (PIP) motifs and have been shown to interact with the two PROLIFERATING CELL NUCLEAR ANTIGEN (PCNA) proteins in Arabidopsis (AtPCNA1 and AtPCNA2)(Raynaud et al., 2006). PCNA interacts with DNA polymerase and serves as a general loading platform for many proteins involved in diverse processes occurring at chromatin(Moldovan et al., 2007). In this report, we show that *ATXR5* and *ATXR6* are critical factors that act in a novel pathway to suppress DNA re-replication, especially in heterochromatic regions of the Arabidopsis genome.

To determine if ATXR5 and/or ATXR6 have a role in DNA replication, we analyzed the DNA content of leaf nuclei by flow cytometry. Leaves are well suited for assessing DNA replication defects as they undergo both mitosis and endoreduplication (genome duplication without mitosis), which is responsible for the widespread polyploidy observed in mature leaf tissue(Galbraith et al., 1991). Nuclei were extracted from mature rosette leaves of wild-type Columbia (Col), *atxr5*, *atxr6,* and *atxr5 atxr6* plants, stained with propidium iodide (PI), and analyzed for DNA content. Col, *atxr5,* and *atxr6* all showed well-resolved populations of 2C, 4C, 8C, 16C, and 32C nuclei (Figure 2-1a). Thus *atxr5* and *atxr6* single mutations do not have an impact on DNA replication in leaves. In contrast, the 8C and 16C peaks for *atxr5 atxr6* were much broader and skewed to the right, suggesting that many 8C and 16C nuclei have higher DNA contents than the corresponding wild-type nuclei (Figure 2-1a). The distribution of nuclei between endoreduplication levels was similar between wild type and *atxr5 atxr6* double mutants (Figure 2-1b), suggesting that the primary defect in *atxr5 atxr6* plants is in the fidelity of S-phase progression rather than the number of rounds of endoreduplication. This is in contrast to mutations previously reported to affect the level of endoreduplication, but not S-phase fidelity(Caro et al., 2008).

Our previous work has shown that ~65% of *atxr5 atxr6* nuclei show significant decondensation of constitutive heterochromatin (i.e., chromocenters) (Figure 2-1c)(Jacob et al., 2009). Consistent with the flow cytometry results showing that the DNA content phenotype of *atxr5 atxr6* is observed most strongly in 8C and 16C nuclei, we found by microscopic analysis of sorted nuclei that the heterochromatic decondensation defect was also more extreme in these higher ploidy nuclei (Figure 2-1c,d). As a control, we also analyzed the DNA content of *decrease in dna methylation1* (*ddm1)* plants which also show very strong chromocenter

decondensation defects, as well as reduced DNA methylation and massively reactivated

transposons(Fransz et al., 2006; Jeddeloh et al., 1999; Soppe et al., 2002). The flow cytometry

profile from extracted nuclei from *ddm1-2* leaves was similar to that of wild-type plants (Figure

2-1a). These results show that the aberrant flow cytometry profiles observed in *atxr5 atxr6* are

not simply a result of chromatin decondensation defects and/or transposon derepression.

To test the hypothesis that there is indeed extra DNA in *atxr5 atxr6* mutants, we used an

Illumina Genome Analyzer II to sequence genomic DNA from sorted nuclei (2C, 4C, 8C, and

16C) of both wild-type and *atxr5 atxr6* plants. A total of 84.9 million uniquely mapping 36-

nucleotide reads were mapped to the *Arabidopsis thaliana* genome allowing up to two

mismatches. We examined the distribution of genomic DNA across all chromosomes by plotting

the density of reads in non-overlapping 100kb bins. For wild type the ratio of 4C, 8C, and 16C

to 2C sequence reads were uniform across the genome, showing the genome is uniformly

endoreduplicated in wild-type Arabidopsis (Figure 2-1e). Strikingly however, in *atxr5 atxr6*

mutants, we observed an enrichment of reads in the pericentromeric heterochromatin in 4C, 8C

and 16C compared to 2C, indicating that heterochromatin is over-replicated in 4C, 8C and 16C

nuclei compared to 2C nuclei (Figure 2-1f), and that the over-replication is more severe in nuclei

with higher ploidy levels.

A comparison of the distribution of reads from *atxr5 atxr6* mutants with wild type

showed that even in the 2C nuclei, *atxr5 atxr6* mutants show over-replication of pericentromeric

heterochromatin, though to a lower extent than in nuclei of higher ploidy levels (Figure 2-2a).

Relative to wild type, *atxr5 atxr6* mutants showed a 2.9%, 11.0%, 29.1%, 28.4% increase in

reads mapping to pericentromeric heterochromatin in 2C, 4C, 8C, and 16C nuclei respectively

(Figure 2-2b). Sites of over-replication were well correlated at the different ploidy levels. For

instance the Pearson correlation between 8C and 16C *atxr5 atxr6* nuclei was 0.84, indicating that the same sites are over-replicating (Figure 2-2c). These results show that *atxr5 atxr6* mutants exhibit over-replication of pericentromeric heterochromatin in both mitotic and endocycling cells, with progressively stronger defects observed in nuclei with higher ploidy levels.

To examine over-replication in *atxr5 atxr6* mutants at higher resolution, sequence reads were grouped and analyzed in 200 base pair non-overlapping bins. We found that over-replication of pericentromeric heterochromatin is the result of the over-replication of many densely spaced, but distinct loci (Figure 2-3a). We also observed localized over-replication in small regions of the euchromatic arms of chromosomes (Figure 2-3b). These small regions of over-replication were highly enriched in transposons and other repeat elements. Using the BLOC algorithm (see Supplementary Methods)(Pauler et al., 2009), we identified 407 sites of over-replication in the arms of *atxr5 atxr6* chromosomes. The over-replicating regions were relatively small; 94% of regions were smaller than 25kb, with a median size of 10.4kb (Figure 2-4a). The majority (80%) overlapped with previously defined H3K9me2 regions, a mark with is strongly correlated with DNA methylation and gene silencing(Bernatavichute et al., 2008) (Figure 2-4b). Thus, the regions that over-replicate in *atxr5 atxr6* mutants primarily consist of transposons and silent elements of the Arabidopsis genome. Over-replication was confirmed by performing qPCR on defined sites (Figure 2-5). Elements that are transcriptionally reactivated in *atxr5 atxr6* mutants (*TSI, Ta3, CACTA*)(Jacob et al., 2009) were found to be over-replicated, suggesting a positive correlation between transposon reactivation and over-replication.

Re-replication is a well-known mechanism by which DNA is known to over-replicate and results when DNA replication is initiated from an origin multiple times during a single S phase(Arias and Walter, 2007). Presumably because recently replicated chromatin is less

compact, secondary replication forks move faster than primary forks, and collisions of the multiple forks result in successively smaller fragments of DNA reiteratively produced from the origin(Gomez, 2008) (Figure 2-3c). This model predicts that sequences in the center of the origin will be the most highly over-replicated and that over-replication should drop off symmetrically on either side of the origin. To determine if over-replication in the *atxr5 atxr6* mutant is consistent with re-replication, plots of sequencing reads averaged over the over-replicated regions were generated (Figure 2-3d). In contrast to wild type, in which sequencing reads were uniformly distributed, *atxr5 atxr6* mutants showed a bilaterally symmetrical distribution of reads, with the highest density of reads in the center of the over-replicated regions (Figure 2-3d). These results suggest that the extra DNA in *atxr5 atxr6* mutants is a result of repeated replication from defined sites.

We next examined whether chromatin or naked DNA is being re-replicated in *atxr5 atxr6* mutants. To test this, we performed chromatin-immunoprecipitation of unmodified histone H3 followed by Illumina sequencing (ChIP-seq) on wild-type and *atxr5 atxr6* mutants. Compared to wild type, H3 ChIP-seq reads in *atxr5 atxr6* mutants were enriched in the pericentromeric heterochromatin to a similar extent as was the input genomic DNA (Figure 2-3e and Figure 2-4c). This result suggests that chromatin (DNA and associated histones) is re-replicated in *atxr5 atxr6* mutants. Our data also suggest that the re-replicated DNA is properly methylated. If the re-replicated DNA was unmethylated, the % methylation in *atxr5 atxr6* mutants would be predicted to be lower than in wild type. However, we have previously shown that the % DNA methylation in *atxr5 atxr6* mutant leaves (where a significant amount of re-replication was observed) was the same as in wild-type(Jacob et al., 2009), which suggests that the re-replicated DNA is properly methylated.   In addition, to determine whether re-replicated DNA was stably

associated with the chromosomes, we performed qPCR on size fractionated DNA (Figure 2-6). We found that the extra DNA could be detected in the high molecular weight DNA fraction, suggesting that at least part of the re-replicated DNA is stably associated with the chromosome. Being associated with chromosomes, rather than being extrachromosomal fragments, may help explain the stability of re-replicating DNA fragments present in 3-4 week leaf cells.

Because ATXR5 and ATXR6 catalyze H3K27me1(Jacob et al., 2009), we wanted to examine whether the spatial distribution of H3K27me1 overlaps with re-replicating regions. Immunolocalization suggests that H3K27me1 is a heterochromatic mark enriched in chromocenters(Fuchs et al., 2006; Jacob et al., 2009; Lindroth et al., 2004; Mathieu et al., 2005). A detailed global map of H3K27me1, however, has not been reported. We therefore profiled H3K27me1 genome-wide using ChIP-seq. Consistent with the re-replication of pericentromeric heterochromatin in *atxr5 atxr6* (Figure 2-1f), we found that H3K27me1 was strongly enriched in pericentromeric heterochromatin (Figure 2-7a). We also observed H3K27me1 in the coding regions of protein-coding genes and found that the amount of H3K27me1 was anticorrelated with gene expression levels (Figure 2-7b, and Figure 2-8a). Together, these results support a role for H3K27me1 in gene silencing.

To gain additional evidence for a correlation between H3K27me1 and re-replication in *atxr5 atxr6* mutants, we examined the dispersed re-replicating regions in the arms of the chromosomes. We found that H3K27me1 ChIP-seq reads were significantly enriched in these regions compared to randomly selected control regions (permutation test, $P<10^{-6}$) (Figure 2-7c). In addition, plots of the ratio of H3K27me1 to H3 ChIP-seq reads averaged over these re-replicating regions showed strong enrichment of H3K27me1, confirming a positive correlation of H3K27me1 with sites that re-replicate in *atxr5 atxr6* mutants (Figure 2-7d and Figure 2-8b).

Given that H3K27me1 levels correlate with the re-replicated regions of *atxr5 atxr6*

mutants, an interesting question concerns the mechanism by which the spatial distribution of

H3K27me1 is established. ATXR5 and ATXR6 both contain PHD domains, which have been

shown in multiple species to mediate interactions with methylated or unmethylated forms of

histone H3(Musselman and Kutateladze, 2009). We performed *in vitro* binding assays with

various H3 peptides using GST-tagged PHD domains of ATXR5 and ATXR6. The PHD

domains of ATXR5 and ATXR6 bound strongly to an unmethylated peptide corresponding to

amino acids 1-21 of H3 (Figure 2-7e). This binding was unaffected by mono-, di-, or

trimethylation at H3K9, however, binding was strongly reduced by increasing levels of H3K4

methylation. Thus the PHD domains of ATXR5 and ATXR6 bound most strongly to H3

unmethylated at K4 (H3K4me0). Consistent with the hypothesis that binding of the ATXR5 and

ATXR6 PHD domains to H3K4me0 chromatin is helping to guide H3K27 monomethylation

activity, we observed a strong anticorrelation between H3K4 methylation(Zhang et al., 2009) and

H3K27me1 within genes and in the genome at large (Figure 2-7f, Figure 2-8c).

Because loss of ATXR5 and ATXR6 leads to lower levels of H3K27me1(Jacob et al.,

2009), it is possible that depletion of this mark is causing re-replication in *atxr5 atxr6* mutants.

One prediction from this model is that the PHD- and SET domains of ATXR5 and ATXR6

would be essential to prevent re-replication, as they are responsible for binding and methylating

H3, respectively (Figure 2-9a). To investigate this, we first created a genomic construct that

expresses *ATXR6* under its own promoter and confirmed that it can rescue (>95% of T1

transformed plants analyzed) the re-replication phenotype of *atxr5 atxr6* plants (Figure 2-9b).

We then made PIP-, PHD-, and SET-mutant *ATXR6* constructs by inserting point mutations

designed to disrupt the activity of each functional element (Figure 2-9a). Yeast-two-hybrid

analysis and *in vitro* histone-peptide-binding and methyltransferase assays were used to confirm disruption of the PIP motif, PHD-, and SET-domain activities, respectively (Figure 2-10). Analysis of T1 plants transformed with each of the mutated *ATXR6* constructs showed that the re-replication phenotype was never rescued by constructs containing the mutated PIP motif, PHD- or SET domains (Figure 2-9b) (n>20). These results show that the PIP motif, PHD and SET domains are all required for ATXR6 activity and suggest that depletion of H3K27me1 in the *atxr5 atxr6* double mutant is likely responsible for the re-replication phenotype. Consistent with this interpretation, we found that the restoration of H3K27me1 levels also required the wild-type PIP motif, PHD and SET domains (Figure 2-11). Furthermore, only the wild-type construct rescued chromatin decondensation and loss of gene silencing at defect seen in *atxr5 atxr6* mutants (Figure 2-9c,d). These results indicate that the three functional elements of ATXR6 contribute to the prevention of re-replication, chromatin decondensation, and loss of gene silencing.

In summary, our results suggest that *ATXR5* and *ATXR6* are components of a novel pathway required to suppress re-replication in Arabidopsis. Remarkably most of the re-replicating sites in *atxr5 atxr6* mutants correspond to silent heterochromatin, which is composed mostly of transposon sequences. It is tempting to speculate that the ATXR5 ATXR6 system may have evolved to suppress excess DNA replication of transposon sequences that would otherwise result in transposon reactivation. Conversely, transposons are remarkable in requiring both the typical repressive modifications such as H3K9me2 and DNA methylation, as well as the novel ATXR5/6 H3K27me1 pathway for transcriptional suppression.

Sequencing files deposited at GEO (accession codes GSE22411 and GSE21673).

Figure 2-1. Heterochromatic DNA is over-produced in *atxr5 atxr6* mutants. **(a)** Flow

cytometry profiles of Col, *atxr5*, *atxr6*, *atxr5 atxr6*, and *ddm1-2* plants. 3000 gated events are

plotted. The number above each peak (Robust CV) indicates the number of fluorescence

intensity units that enclose the central 68% of nuclei for that endoreduplication level. **(b)**

Quantification of nuclei at each ploidy level for samples in Fig. 1a; Col (black), *atxr5* (white),

atxr6 (grey), *atxr5 atxr6* (crosshatched). **(c)** DAPI staining of sorted nuclei from Col and *atxr5*

atxr6 leaves. Scale bar = 10µm. **(d)** Chromocenter decondensation occurs mainly in 8C and 16C

nuclei. 30 nuclei of each ploidy level from three biological replicates were analyzed. White bars

represent wild-type, and black bars represent *atxr5 atxr6*. **(e)** DNA is replicated uniformly in

wild type nuclei during endoreduplication. The log2 ratios of genomic DNA Illumina reads from

wild-type 4C vs 2C, 8C vs 2C and 16C vs 2C are plotted across the chromosomes in 100kb-

sliding windows. Plots of transposable element (TE) abundance (kilobases of transposon

sequence per 100 kilobase genomic DNA) indicate pericentromeric regions. **(f)** Similar analysis

with *atxr5 atxr6* mutants showing an increased proportion of reads in pericentromeric

heterochromatin in higher ploidy nuclei.

Figure 2-2. Characterization of increased DNA copy number in nuclei of different ploidy

levels in *atxr5 atxr6* nuclei compared to wild-type.

(a) The log2 ratios of genomic DNA Illumina reads from *atxr5 atxr6* compared to wild-type 2C

nuclei are plotted across the chromosomes in 100kb-sliding windows.

(**b**) Quantification of the increase in DNA copy number in *atxr5 atxr6* mutants. Percent of total Illumina reads falling into defined pericentromeric regions in nuclei with different ploidy levels in wild type and *atxr5 atxr6* mutants were computed. The percent change in these values in *atxr5 atxr6* mutants compared to wild type is shown.

(**c**) A similar set of regions show increased DNA content in 8C and 16C *atxr5 atxr6* nuclei. Normalized Illumina reads of sorted nuclei from *atxr5 atxr6* mutants were binned in 3 kb windows and log2 ratios of the scores of 16C to 2C bins and 8C to 2C bins along the chromosome were calculated and compared. The P-value was calculated by shuffling the bins and re-computing Pearson correlation coefficients 100,000 times.

Figure 2-3. Increased heterochromatic DNA in *atxr5 atxr6* mutants is consistent with re-replication of chromatin. (**a**) Genome browser view of a region of pericentromeric heterochromatin. Pericentromeric heterochromatin contains densely spaced, ~10kb over-replicating sites. Data are represented as log2 ratios (16C/2C, 8C/2C or 4C/2C) in 200bp bins. H3K9me2 microarray data(Bernatavichute et al., 2008), TAIR8 protein-coding gene (PCG) and transposable element (TE) tracks are also shown on the plus (+) or minus (-) strand of the genome. (**b**) Genome browser view of examples of over-replication in the arms of chromosomes. Three over-replicating regions are shown. (**c**) Model for DNA re-replication (adapted from(Davidson et al., 2006a)). (**d**) Distribution of Illumina reads in re-replicating regions. Plots of the average number of sequence reads +/-5 kilobases relative to the center of over-replicating regions in *atxr5 atxr6* mutants, wild type, or the *atxr5 atxr6* mutants/wild type log2 ratio (plotted in 100bp bins). (**e**) Histone content in re-replicating regions is higher in *atxr5*

atxr6 mutants. Log2 ratios of H3 ChIP-seq reads and input genomic DNA reads in *atxr5 atxr6*

mutants relative to wild-type, plotted over chromosome 3 in 100kb-sliding windows.

Figure 2-4. Heterochromatin is over-replicating in *atxr5 atxr6* mutants

(**a**) Distribution of the sizes of re-replicating regions in the arms of *atxr5 atxr6* mutants. (**b**) The

percentage of high copy number regions in the arms of *atxr5 atxr6* mutants that overlap with

H3K9me2 regions is shown.

 (**c**) Chromosomal views of log2 ratios of H3 ChIP-seq reads to INPUT genomic DNA reads in

atxr5 atxr6 mutants relative to wild type over chromosomes 1,2,4 and 5 are plotted in 100kb

sliding windows (see main text for chromosome 3).

Figure 2-5. Confirmation of re-replicating sites by quantitative PCR

Six re-replicating sites in pericentromeric heterochromatin (*Ta3*, *TSI*, *CACTA*) and in the arms of

chromosomes 1 and 2 (*AT1G44510*, *AT2G04160*, *AT2G16670*) were analyzed by qPCR on

genomic DNA extracted from 16N nuclei of Col (white bars) and *atxr5 atxr6* (black bars).

At1g51800, a non-re-replicating site in the arm of chromosome 1, was used as control. The

averages and standard deviations from three independent experiments are shown. Primers for

Ta3, *TSI*, *CACTA* are previously described(Jacob et al., 2009) and all other primers are listed in

Table 2-1.

Figure 2-6. Size separation of DNA and analysis of re-replicating sites by quantitative PCR

(a) Schematic representation of the size separation of Col and *atxr5 atxr6* DNA.

(b) Three non re-replicating sites (*AT1G31440, AT3G24320* and *AT2G40000*) and four re-

replicating sites (*CACTA, AT1G31355, AT1G34080* and *AT2G29210*) were analyzed by qPCR

on total, high molecular weight and low molecular weight genomic DNA extracted from Col

(white bars) and *atxr5 atxr6* (black bars). *AT1G31440*, a non-re-replicating site in the arm of

chromosome 1, was used as control. The averages and standard deviations from two independent

experiments are shown. All primers are listed in Table 2-1.

Figure 2-7. Genome-wide mapping of H3K27me1 and anticorrelation with H3K4

methylation. **(a)** H3K27me1 is enriched in heterochromatin. The log2 ratios of H3K27me1

reads to H3 ChIP-seq reads in wild-type are plotted across the chromosomes (1 to 5) in 100kb-

sliding windows. **(b)** H3K27me1 is anticorrelated with gene expression level. H3K27me1

ChIP-seq reads normalized to H3 ChIP-seq reads averaged over TAIR8 protein-coding genes.

The bodies of genes are scaled. Three-week-old wild type plants were used for both ChIP-seq

and RNA-seq. **(c)** H3K27me1 is significantly enriched at sites of re-replication in the arms.

Reads per base pair in re-replicating regions were calculated for both H3K27me1 and H3 ChIP-

seq reads, and the ratio was calculated (black bar). Random regions with similar a distribution as

re-replicating regions were generated 100,000 times and the same calculation was performed.

The mean value obtained from random regions are shown (white bar), and the error bars

represent the standard deviation. **(d)** H3K27me1 is enriched in over re-replicating regions. The

log2 ratio of H3K27me1 to H3 reads is plotted +/-20 kilobases relative to the center of re-

replicating regions *atxr5 atxr6* mutants. Data were plotted in 400bp bins and smoothed by taking

the moving average over 6 bins. **(e)** Pull-down assay using purified GST-tagged PHD domains

of ATXR5 and ATXR6 and biotinylated H3 peptides with different methylated lysines.

Interaction between the peptides and the GST-PHD domains was visualized by Western blot

using a GST antibody. **(f)** Analysis of the relationship between H3K27me1 and H3K4 methylation. The log2 ratio of H3K27me1 to H3 is plotted over the boundaries of all H3K4me0/-me1/-me2/-me3 regions in the genome. Data is graphed in 200 base pair bins, and smoothed by taking the moving average over +/-2 bins. The scale for the plots over H3K4me0 is in blue, and the scale for the others is in black.

Figure 2-8. Genome-wide profiling of H3K27me1 reveals that H3K27me1 is a silencing mark that correlates with sites of re-replication in *atxr5 atxr6* mutants, and anticorrelates with H3K4 methylation.

(**a**) H3K27me1 ChIP-seq normalized to H3 ChIP-seq reads plotted over either the 5' end (left) or 3' end (right) of TAIR8 protein-coding genes. Two kilobases upstream or downstream of genes, and 1 kilobase into the genes are shown. Black=all genes, red=genes with the top 10% expression, pink=top10~30% expression, yellow=top30~50% expression, green=50~70% expression, cyan=70~90% expression, blue= lowest 10% expression.

(**b**) Genome browser view of examples of re-replicating regions that are enriched in H3K27me1.

(**c**) H3K27me1 anticorrelates with H3K4 methylation. Each protein-coding gene in the genome was split into regions containing either H3K4me0, -me1, -me2 or -me3, and the density of H3K27me1 and H3 ChIP-seq reads in each of these regions was computed. The % of genes where the ChIP-seq read densities were 2-fold greater in the H3K4 unmethylated regions relative to H3K4 methylated regions were calculated.

Figure 2-9. Functional PHD- and SET domains, and the PIP motif are required for the regulation of DNA replication by *ATXR6*. (a) Structure of ATXR6. The domains (below) and

point mutations (above) made to generate *ATXR6* mutants are represented. **(b, c, d)** Normal

DNA replication as indicated by Robust CV **(b),** chromatin condensation **(c)** and TSI gene

silencing **(d)** are rescued in transgenic *atxr5 atxr6* plants expressing wild-type *ATXR6*, but not

PHD-, SET-, or PIP mutants. All phenotypes were scored on the same four representative

transgenic lines (n>20) generated from each construct.

Figure 2-10. Validation of effect of the point mutations inserted to ATXR6 constructs.

(**a**) Binding of ATXR6 to AtPCNA1 is compromised by mutations in the PIP motif. Interaction

between AtPCNA1 and WT ATXR6 or an ATXR6 mutant (Q92A, I95A, F98A, F99A) in the

PIP motif (ATXR6(pip)) was assessed using the yeast-two-hybrid system (Invitrogen, Carlsbad,

CA). As controls for self-activation, AtPCNA1, ATXR6, and ATXR6(pip) were co-expressed

with empty vectors.

(**b**) Replacement of leucine 49 with tryptophan in the PHD domain of ATXR6 prevents binding

to H3. *In vitro* histone peptide binding assay using purified GST-tagged PHD domains ATXR6

(WT and mutant) and biotinylated H3 peptides with different methylated lysines. Interaction

between the peptides and the GST-PHD domains was visualized by Western blot using a GST

antibody.

(**c**) A point mutation (Y243N) in the SET domain of ATXR6 impairs the methyltransferase

activity. *In vitro* histone methyltransferase assay using an H3K27me1 antibody to detect the

monomethylation of H3.

Figure 2-11. The PHD- and SET domains, and the PIP motif of *ATXR6* are required to restore normal H3K27me1 levels. Immunolocalization of H3K27me1 in nuclei isolated from mature leaves of Col or *atxr5 atxr6* mutants. T1 lines (+ ATXR6) are expressing either wild-type *ATXR6*, or different point mutants in each of the three functional elements (PIP motif, PHD- and SET domains) of the protein.

Table 2-1. Primers used for qPCR

Figure 2-1

Figure 2-2

a

b

% increase in reads falling into pericentromeric heterochromatin
in *atxr5 atxr6* mutants relative to WT

c
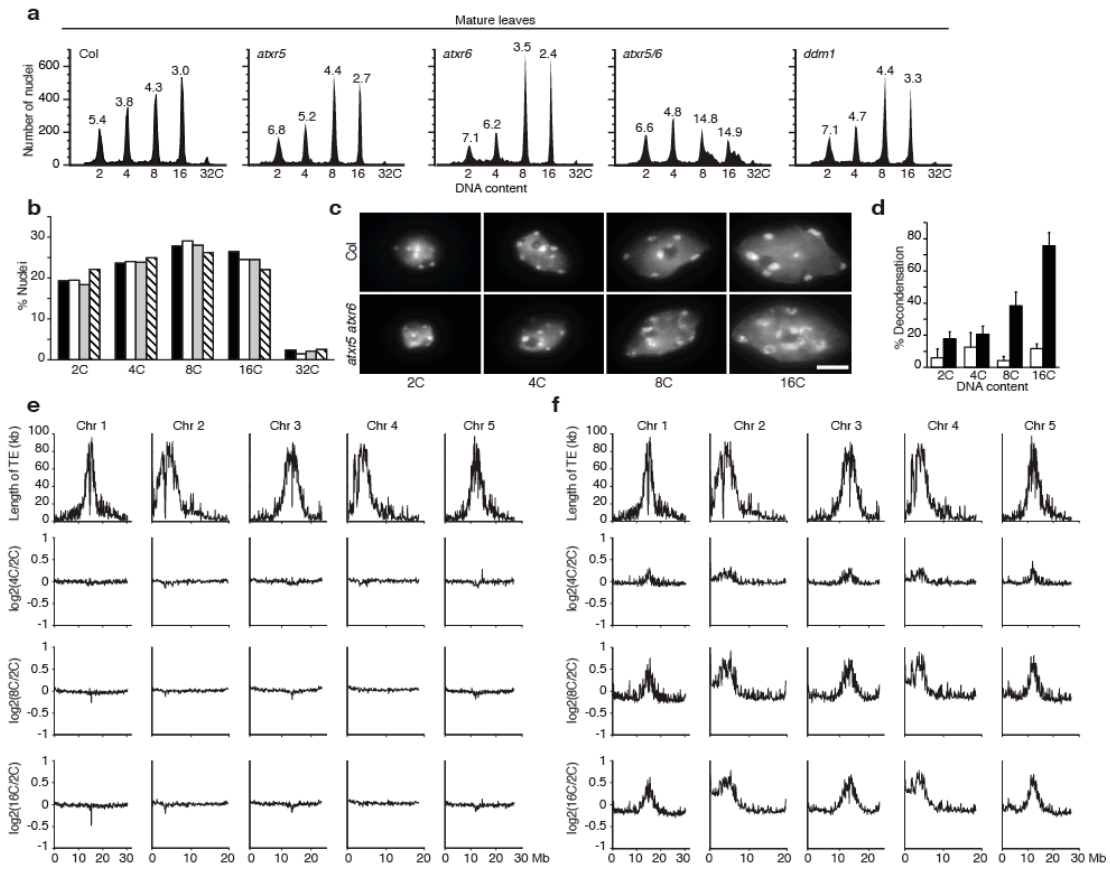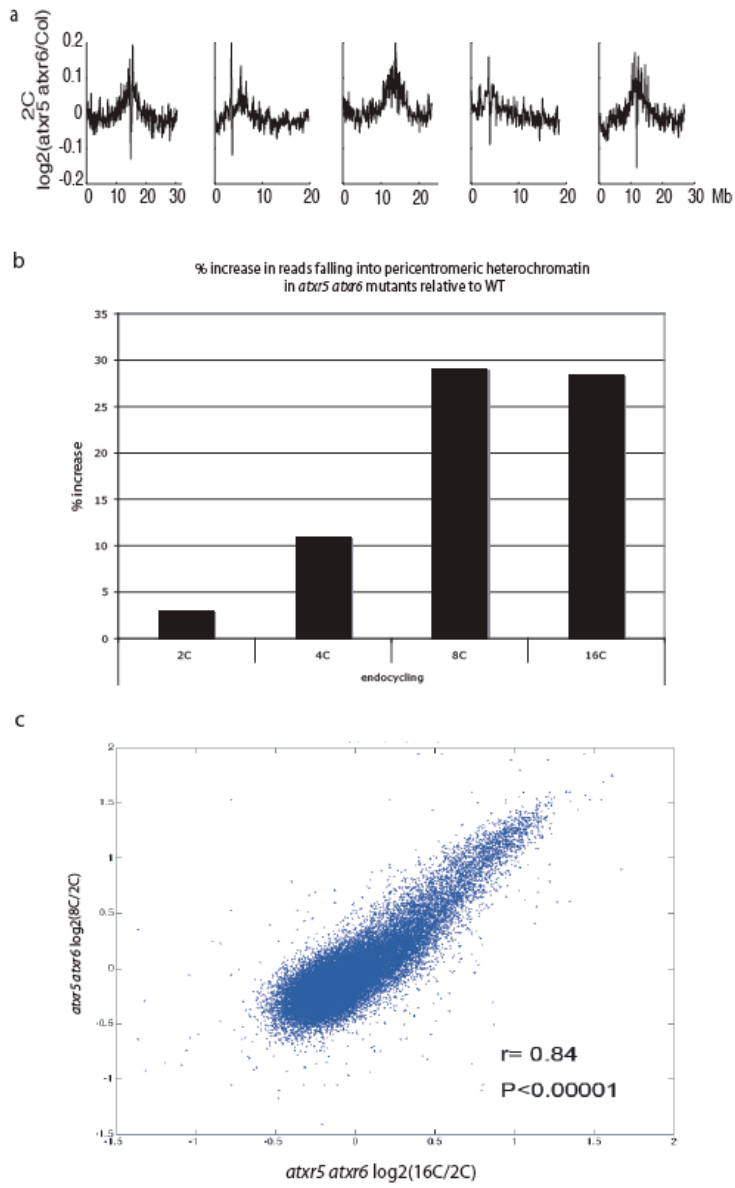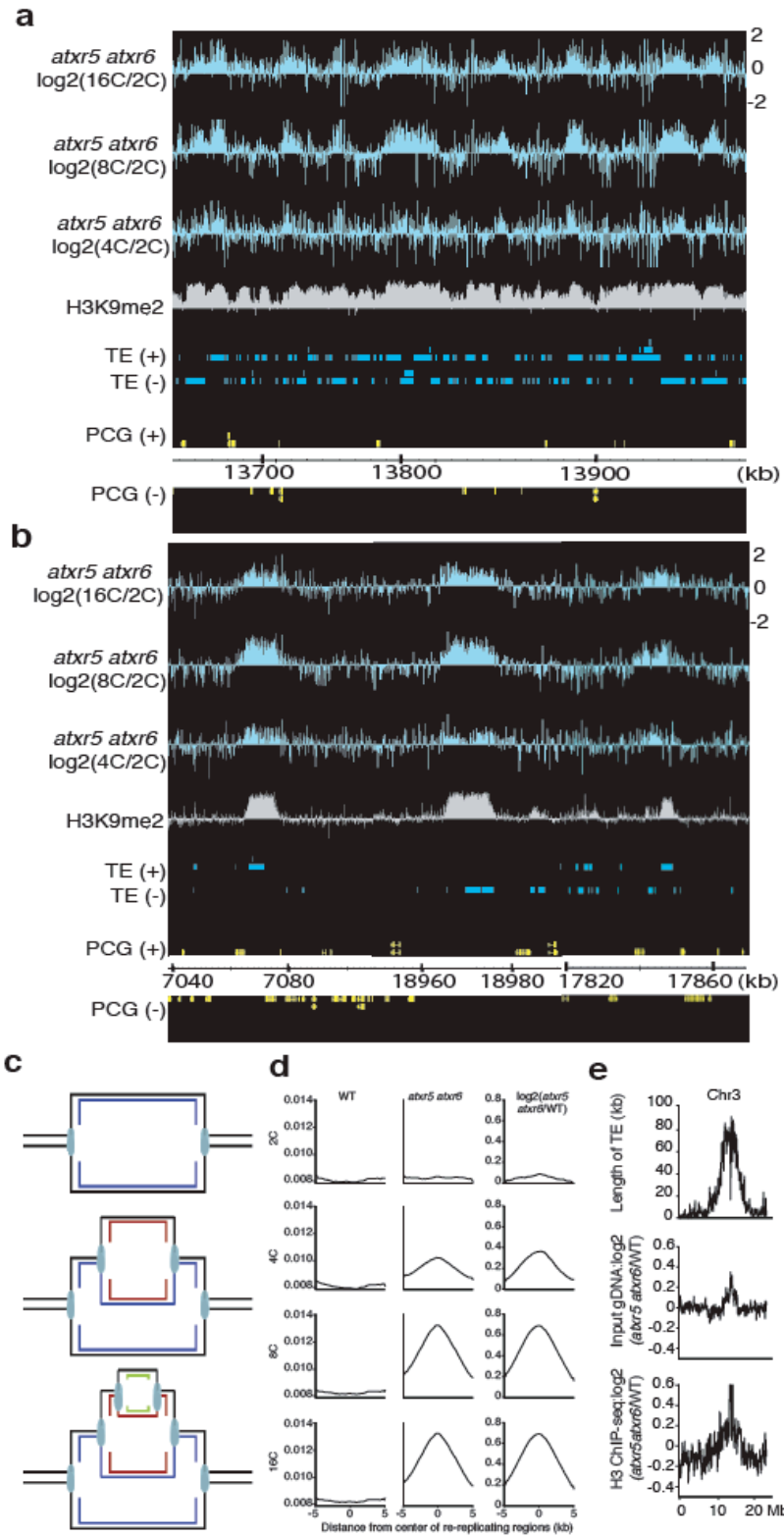
Figure 2-3
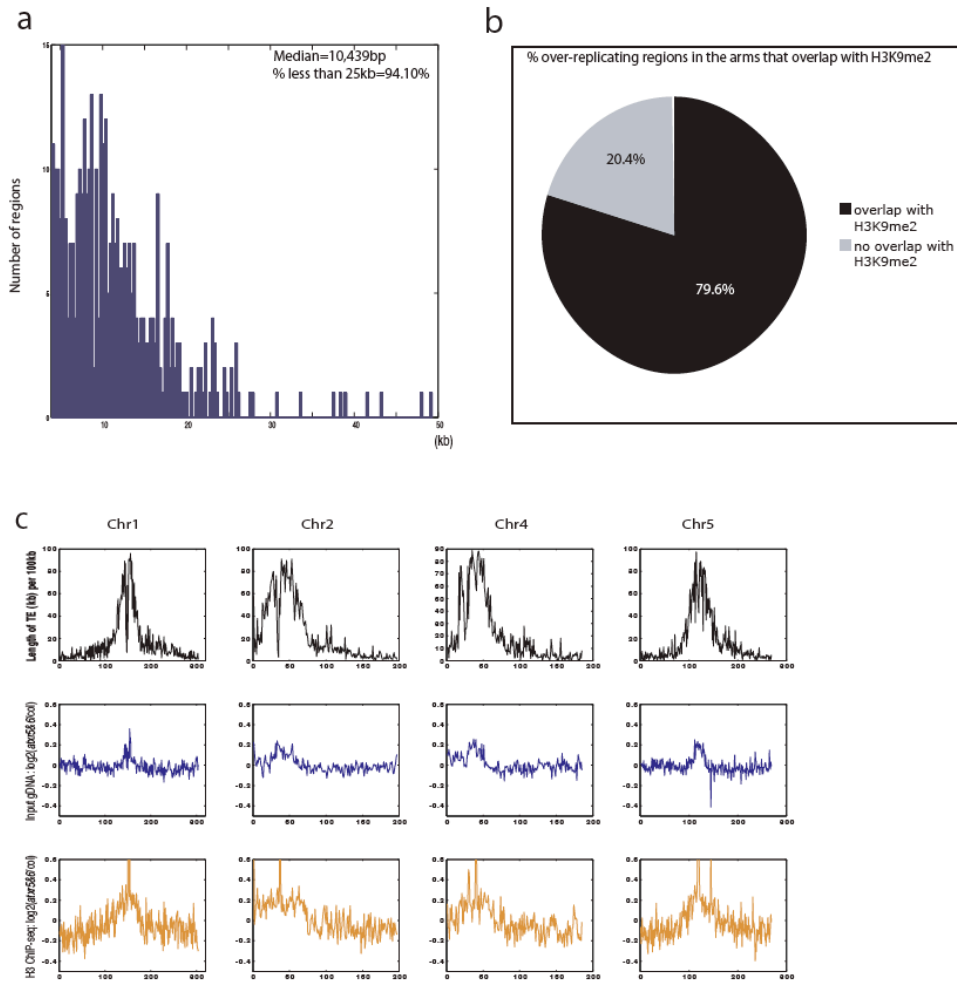
Figure 2-4

Figure 2-4

Figure 2-5
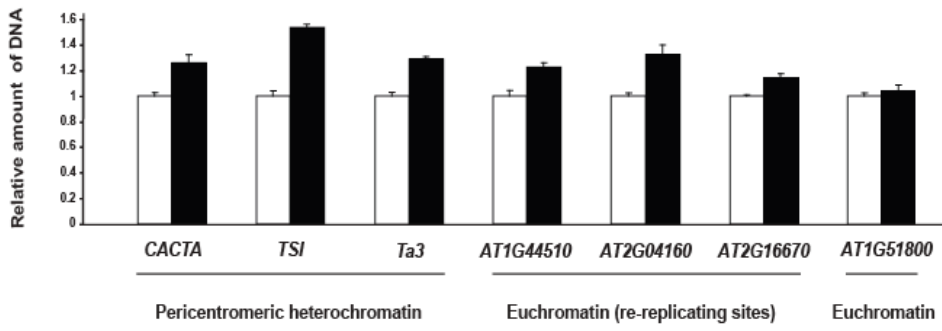
Figure 2-6
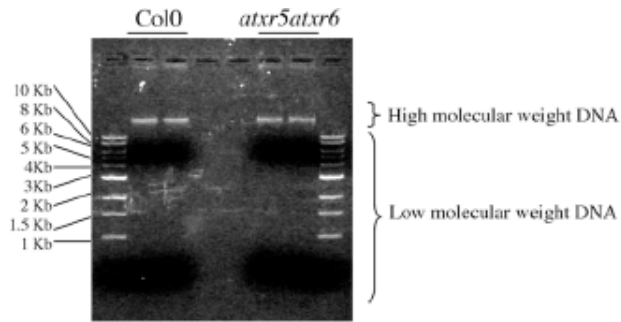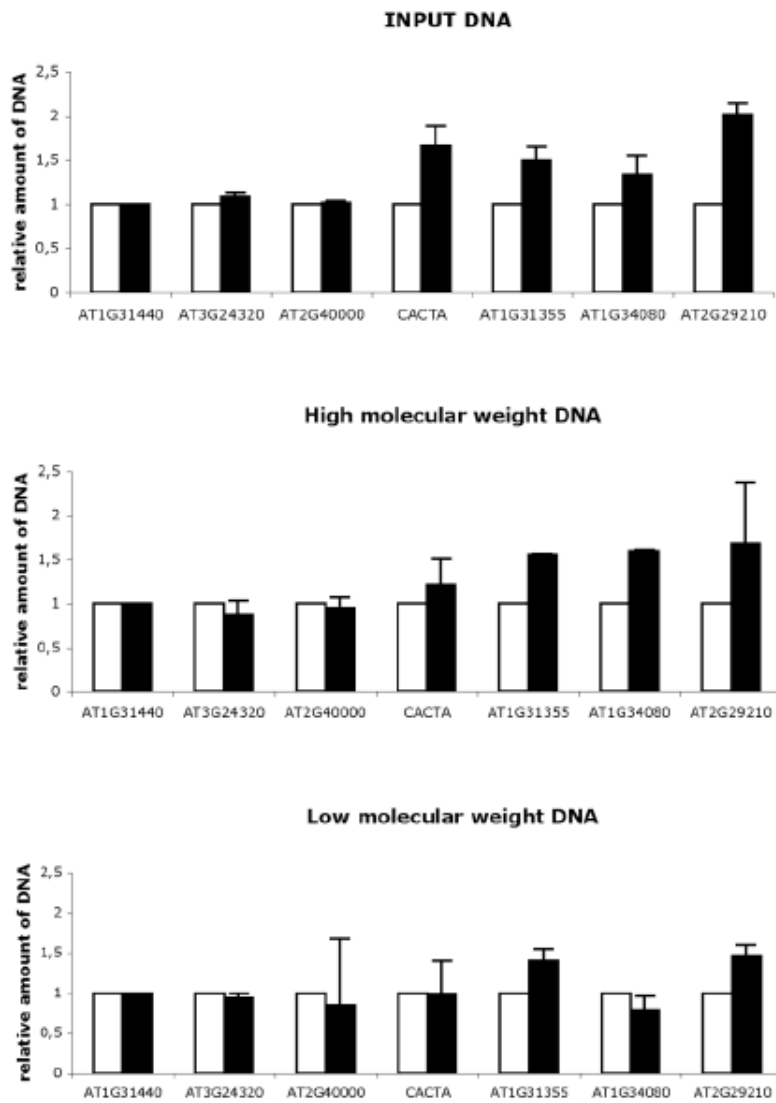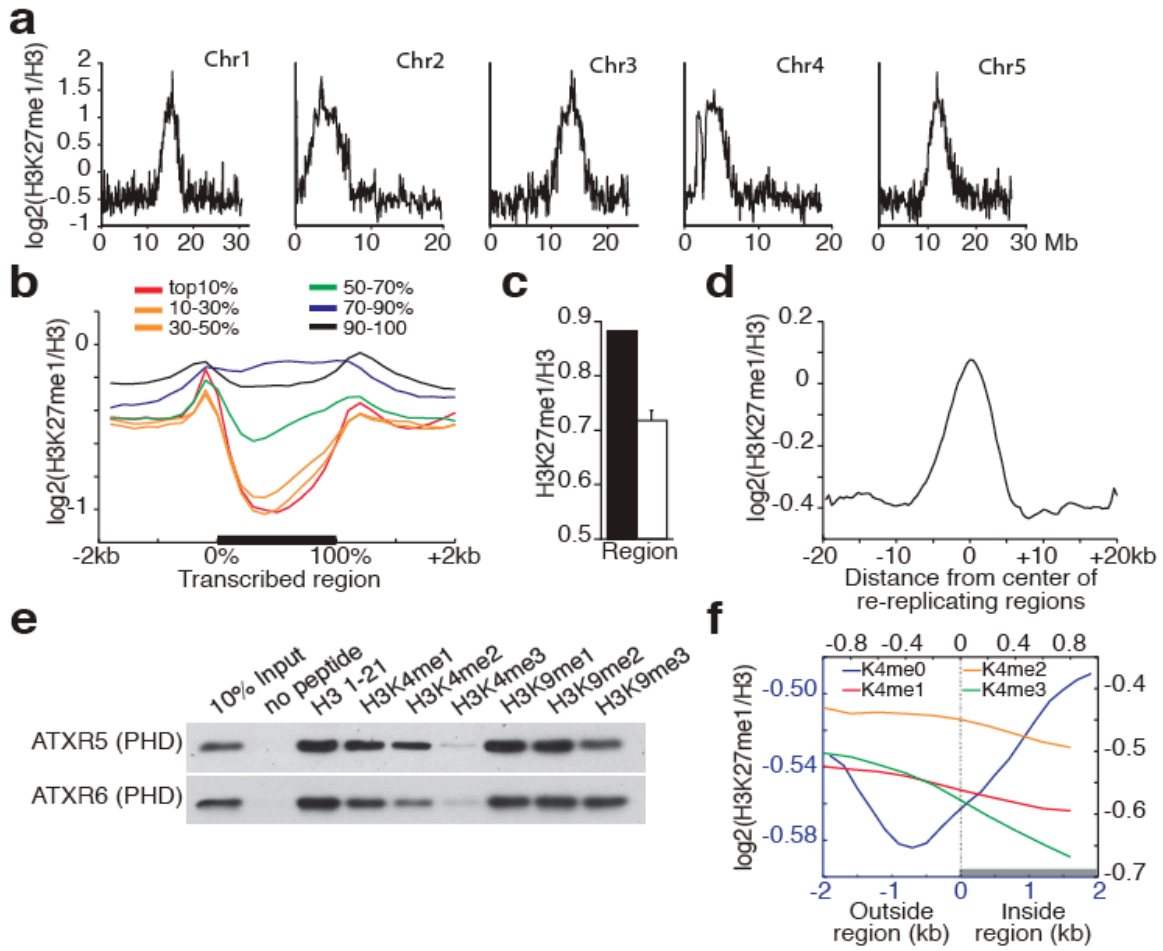
Figure 2-7

Figure 2-8

Figure 2-9

Figure 2-10

a

	-L -T	-L -T -H
pDEST22-AtPCNA1 + pDEST32-ATXR6		
pDEST22-AtPCNA1 + pDEST32-ATXR6(pip)		
pDEST22-AtPCNA1 + pDEST32		
pDEST22 + pDEST32-ATXR6		
pDEST22 + pDEST32-ATXR6(pip)		

b

Lanes: 20% Input, no peptide, H3 1-21, H3K4me1, H3K4me2, H3K4me3, H3K9me1, H3K9me2, H3K9me3

ATXR6

ATXR6-L49W

c

Lanes: no enzyme, ATXR6, ATXR6 Y243N

α-H3K27me1
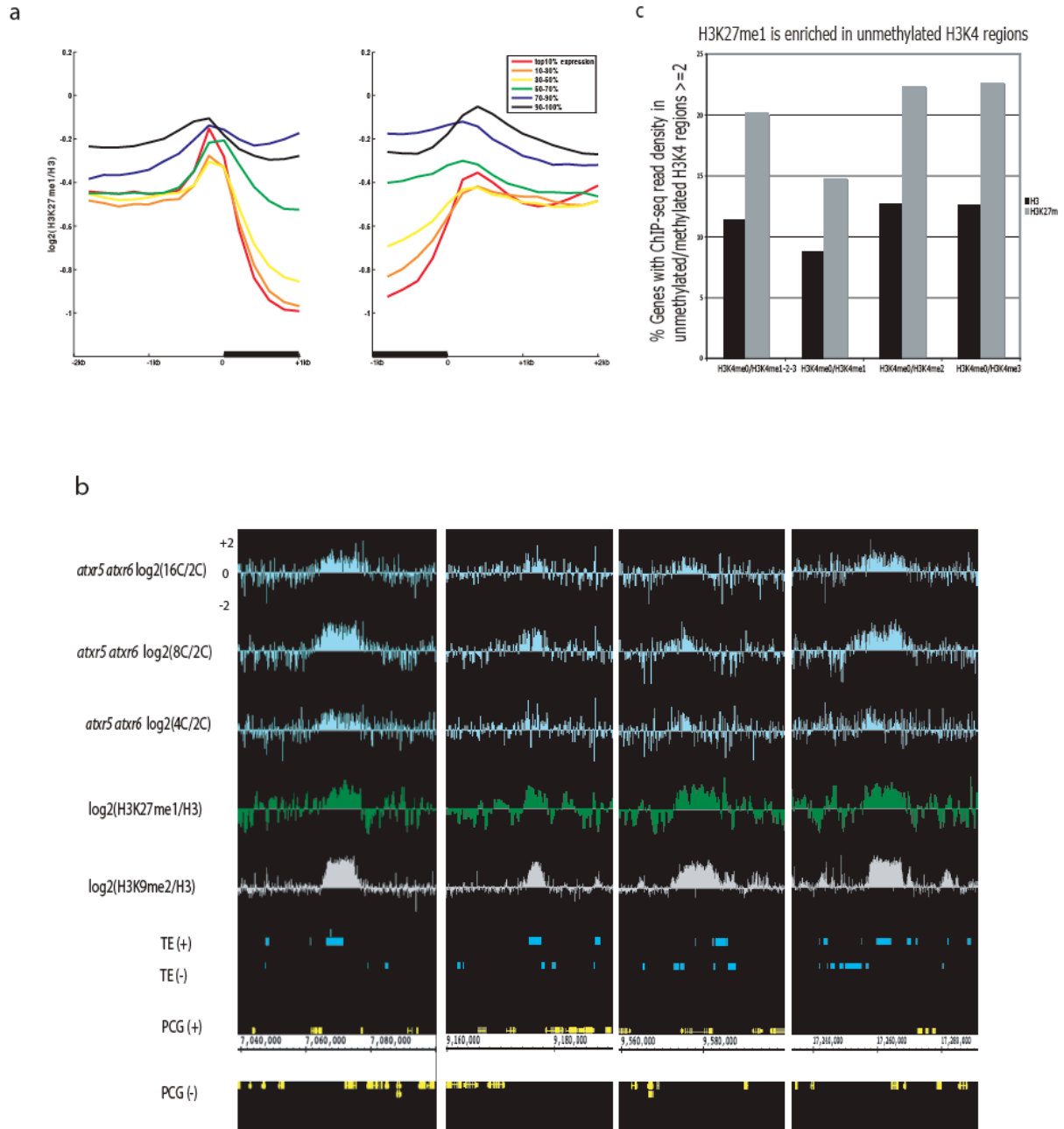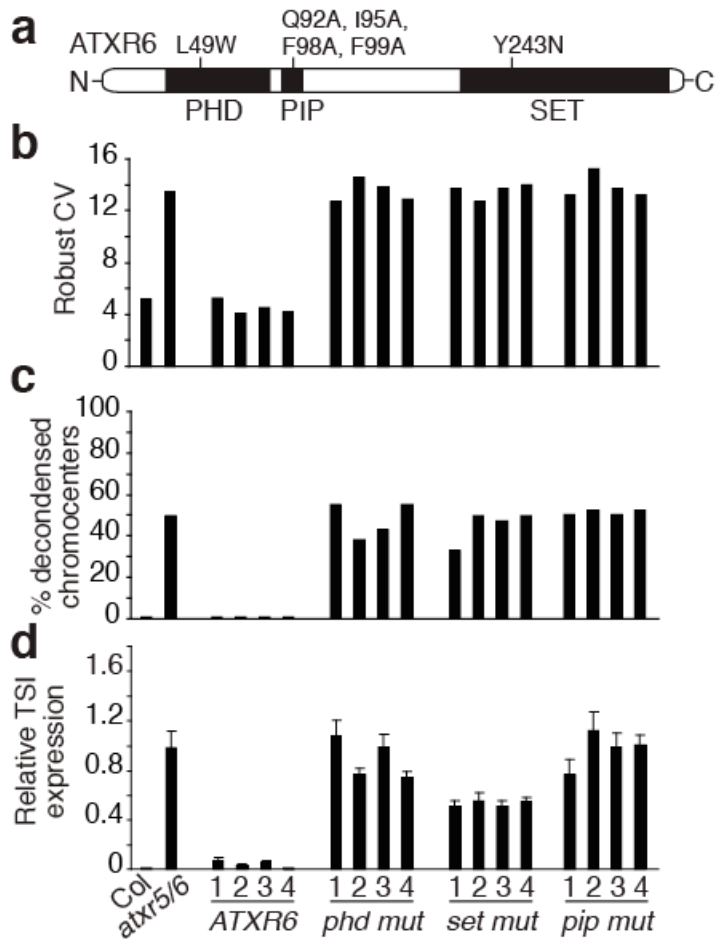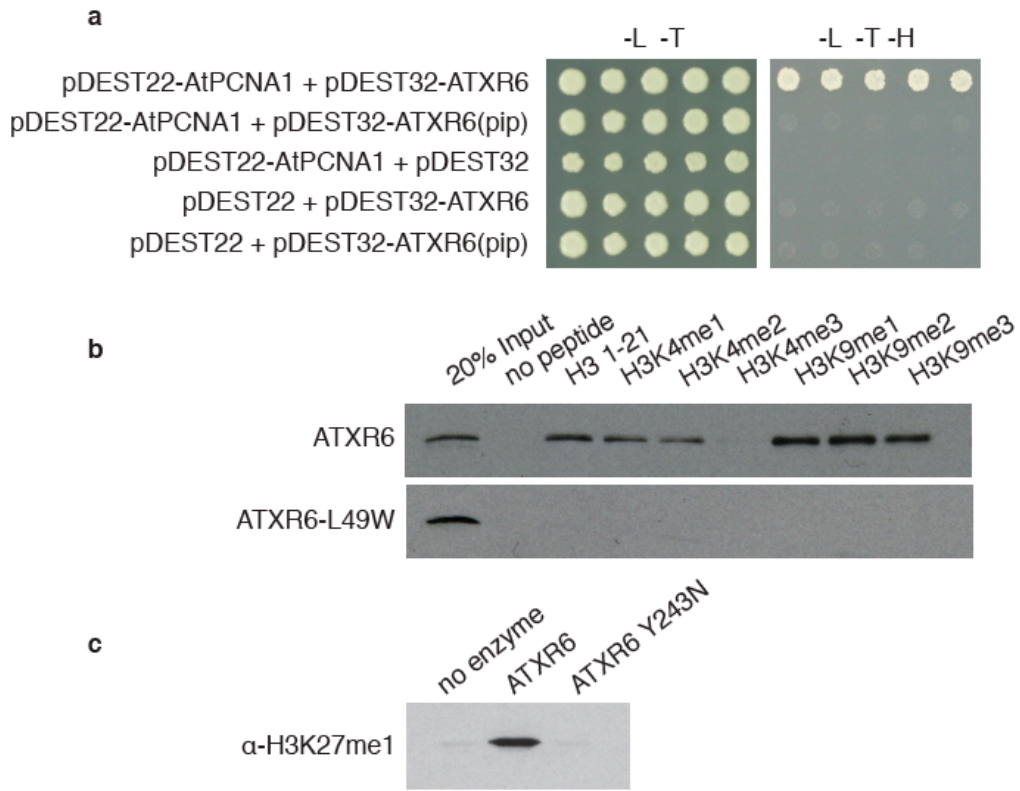
Figure 2-11
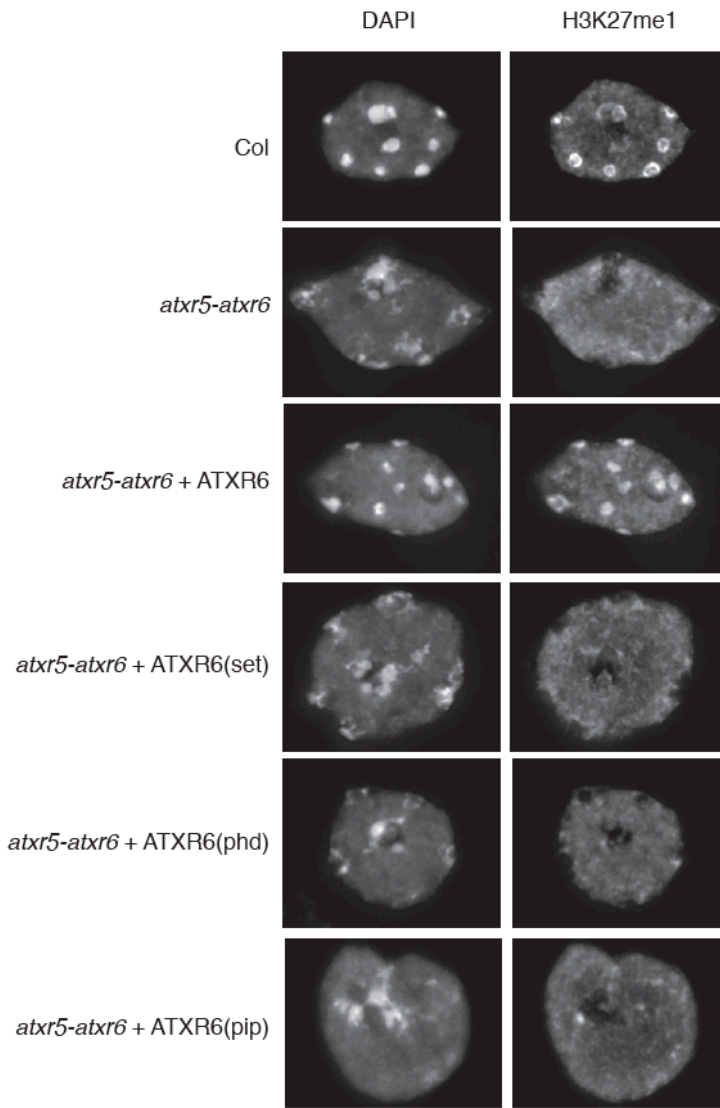
Table 2-1

gene	Primers	Sequence
AT1G44510	2709-CHR1-2-F1	GCCTAGAGGTATAAAGCCGGTTG
	2710-CHR1-2-R1	CCACACCGGTCCTCTGTCTTCAG
AT2G04160	2713-CHR2-1-F1	GTGAGCTGTGTGGCTGGAAGGAC
	2714-CHR2-1-R1	CTTGGATCGCTTGACCCTATAAAG
AT2G16670	2721-CHR2-3-F1	GGCGTCGGAACAACTTCAGTGTC
	2722-CHR2-3-R1	GGGCTCCAATAGTTGATGGGC
At1g51800	2436-1g51800-F2	ACTTTACCTCCTCTGCTCAACGC
	2437-1g51800-R2	CAGTCAATCTTGTTCACGCCG
AT1G31440	JP7583	CCTGTACCAGCCACCTGAAT
	JP7584	GGGAACTCAGTCTCGCTGTC
AT3G24320	JP7220	CAAGCATGGGAAAATAATGC
	JP7221	TTGTTGCAGATACCATGAAGC
AT2G40000	JP7429	GCCGACCAAACGAGATCTATAG
	JP7430	CTTCTCCAGCAAGTAAACATCG
CACTA	JP8158	TGTGTGGAAGGGTCTTGTGGACTT
	JP8159	AACTTACATGTTTGCGGGCACGAG
AT1G31355	JP7581	GGGAAGTGCAGACAGACCTT
	JP7582	AGGAGGCAATTGTGTGAAGC
AT1G34080	JP6479	CAAGGCGAAGTCCTGCTAAC
	JP6480	TCAAGTTCTCGCGGATTTTC
AT2G29210	JP6485	GGTCGCGGTCACCTATTAGA
	JP6486	GAGAACGGTGACGCCTAGC

CHAPTER 3

Genome-wide mapping of Arabidopsis origins of DNA replication and their

associated epigenetic marks

ABSTRACT

Genomic integrity requires faithful chromosome duplication. Origins of replication, the genomic sites where DNA replication initiate, are scattered throughout the genome Their mapping at a genomic scale in multicellular organisms has been challenging. Here we have profiled origins in *Arabidopsis* by high-throughput sequencing of newly-synthesized DNA and identified ~1500 putative origins genome-wide. This was supported by ChIP-chip experiments to identify ORC1 and CDC6 binding sites. Origin activity was validated independently by measuring the abundance of nascent DNA strands. The midpoints of most Arabidopsis origin regions are preferentially located within the 5' half of genes, slightly enriched in G+C, histone H2A.Z, H3K4me2/3 and H4K5ac, and depleted of H3K4me1 and H3K9me2. Our data establish the basis for understanding the epigenetic specification of DNA replication origins in Arabidopsis and have implications for other eukaryotes.

Faithful duplication of the genetic material is crucial to maintain genomic integrity. DNA replication in eukaryotic cells initiates at multiple sites, known as replication origins, which are scattered throughout the genome (Aladjem, 2007; DePamphilis et al., 2006; Huberman and Riggs, 1968). The number of origins ranges from hundreds to thousands depending on the cell type and/or the physiological state (Aladjem, 2007). One of the key steps for understanding replication origin function is whether and how they are specified in the genome. In *S. cerevisiae*, a strict sequence-dependent specification occurs whereby the origin recognition complex (ORC) recognizes an 11bp sequence to define the site of each active replication origin (Bell and Stillman, 1992a; Wyrick et al., 2001). This mechanism appears to be a rather unique situation because a consensus sequence has not been found in other organisms. For example, in *S. pombe*, although origins are associated with A+T-rich stretches they are not specified by a known consensus DNA sequence (Antequera, 2004b; Hayashi et al., 2007).

The identification of the molecular nature of replication origins in multicellular organisms has been elusive and only a handful of them have been analyzed (Costa and Blow, 2007; DePamphilis et al., 2006; Gilbert, 2004; Schepers and Papior, 2010). The large genome size of multicellular eukaryotes, their different developmental strategies and the existence of a diversity of proliferating cell populations have led to increased difficulty in determining origin specification, function and spatio-temporal regulation at a genomic scale (Aladjem, 2007). Local epigenetic modifications can further affect origin selection and usage, e.g. replication timing (Aladjem, 2007; Danis et al., 2004; Lee et al., 2010; MacAlpine et al., 2004). Although attempts to obtain genome-wide maps of replication origins in mammalian cells have been reported (Cadoret et al., 2008; Karnani et al., 2010; Lucas et al., 2007; Sequeira-Mendes et al., 2009), the

molecular features defining replication origins in higher eukaryotes and, in particular, their links to epigenetic modifications still remain largely unknown.

In this study, we have identified replication origins, analyzed their organization, and defined their epigenetic signatures at a high-resolution genome-wide scale in the plant *Arabidopsis thaliana*. Its rather compact genome (~125Mb, ~28,000 protein coding genes), fully sequenced and annotated, and the relatively small amount of repetitive sequences (~17%), largely confined to the pericentromeric areas (Initiative, 2000), make Arabidopsis an excellent system to study origins. Furthermore, the comparison of replication origin features in organisms with very different developmental and growth strategies could shed light into the basic principles governing origin specification and function in eukaryotes. In addition, genome-wide maps of epigenetic marks such as DNA methylation and several histone modifications have already been reported (Bernatavichute et al., 2008; Cokus et al., 2008; Zhang et al., 2009). The use of massive sequencing of short-pulse BrdU-labeled DNA led us to identify ~1500 putative replication origins across the Arabidopsis genome. ORC1 and CDC6 binding regions, which, importantly, are enriched in BrdU-labeled regions, were also identified by chromatin immunoprecipitation and microarray experiments (ChIP-chip). Furthermore, origin activity was validated independently by measurement of nascent DNA strand abundance. Our studies reinforce the idea that some origin features are shared with animal cells whereas others are unique to plants (Fuchs et al., 2006; Sanchez et al., 2008). The Arabidopsis "originome" reported here provides the basis of identifying the key features of eukaryotic replication origins and delineate their possible regulatory mechanisms.

RESULTS

Genome-wide mapping of Arabidopsis DNA replication origins

Functional origins mark the sites where the synthesis of nascent DNA strands occurs. Thus, our strategy was to sequence purified DNA labeled in vivo with a pulse of BrdU and confirm these data with the mapping of pre-RC binding (Figure 3-1). To obtain sufficient amounts of BrdU-labeled DNA, we used Arabidopsis cultures that contain a substantial amount of proliferating cells. We synchronized cells in G0 using sucrose deprivation and labeled them with BrdU a few hours after release from the block when cells are just entering the S-phase (Menges and Murray, 2002, 2006) (Figure 3-1). DNA was extracted, fractionated by CsCl gradient centrifugation and the BrdU-labeled material was purified and used to generate genomic libraries for sequencing using the Solexa (Illumina) technology. We obtained a total of ~4 million high quality reads that uniquely mapped to the Arabidopsis genome. Likewise, a sample of unlabeled DNA was processed as a control (see Methods). This BrdU-seq method rendered a comprehensive list of genomic locations with a significant enrichment in BrdU-labeled DNA strands (Figure 3-2a). To define origin regions using the BrdU-labeled DNA sequencing data we merged BrdU positive regions separated <10kb, as described in Methods (see also Figure 3-3). An alignment of DNA sequences of ±100bp around the midpoint of BrdU-labeled regions did not render any consensus sequence. To corroborate the analysis of BrdU-labeled regions and deal with possible experimental variations, we carried out an independent assay of cell synchronization, BrdU-labeling and CsCl purification followed by massively parallel sequencing. Significantly, 78.2% (p<1.0e-6) of the BrdU-labeled regions overlapped with the regions defined in the previous experiment, supporting the reproducibility of the two independent experiments.

To identify pre-RC binding sites, in the absence of specific antibodies, we used plants

expressing constitutively tagged versions of two pre-RC components, ORC1 (ref. 26) and CDC6

(ref. 27). ORC1- and CDC6-bound DNA fragments were purified by chromatin

immunoprecipitation (ChIP) (Figure 3-1) and hybridized to whole-genome Arabidopsis tiling

arrays to identify their genome-wide binding sites (Figure 3-2a). ORC1 binding was spread over

numerous sites (Figure 3-4) whereas CDC6 binding sites were less abundant (Figure 3-5). First,

we determined the fraction of the BrdU-labeled regions that contained bound pre-RC

components. We found that ~76.7% and 17.0% of BrdU-labeled regions overlapped with ORC1

and CDC6 regions, respectively (midpoint of BrdU region ±2.1 kb, p<0.001; see colocalization

range in Figure 3-2b). More importantly, the midpoints of these regions significantly colocalized

with both ORC1 and CDC6 binding sites within ±2kb regions (Figure 3-2b). Therefore, the 1543

regions rendered by our approach were considered *bona fide* replication origins. They appear

uniformly distributed across the genome, although it is possible to identify clusters of more

closely spaced origins in some genomic locations (Figure 3-6). The number of origins varies for

different chromosomes but they roughly correlate with chromosome size (Figure 3-2c). The

distribution of distances between origin region midpoints gave a median of 51.1 kb, with a mean

of 77.2 kb (Figure 3-2d).

Assessing origin activity by nascent strand abundance

The BrdU-labeled regions identified in our study and the marked colocalization with ORC1-

and CDC6-binding sites strongly support the notion that they represent active DNA replication

origins. To assess origin activity directly we measured the relative abundance of nascent DNA

strands of various putative origin regions relative to adjacent regions in a sample of short DNA

molecules purified by sucrose gradient centrifugation and containing a RNA primer at their 5'
end (Gomez and Antequera, 2008; Prioleau et al., 2003). Thus, origin activity was determined
by real-time PCR methods using primer pairs spanning 5-16kb around putative origin regions. In
all cases analyzed, we could demonstrate a high enrichment of origin sequences in the short
nascent DNA strand sample (Figure 3-7a-c). Importantly, one of the BrdU-labeled regions
included in this analysis was one showing a relatively low CDC6 signal in the ChIP-chip
experiment (Figure 3-7a). In spite of this, it showed a high abundance of nascent DNA strands
measured by qPCR, demonstrating the activity of this region as a functional origin as well as the
robustness of our approach. A control region, lacking BrdU-labeled DNA sequences, did not
show any appreciable enrichment (Figure 3-7d). These data together led us to conclude that the
set of origins identified here provides a solid starting point to define their molecular landscape.

Genomic location of Arabidopsis DNA replication origins

To test whether origins are randomly distributed along the genome or show a preferential
location we estimated origin location relative to various genomic elements. We found that
77.7% and 10.2% of origins colocalized with gene units and transposons, respectively. These
percentages are significantly different from the proportion of the Arabidopsis genome
represented by these elements (Figure 3-8a). Next, we analyzed origin density across genes and
their 5' and 3' upstream regions. We observed that most origins were identified within gene
bodies (Figure 3-8b), but preferentially towards their 5' ends (Figure 3-9). Origin localization to
the bodies of genes did not correlate with gene expression levels (Figure 3-8b), according to
expression data obtained from cell suspensions at the same synchornization time used for BrdU
labeling (Menges et al., 2003). However, highly expressed genes, compared to lowly expressed

genes, tended to have more origins in regions immediately upstream (Wilcoxon ranksum test, p<0.005) or downstream (Wilcoxon ranksum test, P<0.01) of genes (Figure 3-8b).

The body of highly expressed genes in Arabidopsis is enriched in CG methylation whereas the three types of C methylation (CG, CHG and CHH, where H is A, T or C) are highly enriched in the repeat-rich pericentromeric regions of the Arabidopsis genome (Cokus et al., 2008; Zhang et al., 2006). Interestingly, we found a slight decrease in CG methylation levels around origin midpoints compared to regions flanking them (Figure 3-10a). Furthermore, we observed that regions ±0.1kb around the origin midpoints showed higher G+C contents (44.5%), compared to the whole Arabidopsis genome (Figure 3-10b). It is known that the histone variant H2A.Z is preferentially deposited near the 5' end of target genes and anticorrelates with CG methylation (Zilberman et al., 2008). We found a strong correlation between the presence of H2A.Z within ±1kb and the origin midpoints (Figure 3-10c).

The epigenomic landscape of Arabidopsis DNA replication origins

To further determine features defining Arabidopsis replication origins we next sought to profile the landscape of epigenetic histone marks that appear to associate with replication origins. Arabidopsis epigenomics data are already available for dimethylation of histone H3 at lysine 9 (H3K9me2) and for the three methylated forms of H3K4 (Bernatavichute et al., 2008; Zhang et al., 2009). We found that most origins tend to be depleted of H3K4me1 (Figure 3-11a) but are highly enriched in H3K4me2 and H3K4me3 (Figure 3-11b-c). In fact, we observed that H3K4me3 and/or H3K4me2, with or without H3K4me1, appears to be a signature of ~80% of origins associated with genes (Figure 3-11e). This is consistent with the preferential localization of origins in 5' gene body regions observed here and the anticorrelation of these marks and CG

methylation (Zhang et al., 2009). Furthermore, H3K9me2 is highly depleted in most of the origins identified in our study (Figure 3-11d).

A correlation exists between histone hyperacetylation and origin activation in Xenopus (Danis et al., 2004) and Drosophila cells (Aggarwal and Calvi, 2004; Hartl et al., 2007; Schwaiger et al., 2009). Consistent with this, immunofluorescence data obtained in several plant species indicate that increases in histone acetylation occurs during S-phase (Fuchs et al., 2006; Sanchez et al., 2008). Recently, ChIP experiments have revealed that H4K5 and H4K12 (also H4K8 to a lesser extent) but not H4K16 need to be acetylated by the HBO1 histone acetylase at origins in human cells to overcome geminin inhibition and facilitate MCM loading (Miotto and Struhl, 2010). Thus, we profiled H4K5ac over the genome by ChIP-chip and found an enrichment of this mark at the origin midpoint (Figure 3-11f).

DISCUSSION

Initiation of DNA replication in eukaryotes depends on the assembly of pre-replication complexes (pre-RC) in G1 of the cell cycle at certain chromosomal locations and its further activation to initiate DNA replication in S-phase. Both steps must be tightly coordinated to ensure that the genome is duplicated once per cell cycle (DePamphilis et al., 2006). We have found that ORC1 binding sites tend to form clusters, a situation similar to Drosophila cells (MacAlpine et al., 2010) but highly different from that of *S. cerevisiae* (Wyrick et al., 2001). The presence of ORC1 binding sites across the genome may represent not only broad initiation zones with several potential initiation sites but also reflect the function of ORC1 in other processes, e.g. heterochromatin silencing (Pak et al., 1997), transcriptional control (Rusche et

al., 2003; Sanchez and Gutierrez, 2009) or chromatid cohesion (Takahashi et al., 2004). In any case, detection of CDC6 in BrdU regions is highly valuable taking into account the release of CDC6 from the pre-RC once an origin fired (Kim and Kipreos, 2008).

The distribution of distances between origin region midpoints rendered values that fall within the range estimated for other eukaryotes (Cadoret and Prioleau, 2010) and roughly match estimations of replicon size in Arabidopsis (Lee et al., 2010; Van't Hof et al., 1978). It is possible that a fraction of the putative origin regions identified here correspond to elongating forks rather than to initiation events. However, our direct measurements of origin activity by abundance of RNA primer-containing nascent strands support the idea that the "originome" reported here is a *bona fide* list of putative Arabidopsis DNA replication origins. Future analysis should address this point individually. The abundance of origin sequences and the width of the peak of amplified fragments varied for different origins analyzed, suggesting differences in the efficiency of origin usage or in the usage of initiation sequences within an origin region (Cadoret and Prioleau, 2010; Schepers and Papior, 2010).

Interestingly, the location of most Arabidopsis origins is different from other systems in which a large proportion of highly efficient origins are associated with gene promoters or transcriptional start sites (Karnani et al., 2010; MacAlpine et al., 2010; Sequeira-Mendes et al., 2009). We have found that the ±0.1kb region around Arabidopsis DNA replication origins possesses a higher than average G+C content and a slight decrease in CG methylation. Consistent with this observation, early-mid replicons in Arabidopsis chromosome 4 have been also found to be depleted of CG methylation (Lee et al., 2010). One possibility is that in Arabidopsis the relatively high G+C content at origins favors a particular nucleosome organization in these regions. This is reinforced by the colocalization of origins with histone

H2A.Z, which affects nucleosome stability (Jin et al., 2009), and could facilitate pre-RC assembly and/or origin firing. Together, our data show that whereas CG methylation within gene bodies is relevant for gene expression in Arabidopsis (Cokus et al., 2008), it does not seem to be a requirement for origins. Metazoan origins highly correlate with unmethylated CpG islands located at the promoter of active genes or in the proximity to transcriptional start sites (Antequera, 2004b; Cadoret and Prioleau, 2010). While CpG islands are not present in the Arabidopsis genome, our results revealed a conserved trend of having relatively lower CG methylation at origins and show a high correlation between origin activity, a local high G+C content and presence of H2A.Z.

Posttranslational histone modifications can also affect origin specification and function. Most Arabidopsis origins tend to be enriched in H3K4me2 and H3K4me3, as well as in H4K5ac, similar to human origins (Karnani et al., 2010; Miotto and Struhl, 2010). Whether all human origins have the same H4ac pattern, as a consequence of HBO1 activity to overcome geminin inhibition (Miotto and Struhl, 2010), and whether all Arabidopsis origins require H4ac for activation remain open questions for the future. However, the H4 acetylation pattern is of particular relevance due to the presence in Arabidopsis of (i) an HBO1-related acetyltransferase (Earley et al., 2007), (ii) increased tetraH4ac residues around Arabidopsis ORC1-binding sites (Sanchez and Gutierrez, 2009) and (iii) a CDT1-interacting protein, GEM, structurally unrelated to metazoan geminin (Caro et al., 2007; Caro and Gutierrez, 2007). Acetylation in other histone residues may be also relevant for origin function, as suggested by the presence of H3K56ac in early replicons of chromosome 4.

How replication origins are specified in large eukaryotic genomes has been a long-standing question. The association of early-firing origins with transcribed genomic regions has been

reported (Karnani et al., 2007b; Zhou et al., 2005). Origins that have been studied in the 0.4-1%

of mammalian genomes show a preferential association with active promoters that contain CpG

islands (Cadoret et al., 2008; Karnani et al., 2010; Sequeira-Mendes et al., 2009). We have

found that origins located in the upstream regions of genes are preferentially associated with

highly expressed genes. However, the differences in the genomic distribution of CG methylation

pattern in Arabidopsis may contribute to the use of different mechanisms to specify origins. In

fact, a higher proportion of origins in Arabidopsis are located in the 5' half of gene bodies

compared to mammalian cells.

Our work has defined a landscape of epigenetic marks associated with a genome-wide set

of replication origins in Arabidopsis. The midpoints of most origin regions preferentially

colocalize with a significantly higher than average G+C content, but lower CG methylation

level, and are enriched in histone H2A.Z, H3K4me2/3 and acetylated H4K5, and depleted in

H3K4me1 and H3K9me2. Elucidating how epigenetic mechanisms and gene expression

coordinate with DNA replication is of primary importance for understanding these processes in a

genomic and developmental context. The Arabidopsis "originome" reported here provides the

foundation for future studies to identify the mechanisms of origin specification as well as the

regulation and function of DNA replication origins in different eukaryotes.

MATERIALS AND METHODS

Plant Material.

Arabidopsis seedlings (Col-0 ecotype) were grown in MS salts medium supplemented with 1%

(w/v) sucrose and 1% (w/v) agar in a 16h-8h light-dark regime at 22°C. Plants constitutively

expressing the Myc-ORC1a-tagged and the hemagglutinin (HA)-tagged Arabidopsis CDC6a protein have been described (Castellano et al., 2001; Sanchez and Gutierrez, 2009).

Chromatin immunoprecipitation (ChIP) and microarray (chip). ChIP assays of His-ORC1- and HA-CDC6-expressing plants were carried out using 10-day-old plants preincubated with 50 µM MG132 prior to the fixation step. For immunoprecipitation, 10 µl of anti-Myc (SC-40ac; Santa Cruz Biotechnology) or anti-HA (A2095; SIGMA), were incubated with 1mg of protein extract, previously sonicated, to obtain DNA fragments of ~500-1000 bp. After washing, immunocomplexes were incubated at 65 ºC for 6h nad treated with 20 µg of proteinase K and phenol-chloroform extracted. The DNA obtained (ChIP and input) was purified using Affymetrix cDNA cleanup columns. Three biological replicates for each condition were independently hybridized to GeneChip® Arabidopsis Tiling 1.0R Array (Affymetrix). DNA was amplified using Affymetrix Chromatin Immunoprecipitation Assay protocol. 7.5 µg of amplification product were fragmented and labeled using GeneChip® WT Double-stranded DNA terminal labeling kit (Affymetrix). Scanning was performed at 0.7µm resolution using a GeneChip Scanner 3000 7G (Affymetrix Inc, Santa Clara, CA). One of the CDC6 replicates was discarded due to low correlation with the other replicates. The H4K5ac ChIP was performed using an antibody from Millipore (Billerica, MA; 06-759), and the H3 ChIP control was performed using an antibody from Abcam (Cambridge, MA; ab1791), as described (Zhang et al., 2009; Zhang et al., 2007b; Zhang et al., 2007c; Zhang et al., 2006).

Cell culture, BrdU labeling and isolation of nascent strands. *Arabidopsis thaliana* MM2d cells were grown in MS medium supplemented with 3% (w/v) sucrose (Sigma), 0.5 mg ml^{-1}

NAA (Sigma) and 0.05 mg ml^{-1} kinetin (Sigma), pH 5.8, and subcultured every 7 days (Menges et al., 2002). For cell synchronization, a mid-exponential phase culture, was grown in MS medium lacking sucrose during 24 h, then released by changing to the medium supplemented with sucrose and cultured for further 3.5 h (Menges and Murray, 2006).

Newly synthesized DNA was labeled by adding 200 μM of 5-bromo-2'-deoxyuridine (BrdU), 20 μM 5-fluoro-2'-deoxyuridine (FdU) and 10mM hydroxyurea (HU) to the medium for the last 60 min of incubation (Figure 3-1). Cells were collected by filtration and DNA was purified (Soni et al., 1995). Genomic DNA was digested with *Bam*HI and centrifuged in 5.0 ml of a CsCl solution containing 10 mM Tris-HCl (pH 7.5), 1 mM EDTA and 150 mM NaCl (refractive index adjusted to 1.4000, at 25ºC) in a Beckman Vti 65.2 rotor at 50000 r.p.m. (total 227,300 g) for 21 h at 20 ºC. Fractions (100 μl) collected from the bottom were analyzed by immunoblot with anti-BrdU (Becton-Dickinson) to identify the heavy-light (HL) density fractions. BrdU DNA in the HL fractions and genomic DNA in the unlabeled (LL) fractions, used as control, were pooled separately, dialyzed on TE buffer and analyzed by qPCR or massive sequencing.

To isolate short nascent DNA strands, genomic DNA was isolated from *A. thaliana* MM2d cells treated as above, under RNase-free conditions. Purified DNA was denatured by heating and size-fractionated in a seven-step sucrose gradient (5-20%, w/v) by centrifugation at 24,000 rpm in a Beckman SW-40Ti rotor (total 102,300 g) for 20h at 20ºC (Gomez and Antequera, 2008). Fractions (1ml) were collected and aliquots were analyzed by electrophoresis in an alkaline agarose gel to monitor size fractionation. Fractions containing replication intermediates ranging (300-2000 nt in size; fractions 2 through 8) were subjected to polynucleotide kinase treatment and l-exonuclease digestion, which degrades contaminating random sheared DNA and leaves

replication intermediates protected by a 5`RNA-primer (Prioleau et al., 2003). The relative abundance of nascent DNA strands around putative origins was monitored by qPCR.

Library preparation for Illumina sequencing of BrdU DNA. DNA libraries for both BrdU DNA and LL DNA were generated and sequenced on Genome Analyzer II as per the manufacturer instructions (Illumina).

Quantitative real-time PCR. qPCR was performed with a LightCycler Real-time PCR system (Roche) using SYBR Green PCR mix and following manufacturer's instructions. All qPCR reactions were performed at least in duplicate and analyses were carried out using the commercial software. Oligonucleotides used for fragment amplification are listed in Table 3-1.

Data processing and analysis. Sequenced reads were based-called using the standard Illumina software. The reads were trimmed down to 50mer bases from the 3' end, and then mapped to the Arabidopsis genome using SeqMap (Jiang and Wong, 2008). Uniquely mapping reads with up to 3 bp mismatches were used for the analysis, resulting in 3,849,549 BrdU reads and 9,799,171 unlabeled DNA sequences, after collapsing identical reads. The reads were extended so that the data represents the actual DNA fragments of the libraries (130bp, determined from distribution of DNA fragments in the library). For all the analyses, identical reads were collapsed into single reads, and each dataset was normalized to the total number of uniquely mapping reads. BrdU positive regions were defined by using MACS (Zhang et al., 2006), with the unlabeled DNA sample as a control. A Poisson distribution p-value cutoff of 10^{-6}, calculated from the local λ

value, was applied. Then, regions separated within 10kb were combined, a restrictive criterion supported by other origin analysis (Cadoret and Prioleau, 2010; Gregoire et al., 2006).

The CDC6 and ORC1 ChIP data were quantile normalized to genomic DNA using Affymetrix TAS. Intensity peaks were searched using the TileMap algorithm (Ji and Wong, 2005) with hidden Markov model for combining neighboring probes (posterior probability cutoffs of 0.5). Histone modification data were quantile normalized to unmodified H3 ChIP data. Each data set was then normalized so that the mean signal across all probes in the genome was zero. The Affymetrix IGB was used by applying thresholds as previously described (Zhang et al., 2006).

Enrichment of ChIP signals at BrdU midpoints was assessed by first calculating the average scores of probes within ±300 bp of BrdU midpoint of all origin regions and scores in the same number of randomly generated probes in regions 5-10 kb flanking the midpoints. Then, significance of differences were calculated by Wilcoxon ranksum test (two-sided p-values). The H4K5ac data were analyzed as described (Zhang et al., 2009; Zhang et al., 2007b; Zhang et al., 2007c; Zhang et al., 2006).

Illustrations were generated using Adobe Illustrator CS2 and Adobe Photoshop CS3.

Accession codes. The NCBI GEO accession numbers for the datasets generated in this work are GSE21928 (for ORC1 and CDC6 ChIP-chip) and GSE21828 (for BrdU-seq and H4K5ac ChIP-chip).

FIGURE LEGENDS

Figure 3-1. Experimental strategy used to identify replication origins in Arabidopsis cells.

The initial event in pre-replication complex (pre-RC) assembly is the binding of origin

recognition complex (ORC) at certain genomic locations. To determine all possible sites

where ORC can bind in different cell types, we used plants that express constitutively a

Myctagged

ORC1a protein. Chromatin was immunoprecipitated (ChIP) with anti-c-Myc antibodies

and the material used to synthesize probes to hybridize full-genome Arabidopsis tiling

microarrays, as described in Methods. Total genomic DNA of wild type plants was used as a

control to determine the enrichment. Activation of pre-RC initiates when CDC6 protein binds

to a subset of the ORC-bound sites. To determine all potential CDC6-binding sites we used

plants expressing constitutively an HA-tagged CDC6a protein. Samples were processed as

described for ORC1 and the data generated processed as described in Methods.

To determine the locations of putative DNA replication origins we used cultured

Arabidopsis cells because:

(1) the proportion of actively proliferating cells is relatively very small in whole plants, even

in young seedlings,

(2) in the absence of any information on Arabidopsis origins of replication, we wanted to

use a synchronized cell population where we can choose the time for BrdU labeling of

nascent DNA strands. We are aware that with this strategy we may be loosing origins that

are activated at a time not covered by the BrdU pulse,

(3) we wanted to use a treatment with hydroxyurea for two main reasons: (i) this treatment

will reduce the dNTP pool, a necessary condition for an efficient BrdU labeling of nascent

DNA strands, which otherwise will be below our current detection level. (ii) it will slow-down

replication fork movement and will trigger an S-phase checkpoint 1. In that way, we favor that

BrdU incorporation is restricted to sequences around the origins activated during the BrdU

pulse.

Thus, cultured cells were synchronized in G0 by sucrose deprivation, as described 2, and

the purified heavy-light (HL) BrdU-containing DNA subjected to full sequencing using the

Solexa/Illumina technology (see Methods). The genomic location of BrdU-sequences provide

a genomic map of putative DNA replication origins that were active in cultured cells at the

time of BrdU labeling.

Validation of DNA replication origin activity was then carried out by measuring by qPCR

the abundance of nascent DNA strands using a purified sample of size-fractionated, RNA

primer-containing short DNA molecules 3,4 (see also Methods).

Figure 3-2. Identification of DNA replication origins in the Arabidopsis genome. (a)
Representative genome-browser view of a region in chromosome 1. Genes (green) transcribed
from each strand are shown along the chromosome above and below the position scale. The
panels shown in the lower part of the figure correspond to an enlarged region containing a
replication origin, determined as a region enriched for BrdU-labeled DNA strands (light blue)
relative to the unlabeled control DNA (black), together with the ORC1 (red) and CDC6 (dark
blue) binding patterns (posterior probabilities for ORC1 and CDC6 data sets). Origins, e.g. *ori1-
0850* shown here, are named based on their chromosomal location (*ori1-* through *ori5-*)

followed by the four digits that indicate the origin number within each chromosome. They are named consecutively starting at the left tip of each chromosome, i.e. for chromosome 1 where we identified 376 origins, the leftmost origin is *ori1-0010* and the rightmost one is *ori1-3760*. **(b)** The pattern of ORC1 and CDC6 binding over origin regions was obtained by plotting their relative binding signal ±10kb from the origin region midpoint (0) using 50bp-sliding windows (smoothed). The p-values (two-sided) of the difference in the ChIP-chip signals in origins (midpoint ± 300bp), using a two-tailed Welch test, were 7.25e-6 and 1.29e-10 for ORC1 and CDC6, respectively. **(c)** Number of origins relative to chromosomal size. Chromosome size (relative to chromosome 1) was plotted against the number of origins identified in each chromosome (relative to origin number in chromosome 1). The number of origins identified in each chromosome is indicated in parenthesis. **(d)** Distribution of interorigin distances, measured as the distance between the midpoints of two contiguous origins (median = 51.1kb; average = 77.2kb; s.d. = 83.4kb).

Figure 3-3. Size distribution of BrdU-labeled DNA regions.

Origins were defined by the presence of DNA fragments enriched in the BrdU-labeled DNA sample. BrdU positive signals separated within 10kb were combined into a single origin. Here, the size distribution of origin regions is shown (n = 1543; mean = 3224 nt; st. d. = 8274nt; median = 1187nt). The data have been capped at 20kb, for simplicity (min. value = 28nt; max. value = 269861nt).

Figure 3-4. Size distribution of ORC1-bound DNA regions.

ORC1 positive signals were defined as described in Methods using TileMap with HMM

option, with a posterior probability cutoff of 0.5. Here, the size distribution of ORC1-bound regions is shown (n = 27405; mean = 695nt; st. d. = 688nt; median = 494nt). The data have been capped at 5kb, for simplicity (min. value = 52nt; max. value = 10149nt).

Figure 3-5. Size distribution of CDC6-bound DNA regions.

CDC6 positive signals were defined as described in Methods using TileMap with HMM option, with a posterior probability cutoff of 0.5. Here, the size distribution of CDC6-bound regions is shown (n = 2157; mean = 290nt; st. d. = 202nt; median = 245nt). The data have been capped at 1.5kb, for simplicity (min. value = 51nt; max. value = 2272nt).

Figure 3-6. Chromosomal view of origin density.

Origins (midpoints of BrdU regions), indicated here as vertical blue bars, were located across each Arabidopsis chromosome. The position of centromeres is indicated by asterisks (red).

Figure 3-7. DNA replication origin activity determined by nascent DNA strand abundance.

(a-d) Several putative origin-containing regions were chosen for detailed measurement by real-time PCR of nascent strand abundance in a sample of short DNA molecules containing an RNA primer at their 5'end (see Methods). The genomic region under study is indicated at the bottom of each panel and shows the location of genes (green), the ORC1 binding (red) and CDC6 binding (dark blue) signals, and the putative origin location (light blue), defined by direct sequencing of the BrdU-labeled DNA sample (see Methods). DNA fragments (~200bp in length) amplified by primer pairs scanning each region are indicated by the small black rectangles on the X-axis. The coordinates in each chromosome are also indicated at the bottom

of each panel. Results correspond to PCR amplifications using fraction #5 (see Methods). **(a-c)**
Data for origins *ori1-2300*, *ori2-1340* and *ori2-1430.* **(d)** Data for a region used as a negative
control around gene *at4g14700* that lacks BrdU-labeled DNA sequences.

Figure 3-8. Genomic location of Arabidopsis replication origins. (a) Percentage of origins
colocalizing with various genomic elements, as indicated. Numbers in parenthesis indicate the
proportion of the Arabidopsis genome represented by each class. **(b)** Origin densities were
computed for regions upstream, downstream, and within genes of different expression levels (all
genes vs highest 25% vs lowest 25%). Regions 2kb upstream and downstream of genes, as well
as the bodies of genes were each divided into 10 bins, and the origin densities (origins per
10^6bp) were calculated for each bin and represented as boxplots. White lines represent the
median, the edges of the boxes represent the 25th (bottom) and 75th percentiles (top), and the
whiskers stretch out to the minimum and maximum points that fell within the 1.5xIQR range
below the 25th percentile or above the 75th percentile.

Figure 3-9. Ends-analysis of BrdU-seq reads relative to the control unlabeled DNA reads.
Reads were plotted over protein-coding genes (TAIR8) that contained origins in the
bodies or those that had origins within 2kb of the transcriptional start sites (TSS; 0 in left
panel) or the transcription termination sites (TTS; 0 in right panel). Two kb upstream or
downstream of genes, and 2 kb into the genes are shown.

**Figure 3-10. Relationship of Arabidopsis replication origins to CG methylation and histone
H2A.Z. (a)** Relative levels of CG, CHG and CHH methylation were plotted ±10kb relative to

the origin midpoint (0) in 50bp sliding windows (smoothed). Methylation data were published elsewhere (Cokus et al., 2008). **(b)** G+C content (%) of replication origins (blue) and the indicated genomic regions (orange). These values were calculated from the sequence data files available at TAIR web site. **(c)** Density of the histone variant H2A.Z in a ±10kb region relative to the origin midpoint (0) in 50bp sliding windows (smoothed). The genomic distribution of H2A.Z was published elsewhere (Zilberman et al., 2008). The p-value of the difference in the ChIP-chip signals in origins, calculated as in Figure 3-2b (see Methods), was 9.34e-34.

Figure 3-11. Histone modification landscape around replication origins. (a-d) The relative level of the indicated histone mark is plotted ±10kb relative to the center of origins (0) in 50bp sliding windows (smoothed). Data for H3K4me and H3K9me2 were reported elsewhere (Bernatavichute et al., 2008; Zhang et al., 2009). The p-values of the difference in the ChIP-chip signals in origins, calculated as in Fig.1b (see Methods), were 0.86, 3.52e-28, 1.07e-41 and 7.33e-14 for H3K4me1, H3K4me2, H3K4me3 and H3K9me2, respectively. **(e)** Relationship between H3K4 methylation status and the presence of origins. We calculated the fraction of genes containing origins and different combinations of H3K4 methylation, as indicated, and compared it with the fraction of all genes containing the same H3K4me combinations (Zhang et al., 2009). Different classes are ordered with decreasing values of the fraction of "genes with origins". **(f)** The relative level of H4K5ac is plotted ±10kb relative to the center of origins (0) in 50bp sliding windows (smoothed). Calculations are based on the ChIP-chip dataset generated in this work. The p-value of the difference in the ChIP-chip signals in origins, calculated as in Fig.1b (see Methods), was 1.23e-23.

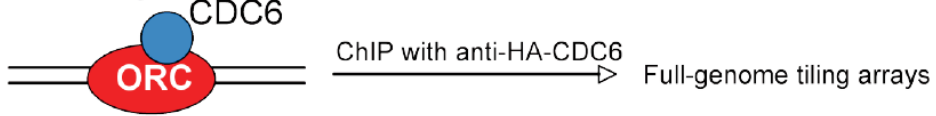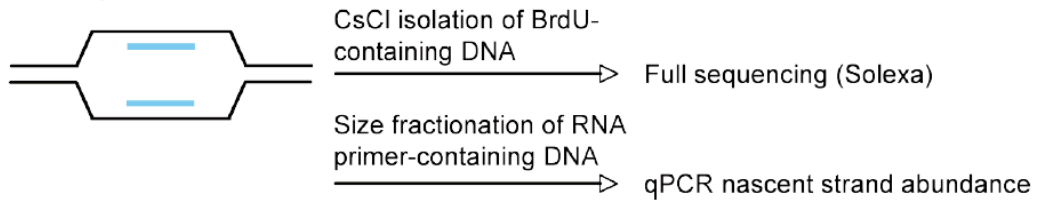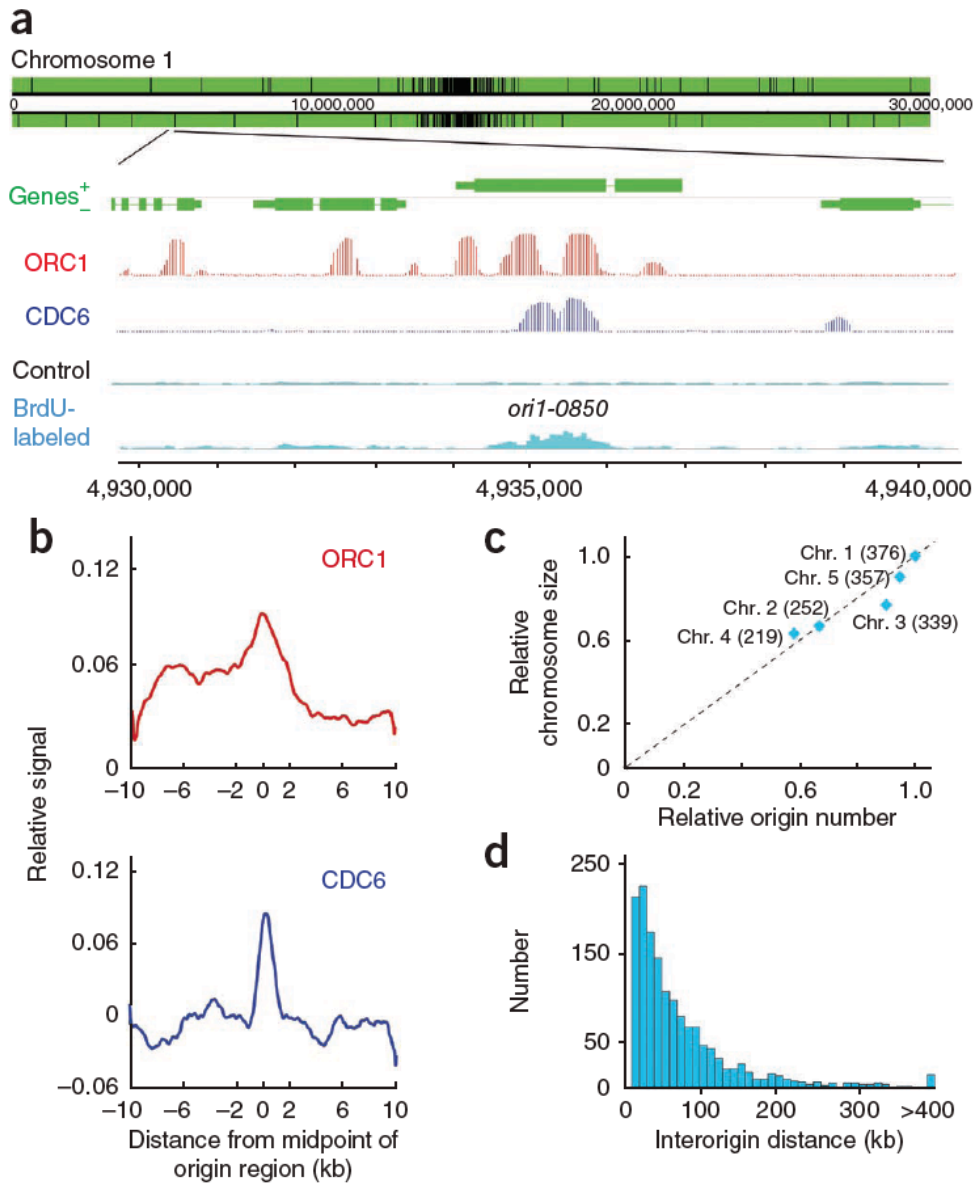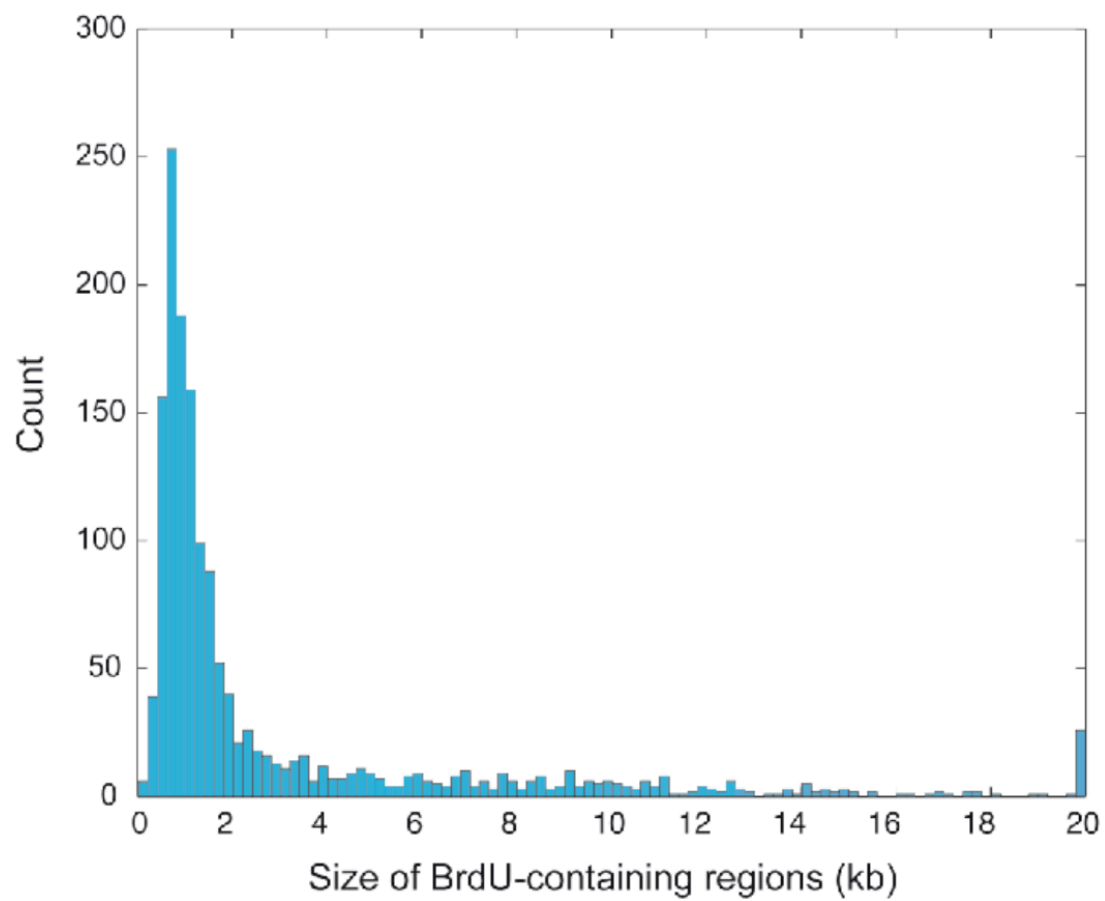**Table 3-1. List of oligonucleotides used in the real-time PCR experiments.**

Figure 3-1
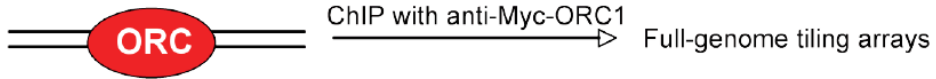
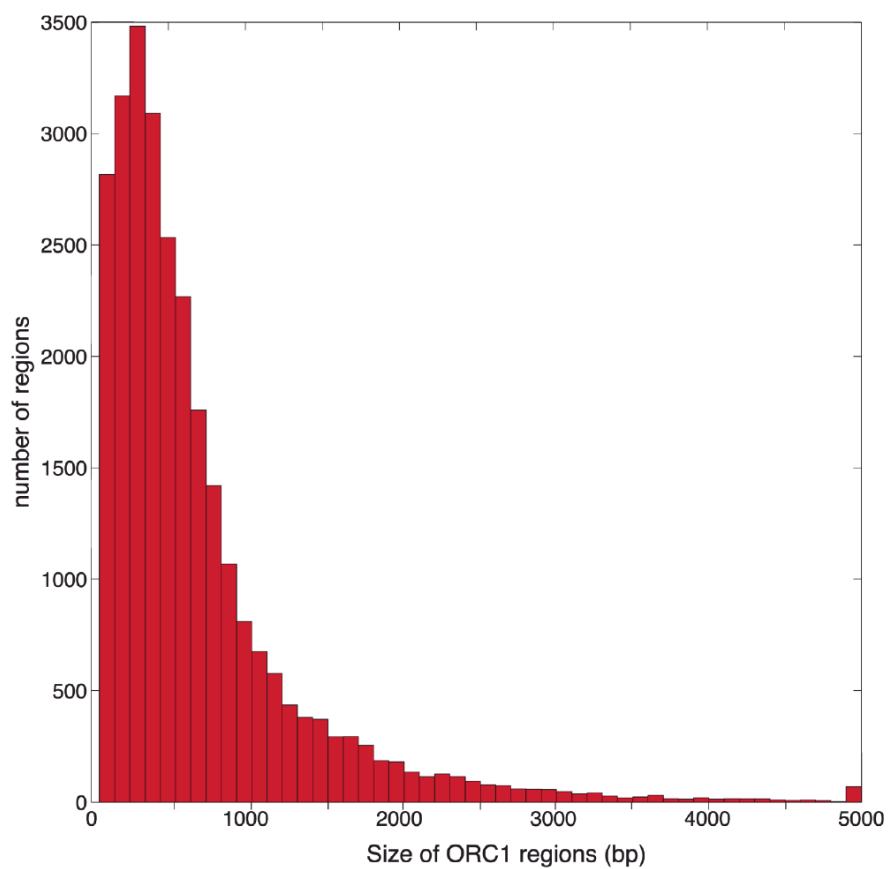Figure 3-2

Figure 3-3

Figure 3-4

Figure 3-5
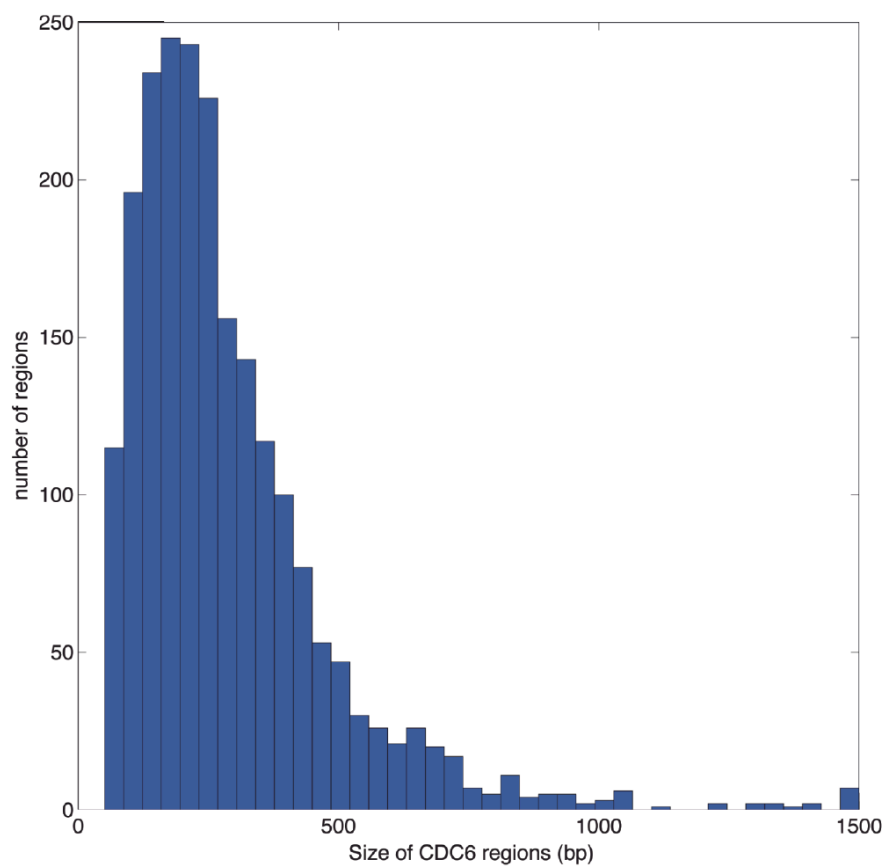
Figure 3-6

Figure 3-7

123
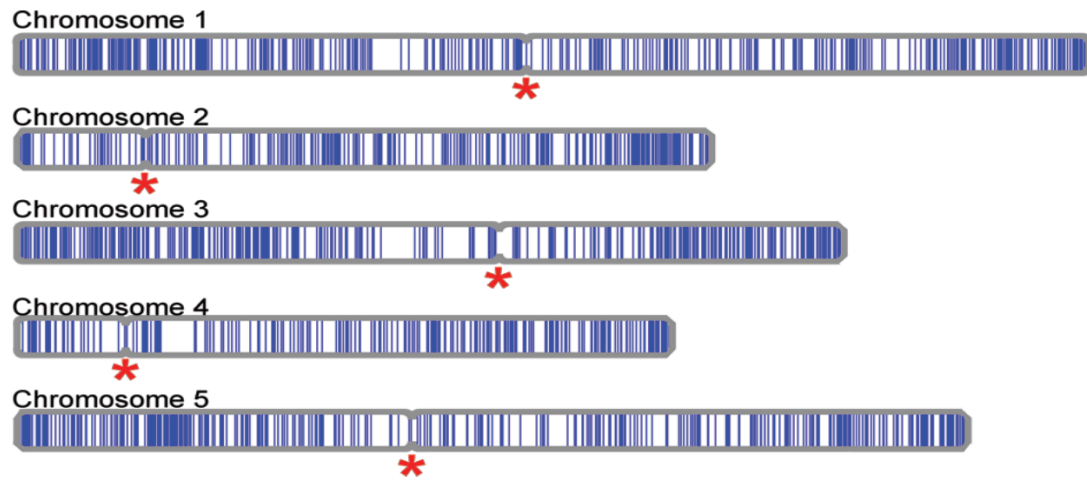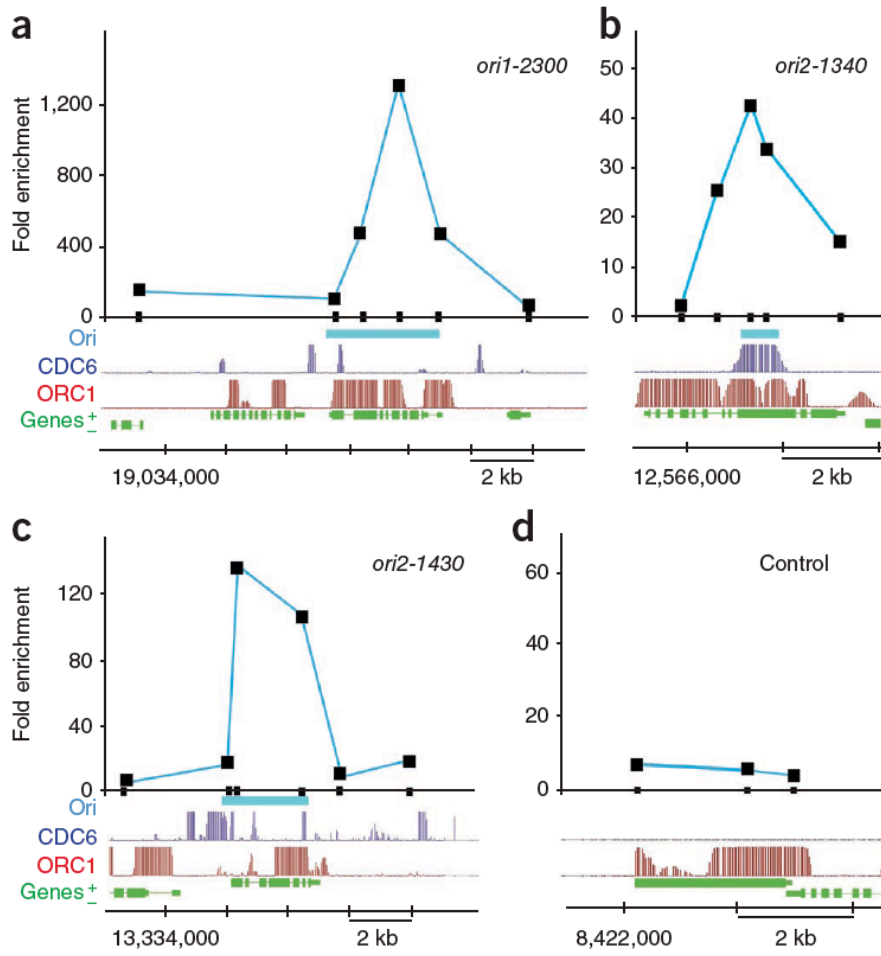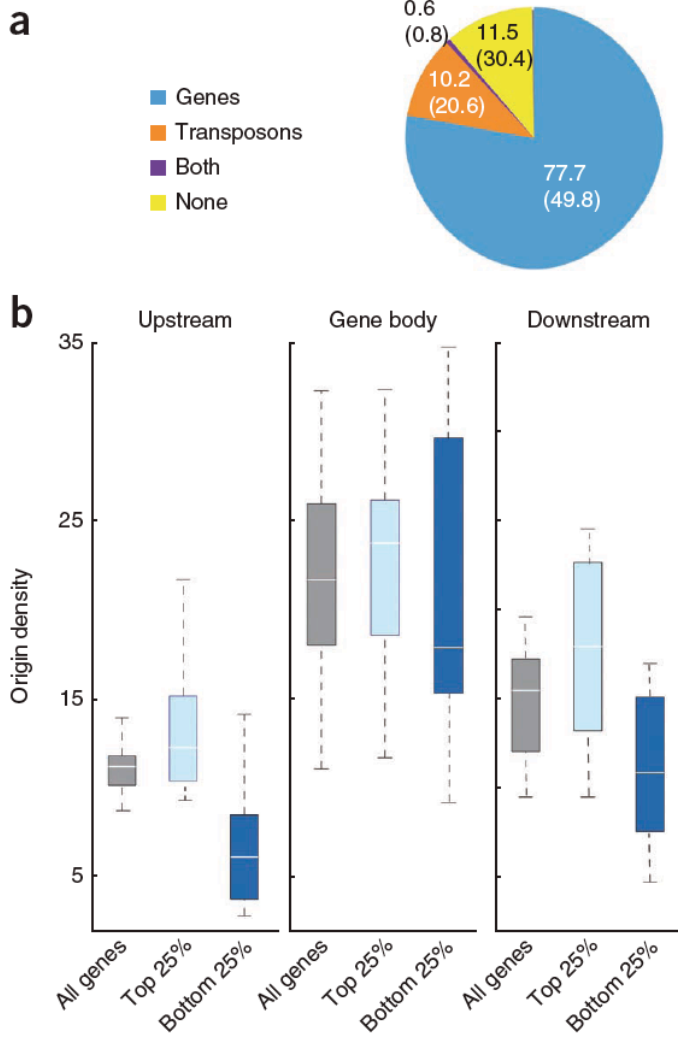
Figure 3-8

Figure 3-9
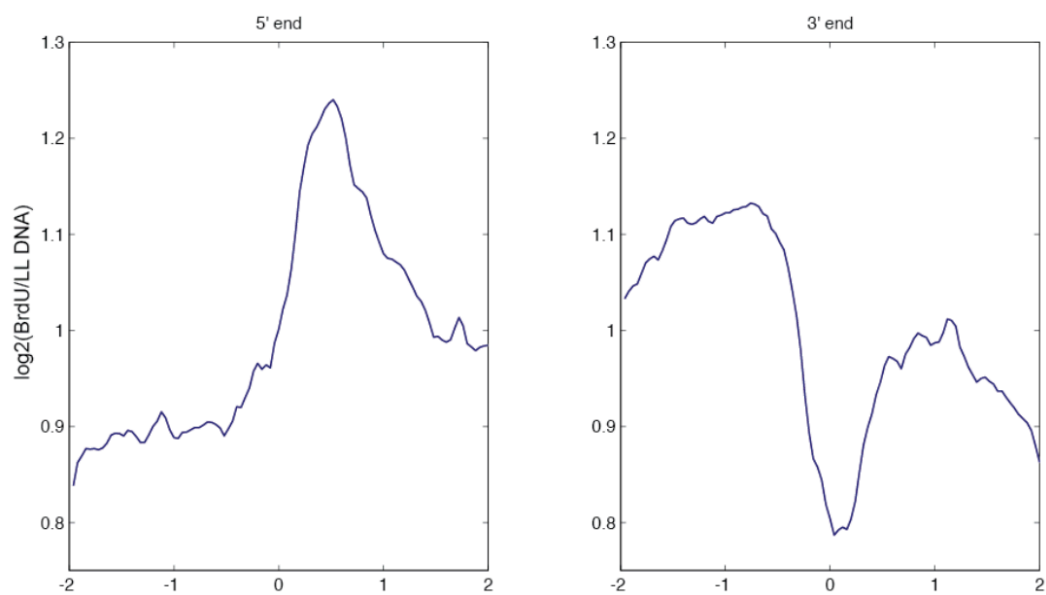
Figure 3-10
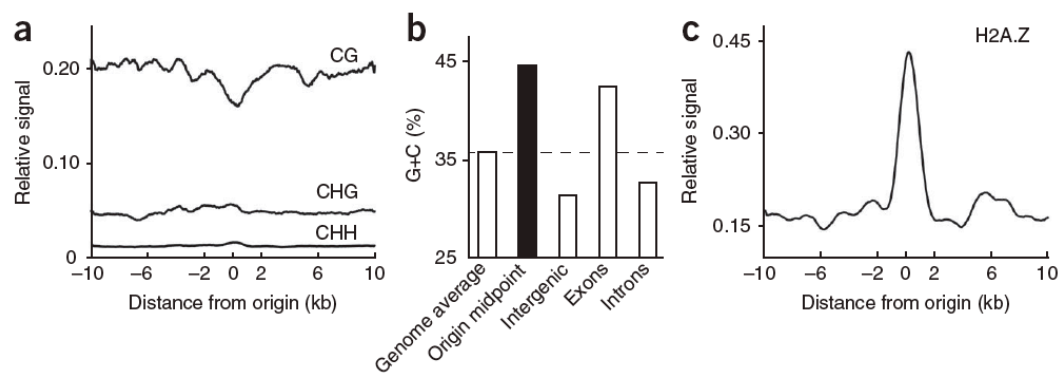
Figure 3-11
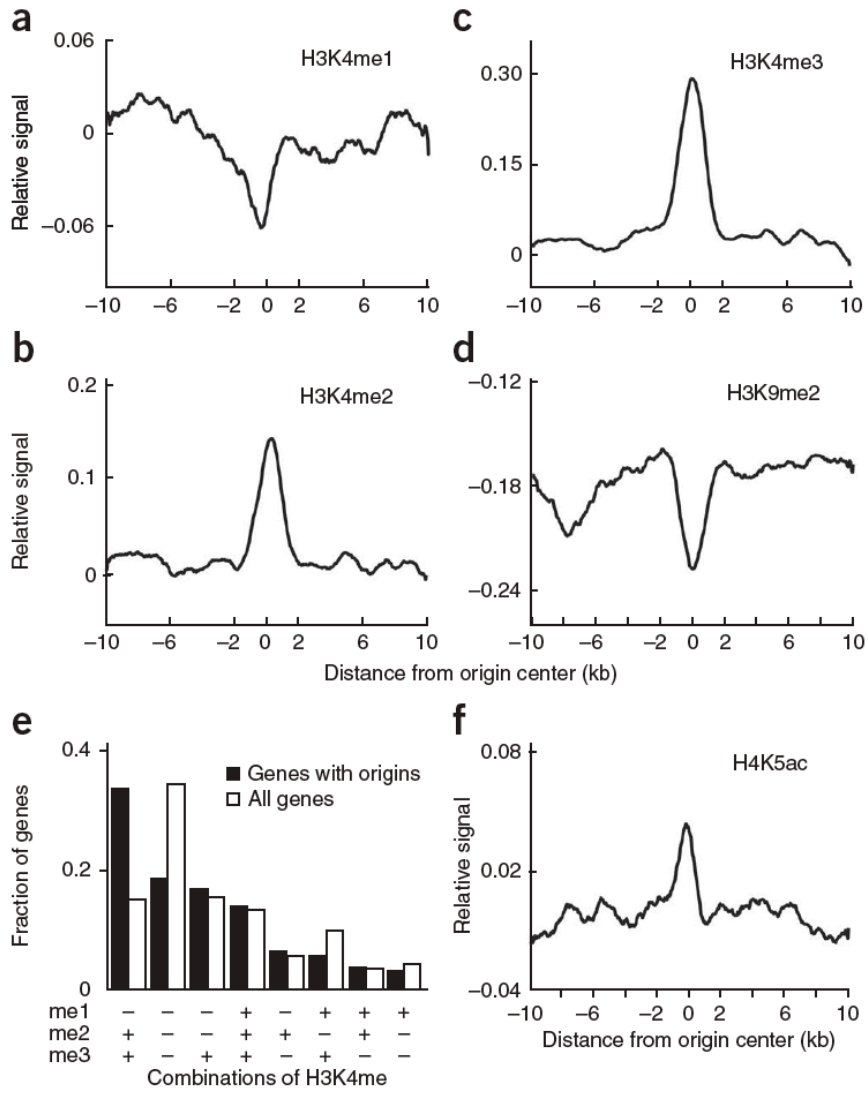
Table 3-1

ori1-2300

1F	AACACAATACCACAAACCAAAG
1R	AGTCAATGGAGTATAGATAGAG
2F	TTCCAATCTAAGCCAAAACTC
2R	ATCAGAATCGTCAGCATCAGC
3F	GTATTATCGCTCATGCTTGTG
3R	TGACAACTAAGCAAAGACAAG
4F	TCAATGGATCCAAATACTCGG
4R	TCAACAAGATTACGGAGGAGG
5F	CTTTACTTGTGCCACTTTTAGA
5R	ATTTTAATTTTATGTTTTGCCACG
6F	GTGGGTTTGAATTTCTGGTAG
6R	TGTACCCAATAAAAAGGAAATG

ori2-1340

1F	TAGCCCATGCCATTGATACA
1R	CTGCTGTGGAACACCACTTG
2F	CATCGCTGTATTCCCTTGTGT
2R	AAACCAAATGCACATGCAAA
3F	GTGCGTAGGTCTCTGTCTC
3R	GTCTAGCAGGTGGTGATGG
4F	CGGCACAGGTCTCCAACAC
4R	TGAGTCAGACCTTCCCCGC
5F	AGGAGAGACATCGTAAAAGG
5R	CAAATAACGCTGGACAAGGG

ori2-1430

1F	AGCAAACCTTTGTTGAGCATTC
1R	CTATGAATAGTTATTGCTAGAGC
2F	ACTGAACAAGCAGCATAAAC
2R	AAGACGACGGAAGAAGAAGA
3F	TTCTTCCGTCGTCTTCATCGTC
3R	CCAGCGGCTTGACGAATTTCTT
4F	GTAAGCC CTCTAACCAC TAAG
4R	CTCCACTTGATGATTTTTTCTC
5F	GCTTAATAGAACATTGCTAA C
5R	CTAGAGAGTCAATAAAGATGAC
6F	AGCATATAATAAGAGCCTCCAG
6R	TGTTAGGCTCGAATTGACGG

At4g14700

1F	TCTTCTCTGAGTTCCAAGGC
1R	CTGGAAGATCGATTGGGTC
2F	AGGTTGGAAGGAATCAATGC
2R	ACTTGAATATGAGGTGCCTG
3F	TCACAAGTGGATCTGTAGATG
3R	AGTTCATCCAAGAAAAGACACG

Chapter 4

Relationship between ATXR5 ATXR6 and DNA methylation in regulating

gene silencing and DNA replication

ABSTRACT

The relationship between epigenetic marks on chromatin and the regulation of DNA replication is poorly understood. Mutation of the H3K27 methyltransferase genes, *ARABIDOPSIS TRITHORAX-RELATED PROTEIN5* (*ATXR5*) and *ATXR6*, result in re-replication (repeated origin firing within the same cell cycle). Here we show that mutations that reduce DNA methylation act to suppress the re-replication phenotype of *atxr5 atxr6* mutants. This suggests that DNA methylation, a mark enriched at the same heterochromatic regions that re-replicate in *atxr5/6* mutants, is required for aberrant re-replication. In contrast, RNA sequencing analyses suggest that ATXR5/6 and DNA methylation cooperatively transcriptionally silence transposable elements (TEs). Hence our results suggest a complex relationship between ATXR5/6 and DNA methylation in the regulation of DNA replication and transcription of TEs.

INTRODUCTION

Faithful DNA replication requires that each origin of replication fire only once per cell cycle. Re-replication has recently been suggested to be an inducer of gene copy number changes and hence threatens genome stability (Green et al., 2010). Multiple mechanisms that prevent re-replication are known (Arias and Walter, 2007), but the regulation of DNA replication at the level of chromatin remains elusive. Especially poorly understood is DNA replication of heterochromatin, which is late replicating in both plants and animals (Birney et al., 2007; Lee et al., 2010; White et al., 2004). In Arabidopsis, heterochromatin is primarily pericentromeric and is enriched in repetitive elements such as transposons that are transcriptionally silenced by epigenetic modifications such as DNA methylation (in both CG and non-CG sequence contexts), H3 lysine 9 dimethylation (H3K9me2) and H3K27me1 (Law and Jacobsen, 2010). The *ATXR5* and ATXR6 methyltransferases catalyze H3K27 monomethylation and function redundantly to suppress over-replication of heterochromatin, likely by inhibiting re-replication (Jacob et al., 2010).

Given the close correlation between sites of re-replication and sites enriched with DNA methylation, we investigated the role of DNA methylation in *atxr5 atxr6*-induced re-replication. We found that loss of DNA methylation suppressed the re-replication phenotype of *atxr5 atxr6* mutants, suggesting a role for DNA methylation in re-replication. We also profiled the transcriptome in different mutants by RNA sequencing (RNA-seq), and found that many TEs are cooperatively silenced by ATXR5/6 and DNA methylation.

RESULTS AND DISCUSSION

Re-replication in *atxr5 atxr6* mutants is closely confined to heterochromatin.

We previously found that most re-replicated sites overlapped with heterochromatin (Jacob et al., 2010), but the extent of the overlap has not been examined. We utilized Illumina sequencing to examine the DNA contents of *atxr5 atxr6* double mutants at the boundaries of previously defined heterochromatic patches in the arms of chromosomes (Bernatavichute et al., 2008). Interestingly we found that DNA content in *atxr5 atxr6* decreased sharply at the boundaries of the heterochromatic patches (Figure 4-1). Moreover, the sizes of re-replicated regions closely tracked the sizes of defined heterochromatin (Figure 4-1), suggesting that re-replication is closely confined to heterochromatin and is unable to spread into flanking euchromatin. The close correlation of re-replication with heterochromatin led us to hypothesize that certain marks of heterochromatin, such as DNA methylation, may be required for the occurrence of aberrant replication.

Loss of DNA methylation suppresses the re-replication defect in *atxr5 atxr6* mutants.

To test the role of DNA methylation in re-replication, we crossed *atxr5 atxr6* mutants to *met1*, *cmt3* and *ddm1* mutants, to reduce CG, non-CG, and both CG and non-CG methylations, respectively. Previous studies indicated that *atxr5 atxr6* mutants themselves show no reduction of DNA methylation (Jacob et al., 2009) and loss of DNA methylation does not affect H3K27me1 levels (Mathieu et al., 2005), suggesting that DNA methylation and H3K27me1 are independent of each other. To confirm that DNA methylation levels were decreased in *ddm1 atxr5 atxr6* triple mutants we performed whole genome bisulfite sequencing (BS-seq) on *atxr5 atxr6* and *ddm1 atxr5 atxr6* mutants. We observed significant loss of DNA methylation in *ddm1*

132

atxr5 atxr6 backgrounds (Figure 4-2).

To compare genomic DNA contents between similar cell types, we sorted and collected 8C nuclei from leaves and sequenced the DNA. We chose 8C nuclei because we previously showed high levels of heterochromatin re-replication in nuclei of this ploidy level as compared to 2C or 4C nuclei (Figure 4-3A) (Jacob et al., 2010). Strikingly, by examining the distribution of sequenced reads across the chromosomes we observed suppression of heterochromatic re-replication in all the mutants compared to *atxr5 atxr6* mutants (Figures 4-4A-D, F). This suggested that factors involved in DNA methylation maintenance are required for re-replication in *atxr5 atxr6* mutants. Consistently, *ddm1* and *met1,* which show the most dramatic losses of DNA methylation (Cokus et al., 2008), most significantly suppressed re-replication (Figure 4-4B, C, F). The relatively weak but reproducible suppression in *cmt3 atxr5 atxr6* mutants could be explained by the fact that non-CG sites are relatively lowly methylated (<7%) in Arabidopsis (Cokus et al., 2008).

We next tested the role of H3K9me2 in regulating re-replication by generating *kyp suvh5 suvh6 atxr5 atxr6* quintuple mutants. Because *kyp suvh5 suvh6* mutants are depleted in both H3K9me2 and non-CG methylation, we reasoned that if H3K9me2 played a dominant role in regulating re-replication, *kyp suvh5 suvh6 atxr5 atxr6* mutants should exhibit a stronger degree of suppression of re-replication than do *cmt3 atxr5 atxr6* mutants, which lose non-CG methylation but retain a significant amount of H3K9me2 (Inagaki et al., 2010). However, we observed a very similar degree of suppression of re-replication in *kyp suvh5 suvh6 atxr5 atxr6* mutants compared to *cmt3 atxr5 atxr6* mutants (Figure 4-4E,F, Figure 4-5). Hence the reduction of DNA re-replication

observed in *kyp suvh5 suvh6 atxr5 atxr6* mutants is likely due to losses of DNA methylation

rather than losses of H3K9me2. We also crossed *atxr5 atxr6* to *drm1 drm2* double mutants.

DRM1 and DRM2 maintain asymmetric methylation at a subset of cytosines. Consistent with the

fact that *drm1 drm2* mutants show only limited reductions in DNA methylation genome-wide

(Cokus et al., 2008), we did not observe significant suppression in *drm1 drm2 atxr5 atxr6*

quadruple mutants (Figure 4-6).

Finally, we crossed *atxr5 atxr6* to *mom1*, which exhibits transcriptional derepression of TEs

without altering DNA methylation (Amedeo et al., 2000). We did not observe significant

suppression of re-replication in *mom1 atxr5 atxr6* mutants (Figure 4-7), further supporting our

hypothesis that it is the loss of DNA methylation that causes suppression of re-replication. Flow

cytometry analyses on multiple biological replicates of all the mutants confirmed the results

obtained by DNA sequencing (Figure 4-3B). It should be noted that DNA methylation single

mutants in wild type ATXR5 ATXR6 backgrounds did not cause significant changes in DNA

content (Figure 4-3C). In sum, these results suggest that DNA methylation plays a role in the

induction of re-replication in *atxr5 atxr6*, consistent with the hypothesis that DNA methylation

promotes DNA replication in heterochromatin.

Relationship between transcriptional derepression and re-replication in *atxr5 atxr6*
mutants.

Analyses at a few loci have shown that *atxr5 atxr6* mutants exhibit transcriptional derepression

of certain TEs (Jacob et al., 2009). A possible explanation for the TE reactivation could be that

more permissive chromatin assembled on additional DNA copies of TEs resulting from re-

replication might allow for transcription to occur. Conversely, it is also possible that the transcriptional derepression in *atxr5 atxr6* mutants may in some way be causing the re-replication defect. To examine the relationship between heterochromatin re-replication and transposon derepression in *atxr5 atxr6* mutants, we analyzed the transcriptome of *atxr5 atxr6* mutants by performing RNA-seq on cotyledons, which we found show significant re-replication (Figure 4-8). We defined 100 TEs that were consistently upregulated in biological replicates by applying stringent thresholds (see Materials and Methods). These TEs were highly methylated in wild type, and did not lose DNA methylation in *atxr5 atxr6* (Figure 4-9). In addition we sequenced genomic DNA from cotyledons of *atxr5 atxr6* mutants, and defined re-replicated TEs (see Materials and Methods). Importantly, a subset of TEs that were transcriptionally derepressed was not re-replicated (Figure 4-10, Figure 4-11). The presence of TEs derepressed in *atxr5 atxr6* mutants that were not overlapping with re-replicated regions suggests that TE reactivation is not due to increased DNA copy numbers at these loci. Re-replicated TEs were much more numerous than those transcriptionally derepressed in *atxr5 atxr6* mutants, where only 1.4% of TEs that showed re-replication also showed derepression. These results suggest that heterochromatin re-replication and transposon derepression are likely two separate phenomena in *atxr5 atxr6* mutants, which is consistent with the observations that DNA methyltransferase mutants do not act as enhancers of the DNA replication defects in *atxr5 atxr6* mutants, as would be expected if there were a simple relationship between transcriptional derepression and DNA replication.

Comparison of TEs regulated by ATXR5/6 and DNA methylation.

We next examined the relationship between TEs derepressed in *atxr5 atxr6* mutants and DNA methylation mutants. Again, TEs were defined to be derepressed using stringent thresholds. We

found significant overlap between TEs derepressed in *ddm1* and *met1* mutants (Figure 4-12A).

And consistent with the dependence of non-CG methylation on H3K9 methylation, we found that

all TEs derepressed in *cmt3* mutants overlapped with those derepressed in *kyp suvh5 suvh6*

mutants (Figure 4-10B). Most TEs derepressed in *atxr5 atxr6* mutants overlapped with those

derepressed in *ddm1* and *met1* mutants (Figure 4-10B, Figures 4-12B and 4-13), however there

was little overlap with those derepressed in *cmt3* and *kyp suvh5 suvh6* mutants (Figure 4-10B).

Hence ATXR5/6 and H3K9me2/non-CG methylation generally regulate different TEs at a

transcriptional level.

TEs regulated by ATXR5/6 were over-represented by LTR/Gypsy type TEs, whereas TEs

regulated by KYP SUVH5 SUVH6 and CMT3 were over-represented by LTR/Copia and

LINE/L1 type TEs (Figure 4-10D). This was not necessarily predicted based on methylation

levels since it was not the case that certain types of TEs were preferentially CG or non-CG

methylated (Figure 4-14). Hence different silencing pathways tend to regulate specific types of

TEs. Notably, a large subset of TEs became reactivated only when combining *atxr5 atxr6*

mutants with DNA methylation mutants (Figure 4-10C, Figure 4-12D). Hence, a relatively large

proportion of TEs are cooperatively silenced by ATXR5/6 and DNA methylation. It is not clear

why *ddm1 atxr5 atxr6* mutants did not show as many additional TEs derepressed as *met1 atxr5*

atxr6 mutants. We also found that the combination of *mom1* with *atxr5 atxr6* mutants caused

activation of many additional TEs (Figures 4-10C and 4-12C).

Over-expression of DNA repair genes in *atxr5 atxr6* mutants.

An additional insight from the transcriptome of *atxr5 atxr6* mutants was that genes in the homologous recombination (HR) DNA repair pathway were over-expressed (Figure 4-10E). This was confirmed by quantitative RT-PCR analyses (Figure 4-15). None of the other tested mutants showed this effect (Figure 4-10E), suggesting that induction of these HR genes is likely due to re-replication. Furthermore, the *ddm1* mutant which caused the most significant suppression of re-replication in the *atxr5 atxr6* background, also caused a significant suppression of the expression of these HR genes (Figure 4-16), further supporting the hypothesis that these HR genes are up-regulated in response to DNA damage caused by re-replication.

In summary, our results show that DNA methylation is required for the heterochromatic re-replication defect in *atxr5 atxr6*, suggesting that DNA methylation positively regulates DNA replication in heterochromatin. We further find that DNA methylation and ATXR5/6 act synergistically in the transcriptional suppression of transposons. The molecular mechanisms causing these relationships are unclear, and it may be that losing DNA methylation causes mobilization of unknown pathways that result in suppression of re-replication. Our findings suggest a complex interplay between epigenetic marks in regulating DNA replication and transcriptional silencing in heterochromatin.

MATERIALS AND METHODS

Plant material

All mutant lines in this study were in the Columbia background. Previously characterized mutant alleles were used for crosses: *cmt3-11*, *ddm1-2*, *met1-3*, *mom1-2*, *atxr5 atxr6* (Jacob et al., 2009) and *kyp suvh5 suvh6* (Ebbs and Bender, 2006). *met1-3* mutants in both wild-type and *atxr5 atxr6* mutant backgrounds used in this study were second generation homozygous lines. *ddm1-2* mutants in both wild-type and *atxr5 atxr6* mutant backgrounds were sixth generation homozygous lines. Plants were grown under continuous light.

Flow cytometry

For sorting nuclei: One   gram of mature rosette leaves were collected from 3-4-week-old plants, chopped in 0.5   ml of filtered Galbraith buffer, and stained with propidium iodide. A BD FACS Aria II in the UCLA Jonsson Comprehensive Cancer Center (JCCC) Flow Cytometry Core Facility was used to sort the nuclei. For sequencing, 7,000-9,000 8C nuclei of each sample were collected, and purified DNA with Picopure purification kit (Arcturus) following manufacturer instructions.

For generating flow cytometry profiles: Three mature rosette leaves were pooled from separate 4-week-old plants, chopped in 2   ml of filtered Galbraith buffer, and stained with a solution of propidium iodide and RNAse A. A BD FACScan flow cytometer in the UCLA JCCC Flow Cytometry Core Facility was used to generate the FACS profiles. For each sample at least

10,000 nuclei were analyzed, and widths of peaks (coefficient of variation values) (Jacob et al., 2010) were calculated using Cyflogic analysis software (http://www.cyflogic.com).

Illumina genomic library preparation

Genomic DNA was sonicated to 200bp with a Covaris S2, and Illumina libraries were generated following manufacturer instructions. The libraries were sequenced using Illumina Genome Analyzer II following manufacturer instructions.

Illumina mRNA-seq library preparation

RNA-seq experiments were performed in two biological replicates for each genotype. 0.1g of tissue was ground in Trizol. Total RNA were treated with DNaseI (Roche), and cleaned up with phenol-chlorophorm and precipitated with ethanol. Libraries were generated and sequenced following manufacturer instructions (Illumina). For verification experiments, the same RNA extraction protocol was used. Single stranded cDNA was synthesized using polyA primers and Superscript II (Invitrogen). For quantitative PCR analysis, cDNA were amplified with iQ SYBR Green Supermix (Biorad) using primers previously described (Ramirez-Parra and Gutierrez, 2007) and *ACTIN* gene was used as an internal control. Primers are listed in Table 4-1.

Whole genome bisulfite sequencing (BS-seq) library generation

0.5~1ug of genomic DNA was used to generate BS-seq libraries. Libraries were generated as previously described (Cokus et al., 2008).

Illumina read alignment and analysis

139

Genomic DNA sequenced reads were base-called using the standard Illumina software. We used Bowtie (Langmead et al., 2009) to uniquely align the reads to the *Arabidopsis thaliana* genome (TAIR8), allowing up to 2 mismatches. For all genomic libraries, reads mapping to identical positions were collapsed into single reads.

Method for defining re-replicated regions: Genome was tiled into 100bp bins, and scores of reads for each bin were computed before the log2 ratio of *atxr5 atxr6* to wild-type was taken. Score= (# reads + c)/0.1kb/(# million mapping reads), where c is a pseudocount defined by (# million mapping reads)/10. In this case, c_{WT}=6.8 and $c_{atxr5atxr6}$=5.3. Pseudocounts were used to avoid divisions by zero. Next, the genome was tiled into 1kb bins (500bp overlap), and Z-scores of log2(*atxr5 atxr6* / wild-type) were computed. A Z>2 cutoff was applied and regions within 500bp were merged. Transposable elements (TEs) overlapping with these regions by 1bp were considered to be re-replicating in *atxr5 atxr6* double mutants.

Both gene and TE expression in the RNA-seq data was measured by calculating reads per kilobase per million mapped reads (RPKM) (Mortazavi et al., 2008). P-values to detect differential expression were calculated by Fisher's exact test and *Benjamini*-Hochberg corrected (Benjamini, 1995) for multiple testing. TEs upregulated in wild-type and mutants were defined by mutant/wild-type>4 and P<0.01. To avoid divisions by zero, TEs with expression levels of zero were assigned the lowest non-zero TE expression value in each sample. Only TEs defined as upregulated in all biological replicates were considered as being derepressed.

BS-seq data was mapped to the TAIR8 genome by BS Seeker (Chen et al., 2010) by allowing up to 2 mismatches. Methylation levels were computed by calculating #C/(#C+#T).

All sequencing data have been deposited at Gene Expression Omnibus (GEO) (accession number GSE38286).

FIGURE LEGENDS

Figure 4-1. Heterochromatin is specifically re-replicated in *atxr5 atxr6* double mutants.

Top panels show enrichment of DNA methylation (CG, CHG, CHH, where H=A,T or C)(Cokus et al., 2008), transposable element (TE) densities (TE per base-pair) and H3K9me2 (Bernatavichute et al., 2008) over the boundaries of heterochromatic regions of indicated sizes. Values were plotted +/-10 kilobase from the boundary of heterochromatin in 500bp bins. x=0 is the heterochromatin boundary, x<0 is outside the region, and x>0 is into the region. Bottom panels show the distribution of DNA contents from *atxr5 atxr6* mutants relative to wild-type (log2 ratios) (Jacob et al., 2010). Heterochromatic regions were defined using previously characterized H3K9me2 regions (Bernatavichute et al., 2008). Plots in both top and bottom panels were smoothed by taking the moving average over +/-1 bins and +/-3 bins, respectively.

Figure 4-2. Genome-wide comparison of losses of DNA methylation in *ddm1* mutants in wild-type and *atxr5 atxr6* backgrounds.

Whole genome bisulfite sequencing (BS-seq) was used to calculate DNA methylation levels in different mutants. DNA methylation levels across all five chromosomes, as well as average DNA methylation levels over TEs and genes were plotted. Plots were smoothed triangularly ($bin_i = 0.25 \times bin_{i-1} + 0.5 \times bin_i + 0.25 \times bin_{i+1}$) once for chromosomal views.

Figure 4-3. Flow cytometry profiles of different mutants.

A. Examples of flow cytometry profiles.

B. Flow cytometry profiles of different mutants. To quantify re-replication, coefficient of variation (CV) values (Jacob et al., 2010) of peaks defined by fluoresence intensity of 8C nuclei (i.e. the widths of peaks) were calculated. Data represented as mean ± SD for triplicates.

C. CV values for DNA methylation mutants in absence of *atxr5 atxr6* mutations.

Figure 4-4. Relationship between ATXR5/6 and DNA methylation in regulating DNA replication in heterochromatin.

A. Chromosomal distribution of transposable element (TE) density, DNA methylation (Zhang et al., 2006) and H3K9me2 (Bernatavichute et al., 2008) data are presented to mark the locations of pericentromeric heterochromatin.

B. Chromosomal views of the log2 ratio of genomic DNA reads of *atxr5 atxr6* mutants to wild type (WT) are shown in black, and the log2 ratio of *ddm1 atxr5 atxr6* mutants to WT are shown in red.

C. *met1 atxr5 atxr6* mutants,

D. *cmt3 atxr5 atxr6* mutants, and

E. *kyp suvh5 suvh6 atxr5 atxr6* mutants.

F. Quantitation of reads in heterochromatin in mutants. Fraction of reads falling into previously defined pericentromeric heterochromatin(Bernatavichute et al., 2008) was calculated. *$P<10^{-5}$ relative to *atxr5 atxr6* mutants.

Figure 4-5. Chromosomal view comparison between *kyp suvh5 suvh6 atxr5 atxr6* quintuple mutants and *cmt3 atxr5 atxr6* triple mutants.

The normalized density of reads (reads/ base/ million uniquely mapping reads) were calculated in 100 kilobase bins, and smoothed triangularly ten times in order to superimpose the two profiles.

Figure 4-6. Mutations in DRM1 DRM2 do not suppress re-replication. Genomic DNA content was measured by FACS (%CV). CV values were normalized to wild type.

Figure 4-7. Mutation in MOM1 does not suppress re-replication. Genomic DNA content was measured by FACS (%CV). CV values were normalized to wild type.

Figure 4-8. Cotyledons are re-replicated in *atxr5 atxr6*. Chromosomal views of genomic DNA sequencing reads in *atxr5 atxr6* vs WT.

Figure 4-9. DNA methylation over TEs derepressed in *atxr5 atxr6* double mutants. Average DNA methylation levels (measured by BS-seq) over all TEs and TEs upregulated in *atxr5 atxr6* double mutants. Plots were smoothed triangularly three times. TSS= transcription start sites, TTS= transcription termination sites.

Figure 4-10. Relationship between ATXR5/6 and DNA methylation in transcriptionally silencing TEs.

A. Overlap of re-replicated TEs in *atxr5 atxr6* mutants, with transcriptionally reactivated TEs in *atxr5 atxr6* mutants.

B. Overlap of TEs transcriptionally derepressed in *atxr5 atxr6* mutants and DNA methylation mutants.

C. Overlap of TEs derepressed in indicated mutants.

D. TE families of derepressed TEs in indicated mutants.

E. Normalized expression levels of key genes in homologous recombination (HR) and non-homologous end-joining (NHEJ) DNA repair pathways.

Figure 4-11. DNA contents in TEs defined to be transcriptionally derepressed but not re-replicated in *atxr5 atxr6* mutants. Cotyledon genomic DNA reads per kilobase TE length per million mapping reads in WT and *atxr5 atxr6* were calculated for the 13 non-re-replicating TEs indicated in Figure 4A. Significance was assessed by Wilcoxon ranksum test.

Figure 4-12. Overlap of TEs derepressed in different mutants.

Figure 4-13. Genome-browser view examples of TE derepression. Normalized expression values (RPKM) were calculated in 20bp non-overlapping bins. TEs are also shown.

Figure 4-14. DNA methylation over different TE families. Average wild-type DNA methylation levels (measured by BS-seq) over indicated classes of TEs. Plots were smoothed triangularly three times.

Figure 4-15. Quantitative RT-PCR analyses on homologous recombination repair genes. The values were normalized to *ACTIN* gene (Q=2^-ΔΔCt). Error bars represent the standard deviation.

Figure 4-16. Suppression of over-expression of HR genes in *ddm1 atxr5 atxr6* mutants.

Table 4-1. Primers used for quantitative PCR experiments.

Figure 4-1

Figure 4-2

CG CHG CHH methylation levels
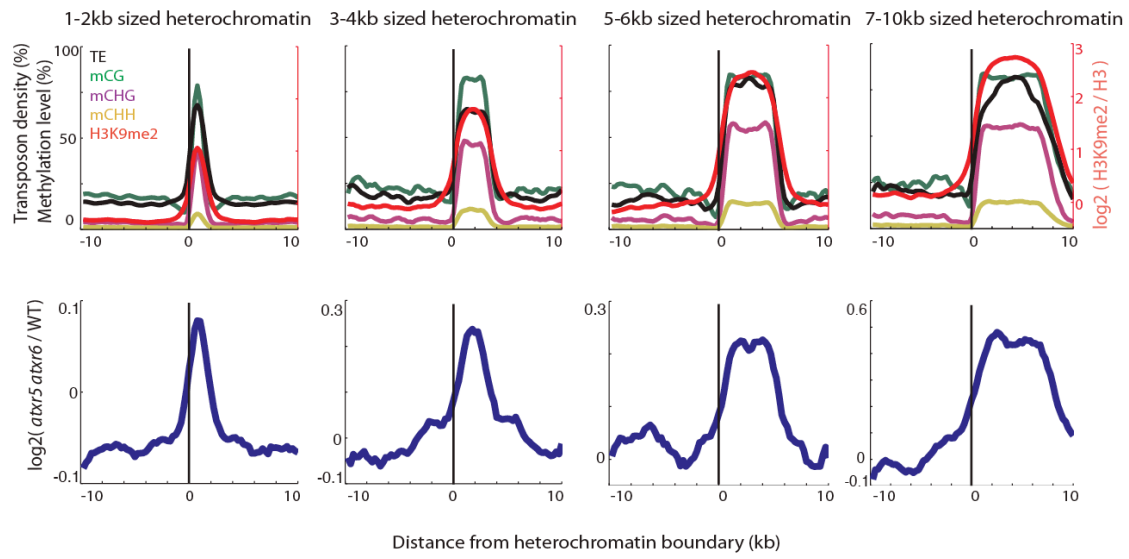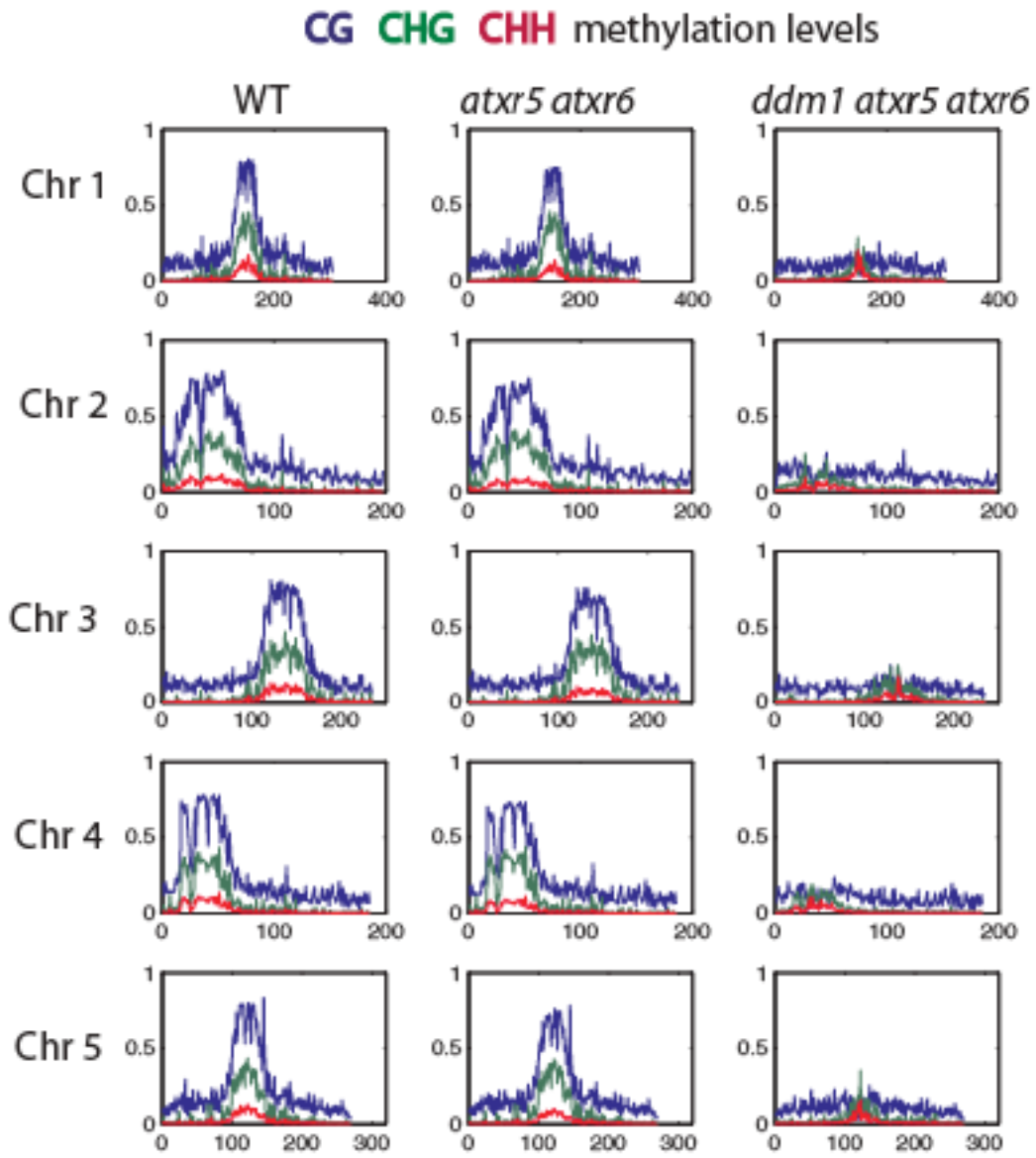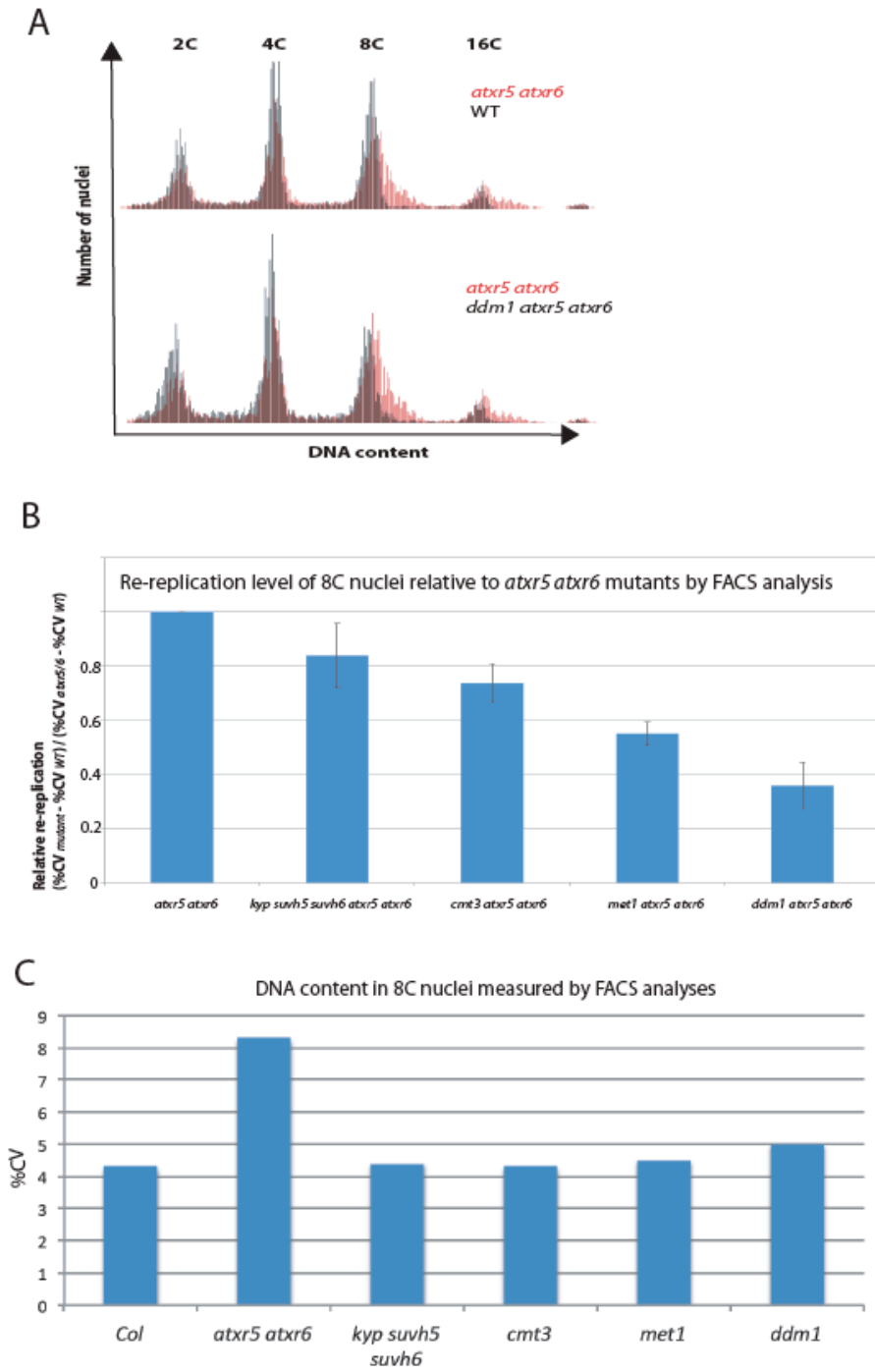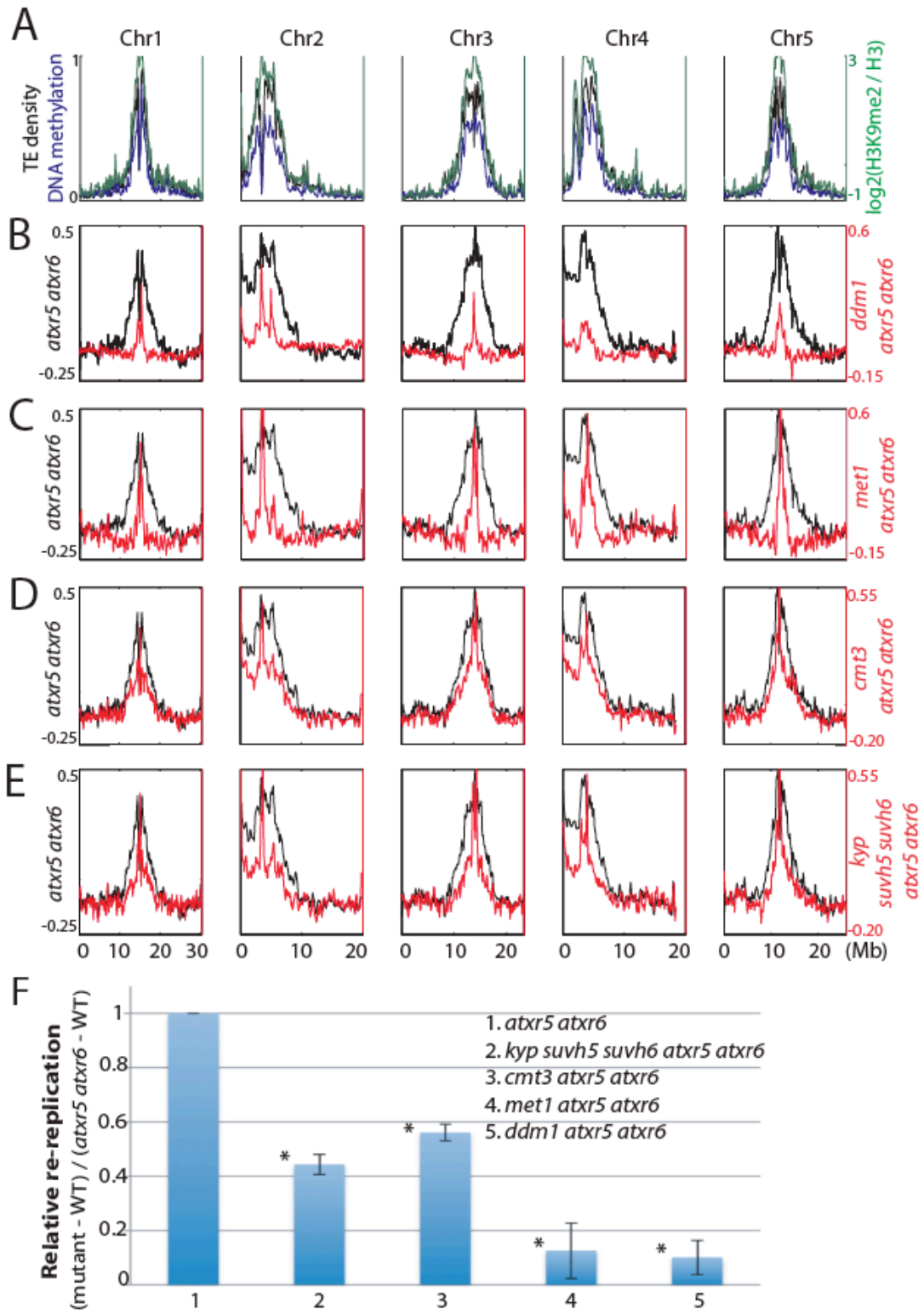
Figure 4-3

Figure 4-4

Figure 4-5

Figure 4-6

DNA content in 8C nuclei
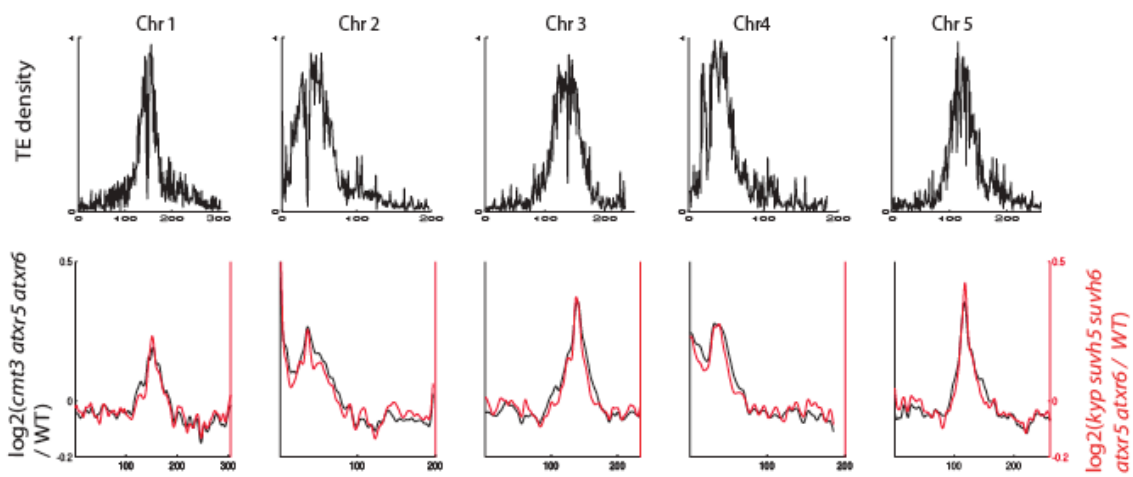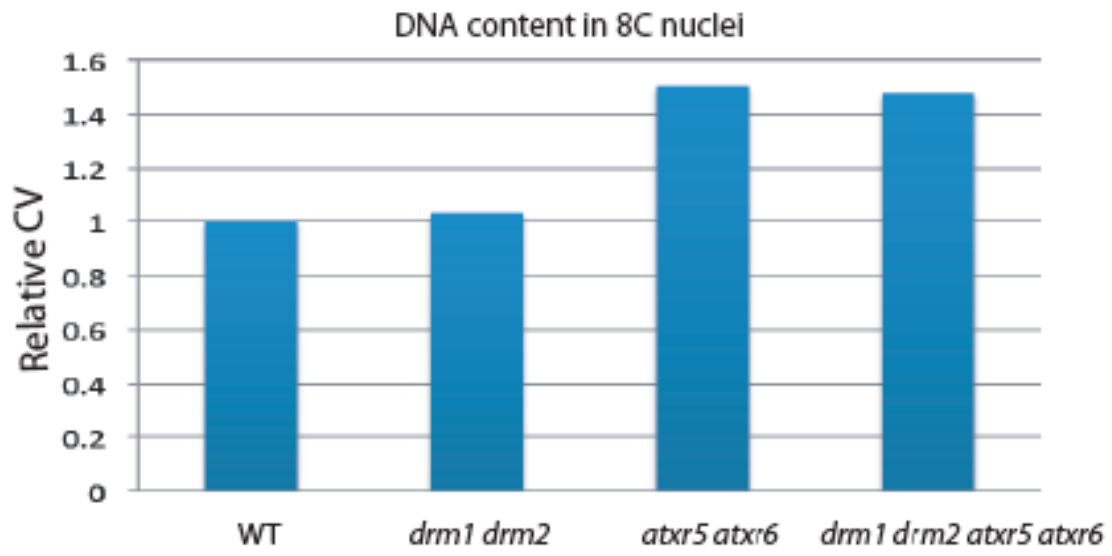
Figure 4-7

Figure 4-8

Figure 4-9

All TEs   TEs derepressed in *atxr5 atxr6* mutants

WT   *atxr5 atxr6*   WT   *atxr5 atxr6*

DNA methylation level

TSS   TTS   TSS   TTS   TSS   TTS   TSS   TTS

CG   CHG   CHH methylation levels

Figure 4-10

Figure 4-11

Normalized read densities within TEs that are
derepressed but not re-replicated in *atxr5 atxr6*
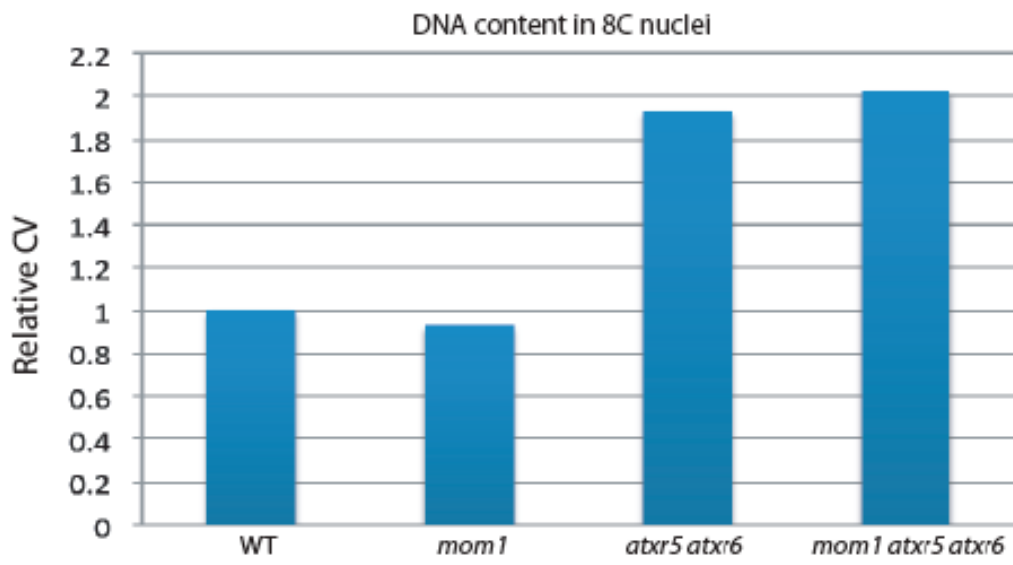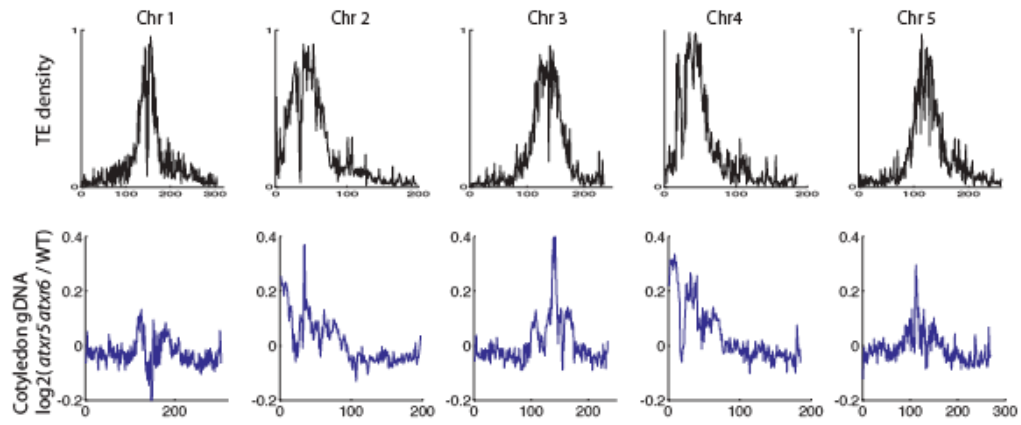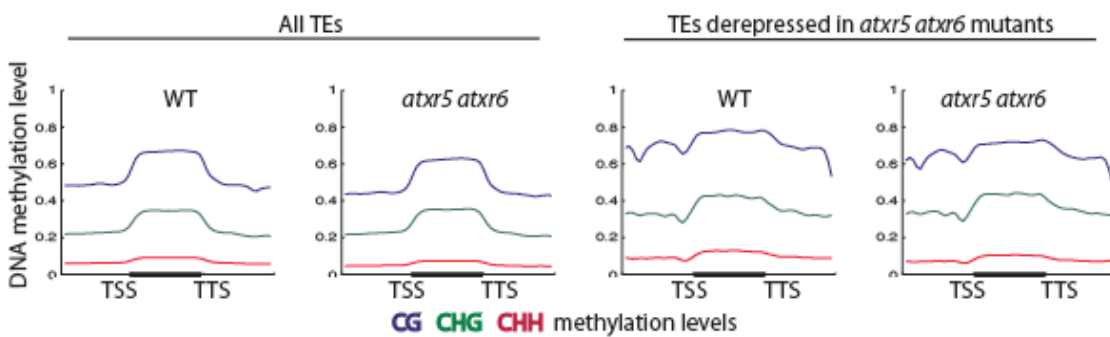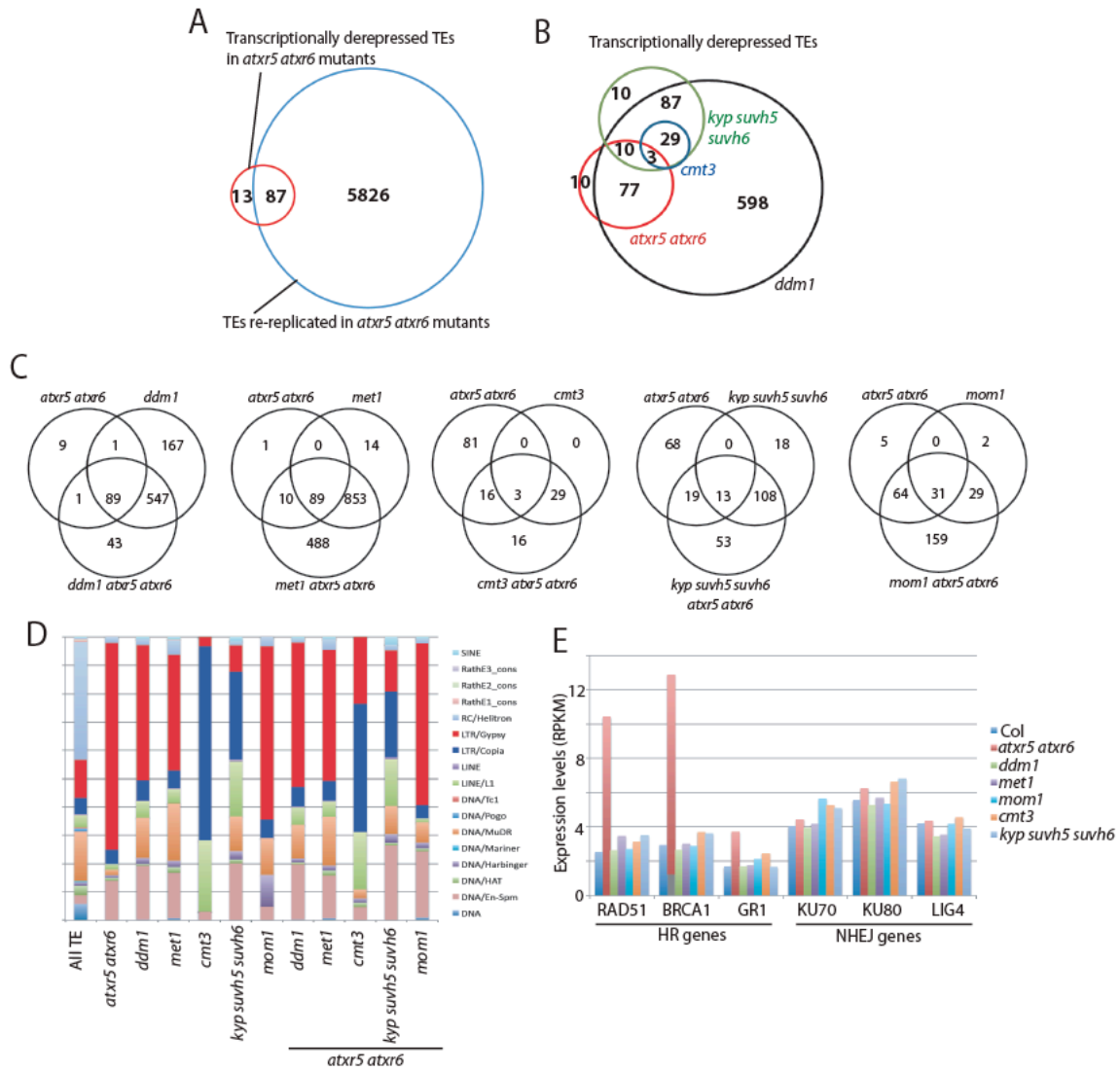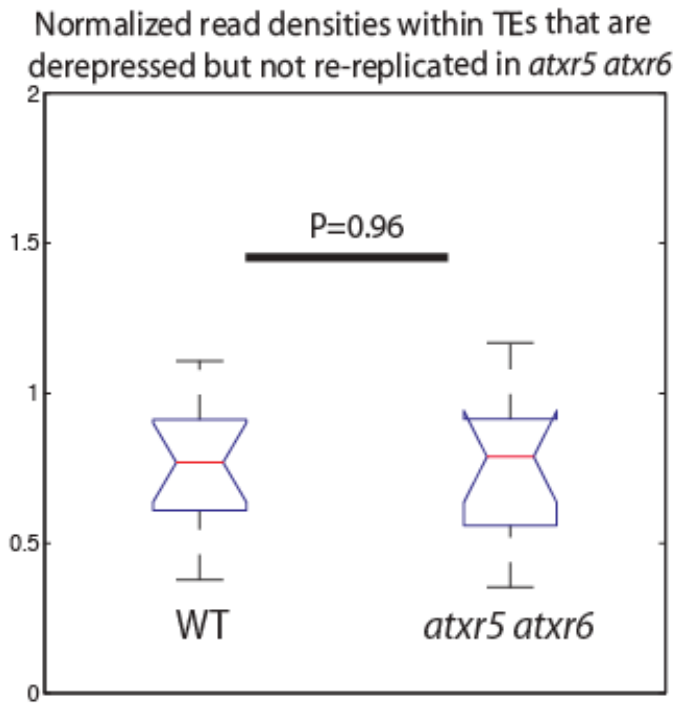
Figure 4-12

Figure 4-13

Figure 4-14

CG CHG CHH methylation levels
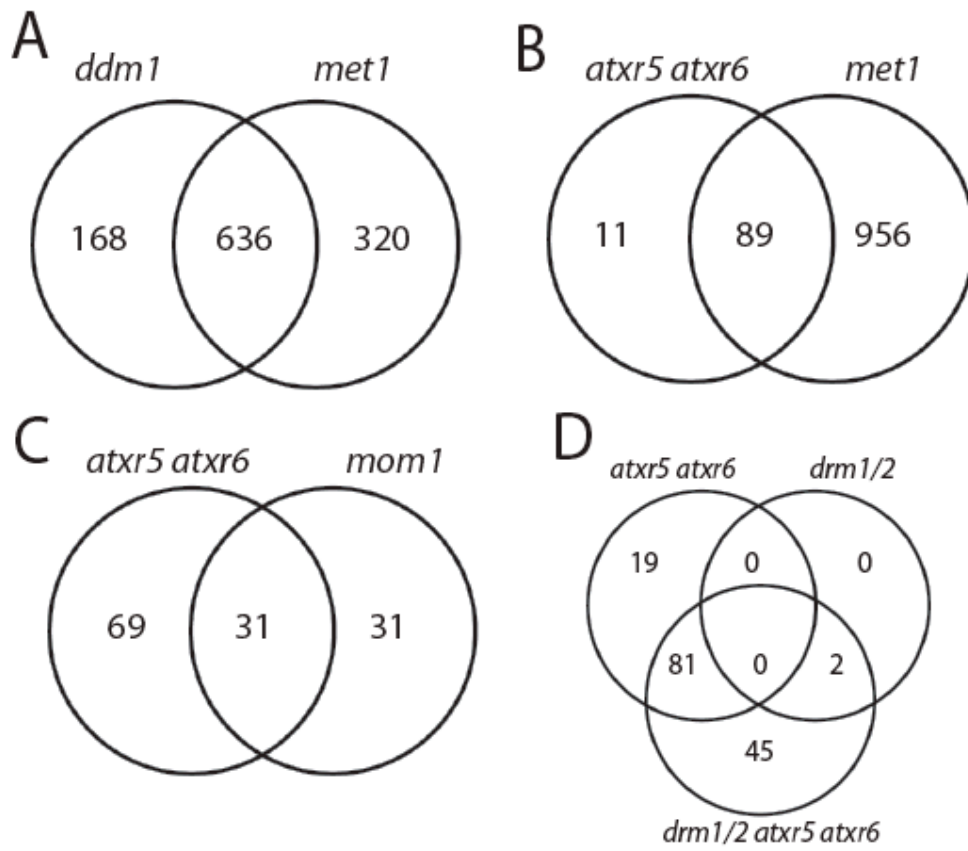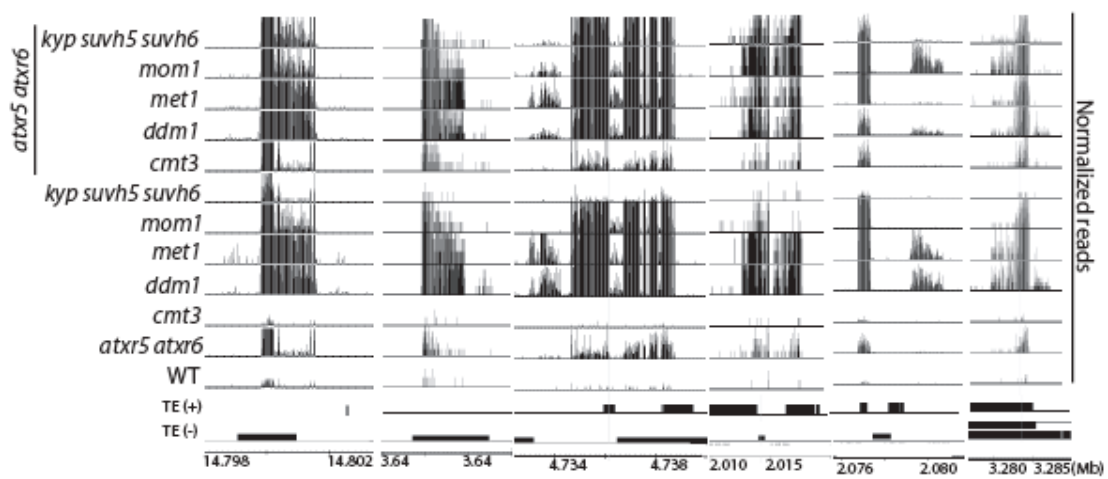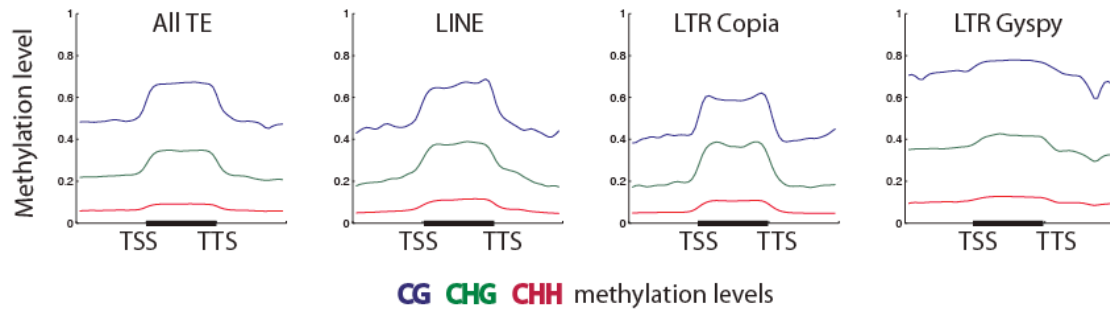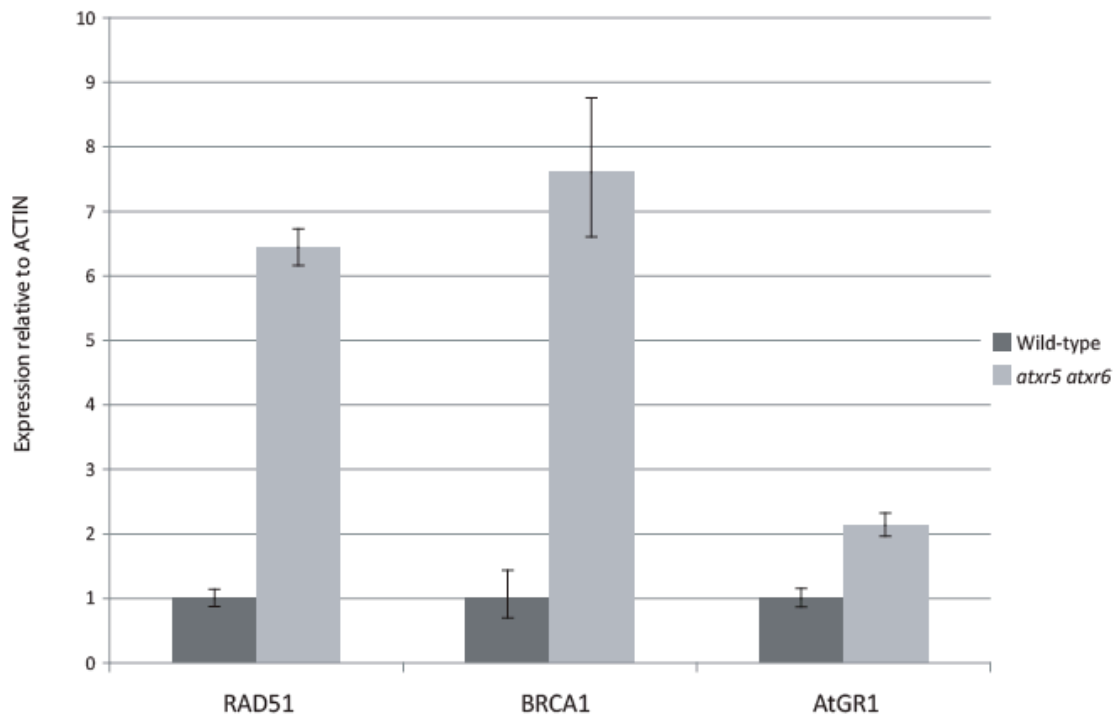
Figure 4-15

Figure 4-16
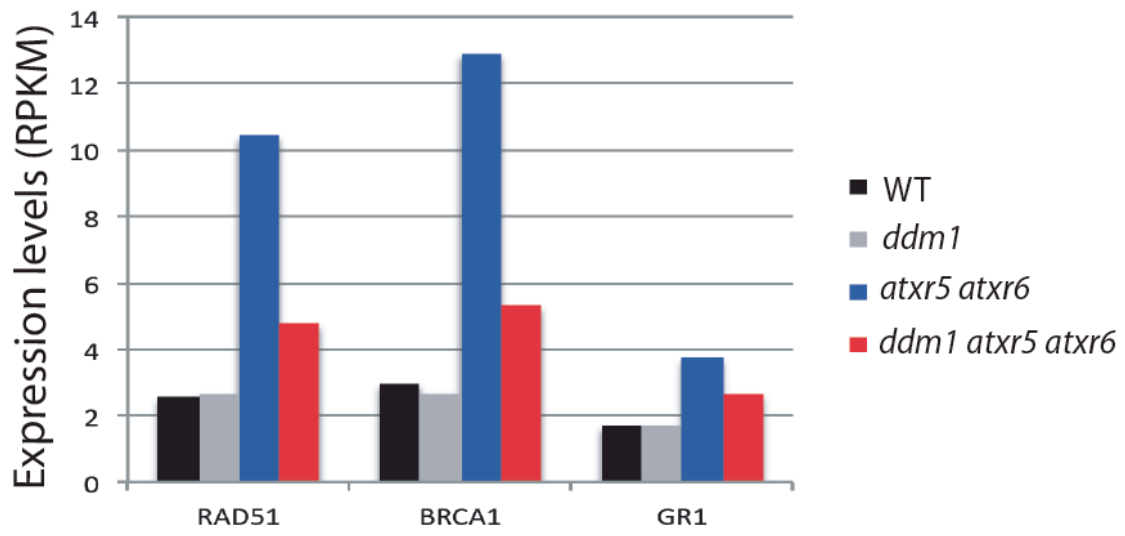
Table 4-1

gene	name	primers	sequence
AT5G09810	ACTIN	JP2452	TCGTGGTGGTGAGTTTGTTAC
		JP2453	CAGCATCATCACAAGCATCC
AT4G21070	BRCA1	JP6523	CATGTGCCTTTTGTCAGTGTTC
		JP6524	TGGAGCCCATTCAGCACAGTTT
AT5G20850	RAD51	JP7554	ACAGAGTTATATGGTGAGTTCC
		JP7487	ATCCTTGCAATCTGTTACACC
AT3G52115	AtGR1	JP7930	AGTCCCAACATACTTTCTCTGC
		JP8329	CTATACCTTTCAAATTCTCCCTCTC

REFERENCES

Abbas, T., Shibata, E., Park, J., Jha, S., Karnani, N., and Dutta, A. (2010). CRL4(Cdt2) regulates cell proliferation and histone gene expression by targeting PR-Set7/Set8 for degradation. Mol Cell *40*, 9-21.

Aggarwal, B.D., and Calvi, B.R. (2004). Chromatin regulates origin activity in Drosophila follicle cells. Nature *430*, 372-376.

Aladjem, M.I. (2007). Replication in context: dynamic regulation of DNA replication patterns in metazoans. Nat Rev Genet *8*, 588-600.

Amedeo, P., Habu, Y., Afsar, K., Mittelsten Scheid, O., and Paszkowski, J. (2000). Disruption of the plant gene MOM releases transcriptional silencing of methylated genes. Nature *405*, 203-206.

Antequera, F. (2004a). Genomic specification and epigenetic regulation of eukaryotic DNA replication origins. EMBO J *23*, 4365-4370.

Antequera, F. (2004b). Genomic specification and epigenetic regulation of eukaryotic DNA replication origins. The EMBO journal *23*, 4365-4370.

Archambault, V., Ikui, A.E., Drapkin, B.J., and Cross, F.R. (2005). Disruption of mechanisms that prevent rereplication triggers a DNA damage response. Mol Cell Biol *25*, 6707-6721.

Arias, E.E., and Walter, J.C. (2007). Strength in numbers: preventing rereplication via multiple mechanisms in eukaryotic cells. Genes Dev *21*, 497-518.

Aufsatz, W., Mette, M.F., van der Winden, J., Matzke, M., and Matzke, A.J. (2002). HDA6, a putative histone deacetylase needed to enhance DNA methylation induced by double-stranded RNA. Embo J *21*, 6832-6841.

Ausin, I., Mockler, T.C., Chory, J., and Jacobsen, S.E. (2009). IDN1 and IDN2 are required for de novo DNA methylation in Arabidopsis thaliana. Nat Struct Mol Biol *16*, 1325-1327.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Bartke, T., Vermeulen, M., Xhemalce, B., Robson, S.C., Mann, M., and Kouzarides, T. (2010). Nucleosome-interacting proteins regulated by DNA and histone methylation. Cell *143*, 470-484.

Baumbusch, L.O., Thorstensen, T., Krauss, V., Fischer, A., Naumann, K., Assalkhou, R., Schulz, I., Reuter, G., and Aalen, R.B. (2001). The Arabidopsis thaliana genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes. Nucleic Acids Res *29*, 4319-4333.

Baurle, I., Smith, L., Baulcombe, D.C., and Dean, C. (2007). Widespread role for the flowering-time regulators FCA and FPA in RNA-mediated chromatin silencing. Science *318*, 109-112.

Beck, D.B., Burton, A., Oda, H., Ziegler-Birling, C., Torres-Padilla, M.E., and Reinberg, D. (2012). The role of PR-Set7 in replication licensing depends on Suv4-20h. Genes Dev *26*, 2580-2589.

Becker, C., Hagmann, J., Muller, J., Koenig, D., Stegle, O., Borgwardt, K., and Weigel, D. (2011). Spontaneous epigenetic variation in the Arabidopsis thaliana methylome. Nature *480*, 245-249.

Bell, S.P., and Dutta, A. (2002). DNA replication in eukaryotic cells. Annu Rev Biochem *71*, 333-374.

Bell, S.P., and Stillman, B. (1992a). ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. Nature *357*, 128-134.

Bell, S.P., and Stillman, B. (1992b). ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. Nature *357*, 128-134.

Benjamini, Y., Hochberg, Y. (1995). Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series *B 57: 12*.

Berbenetz, N.M., Nislow, C., and Brown, G.W. (2010). Diversity of eukaryotic DNA replication origins revealed by genome-wide analysis of chromatin structure. PLoS Genet *6*.

Bernatavichute, Y.V., Zhang, X., Cokus, S., Pellegrini, M., and Jacobsen, S.E. (2008). Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in Arabidopsis thaliana. PLoS One *3*, e3156.

Berr, A., McCallum, E.J., Menard, R., Meyer, D., Fuchs, J., Dong, A., and Shen, W.H. (2010). Arabidopsis SET DOMAIN GROUP2 is required for H3K4 trimethylation and is crucial for both sporophyte and gametophyte development. Plant Cell *22*, 3232-3248.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes Dev *16*, 6-21.

Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Thurman, R.E., *et al*. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature *447*, 799-816.

Blow, J.J., and Dutta, A. (2005). Preventing re-replication of chromosomal DNA. Nat Rev Mol Cell Biol *6*, 476-486.

Blow, J.J., and Gillespie, P.J. (2008). Replication licensing and cancer--a fatal entanglement? Nat Rev Cancer *8*, 799-806.

Blow, J.J., and Laskey, R.A. (1988). A role for the nuclear envelope in controlling DNA replication within the cell cycle. Nature *332*, 546-548.

Cadoret, J.C., Meisch, F., Hassan-Zadeh, V., Luyten, I., Guillet, C., Duret, L., Quesneville, H., and Prioleau, M.N. (2008). Genome-wide studies highlight indirect links between human replication origins and gene regulation. Proc Natl Acad Sci U S A *105*, 15837-15842.

Cadoret, J.C., and Prioleau, M.N. (2010). Genome-wide approaches to determining origin distribution. Chromosome Res, 79-89.

Cao, X., and Jacobsen, S.E. (2002a). Locus-specific control of asymmetric and CpNpG methylation by the DRM and CMT3 methyltransferase genes. Proc Natl Acad Sci U S A *99 Suppl 4*, 16491-16498.

Cao, X., and Jacobsen, S.E. (2002b). Role of the arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing. Curr Biol *12*, 1138-1144.

Caro, E., Castellano, M.M., and Gutierrez, C. (2007). A chromatin link that couples cell division to root epidermis patterning in Arabidopsis. Nature *447*, 213-217.

Caro, E., Desvoyes, B., Ramirez-Parra, E., Sanchez, M.P., and Gutierrez, C. (2008). Endoreduplication control during plant development. SEB Exp Biol Ser *59*, 167-187.

Caro, E., and Gutierrez, C. (2007). A green GEM: intriguing analogies with animal geminin. Trends in cell biology *17*, 580-585.

Castellano, M.M., del Pozo, J.C., Ramirez-Parra, E., Brown, S., and Gutierrez, C. (2001). Expression and stability of Arabidopsis CDC6 are associated with endoreplication. Plant Cell *13*, 2671-2686.

Cayrou, C., Coulombe, P., Vigneron, A., Stanojcic, S., Ganier, O., Peiffer, I., Rivals, E., Puy, A., Laurent-Chabalier, S., Desprat, R., *et al*. (2011). Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. Genome Res *21*, 1438-1449.

Cedar, H., and Bergman, Y. (2009). Linking DNA methylation and histone modification: patterns and paradigms. Nat Rev Genet *10*, 295-304.

Chan, S.W., Zilberman, D., Xie, Z., Johansen, L.K., Carrington, J.C., and Jacobsen, S.E. (2004). RNA silencing genes control de novo DNA methylation. Science *303*, 1336.

Chen, P.Y., Cokus, S.J., and Pellegrini, M. (2010). BS Seeker: precise mapping for bisulfite sequencing. BMC Bioinformatics *11*, 203.

Chen, P.Y., Feng, S., Joo, J.W., Jacobsen, S.E., and Pellegrini, M. (2011). A comparative analysis of DNA methylation across human embryonic stem cell lines. Genome Biol *12*, R62.

Chodavarapu, R.K., Feng, S., Bernatavichute, Y.V., Chen, P.Y., Stroud, H., Yu, Y., Hetzel, J.A., Kuo, F., Kim, J., Cokus, S.J*., et al*. (2010). Relationship between nucleosome positioning and DNA methylation. Nature *466*, 388-392.

Cokus, S.J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C.D., Pradhan, S., Nelson, S.F., Pellegrini, M., and Jacobsen, S.E. (2008). Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature *452*, 215-219.

Costa, S., and Blow, J.J. (2007). The elusive determinants of replication origins. EMBO Rep *8*, 332-334.

Danis, E., Brodolin, K., Menut, S., Maiorano, D., Girard-Reydet, C., and Mechali, M. (2004). Specification of a DNA replication origin by a transcription complex. Nat Cell Biol *6*, 721-730.

Davidson, I.F., Li, A., and Blow, J.J. (2006a). Deregulated replication licensing causes DNA fragmentation consistent with head-to-tail fork collision. Mol Cell *24*, 433-443.

Davidson, I.F., Li, A., and Blow, J.J. (2006b). Deregulated replication licensing causes DNA fragmentation consistent with head-to-tail fork collision. Mol Cell *24*, 433-443.

de la Paz Sanchez, M.P., and Gutierrez, C. (2009). Arabidopsis ORC1 is a PHD-containing H3K4me3 effector that regulates transcription. Proc Natl Acad Sci USA *106*, 2065-2070.

Deleris, A., Stroud, H., Bernatavichute, Y., Johnson, E., Klein, G., Schubert, D., and Jacobsen, S.E. (2012). Loss of the DNA methyltransferase MET1 Induces H3K9 hypermethylation at PcG target genes and redistribution of H3K27 trimethylation to transposons in Arabidopsis thaliana. PLoS Genet *8*, e1003062.

DePamphilis, M.L., Blow, J.J., Ghosh, S., Saha, T., Noguchi, K., and Vassilev, A. (2006). Regulating the licensing of DNA replication origins in metazoa. Curr Opin Cell Biol *18*, 231-239.

Du, J., Zhong, X., Bernatavichute, Y.V., Stroud, H., Feng, S., Caro, E., Vashisht, A.A., Terragni, J., Chin, H.G., Tu, A*., et al*. (2012). Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. Cell *151*, 167-180.

Earley, K.W., Pontvianne, F., Wierzbicki, A.T., Blevins, T., Tucker, S., Costa-Nunes, P., Pontes, O., and Pikaard, C.S. (2010). Mechanisms of HDA6-mediated rRNA gene silencing: suppression of intergenic Pol II transcription and differential effects on maintenance versus siRNA-directed cytosine methylation. Genes Dev *24*, 1119-1132.

Earley, K.W., Shook, M.S., Brower-Toland, B., Hicks, L., and Pikaard, C.S. (2007). In vitro specificities of Arabidopsis co-activator histone acetyltransferases: implications for histone hyperacetylation in gene activation. Plant J *52*, 615-626.

Eaton, M.L., Galani, K., Kang, S., Bell, S.P., and MacAlpine, D.M. (2010). Conserved nucleosome positioning defines replication origins. Genes Dev *24*, 748-753.

Ebbs, M.L., and Bender, J. (2006). Locus-specific control of DNA methylation by the Arabidopsis SUVH5 histone methyltransferase. Plant Cell *18*, 1166-1176.

Feng, S., Cokus, S.J., Zhang, X., Chen, P.Y., Bostick, M., Goll, M.G., Hetzel, J., Jain, J., Strauss, S.H., Halpern, M.E.*, et al*. (2010). Conservation and divergence of methylation patterning in plants and animals. Proc Natl Acad Sci U S A *107*, 8689-8694.

Feng, S., and Jacobsen, S.E. (2011). Epigenetic modifications in plants: an evolutionary perspective. Current opinion in plant biology *14*, 179-186.

Feng, S., Rubbi, L., Jacobsen, S.E., and Pellegrini, M. (2011). Determining DNA methylation profiles using sequencing. Methods Mol Biol *733*, 223-238.

Finnegan, E.J., Peacock, W.J., and Dennis, E.S. (1996). Reduced DNA methylation in Arabidopsis thaliana results in abnormal plant development. Proc Natl Acad Sci U S A *93*, 8449-8454.

Fransz, P., ten Hoopen, R., and Tessadori, F. (2006). Composition and formation of heterochromatin in Arabidopsis thaliana. Chromosome Res *14*, 71-82.

Fuchs, J., Demidov, D., Houben, A., and Schubert, I. (2006). Chromosomal histone modification patterns--from conservation to diversity. Trends Plant Sci *11*, 199-208.

Galbraith, D.W., Harkins, K.R., and Knapp, S. (1991). Systemic Endopolyploidy in Arabidopsis thaliana. Plant Physiol *96*, 985-989.

Gilbert, D.M. (2004). In search of the holy replicator. Nature reviews *5*, 848-855.

Gomez, M. (2008). Controlled rereplication at DNA replication origins. Cell Cycle *7*, 1313-1314.

Gomez, M., and Antequera, F. (2008). Overreplication of short DNA regions during S phase in human cells. Genes Dev *22*, 375-385.

Goren, A., Tabib, A., Hecht, M., and Cedar, H. (2008). DNA replication timing of the human beta-globin domain is controlled by histone modification at the origin. Genes Dev *22*, 1319-1324.

Green, B.M., Finn, K.J., and Li, J.J. (2010). Loss of DNA replication control is a potent inducer of gene amplification. Science *329*, 943-946.

Green, B.M., and Li, J.J. (2005). Loss of rereplication control in Saccharomyces cerevisiae results in extensive DNA damage. Mol Biol Cell *16*, 421-432.

Greenberg, M.V., Ausin, I., Chan, S.W., Cokus, S.J., Cuperus, J.T., Feng, S., Law, J.A., Chu, C., Pellegrini, M., Carrington, J.C.*, et al*. (2011). Identification of genes required for de novo DNA methylation in Arabidopsis. Epigenetics : official journal of the DNA Methylation Society *6*, 344-354.

Gregoire, D., Brodolin, K., and Mechali, M. (2006). HoxB domain induction silences DNA replication origins in the locus and specifies a single origin at its boundary. EMBO Rep *7*, 812-816.

Grewal, S.I., and Elgin, S.C. (2002). Heterochromatin: new possibilities for the inheritance of structure. Curr Opin Genet Dev *12*, 178-187.

Grewal, S.I., and Elgin, S.C. (2007). Transcription and RNA interference in the formation of heterochromatin. Nature *447*, 399-406.

Gu, X., Jiang, D., Yang, W., Jacob, Y., Michaels, S.D., and He, Y. (2011). Arabidopsis homologs of retinoblastoma-associated protein 46/48 associate with a histone deacetylase to act redundantly in chromatin silencing. PLoS Genet *7*, e1002366.

Guo, L., Yu, Y., Law, J.A., and Zhang, X. (2010). SET DOMAIN GROUP2 is the major histone H3 lysine [corrected] 4 trimethyltransferase in Arabidopsis. Proc Natl Acad Sci U S A *107*, 18557-18562.

Haag, J.R., and Pikaard, C.S. (2011). Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. Nat Rev Mol Cell Biol *12*, 483-492.

Hartl, T., Boswell, C., Orr-Weaver, T.L., and Bosco, G. (2007). Developmentally regulated histone modifications in Drosophila follicle cells: initiation of gene amplification is associated with histone H3 and H4 hyperacetylation and H1 phosphorylation. Chromosoma *116*, 197-214.

Harvey, K.J., and Newport, J. (2003). CpG methylation of DNA restricts prereplication complex assembly in Xenopus egg extracts. Mol Cell Biol *23*, 6769-6779.

Hayashi, M., Katou, Y., Itoh, T., Tazumi, A., Yamada, Y., Takahashi, T., Nakagawa, T., Shirahige, K., and Masukata, H. (2007). Genome-wide localization of pre-RC sites and identification of replication origins in fission yeast. Embo J *26*, 1327-1339.

He, X.J., Hsu, Y.F., Pontes, O., Zhu, J., Lu, J., Bressan, R.A., Pikaard, C., Wang, C.S., and Zhu, J.K. (2009). NRPD4, a protein related to the RPB4 subunit of RNA polymerase II, is a component of RNA polymerases IV and V and is required for RNA-directed DNA methylation. Genes Dev *23*, 318-330.

Henderson, I.R., Zhang, X., Lu, C., Johnson, L., Meyers, B.C., Green, P.J., and Jacobsen, S.E. (2006). Dissecting Arabidopsis thaliana DICER function in small RNA processing, gene silencing and DNA methylation patterning. Nat Genet *38*, 721-725.

Hiratani, I., Ryba, T., Itoh, M., Yokochi, T., Schwaiger, M., Chang, C.W., Lyou, Y., Townes, T.M., Schubeler, D., and Gilbert, D.M. (2008). Global reorganization of replication domains during embryonic stem cell differentiation. PLoS Biol *6*, e245.

Hon, G.C., Hawkins, R.D., Caballero, O.L., Lo, C., Lister, R., Pelizzola, M., Valsesia, A., Ye, Z., Kuan, S., Edsall, L.E., *et al*. (2012). Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. Genome Res *22*, 246-258.

Huberman, J.A., and Riggs, A.D. (1968). On the mechanism of DNA replication in mammalian chromosomes. J Mol Biol *32*, 327-341.

Inagaki, S., Miura-Kamio, A., Nakamura, Y., Lu, F., Cui, X., Cao, X., Kimura, H., Saze, H., and Kakutani, T. (2010). Autocatalytic differentiation of epigenetic modifications within the Arabidopsis genome. Embo J *29*, 3496-3506.

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

Jackson, J.P., Lindroth, A.M., Cao, X., and Jacobsen, S.E. (2002). Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature *416*, 556-560.

Jacob, Y., Feng, S., LeBlanc, C.A., Bernatavichute, Y.V., Stroud, H., Cokus, S., Johnson, L.M., Pellegrini, M., Jacobsen, S.E., and Michaels, S.D. (2009). ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing. Nat Struct Mol Biol *16*, 763-768.

Jacob, Y., Stroud, H., Leblanc, C., Feng, S., Zhuo, L., Caro, E., Hassel, C., Gutierrez, C., Michaels, S.D., and Jacobsen, S.E. (2010). Regulation of heterochromatic DNA replication by histone H3 lysine 27 methyltransferases. Nature *466*, 987-991.

Jeddeloh, J.A., Stokes, T.L., and Richards, E.J. (1999). Maintenance of genomic methylation requires a SWI2/SNF2-like protein. Nat Genet *22*, 94-97.

Ji, H., and Wong, W.H. (2005). TileMap: create chromosomal map of tiling array hybridizations. Bioinformatics *21*, 3629-3636.

Jiang, H., and Wong, W.H. (2008). SeqMap: mapping massive amount of oligonucleotides to the genome. Bioinformatics *24*, 2395-2396.

Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K., and Felsenfeld, G. (2009). H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. Nat Genet *41*, 941-945.

Jin, J., Arias, E.E., Chen, J., Harper, J.W., and Walter, J.C. (2006). A family of diverse Cul4-Ddb1-interacting proteins includes Cdt2, which is required for S phase destruction of the replication factor Cdt1. Mol Cell *23*, 709-721.

Johnson, L., Cao, X., and Jacobsen, S. (2002). Interplay between two epigenetic marks. DNA methylation and histone H3 lysine 9 methylation. Curr Biol *12*, 1360-1367.

Johnson, L.M., Bostick, M., Zhang, X., Kraft, E., Henderson, I., Callis, J., and Jacobsen, S.E. (2007). The SRA methyl-cytosine-binding domain links DNA and histone methylation. Curr Biol *17*, 379-384.

Jorgensen, S., Elvers, I., Trelle, M.B., Menzel, T., Eskildsen, M., Jensen, O.N., Helleday, T., Helin, K., and Sorensen, C.S. (2007). The histone methyltransferase SET8 is required for S-phase progression. J Cell Biol *179*, 1337-1345.

Kakutani, T., Jeddeloh, J.A., Flowers, S.K., Munakata, K., and Richards, E.J. (1996). Developmental abnormalities and epimutations associated with DNA hypomethylation mutations. Proc Natl Acad Sci U S A *93*, 12406-12411.

Kaplan, T., Liu, C.L., Erkmann, J.A., Holik, J., Grunstein, M., Kaufman, P.D., Friedman, N., and Rando, O.J. (2008). Cell cycle- and chaperone-mediated regulation of H3K56ac incorporation in yeast. PLoS Genet *4*, e1000270.

Karnani, N., Taylor, C., Malhotra, A., and Dutta, A. (2007a). Pan-S replication patterns and chromosomal domains defined by genome-tiling arrays of ENCODE genomic areas. Genome Res *17*, 865-876.

Karnani, N., Taylor, C., Malhotra, A., and Dutta, A. (2007b). Pan-S replication patterns and chromosomal domains defined by genome-tiling arrays of ENCODE genomic areas. Genome Res *17*, 865-876.

171

Karnani, N., Taylor, C.M., Malhotra, A., and Dutta, A. (2010). Genomic study of replication initiation in human chromosomes reveals the influence of transcription regulation and chromatin structure on origin selection. Mol Biol Cell *21*, 393-404.

Kim, J., and Kipreos, E.T. (2008). Control of the Cdc6 replication licensing factor in metazoa: the role of nuclear export and the CUL4 ubiquitin ligase. Cell Cycle *7*, 146-150.

Kim, S.M., Dubey, D.D., and Huberman, J.A. (2003). Early-replicating heterochromatin. Genes Dev *17*, 330-335.

Kornberg, R.D., and Lorch, Y. (1999). Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. Cell *98*, 285-294.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693-705.

Kuo, A.J., Song, J., Cheung, P., Ishibe-Murakami, S., Yamazoe, S., Chen, J.K., Patel, D.J., and Gozani, O. (2012). The BAH domain of ORC1 links H4K20me2 to DNA replication licensing and Meier-Gorlin syndrome. Nature *484*, 115-119.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W.*, et al*. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

Laurent, L., Wong, E., Li, G., Huynh, T., Tsirigos, A., Ong, C.T., Low, H.M., Kin Sung, K.W., Rigoutsos, I., Loring, J.*, et al*. (2010). Dynamic changes in the human methylome during differentiation. Genome Res *20*, 320-331.

Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet *11*, 204-220.

Law, J.A., Vashisht, A.A., Wohlschlegel, J.A., and Jacobsen, S.E. (2011). SHH1, a homeodomain protein required for DNA methylation, as well as RDR2, RDM4, and chromatin remodeling factors, associate with RNA polymerase IV. PLoS Genet *7*, e1002195.

Lee, T.F., Gurazada, S.G., Zhai, J., Li, S., Simon, S.A., Matzke, M.A., Chen, X., and Meyers, B.C. (2012). RNA polymerase V-dependent small RNAs in Arabidopsis originate from small, intergenic loci including most SINE repeats. Epigenetics : official journal of the DNA Methylation Society *7*.

Lee, T.J., Pascuzzi, P.E., Settlage, S.B., Shultz, R.W., Tanurdzic, M., Rabinowicz, P.D., Menges, M., Zheng, P., Main, D., Murray, J.A.*, et al*. (2010). Arabidopsis thaliana chromosome 4 replicates in two phases that correlate with chromatin state. PLoS Genet *6*, e1000982.

Li, B., Carey, M., and Workman, J.L. (2007). The role of chromatin during transcription. Cell *128*, 707-719.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O.*, et al*. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289-293.

Lindroth, A.M., Cao, X., Jackson, J.P., Zilberman, D., McCallum, C.M., Henikoff, S., and Jacobsen, S.E. (2001). Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. Science *292*, 2077-2080.

Lindroth, A.M., Shultis, D., Jasencakova, Z., Fuchs, J., Johnson, L., Schubert, D., Patnaik, D., Pradhan, S., Goodrich, J., Schubert, I.*, et al*. (2004). Dual histone H3 methylation marks at lysines 9 and 27 required for interaction with CHROMOMETHYLASE3. Embo J *23*, 4286-4296.

Lippman, Z., Gendrel, A.V., Black, M., Vaughn, M.W., Dedhia, N., McCombie, W.R., Lavine, K., Mittal, V., May, B., Kasschau, K.D.*, et al*. (2004). Role of transposable elements in heterochromatin and epigenetic control. Nature *430*, 471-476.

Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., and Ecker, J.R. (2008). Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell *133*, 523-536.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.M.*, et al*. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature *462*, 315-322.

Liu, X., Yu, C.W., Duan, J., Luo, M., Wang, K., Tian, G., Cui, Y., and Wu, K. (2012). HDA6 directly interacts with DNA methyltransferase MET1 and maintains transposable element silencing in Arabidopsis. Plant Physiol *158*, 119-129.

Lovejoy, C.A., Lock, K., Yenamandra, A., and Cortez, D. (2006). DDB1 maintains genome integrity through regulation of Cdt1. Mol Cell Biol *26*, 7977-7990.

Lu, F., Cui, X., Zhang, S., Jenuwein, T., and Cao, X. (2011). Arabidopsis REF6 is a histone H3 lysine 27 demethylase. Nat Genet *43*, 715-719.

Lucas, I., Palakodeti, A., Jiang, Y., Young, D.J., Jiang, N., Fernald, A.A., and Le Beau, M.M. (2007). High-throughput mapping of origins of replication in human cells. EMBO Rep *8*, 770-777.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.

MacAlpine, D.M., Rodriguez, H.K., and Bell, S.P. (2004). Coordination of replication and transcription along a Drosophila chromosome. Genes & Dev *18*, 3094-3105.

MacAlpine, H.K., Gordan, R., Powell, S.K., Hartemink, A.J., and MacAlpine, D.M. (2010). Drosophila ORC localizes to open chromatin and marks sites of cohesin complex loading. Genome Res *20*, 201-211.

Malagnac, F., Bartee, L., and Bender, J. (2002). An Arabidopsis SET domain protein required for maintenance but not establishment of DNA methylation. Embo J *21*, 6842-6852.

Martin, M.M., Ryan, M., Kim, R., Zakas, A.L., Fu, H., Lin, C.M., Reinhold, W.C., Davis, S.R., Bilke, S., Liu, H.*, et al*. (2011). Genome-wide depletion of replication initiation events in highly transcribed regions. Genome Res *21*, 1822-1832.

Mathieu, O., Probst, A.V., and Paszkowski, J. (2005). Distinct regulation of histone H3 methylation at lysines 27 and 9 by CpG methylation in Arabidopsis. Embo J *24*, 2783-2791.

Mathieu, O., Reinders, J., Caikovski, M., Smathajitt, C., and Paszkowski, J. (2007). Transgenerational stability of the Arabidopsis epigenome is coordinated by CG methylation. Cell *130*, 851-862.

Mechali, M. (2010). Eukaryotic DNA replication origins: many choices for appropriate answers. Nat Rev Mol Cell Biol *11*, 728-738.

Meissner, A. (2010). Epigenetic modifications in pluripotent and differentiated cells. Nat Biotechnol *28*, 1079-1088.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B.*, et al*. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766-770.

Melixetian, M., Ballabeni, A., Masiero, L., Gasparini, P., Zamponi, R., Bartek, J., Lukas, J., and Helin, K. (2004). Loss of Geminin induces rereplication in the presence of functional p53. J Cell Biol *165*, 473-482.

Menges, M., Hennig, L., Gruissem, W., and Murray, J.A. (2002). Cell cycle-regulated gene expression in Arabidopsis. J Biol Chem *277*, 41987-42002.

Menges, M., Hennig, L., Gruissem, W., and Murray, J.A. (2003). Genome-wide gene expression in an Arabidopsis cell suspension. Plant Mol Biol *53*, 423-442.

Menges, M., and Murray, J.A. (2002). Synchronous Arabidopsis suspension cultures for analysis of cell-cycle gene activity. Plant J *30*, 203-212.

Menges, M., and Murray, J.A. (2006). Synchronization, transformation, and cryopreservation of suspension-cultured cells. Methods in molecular biology (Clifton, NJ *323*, 45-61.

Mesner, L.D., Valsakumar, V., Karnani, N., Dutta, A., Hamlin, J.L., and Bekiranov, S. (2011). Bubble-chip analysis of human origin distributions demonstrates on a genomic scale significant clustering into zones and significant association with transcription. Genome Res *21*, 377-389.

Miotto, B., and Struhl, K. (2010). HBO1 Histone Acetylase Activity Is Essential for DNA Replication Licensing and Inhibited by Geminin. Mol Cell *37*, 57-66.

Mittelsten Scheid, O., Probst, A.V., Afsar, K., and Paszkowski, J. (2002). Two regulatory levels of transcriptional gene silencing in Arabidopsis. Proc Natl Acad Sci U S A *99*, 13659-13662.

Miura, A., Nakamura, M., Inagaki, S., Kobayashi, A., Saze, H., and Kakutani, T. (2009). An Arabidopsis jmjC domain protein protects transcribed genes from DNA methylation at CHG sites. Embo J *28*, 1078-1086.

Moldovan, G.L., Pfander, B., and Jentsch, S. (2007). PCNA, the maestro of the replication fork. Cell *129*, 665-679.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods *5*, 621-628.

Musselman, C.A., and Kutateladze, T.G. (2009). PHD fingers: epigenetic effectors and potential drug targets. Mol Interv *9*, 314-323.

Nanty, L., Carbajosa, G., Heap, G.A., Ratnieks, F., van Heel, D.A., Down, T.A., and Rakyan, V.K. (2011). Comparative methylomics reveals gene-body H3K36me3 in Drosophila predicts DNA methylation and CpG landscapes in other invertebrates. Genome Res *21*, 1841-1850.

Obayashi, T., Hayashi, S., Saeki, M., Ohta, H., and Kinoshita, K. (2009). ATTED-II provides coexpressed gene networks for Arabidopsis. Nucleic Acids Res *37*, D987-991.

Onodera, Y., Nakagawa, K., Haag, J.R., Pikaard, D., Mikami, T., Ream, T., Ito, Y., and Pikaard, C.S. (2008). Sex-biased lethality or transmission of defective transcription machinery in Arabidopsis. Genetics *180*, 207-218.

Pak, D.T., Pflumm, M., Chesnokov, I., Huang, D.W., Kellum, R., Marr, J., Romanowski, P., and Botchan, M.R. (1997). Association of the origin recognition complex with heterochromatin and HP1 in higher eukaryotes. Cell *91*, 311-323.

Pauler, F.M., Sloane, M.A., Huang, R., Regha, K., Koerner, M.V., Tamir, I., Sommer, A., Aszodi, A., Jenuwein, T., and Barlow, D.P. (2009). H3K27me3 forms BLOCs over silent genes and intergenic regions and specifies a histone banding pattern on a mouse autosomal chromosome. Genome Res *19*, 221-233.

Penterman, J., Zilberman, D., Huh, J.H., Ballinger, T., Henikoff, S., and Fischer, R.L. (2007). DNA demethylation in the Arabidopsis genome. Proc Natl Acad Sci U S A *104*, 6752-6757.

Pontvianne, F., Blevins, T., and Pikaard, C.S. (2010). Arabidopsis Histone Lysine Methyltransferases. Adv Bot Res *53*, 1-22.

Prasanth, S.G., Shen, Z., Prasanth, K.V., and Stillman, B. (2010). Human origin recognition complex is essential for HP1 binding to chromatin and heterochromatin organization. Proc Natl Acad Sci U S A *107*, 15093-15098.

Prioleau, M.N., Gendron, M.C., and Hyrien, O. (2003). Replication of the chicken beta-globin locus: early-firing origins at the 5' HS4 insulator and the rho- and betaA-globin genes show opposite epigenetic modifications. Mol Cell Biol *23*, 3536-3549.

Ptashne, M., and Gann, A. (2002). Genes and Signals. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Raghuraman, M.K., Winzeler, E.A., Collingwood, D., Hunt, S., Wodicka, L., Conway, A., Lockhart, D.J., Davis, R.W., Brewer, B.J., and Fangman, W.L. (2001). Replication dynamics of the yeast genome. Science *294*, 115-121.

Ramirez-Parra, E., and Gutierrez, C. (2007). E2F regulates FASCIATA1, a chromatin assembly gene whose loss switches on the endocycle and activates gene expression by changing the epigenetic status. Plant Physiol *144*, 105-120.

Raynaud, C., Sozzani, R., Glab, N., Domenichini, S., Perennes, C., Cella, R., Kondorosi, E., and Bergounioux, C. (2006). Two cell-cycle regulated SET-domain proteins interact with proliferating cell nuclear antigen (PCNA) in Arabidopsis. Plant J *47*, 395-407.

Rein, T., Kobayashi, T., Malott, M., Leffak, M., and DePamphilis, M.L. (1999). DNA methylation at mammalian replication origins. J Biol Chem *274*, 25792-25800.

Remus, D., Beall, E.L., and Botchan, M.R. (2004). DNA topology, not DNA sequence, is a critical determinant for Drosophila ORC-DNA binding. Embo J *23*, 897-907.

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

Rusche, L.N., Kirchmaier, A.L., and Rine, J. (2003). The establishment, inheritance, and function of silenced chromatin in Saccharomyces cerevisiae. Annu Rev Biochem *72*, 481-516.

Sanchez, M., Caro, E., Desvoyes, B., Ramirez-Parra, E., and Gutierrez, C. (2008). Chromatin dynamics during the plant cell cycle. Seminars in cell & developmental biology *19*, 537- 546.

Sanchez, M.P., and Gutierrez, C. (2009). Arabidopsis ORC1 is a PHD-containing H3K4me3 effector that regulates transcription. Proc Natl Acad Sci U S A *106*, 2065-2070.

SanMiguel, P., Tikhonov, A., Jin, Y.K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P.S., Edwards, K.J., Lee, M., Avramova, Z., *et al.* (1996). Nested retrotransposons in the intergenic regions of the maize genome. Science *274*, 765-768.

Schepers, A., and Papior, P. (2010). Why are we where we are? Understanding replication origins and initiation sites in eukaryotes using ChIP-approaches. Chromosome Res, 63-77.

Schmitz, R.J., Schultz, M.D., Lewsey, M.G., O'Malley, R.C., Urich, M.A., Libiger, O., Schork, N.J., and Ecker, J.R. (2011). Transgenerational epigenetic instability is a source of novel methylation variants. Science *334*, 369-373.

Schonrock, N., Exner, V., Probst, A., Gruissem, W., and Hennig, L. (2006). Functional genomic analysis of CAF-1 mutants in Arabidopsis thaliana. J Biol Chem *281*, 9560-9568.

Schotta, G., Ebert, A., and Reuter, G. (2003). SU(VAR)3-9 is a conserved key function in heterochromatic gene silencing. Genetica *117*, 149-158.

Schwaiger, M., Stadler, M.B., Bell, O., Kohler, H., Oakeley, E.J., and Schubeler, D. (2009). Chromatin state marks cell-type- and gender-specific replication of the Drosophila genome. Genes Dev *23*, 589-601.

Sequeira-Mendes, J., Diaz-Uriarte, R., Apedaile, A., Huntley, D., Brockdorff, N., and Gomez, M. (2009). Transcription initiation activity sets replication origin efficiency in mammalian cells. PLoS Genet *5*, e1000446.

Shultz, R.W., Tatineni, V.M., Hanley-Bowdoin, L., and Thompson, W.F. (2007). Genome-wide analysis of the core DNA replication machinery in the higher plants Arabidopsis and rice. Plant Physiol *144*, 1697-1714.

Soni, R., Carmichael, J.P., Shah, Z.H., and Murray, J.A. (1995). A family of cyclin D homologs from plants differentially controlled by growth regulators and containing the conserved retinoblastoma protein interaction motif. Plant Cell *7*, 85-103.

Soppe, W.J., Jasencakova, Z., Houben, A., Kakutani, T., Meister, A., Huang, M.S., Jacobsen, S.E., Schubert, I., and Fransz, P.F. (2002). DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in Arabidopsis. Embo J *21*, 6549-6559.

Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. Nature *403*, 41-45.

Stroud, H., Ding, B., Simon, S.A., Feng, S., Bellizzi, M., Pellegrini, M., Wang, G.L., Meyers, B.C., and Jacobsen, S.E. (2013a). Plants regenerated from tissue culture contain stable epigenome changes in rice. elife *2*, e00354.

Stroud, H., Greenberg, M.V., Feng, S., Bernatavichute, Y.V., and Jacobsen, S.E. (2013b). Comprehensive Analysis of Silencing Mutants Reveals Complex Regulation of the Arabidopsis Methylome. Cell.

Stroud, H., Hale, C.J., Feng, S., Caro, E., Jacob, Y., Michaels, S.D., and Jacobsen, S.E. (2012a). DNA methyltransferases are required to induce heterochromatic re-replication in Arabidopsis. PLoS Genet *8*, e1002808.

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

Tada, S., Li, A., Maiorano, D., Mechali, M., and Blow, J.J. (2001). Repression of origin assembly in metaphase depends on inhibition of RLF-B/Cdt1 by geminin. Nat Cell Biol *3*, 107-113.

Takahashi, T.S., Yiu, P., Chou, M.F., Gygi, S., and Walter, J.C. (2004). Recruitment of Xenopus Scc2 and cohesin to chromatin requires the pre-replication complex. Nat Cell Biol *6*, 991-996.

Tardat, M., Brustel, J., Kirsh, O., Lefevbre, C., Callanan, M., Sardet, C., and Julien, E. (2010). The histone H4 Lys 20 methyltransferase PR-Set7 regulates replication origins in mammalian cells. Nat Cell Biol *12*, 1086-1093.

Tardat, M., Murr, R., Herceg, Z., Sardet, C., and Julien, E. (2007). PR-Set7-dependent lysine methylation ensures genome replication and stability through S phase. J Cell Biol *179*, 1413-1426.

Tariq, M., Saze, H., Probst, A.V., Lichota, J., Habu, Y., and Paszkowski, J. (2003). Erasure of CpG methylation in Arabidopsis alters patterns of histone H3 methylation in heterochromatin. Proc Natl Acad Sci U S A *100*, 8823-8827.

Teixeira, F.K., Heredia, F., Sarazin, A., Roudier, F., Boccara, M., Ciaudo, C., Cruaud, C., Poulain, J., Berdasco, M., Fraga, M.F.*, et al*. (2009). A role for RNAi in the selective correction of DNA methylation defects. Science *323*, 1600-1604.

TheArabidopsisGenomeInitiative (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature *408*, 796-815.

Thommes, P., and Blow, J.J. (1997). The DNA replication licensing system. Cancer Surv *29*, 75-90.

To, T.K., Kim, J.M., Matsui, A., Kurihara, Y., Morosawa, T., Ishida, J., Tanaka, M., Endo, T., Kakutani, T., Toyoda, T.*, et al*. (2011). Arabidopsis HDA6 regulates locus-directed heterochromatin silencing in cooperation with MET1. PLoS Genet *7*, e1002055.

Tran, R.K., Zilberman, D., de Bustos, C., Ditt, R.F., Henikoff, J.G., Lindroth, A.M., Delrow, J., Boyle, T., Kwong, S., Bryson, T.D.*, et al*. (2005). Chromatin and siRNA pathways cooperate to maintain DNA methylation of small transposable elements in Arabidopsis. Genome Biol *6*, R90.

Turck, F., Roudier, F., Farrona, S., Martin-Magniette, M.L., Guillaume, E., Buisine, N., Gagnot, S., Martienssen, R.A., Coupland, G., and Colot, V. (2007). Arabidopsis TFL2/LHP1 specifically associates with genes marked by trimethylation of histone H3 lysine 27. PLoS Genet *3*, e86.

Van't Hof, J., Kuniyuki, A., and Bjerknes, C.A. (1978). The size and number of replicon families of chromosomal DNA of Arabidopsis thaliana. Chromosoma *68*, 269-285.

Vashee, S., Cvetic, C., Lu, W., Simancek, P., Kelly, T.J., and Walter, J.C. (2003). Sequence-independent DNA binding and replication initiation by the human origin recognition complex. Genes Dev *17*, 1894-1908.

Vaziri, C., Saxena, S., Jeon, Y., Lee, C., Murata, K., Machida, Y., Wagle, N., Hwang, D.S., and Dutta, A. (2003). A p53-dependent checkpoint pathway prevents rereplication. Mol Cell *11*, 997-1008.

Vermeulen, M., Eberl, H.C., Matarese, F., Marks, H., Denissov, S., Butter, F., Lee, K.K., Olsen, J.V., Hyman, A.A., Stunnenberg, H.G.*, et al*. (2010). Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. Cell *142*, 967-980.

Voinnet, O. (2008). Use, tolerance and avoidance of amplified RNA silencing by plants. Trends Plant Sci *13*, 317-328.

Vongs, A., Kakutani, T., Martienssen, R.A., and Richards, E.J. (1993). Arabidopsis thaliana DNA methylation mutants. Science *260*, 1926-1928.

Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., *et al*. (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520-562.

Weinhofer, I., Hehenberger, E., Roszak, P., Hennig, L., and Kohler, C. (2010). H3K27me3 profiling of the endosperm implies exclusion of polycomb group protein targeting by DNA methylation. PLoS Genet *6*.

White, E.J., Emanuelsson, O., Scalzo, D., Royce, T., Kosak, S., Oakeley, E.J., Weissman, S., Gerstein, M., Groudine, M., Snyder, M., *et al*. (2004). DNA replication-timing analysis of human chromosome 22 at high resolution and different developmental states. Proc Natl Acad Sci U S A *101*, 17771-17776.

Wohlschlegel, J.A., Dwyer, B.T., Dhar, S.K., Cvetic, C., Walter, J.C., and Dutta, A. (2000). Inhibition of eukaryotic DNA replication by geminin binding to Cdt1. Science *290*, 2309-2312.

Woo, H.R., Dittmer, T.A., and Richards, E.J. (2008). Three SRA-domain methylcytosine-binding proteins cooperate to maintain global CpG methylation and epigenetic silencing in Arabidopsis. PLoS Genet *4*, e1000156.

Wyrick, J.J. (2001). Genome-wide distribution of ORC and MCM proteins in S. cerevisiae: high-resolution mapping of replication origins. Science *294*, 2357-2360.

Wyrick, J.J., Aparicio, J.G., Chen, T., Barnett, J.D., Jennings, E.G., Young, R.A., Bell, S.P., and Aparicio, O.M. (2001). Genome-wide distribution of ORC and MCM proteins in S. cerevisiae: high-resolution mapping of replication origins. Science *294*, 2357-2360.

Xie, W., Barr, C.L., Kim, A., Yue, F., Lee, A.Y., Eubanks, J., Dempster, E.L., and Ren, B. (2012). Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. Cell *148*, 816-831.

Xu, J., Yanagisawa, Y., Tsankov, A.M., Hart, C., Aoki, K., Kommajosyula, N., Steinmann, K.E., Bochicchio, J., Russ, C., Regev, A., *et al*. (2012). Genome-wide identification and characterization of replication origins by deep sequencing. Genome Biol *13*, R27.

Xu, L., Zhao, Z., Dong, A., Soubigou-Taconnat, L., Renou, J.P., Steinmetz, A., and Shen, W.H. (2008). Di- and tri- but not monomethylation on histone H3 lysine 36 marks active transcription of genes involved in flowering time regulation and other processes in Arabidopsis thaliana. Mol Cell Biol *28*, 1348-1360.

Yin, S., Deng, W., Hu, L., and Kong, X. (2009). The impact of nucleosome positioning on the organization of replication origins in eukaryotes. Biochem Biophys Res Commun *385*, 363-368.

Zemach, A., Kim, M.Y., Hsieh, P.H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., Harmer, S.L., and Zilberman, D. (2013). The Arabidopsis Nucleosome Remodeler DDM1 Allows DNA Methyltransferases to Access H1-Containing Heterochromatin. Cell *153*, 193-205.

Zemach, A., McDaniel, I.E., Silva, P., and Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science *328*, 916-919.

Zhang, K., Sridhar, V.V., Zhu, J., Kapoor, A., and Zhu, J.K. (2007a). Distinctive core histone post-translational modification patterns in Arabidopsis thaliana. PLoS One *2*, e1210.

Zhang, X., Bernatavichute, Y.V., Cokus, S., Pellegrini, M., and Jacobsen, S.E. (2009). Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in Arabidopsis thaliana. Genome Biol *10*, R62.

Zhang, X., Clarenz, O., Cokus, S., Bernatavichute, Y.V., Pellegrini, M., Goodrich, J., and Jacobsen, S.E. (2007b). Whole-genome analysis of histone H3 lysine 27 trimethylation in Arabidopsis. PLoS Biol *5*, e129.

Zhang, X., Germann, S., Blus, B.J., Khorasanizadeh, S., Gaudin, V., and Jacobsen, S.E. (2007c). The Arabidopsis LHP1 protein colocalizes with histone H3 Lys27 trimethylation. Nat Struct Mol Biol *14*, 869-871.

Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S.W., Chen, H., Henderson, I.R., Shinn, P., Pellegrini, M., Jacobsen, S.E., *et al*. (2006). Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. Cell *126*, 1189-1201.

Zheng, B., Wang, Z., Li, S., Yu, B., Liu, J.Y., and Chen, X. (2009). Intergenic transcription by RNA polymerase II coordinates Pol IV and Pol V in siRNA-directed transcriptional gene silencing in Arabidopsis. Genes Dev *23*, 2850-2860.

Zheng, X., Zhu, J., Kapoor, A., and Zhu, J.K. (2007). Role of Arabidopsis AGO6 in siRNA accumulation, DNA methylation and transcriptional gene silencing. Embo J *26*, 1691-1701.

Zhong, X., Hale, C.J., Law, J.A., Johnson, L.M., Feng, S., Tu, A., and Jacobsen, S.E. (2012). DDR complex facilitates global association of RNA polymerase V to promoters and evolutionarily young transposons. Nat Struct Mol Biol *19*, 870-875.

Zhou, J., Chau, C.M., Deng, Z., Shiekhattar, R., Spindler, M.P., Schepers, A., and Lieberman, P.M. (2005). Cell cycle regulation of chromatin at an origin of DNA replication. Embo J *24*, 1406-1417.

Zhu, W., Chen, Y., and Dutta, A. (2004). Rereplication by depletion of geminin is seen regardless of p53 status and activates a G2/M checkpoint. Mol Cell Biol *24*, 7140-7150.

Zhu, W., and Dutta, A. (2006). An ATR- and BRCA1-mediated Fanconi anemia pathway is required for activating the G2/M checkpoint and DNA damage repair upon rereplication. Mol Cell Biol *26*, 4601-4611.

Zilberman, D., Coleman-Derr, D., Ballinger, T., and Henikoff, S. (2008). Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. Nature *456*, 125-129.

Zilberman, D., Gehring, M., Tran, R.K., Ballinger, T., and Henikoff, S. (2007). Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. Nat Genet *39*, 61-69.