CellPress

Molecular Cell
Article

Genome-wide Hi-C Analyses in Wild-Type and Mutants Reveal High-Resolution Chromatin Interactions in *Arabidopsis*

Suhua Feng,[1,2,3,5] Shawn J. Cokus,[1,5] Veit Schubert,[4] Jixian Zhai,[1] Matteo Pellegrini,[1] and Steven E. Jacobsen[1,2,3,*]
[1]Department of Molecular, Cell and Developmental Biology
[2]Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research
[3]Howard Hughes Medical Institute
University of California, Los Angeles, Los Angeles, CA 90095, USA
[4]Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), 06466 Gatersleben, Germany
[5]Co-first author
*Correspondence: jacobsen@ucla.edu
http://dx.doi.org/10.1016/j.molcel.2014.07.008

SUMMARY

Chromosomes form 3D structures that are critical to the regulation of cellular and genetic processes. Here, we present a study of global chromatin interaction patterns in *Arabidopsis thaliana*. Our genome-wide approach confirmed interactions that were previously observed by other methods as well as uncovered long-range interactions such as those among small heterochromatic regions embedded in euchromatic arms. We also found that interactions are correlated with various epigenetic marks that are localized in active or silenced chromatin. *Arabidopsis* chromosomes do not contain large local interactive domains that resemble the topological domains described in animals but, instead, contain relatively small interactive regions scattered around the genome that contain H3K27me3 or H3K9me2. We generated interaction maps in mutants that are defective in specific epigenetic pathways and found altered interaction patterns that correlate with changes in the epigenome. These analyses provide further insights into molecular mechanisms of epigenetic regulation of the genome.

INTRODUCTION

Spatial organization of the genome and higher-order chromosome structures can be studied by a series of chromosome conformation capture (3C)-based approaches (Dekker et al., 2013). Eukaryotic chromosomes are organized in three dimensions inside the nucleus (Dixon et al., 2012; Gibcus and Dekker, 2013; Jin et al., 2013; Lieberman-Aiden et al., 2009; Sexton et al., 2012). Folding of chromosomes leads different regions of chromatin to interact with each other, which often bears important functional significance such as maintaining genome integrity, compartmentalizing silent chromatin, regulating gene expression, regulating DNA replication, and forming highly interactive local domains (Dixon et al., 2012; Jin et al., 2013; Sexton et al., 2012; Zhang et al., 2012). Chromatin-interaction studies have uncovered looping interactions in genes, such as those between enhancers and CTCF-binding sites with their target promoters, as well as interactions among centromeres, telomeres, early origins of replication, and chromosomal breakpoints (Crevillén et al., 2013; Dekker et al., 2013; Dixon et al., 2012; Duan et al., 2010; Jin et al., 2013; Li et al., 2012; Sanyal et al., 2012; Shen et al., 2012). Highly expressed and transcriptionally related genes have also been shown to interact with each other (Dekker et al., 2013; Gibcus and Dekker, 2013; Osborne et al., 2004; Tanizawa et al., 2010). Interestingly, mammalian chromosomes often have subchromosomal compartments preferentially enriched for either gene-rich transcriptionally active or gene-poor transcriptionally silent chromatin regions (Gibcus and Dekker, 2013; Lieberman-Aiden et al., 2009).

Transcriptionally active and silent chromatin regions are distinguished by distinct epigenetic marks such as DNA methylation and particular histone modifications (Feng and Jacobsen, 2011). For example, in *Arabidopsis*, silent heterochromatin is marked by histone H3 lysine 9 dimethylation (H3K9me2) and H3K27 mono-methylation (H3K27me1) (Bernatavichute et al., 2008; Jacob et al., 2009), whereas H3K4 trimethylation (H3K4me3) marks actively transcribed regions and is absent from heterochromatin (Zhang et al., 2009). Constitutively highly expressed protein-coding genes are also modified by DNA methylation occurring in the CG dinucleotide context within the transcribed region (gene body methylation) (Feng et al., 2010; Zemach et al., 2010). On the other hand, H3K27me3 (mediated by Polycomb pathway) marks silent genes and genes with tissue-specific expression patterns and is highly anticorrelated with gene body methylation (Zhang et al., 2007). DNA methylation in non-CG sequence contexts (CHG and CHH, where H = A, C, or T) is often found on repetitive DNA, such as transposable elements (“TEs”) that form heterochromatin (Feng and Jacobsen, 2011). These observations raise the possibility that epigenetic marks might influence the architecture and interaction patterns of the chromosomes. This idea is consistent with a recent study that employed a circular chromosome conformation capture (4C) technique in *Arabidopsis* and found that interacting regions tend to have similar epigenetic landscapes (Grob et al., 2013).

Chromosomes from mammals, *Drosophila*, and bacteria form large local interactive domains, termed "topological," "physical," and "chromosome-interaction" domains, respectively (Dixon et al., 2012; Le et al., 2013; Sexton et al., 2012). Although the sizes of these domains vary in different organisms, it is common in eukaryotes (mammals and *Drosophila*) that these domains correlate extensively with active or repressive epigenetic modifications (Dixon et al., 2012; Sexton et al., 2012). Furthermore, the boundaries of topological and physical domains are often marked by the insulator binding protein CTCF, as well as other factors such as actively transcribed genes or certain types of TEs (Dixon et al., 2012; Sexton et al., 2012). So far, it remains unclear whether chromosomes of the model plant *Arabidopsis* also contain well-defined regular domain structures or whether they have unique folding and interaction patterns.

In this study, we adapted genome-wide Hi-C (Lieberman-Aiden et al., 2009) to study chromatin interactions in *Arabidopsis*. Our data sets in wild-type (WT) and various epigenetic mutants provide a framework for the further understanding of the nuclear organization of plant genomes and the relationship between epigenetic marks and chromosome architectures.

RESULTS AND DISCUSSION

High-Resolution Maps of Chromatin Interaction in *Arabidopsis*

We previously developed a Hi-C protocol suitable for analysis of the genome of *Arabidopsis thaliana* (L.) Heynh (Moissiard et al., 2012). Using this protocol, we generated Hi-C libraries for WT *Arabidopsis*, as well as from a suite of mutants defective in gene silencing, DNA methylation, or specific histone modifications. We obtained ~41–66 million usable paired-end reads from each library, which given the relatively small size of the *Arabidopsis* genome (~125 million bases) (Arabidopsis Genome Initiative, 2000) and the dynamic smoothing process we applied, permits analysis of interaction at high resolution. Sequencing data were first processed using previously described analysis filters and pipelines (Lieberman-Aiden et al., 2009; Moissiard et al., 2012) with modifications and then plotted in 2D matrices to present interaction tendency between any two locations in the *Arabidopsis* genome (Figure 1A; File S1 available online) (see Supplemental Experimental Procedures). We also generated one-dimensional plots depicting the general relationship between interaction tendency and genomic distance (Figure S1A).

General Patterns of Chromatin Interaction Exhibited by the *Arabidopsis* Genome

A prominent feature of the WT *Arabidopsis* genome is that it is generally not segmented into large local adjacent interactive domains as has been previously described in several other organisms (Figure 1A) (Dixon et al., 2012; Le et al., 2013; Sexton et al., 2012), although a small number of interactive domains do exist (see later sections). To ensure that this difference was not due to differences in our analysis methods, we applied our pipeline to a previously published mouse Hi-C data set (Dixon et al., 2012) and readily observed topological domains that have been described (Figure S1B). The boundaries of topological domains are highly correlated with the binding of insulator protein CTCF (Dixon et al., 2012; Sexton et al., 2012), suggesting that CTCF may play a central role in defining the borders of topological domains. The lack of CTCF in plants (Heger et al., 2012) may help explain the lack of regular large topological domains and suggests that nuclear architecture of plants is significantly different from that of animals.

The strongest interactions we observed were exhibited by the blocks of pericentromeric heterochromatin, both among sequences within the same pericentromere and between sequences of different pericentromeres (Figure 1A) (Moissiard et al., 2012). This is consistent with previous studies showing that *Arabidopsis* has well-defined chromocenters formed by pericentromeric heterochromatin, these being visible by light microscopy, and that homologous pericentromeric regions frequently associate with each other in DNA fluorescence in situ hybridization (FISH) assays (Fransz et al., 2002; Schubert et al., 2012). We detect interaction between all pairs of heterologous pericentromeres (Figure 1A). Also, consistent with a previous 4C study in *Arabidopsis* (Grob et al., 2013), the interaction of pericentromeres includes the heterochromatic knob on the short arm of chromosome 4, which is close to the pericentromere (Figure 1A) (Arabidopsis Genome Initiative, 2000). The knob is composed of heterochromatin regions that were originally derived from the pericentromere, and therefore, the interaction of the knob with pericentromeres is analogous to the interactions between the pericentromeres. The region between the knob and the pericentromere is not included in these strong interactions, indicating the presence of a euchromatic loop (Figure 1A), consistent with earlier FISH studies (Fransz et al., 2002). Moreover, pericentromeres have very little interaction with regions outside of the pericentromeres or the knob, except for the nucleolar-organizing regions ("NORs"; see below) (Figure 1A) (Moissiard et al., 2012).

We also detected strong interactions among telomeres (Figure 1A) (Moissiard et al., 2012), supported by previous DNA FISH assays (Schubert et al., 2012), and consistent with the "telomere bouquet model" in which telomeres cluster around the nucleolus (Harper et al., 2004; Scherthan, 2007). Figures 1A and S1A show that telomere interactions take place among all telomeres of all chromosomes, suggesting that all telomeres cluster together randomly. The telomeres on the short arms of chromosomes 2 and 4 take part in these interactions but interact more weakly than the other telomeres (Figure 1A). The short arms of chromosomes 2 and 4 contain the 45S rDNA containing NORs, which are very close to the telomeres (Arabidopsis Genome Initiative, 2000; Fransz et al., 2002). By DNA FISH, the two NORs not only interact with each other but also colocalize with the chromocenters of chromosomes 2 and 4 (Fransz et al., 2002), as seen with Hi-C (Figures 1A and 1B). The strong association of the NORs and chromocenters likely prevents the adjacent telomeres from strongly interacting with other telomeres. Although not seen in Figure 1A (due to genomic distance-related data modeling; see Supplemental Experimental Procedures), the strong interaction between the two telomeres of each single chromosome is clearly evident both from the general relationship between interactivity and genomic distance (Figure S1A) and 2D plots in which the average distance-related interactivity has not been removed (Figure S1C). The interactive domains at the ends of chromosomes extend much further than

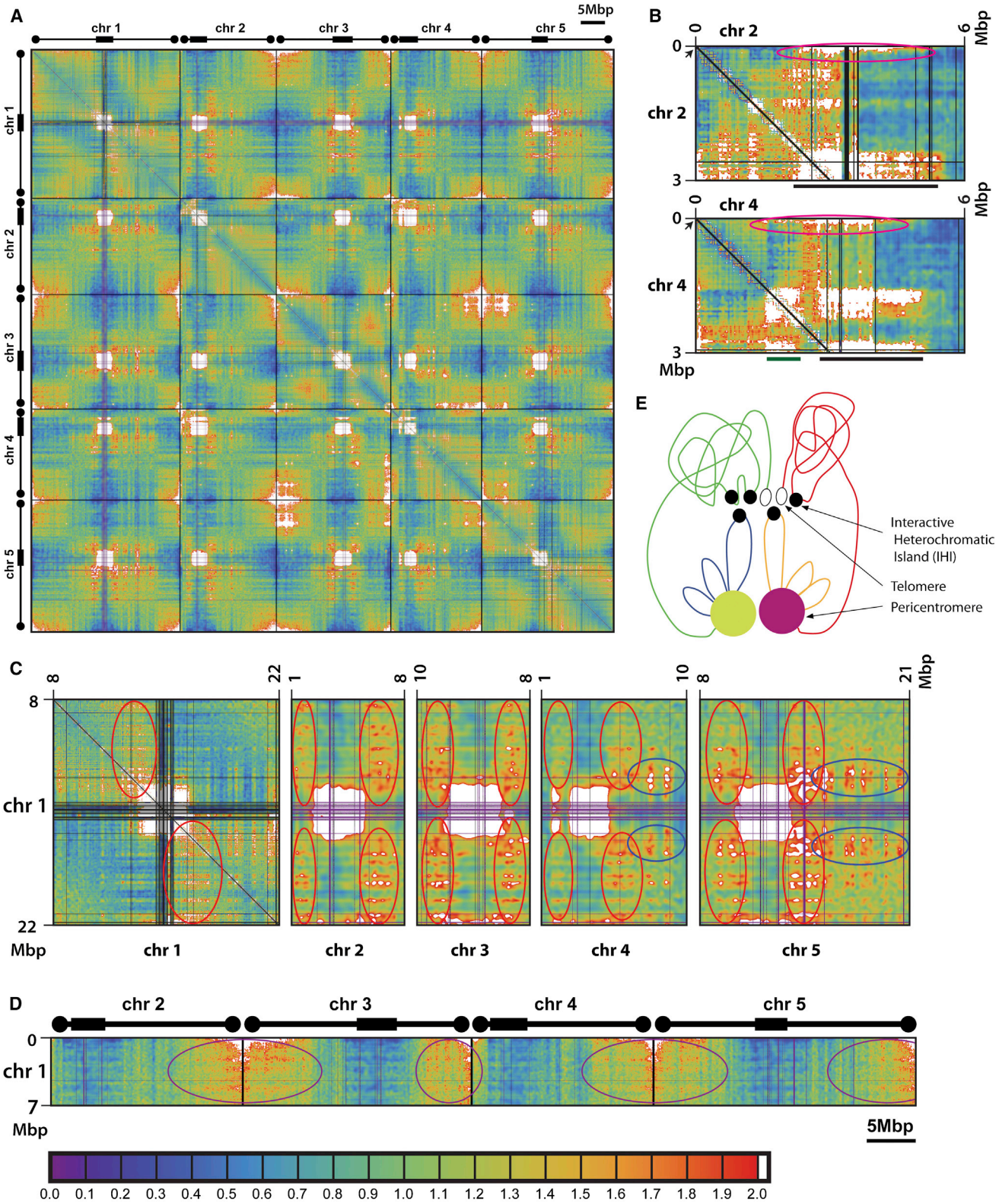


Figure 1. Chromatin-Interaction Patterns in *Arabidopsis*

(A) 2D interaction map of WT. The five *Arabidopsis* chromosomes (chr) are shown from left to right and top to bottom. Chromosomes are separated by thin black bars. Thick boxes mark approximate positions of the pericentromeric regions, and circles mark approximate positions of the telomeric regions. Blue to green to

(legend continued on next page)

the telomere territories themselves (Figure 1A), in line with previous findings showing that distal regions of chromosome arms interact more frequently than regions close to the centromeres (Grob et al., 2013; Schubert et al., 2012).

A previous DNA FISH-based model of *Arabidopsis* nuclear architecture hypothesized that regions of the euchromatic arms associate with pericentromeric heterochromatin, with euchromatic loops emanating from the chromocenters (Fransz et al., 2002). Our data support this model by revealing that there are regions in the euchromatic arms that frequently interact with the peripheral areas of pericentromeric heterochromatin, both intra- and interchromosomally (Figures 1A and 1C). However, we found that the euchromatic regions that participate in these interactions are those in the roughly one-half of the chromosome arm that is closest to the pericentromere (Figures 1A and 1C). We also found that for a given chromosome, euchromatic-pericentromeric interactions are higher between sequences on a given euchromatic arm with the half of the pericentromere that it is adjacent to (Figure 1C, chromosome 1). This preference might contribute to the observation of a generally higher level of interaction between the chromatin regions on one side of a centromere with other chromatin on that same side, as opposed to interaction with chromatin regions on the opposite sides of the centromere (Figure S1C), which is supported by FISH studies showing that the two euchromatic arms of the same chromosome tend not to intermingle (Berr and Schubert, 2007). Interestingly, these same regions tend to interact with the areas adjacent to the pericentromeres on all of the other chromosomes as well, and in this case, they interact with chromatin regions on both sides of the centromere (Figure 1C, chromosomes 2–5). This could be explained by the observation that all of the pericentromeric regions interact strongly (Figure 1A), which would bring these euchromatic sequences in closer proximity to each other.

In sharp contrast to the interactions of the proximal half of the euchromatic arms, the distal half of the arms preferentially interacts with the distal half of the other chromosome arms (Figures 1A, 1D, and S1C), and sharp boundaries between these two halves of the arms are frequently apparent (e.g., arms of chromosome 1 with those of chromosome 3 in Figure 1A). These distal interactions mirror, and may be in part driven by, the telomere interactions described above.

The *Arabidopsis* genome thus shows a unique nuclear organization that is different from that of animal genomes previously studied. *Arabidopsis* shows a lack of regular large topological domains but prominent interaction patterns involving pericentromeres, telomeres, and very large areas of the euchromatic arms partitioned into a centromere proximal region that interacts with itself and the pericentromeric regions, and a distal region that interacts with itself and the telomeric regions (Figure 1E). This is further supported by a hierarchical clustering analysis (Figure S1D).

Interactive Heterochromatic Islands Formed by Multiple Loci in *Arabidopsis*

Inspection of the interaction maps revealed a number of high-intensity off-diagonal punctate signals, indicating strong interactions between loci far apart in the primary DNA sequence (Figure 1A). Prominent examples include two loci near the beginning of chromosome 3, which can also be readily detected by 3C analyses (Figures 2A and S2A; Table S1). DNA FISH further confirmed a high frequency of interaction between these two loci (Figure 2B; Table S1B). We found that regions participating in these interactions all contain small patches of heterochromatin (average size ~7 kbp) that exist in the otherwise euchromatic arms. They are marked, for instance, by H3K9me2 and contain numerous TE-related repetitive sequences but are flanked by expressed protein-coding genes (Figures S2B–S2E). We therefore term these structures "interactive heterochromatic islands" ("IHIs"). The euchromatic arms also contain many additional similar patches of heterochromatin, some larger in size, but do not show these long-range interactions. Although the IHI-interacting regions all contained these heterochromatin patches, the interacting region is much larger than the patches themselves, ranging from 200 to 1,600 kbp (Table S1A). It is clear, though, that the peaks of highest interaction intensity overlap small heterochromatin patches (sharp H3K9me2 peaks) within the IHIs (Figures S2B–S2F), suggesting that the interactions of IHIs are possibly mediated through the heterochromatin patches. Nonetheless, Figure S2F also suggests that heterochromatin patches alone are not sufficient to cause chromatin interaction because there are peaks of H3K9me2 within the IHIs that do not overlap with peaks of interaction. Intriguingly, all IHIs showed interaction with all other IHIs (Figures 2A, 2C, and 2D). Because our Hi-C data are generated from large collections of cells, it is not possible to tell if all IHIs are clustered together simultaneously or whether only a fraction of these interactions exist at any given moment in a given single cell.

yellow to orange to red is weak-to-strong interaction tendency, with white for very strong interaction beyond a threshold. Light-gray indicates areas withheld from analysis due to, for example, problematic 50-mer mapping.

(B) Selected detail of WT interaction map. Chromosomal coordinates are labeled on the top and left sides. Pericentromeres of both chromosomes and the knob of chromosome 4 are labeled by black and green bars at the bottom, respectively. Locations of the NORs are indicated by black arrows. Pink circles indicate interaction of NORs with pericentromeres and the knob of chromosome 4.

(C) Selected detail of WT interaction map illustrating the interaction between proximal euchromatic arms and peripheral areas of pericentromeric heterochromatin. Red circles indicate interaction of the arms of chromosome 1 with pericentromeres of all five chromosomes. Blue circles indicate interaction of the arms of chromosomes 4 and 5 with pericentromere of chromosome 1.

(D) Selected detail of WT interaction map illustrating the interaction of a distal euchromatic arm of chromosome 1 with the distal euchromatic arms of chromosomes 2–5. Purple circles indicate areas of strong interaction.

(E) Working model of typical major chromatin-interaction patterns in *Arabidopsis*. The diagram depicts the pericentromeres of chromosomes 3 (yellow) and 5 (pink), along with the long arm of chromosome 3 (proximal chromatins in blue and distal chromatins in green) and the short arm of chromosome 5 (proximal chromatins in orange and distal chromatins in red). IHIs are indicated by closed black circles, and the telomeres are indicated by open black circles. See text for details.

Color bar for (A)–(D) is shown at the bottom of the figure. See Supplemental Experimental Procedures for details. See also Figure S1 and File S1.

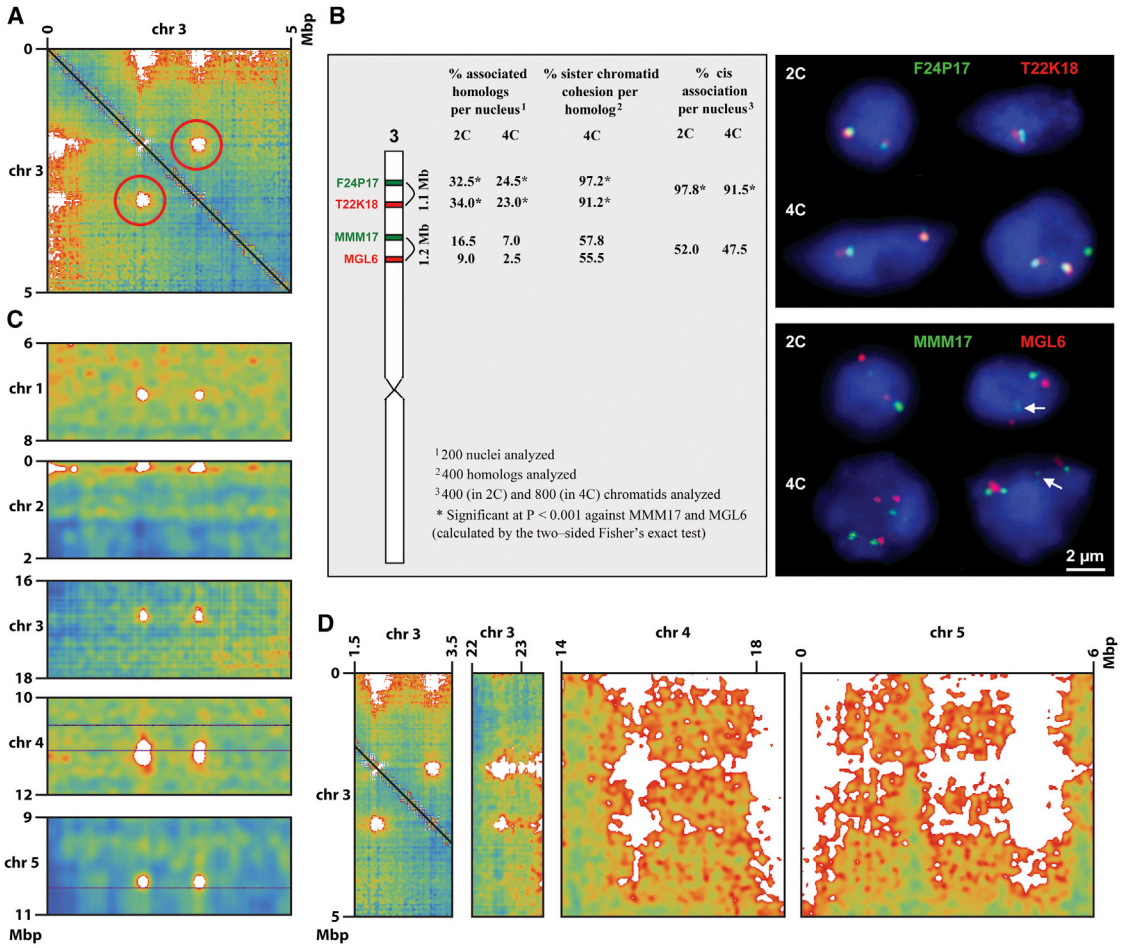


Figure 2. IHIs in *Arabidopsis*

(A) Two loci on chromosome 3 interact strongly with each other. Chromosomal coordinates are labeled on the top and left sides. Red circles indicate the interaction (interaction maps are symmetric about their main diagonals, and so it appears twice).

(B) DNA FISH analysis of the two loci on chromosome 3 from (A). The left subpanel diagrams chromosome 3 and the approximate positions of bacterial artificial chromosomes (BACs) used, along with a statistical summary of the results (*$p < 0.001$ via two-sided Fisher's exact test) from 200 nuclei. Here, "homologs" refer to the same regions on homologous chromosomes, and "*cis* association" refers to the association between different regions on the same chromosome. Representative images of 2C and 4C nuclei show different configurations of associated and/or separated chromatin segments of the frequently associated BACs F24P17 and T22K18 (upper panel) in comparison to configurations in nuclei labeled by BACs MMM17 and MGL6 as a control (lower panel). Chromatin elongation in 2C and 4C nuclei (indicated by white arrows) and sister chromatid separation in 4C may lead to more than two or four signals per nucleus in 2C and 4C nuclei, respectively.

(C) Interaction of the two loci on chromosome 3 from (A) with five other loci, one from each chromosome.

(D) Interaction of the two loci on chromosome 3 from (A) with loci toward the beginning of chromosome 5 and the ends of chromosomes 3 and 4.

Color scales in (A), (C), and (D) are the same as in Figure 1A. See also Figure S2 and Table S1.

Curiously, although IHIs have heterochromatic features similar to sequences in pericentromeric heterochromatin, they do not interact with pericentromeric regions. Instead, we found that IHIs frequently interact with telomeric and subtelomeric regions (Figures 1A, 2A, and 2D). In addition, the broad regions surrounding IHIs show a higher level of interchromosomal interactions, especially for those in the distal half of the euchromatic arms (Figures 1A and 2D). Although the function of IHI interactions is not known, it is possible that they strengthen the interactions of larger domains, such as the telomere-containing distal halves of the euchromatic chromosomes (Figure 1E). It is interesting that the dynamics and participants of the IHI-interaction network are altered in several epigenetic mutants (see below), suggesting that epigenetic marks are in part regulating these interactions. Together, these findings suggest that a network of interactions takes place among a small set of epigenetically silent regions and the telomeres, forming a previously unappreciated level of complexity of interaction within the *Arabidopsis* nucleus.

Comparisons of Genomic Features and Chromatin Interactions

We assessed connections between chromatin-interaction patterns and various genomic data sets, including histone modifications (chromatin immunoprecipitation sequencing

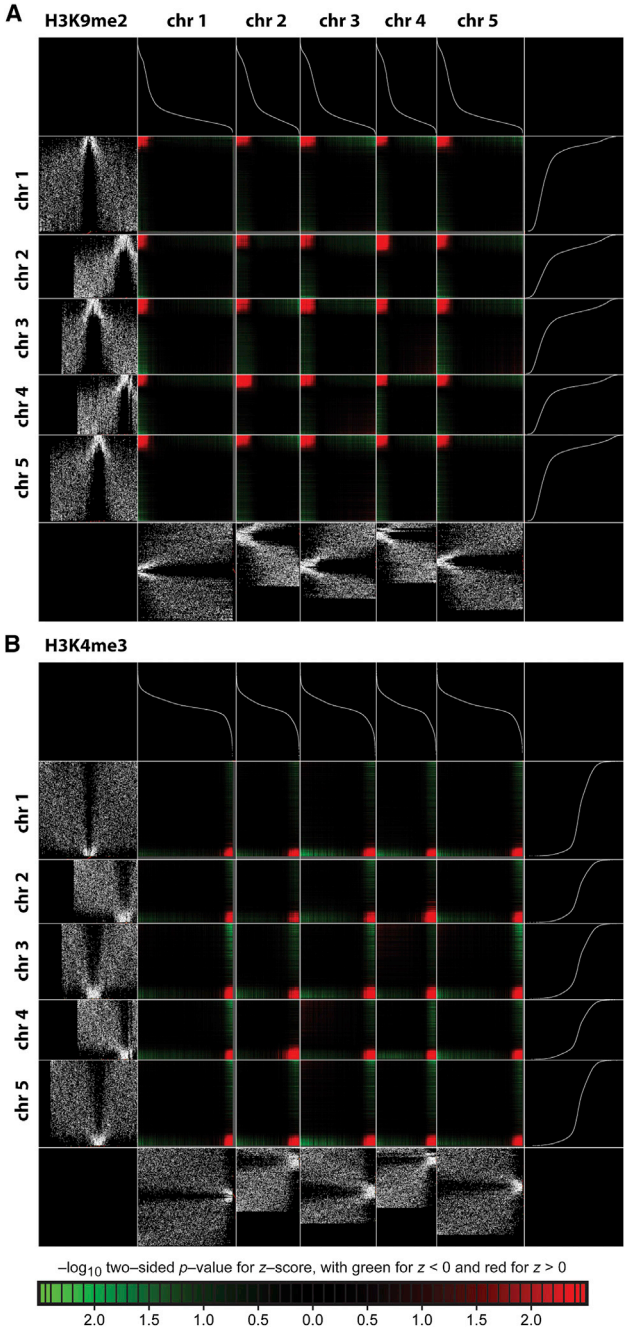


Figure 3. Genomic Features and Chromatin Interactions

(A) 2D interaction map of WT, except with chromosomal positions (2.5 kbp bins) permuted within each chromosome based on intensity of H3K9me2 signals. Panels on the left and bottom indicate the chromosomal positions of the loci in the map. Panels on the right and top indicate the H3K9me2 modification level of the corresponding loci. Regions with the highest and lowest H3K9me2 levels are clustered toward the top-left corner and bottom-right corner, respectively, of each permuted map. Red indicates higher interaction, and green indicates lower interaction than average.

(B) 2D interaction map of WT as in (A), except with chromosomal positions permuted based on intensity of H3K4me3 signals.

Color bar is shown at the bottom of the figure. See Supplemental Experimental Procedures for details. See also Figure S3 and File S1.

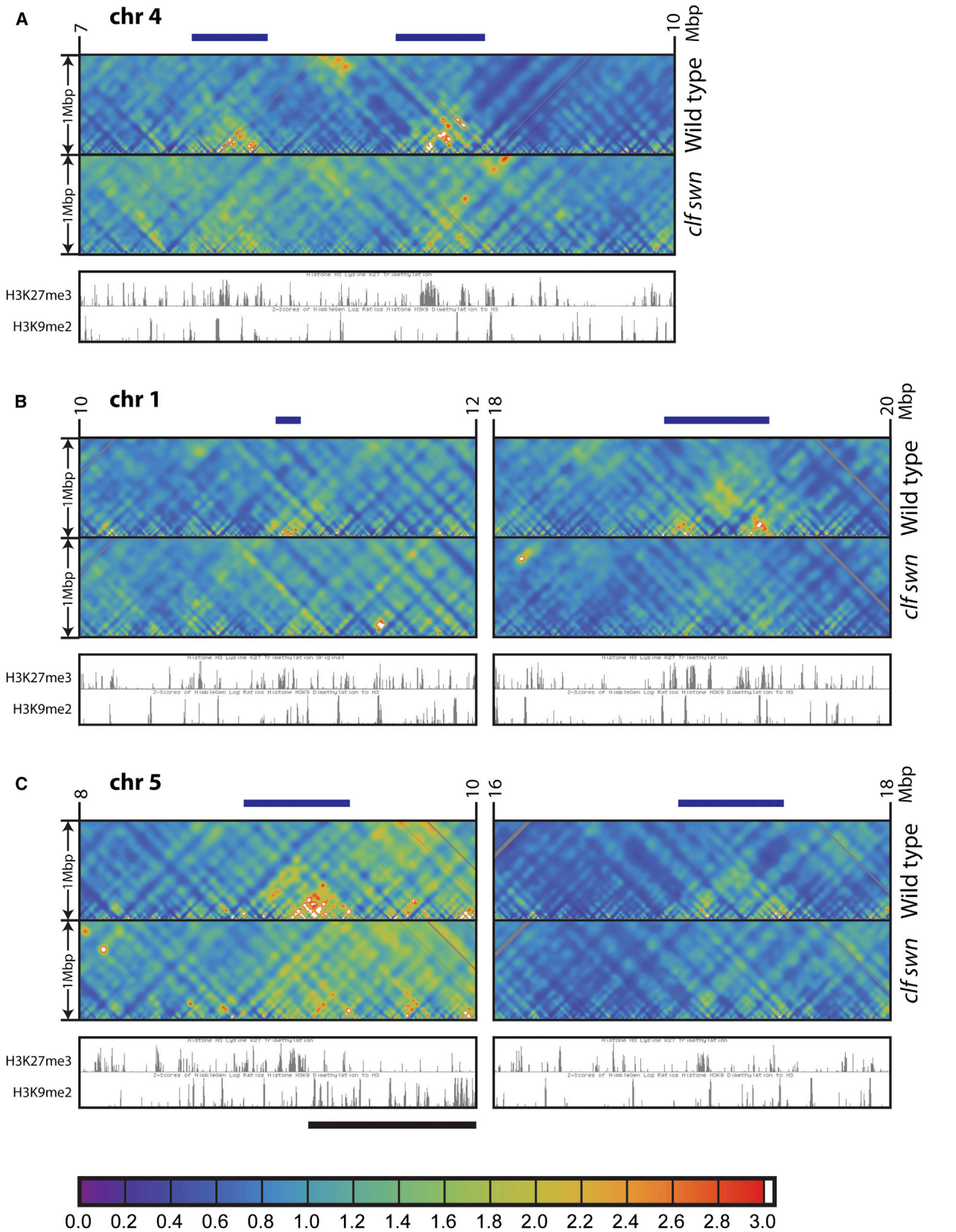
[ChIP-seq] and ChIP with DNA microarray), DNA methylation (whole-genome bisulfite sequencing), and RNA abundance (RNA-seq) (all data utilized are displayed at http://genomes.mcdb.ucla.edu/AthBSseq/ and were produced from similar tissues as the Hi-C data) (Bernatavichute et al., 2008; Jacob et al., 2009; Stroud et al., 2013, 2014; Zhang et al., 2007, 2009). We permuted 2.5 kbp binned copies of the *Arabidopsis* genome based on bin average signal intensity for each genomic data set and then rendered the interaction map in each new order. Thus, regions with the highest signal intensity for a given genomic feature (e.g., H3K9me2) are clustered to the top left of each permuted map and those with low intensity toward the bottom right (Figures 3 and S3; File S1). This allows for easy visualization of increased or decreased chromatin interactions in regions enriched or depleted for particular genomic features.

Most repressive epigenetic marks, including DNA methylation, H3K9me2, and H3K27me1, strongly colocalized with the highest levels of chromatin interactions (upper-left corners in Figures 3A and S3C; File S1). This is consistent with our finding that the strongest chromatin interactions exist among and between pericentromeric regions that are enriched in these marks (Figures 1A, 3A, and S3C; File S1). We also observed lower interaction between the heavily DNA, H3K9 di-, or H3K27 monomethylated regions with the rest of each chromosome (Figures 3A and S3C; File S1). This likely reflects the association of pericentromeric regions with themselves and avoidance of the rest of the genome (Figure 1A). We also found that if we exclude previously defined pericentromeric regions (Bernatavichute et al., 2008) from the permutation analysis, there is no clear association of chromatin-interaction strength with either highly or lowly H3K9m2-modified regions (Figure S3D), suggesting that the effect observed in Figure 3A is primarily the result of pericentromere interactions.

Marks of active chromatin, such as H3K4 mono-, di-, and trimethylation, showed antilocalization with the brightest Hi-C signals (Figure 3B; File S1), suggesting that transcriptional-active regions do not show unusually high interaction among themselves. Consistently, these marks almost exclusively exist outside of the pericentromeric regions that show the strongest interactions in the genome (Figure 3B; File S1). We also examined the relationship between mRNA-seq-derived gene expression levels and chromatin interactivity. Notably, we did not see a correlation of chromatin interactivity with genes of the highest level of RNA-seq expression, even when pericentromeric regions were excluded (Figure S3E), suggesting a lack of clustering of the most actively transcribed genes generally. This result is consistent with our recent observation that plant RNA Polymerase II, while excluded from the heterochromatic regions, exhibits a homogeneous distribution pattern within the euchromatic regions of interphase nuclei (Schubert, 2014).

Local Interactive Domains Correlated with Certain H3K27me3 and H3K9me2 Regions

To investigate interactive domains on a short-to-medium distance scale (up to a few megabase pairs), we extracted strips of the interaction maps near the main diagonal and rendered



(legend on next page)

them as detailed local interaction maps (see Figure S4A for chromosome 3 of WT, and File S1 for a complete set). From these, we identified a small number of interactive domains scattered throughout the genome, which are correlated with certain regions marked with either of the two histone modifications: H3K27me3 or H3K9me2.

H3K27me3 is a silencing mark of the Polycomb system and is present on about 17% of protein-coding genes (Turck et al., 2007; Zhang et al., 2007). As opposed to the frequent clustering of Polycomb-regulated genes known in animals (e.g., Hox gene clusters) (Schwartz et al., 2006; Tolhuis et al., 2006), H3K27me3 regions are more evenly scattered throughout the *Arabidopsis* genome and are generally restricted to the transcribed regions of single genes, with only a few H3K27me3 regions clustered together, especially in the case of tandemly repeated homologous genes (Turck et al., 2007; Zhang et al., 2007). Recent work by Rosa et al. (2013) has shown that the two *FLOWERING LOCUS C* (a well-known target of the Polycomb/H3K27me3 pathway) alleles on homologous chromosomes cluster together, suggesting long-range interactions. Furthermore, in *Drosophila*, Polycomb/H3K27me3 target regions have been shown to interact with each other (Sexton et al., 2012). We identified two heavily H3K27me3-marked regions on chromosome 4 of WT, lying within ~1 Mbp of each other, which show strong interaction within the domains (Figure 4A, WT) as well as a number of other similar regions (Figures 4B and 4C, WT). These regions correspond to unusually large clusters of adjacent H3K27me3-marked genes (Figures 4 and S4B). On the other hand, we did not observe a higher chromatin interactivity for H3K27me3-marked regions generally in the genome (Figure S3F; File S1). Therefore, the H3K27me3-interactive domains are apparently limited to a small number of sites in the genome, where H3K27me3-marked genes are clustered.

As discussed above, pericentromeric regions marked by H3K9me2 interact extensively within and among each other (Figure 3A), readily visualized in local interaction maps as large domains near the centromeres of each chromosome (Figures S4A and S4C; File S1). In addition, we observed a number of small patches of H3K9me2 outside of the pericentromeric regions (Bernatavichute et al., 2008) that show a high level of interactivity within the domain (Figures S4A and S4C). All of the most prominent H3K9me2-containing small local interactive domains correspond to the IHIs described above; however, not all IHIs form locally interacting domains of this type (Figures 2, S4A, and S4C; Table S1A). Indeed, the IHIs form stronger interactions among each other than they do with other areas of chromatin in the vicinity of the IHI (Figures 1 and 2). This suggests that IHIs, unlike pericentromeric regions, do not necessarily form strong local interactive domains but, instead, have a relatively high level of interactivity with other IHIs and the telomere regions.

Epigenetic Mutants Show Altered Chromatin-Interaction Patterns

We performed Hi-C in a number of mutants affecting epigenetic regulation, including the *curly leaf* (*clf*) *swinger* (*swn*) double mutant, *arabidopsis thaliana microrchidia 6* (*atmorc6*), *morpheus' molecule 1* (*mom1*), *decrease in dna methylation 1* (*ddm1*), *methyltransferase 1* (*met1*), *chromomethylase 3* (*cmt3*), and the *su(var)3-9 homolog 4* (*suvh4*) *suvh5 suvh6* triple mutant. We generated 2D interaction maps (Figures 5A and S5; Files S2, S3, S4, and S5), local interaction maps (File S6), and histograms and 2D plots of interaction differences between each mutant and WT (Figures 6 and S6; Files S2, S3, S4, and S5). Histograms suggest that the overall differences between mutants and WT are readily detectable and are visibly larger than the difference between two WT replicates (Figure S6B).

CLF and SWN are two enhancer of zeste (Ez) homologs that control H3K27me3, and the *clf swn* double mutant lacks virtually all H3K27me3 (Lafos et al., 2011). We analyzed the local interactive domains consisting of clustered H3K27me3 genes described above and found that the interaction within these domains was dramatically reduced or eliminated in the double mutant (Figure 4; Files S1 and S6). Hence, H3K27me3 may act directly or indirectly to regulate the interactivity of these domains. Although the functions of the clustering of these H3K27me3-marked genes and their higher level of local interactivity are unknown, the change in the behavior of these domains in *clf swn* suggests regulation of these interactions at an epigenetic level.

AtMORC6 is required for heterochromatin condensation, and we have previously shown a low-resolution Hi-C comparison of WT versus *atmorc6* (Moissiard et al., 2012). We repeated Hi-C analysis of *atmorc6* at higher coverage and resolution. As we showed previously, *atmorc6* exhibits decreased association of pericentromeric heterochromatin with itself as well as elevated interaction between pericentromeric heterochromatin and euchromatic arms (Figure 6), consistent with decondensed chromocenters and derepressed transposon expression in *atmorc6* (Moissiard et al., 2012). Interestingly, the *atmorc6* mutant clearly shows enhanced interaction among telomeres when compared to WT, and the interactive chromatin regions at the ends of chromosomes extend much further from the telomere territories into euchromatic arms than in WT (Figures 1A, 5A, and 6). In addition, *atmorc6* affects the above-mentioned interactions involving the distal and proximal halves of the euchromatic chromosome arms, such that the proximal interaction is enhanced and the distal interaction is reduced in the *atmorc6* mutant (Figure 6). Thus, AtMORC6 appears to regulate large-scale nuclear organization and shifts the balance of interaction of euchromatic arms with themselves and with pericentromeres. Moreover, AtMORC6 affects the interactions of the IHIs: compared to WT, *atmorc6* shows enhanced interaction

Figure 4. Local Interactive Domains in *Arabidopsis*

(A–C) Local interaction maps of a selected 3 Mbp region in chromosome 4 (A) and 2 Mbp regions in chromosomes 1 (B) and 5 (C) from WT and *clf swn* double mutant, to a distance of 1 Mbp (see y axis). Chromosomal coordinates are labeled on top. H3K27me3 and H3K9me2 tracks shown are from WT on the UCSC Genome Browser. The local interactive domains are labeled by dark-blue bars on tops of plots. The left subpanel of (C) is close to pericentromeric heterochromatin, and thus, a large interactive domain overlapping with strong H3K9me2 signals appears in the right half (black bar). Color bar (same as Figure S1D) is shown at the bottom of the figure. See Supplemental Experimental Procedures for details. See also Figure S4 and Files S1 and S6.

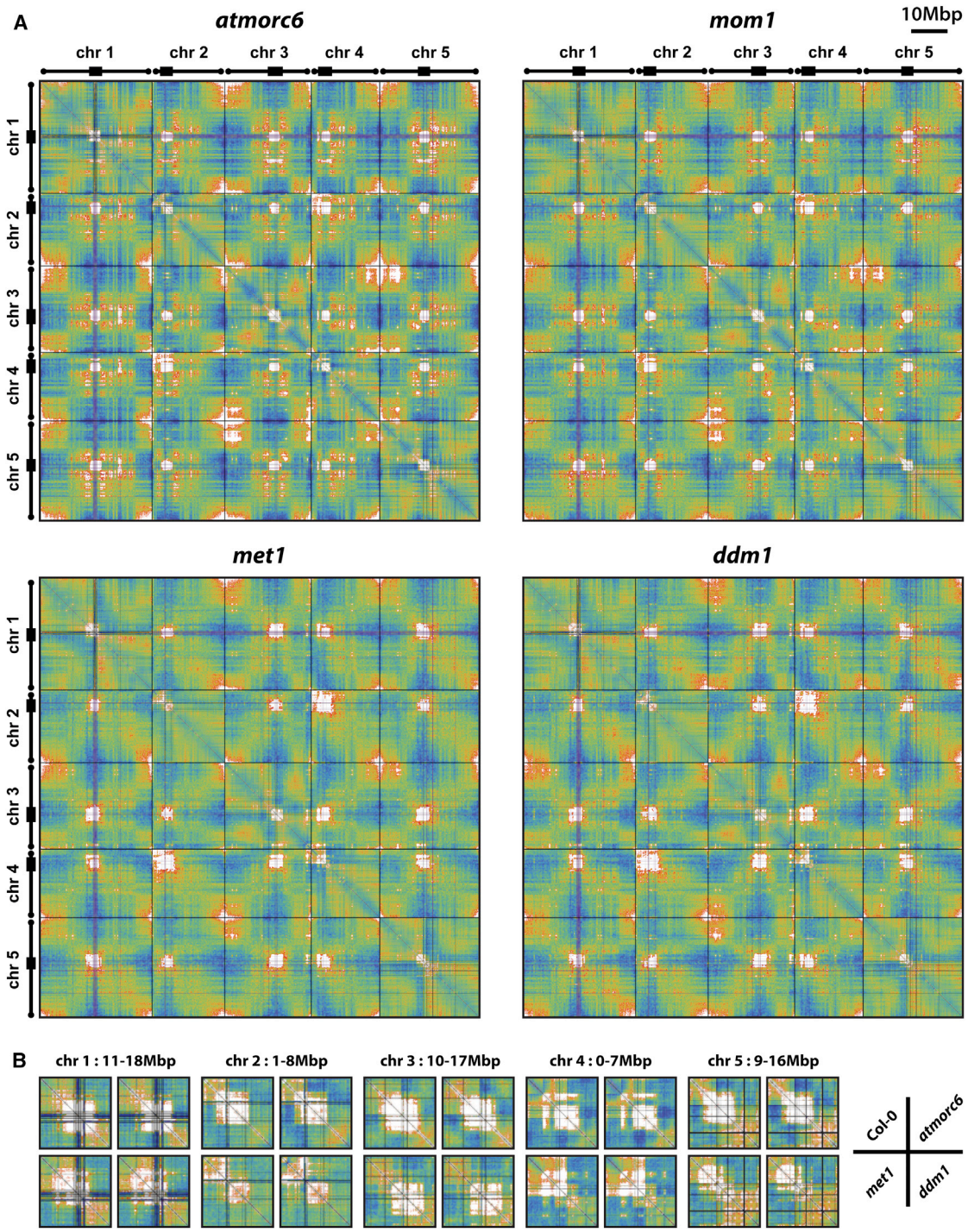


Figure 5. Chromatin-Interaction Patterns in Mutants Affecting Epigenetic Processes

(A) 2D interaction maps of *atmorc6*, *mom1*, *met1*, and *ddm1* generated in the same way as Figure 1A.

(B) Detail of selected pericentromeric interactions across WT and mutants. The region selected for each comparison is indicated on top. Color scales are the same as in Figure 1A. See also Figure S5 and Files S2, S3, S4, and S5.

among IHIs (Figure 7A; File S3). This is particularly interesting in light of the fact that both IHIs and pericentromeric regions are characterized by high levels of H3K9me2 and transposon sequences, yet *atmorc6* reduces the interaction among pericentromeres while increasing the interactions among IHIs (Figures 6 and 7A).

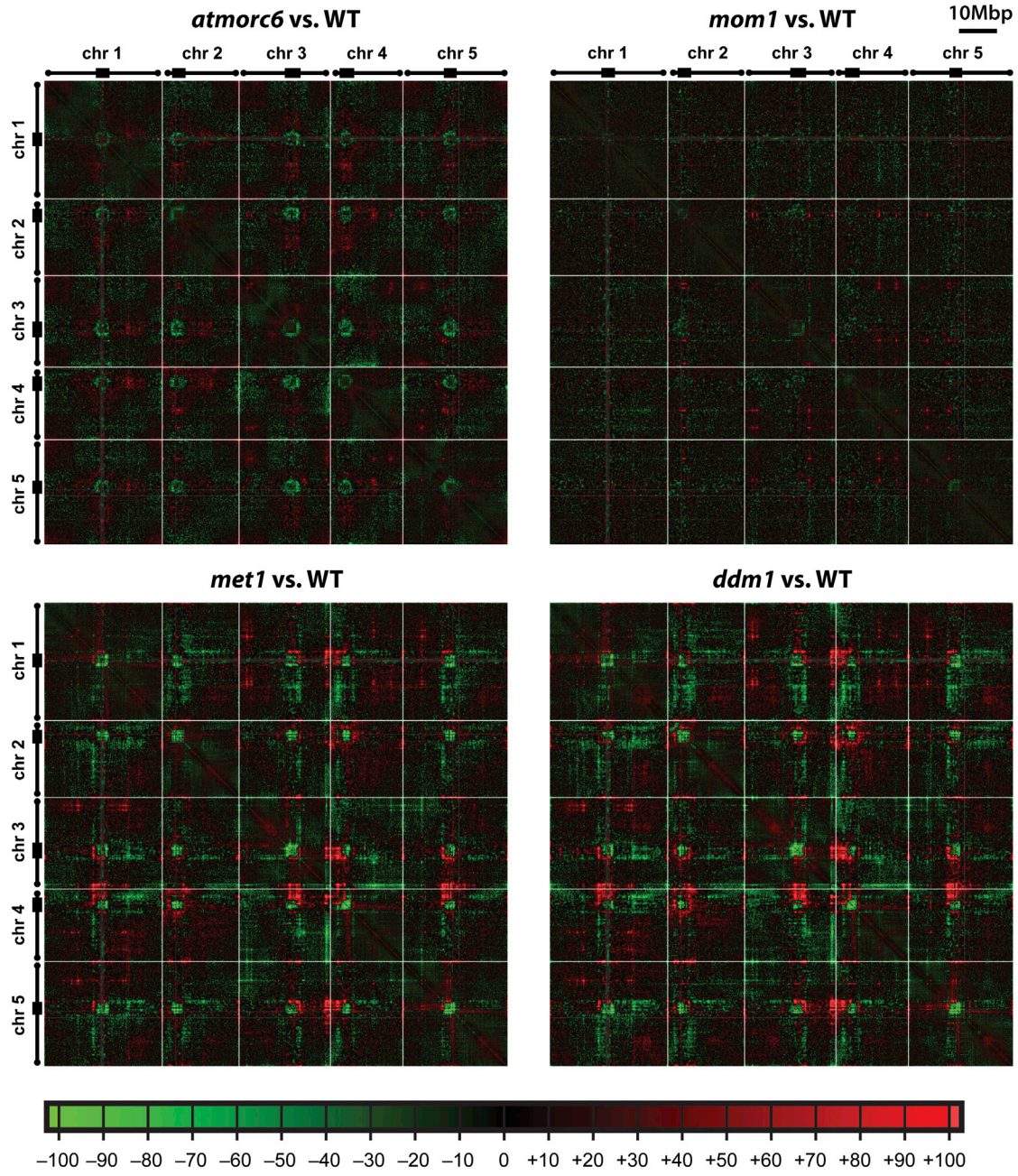


Figure 6. Comparison of Interaction Patterns across Mutants and WT

Chromatin-interaction patterns of *atmorc6*, *mom1*, *met1*, and *ddm1* versus WT. Colors show difference of mutant relative to mean of WT and mutant ("percent difference"); black is no change, and brightest red (mutant higher) and brightest green (mutant lower) are 100%. Chromosome labels and gray are as in Figure 1A. Color bar is shown at the bottom of the figure. See Supplemental Experimental Procedures for details. See also Figure S6 and Files S2, S3, S4, and S5.

Similar to *AtMORC6*, *MOM1* is another *Arabidopsis* gene that represses genes and TEs without altering DNA methylation (Amedeo et al., 2000). Interestingly, several loss-of-function *mom1* alleles were recovered from the same genetic screen that identified *atmorc6* (see Supplemental Experimental Procedures), again implicating a functional resemblance of the *MOM1* and *AtMORC6* genes. We found very little difference in the pattern of chromatin interaction of *mom1* versus WT, and we did not observe a decrease in interaction within and between pericentromeric regions (Figure 6; File S3). These results are consistent with previous DNA FISH observations showing that *mom1* does not show decondensed chromocenters (Probst et al., 2003) and suggest that although MOM1 and AtMORC6 share some similarities, they likely employ different mechanisms in regulating transcriptional gene silencing.

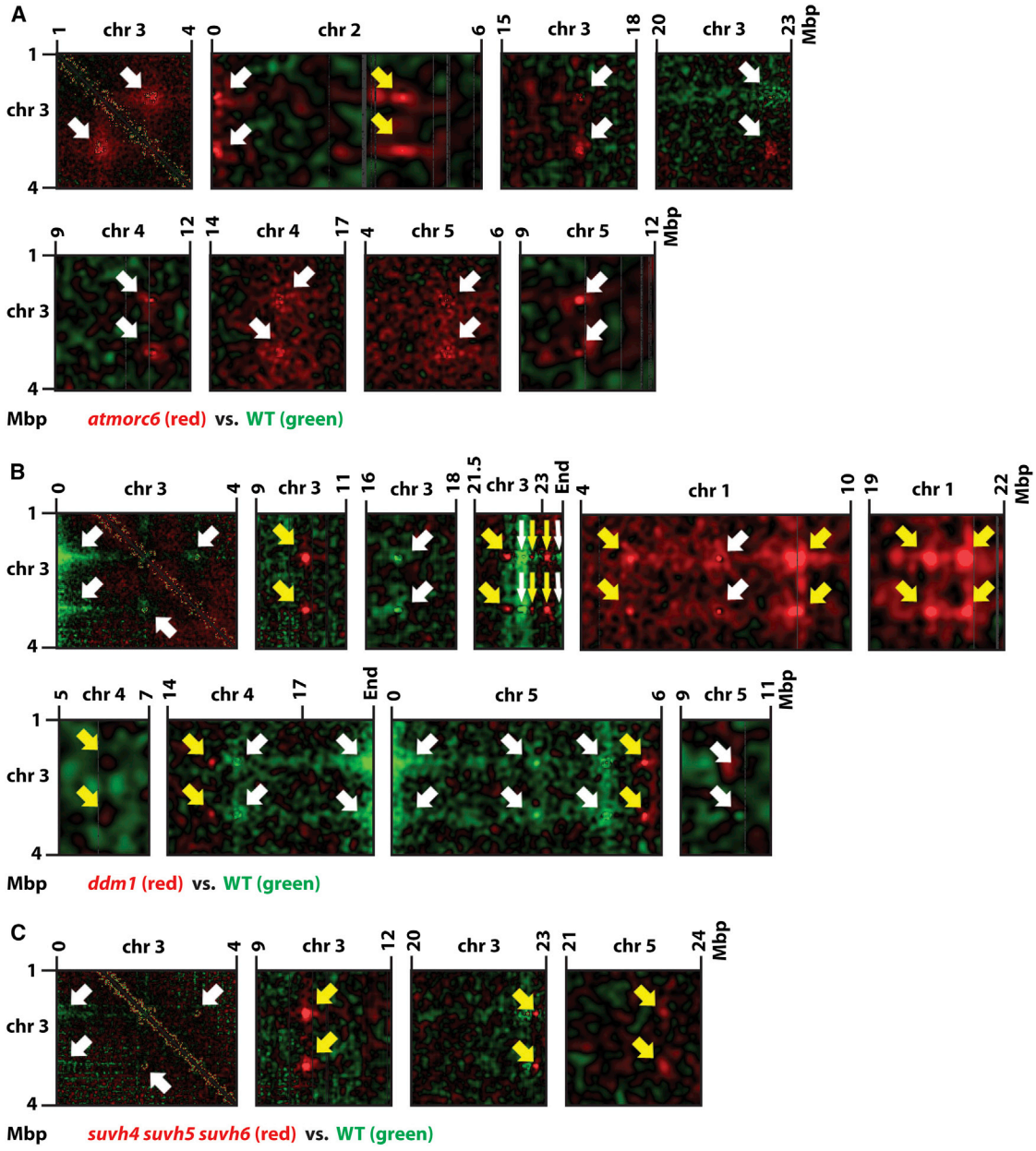


Figure 7. Dynamics of IHIs in Mutants

(A–C) Details of plots in the style of Figure 6 (with mutant higher/lower red/green) showing changes in IHI interaction in *atmorc1* (A), *ddm1* (B), and *suvh4 suvh5 suvh6* (C). Chromosomal coordinates are labeled on the top and left sides. White arrows indicate IHIs that are also found in WT (see Figures 2A, 2C, and 2D). Yellow arrows indicate IHIs only found in the mutants. Color scales are the same as in Figure 6. See also Figure S7, Table S1, and Files S3, S4, and S5.

Because chromatin interaction is positively correlated with repressive epigenetic marks (see above), we sought to examine whether reducing DNA methylation and H3K9me2 levels would impact chromatin interaction in the regions with these marks. We therefore performed Hi-C analysis in *met1*, *ddm1, cmt3*, and *suvh4 suvh5 suvh6* triple mutants. The 2D interaction maps and comparative maps revealed striking changes in chromatin-interaction patterns, particularly in *met1* and *ddm1* (Figures 5A and 6; File S4), two mutants that reduce DNA methylation in all sequence contexts (Stroud et al., 2013). First, both *met1* and *ddm1* resemble *atmorc6* in having less interaction of pericentromeric regions with themselves and more interaction of pericentromeres with euchromatic chromosome arms. However, the degree of change displayed by each of *met1* and *ddm1* is even higher than *atmorc6*. This is consistent with previous microscopic studies showing that *met1* and *ddm1* mutants have decondensed chromocenters and that *DDM1* and *MET1* are both required for the repression of a larger number of genes and TEs

than *AtMORC6* (Moissiard et al., 2012; Probst et al., 2003; Soppe et al., 2002; Stroud et al., 2012). The most drastic changes in pericentromere association take place on chromosomes 2 and 4, whereas areas of pericentromeric heterochromatin on these that interact with NORs still interact with NORs in the mutants; other areas of the pericentromeres show very little interaction with adjacent areas, especially for chromosome 4 (Figure 5B). We also observed a loss of interactivity of internal regions of chromocenters, such as on chromosome 5, suggesting the formation of loops that no longer participate in the main block of pericentromeric interactions (Figure 5B). These results are consistent with previous work showing that the visible chromocenters are smaller in *met1* and *ddm1* mutants (Soppe et al., 2002). The effects of *met1* and *ddm1* on the pericentromeric regions are very different from the effect of the *atmorc6* mutant, which shows decreased interaction for all pericentromeric regions but maintains the same regions in the generally still-strongly interacting pericentromeres (Figure 5B). This is also consistent with the different morphology of the chromocenters; *atmorc6* mutants show larger and more diffused chromocenters, whereas *met1* and *ddm1* show smaller but still very punctate chromocenters (Moissiard et al., 2012; Soppe et al., 2002). We also observed another unique and striking phenotype in both *met1* and *ddm1*—a shift in the locations of the chromatin regions that participate in strong interactions within pericentromeric regions—suggesting that the regions that fold into the chromocenters have changed in these mutants (Figure 5B). This phenomenon is possibly due to a combinatorial effect of reduction of DNA methylation in pericentromeres of *met1* and *ddm1* (Stroud et al., 2013) together with changes in the histone methylation landscapes in the pericentromeres of these mutants (Deleris et al., 2012; Mathieu et al., 2005). We also observed an interesting behavior of IHIs in *met1* and *ddm1*, with the IHIs defined in WT showing slightly decreased mutual interaction in *met1* and *ddm1*. However, this slight decrease is perhaps the result of recruitment of an increased number of participating IHI loci in the mutants (Figure 7B; Tables S1A and S1C; File S4). The newly appearing IHI loci resemble the previously described loci in WT in that they are centered on small patches of heterochromatin in the otherwise euchromatic arms (Figures S7A–S7D). A comparison of WT and mutant expression profiles did not reveal a clear correlation between reactivation of transposons and recruitment of new IHI loci in the mutants (Figure S7E); therefore, the mechanism by which the new IHIs get recruited is unknown. Together, these results demonstrate dramatic alteration of chromatin interactions in the *met1* and *ddm1* mutants, suggesting that DNA methylation is a major epigenetic determinant of the natural nuclear architecture of chromatin.

SUVH4, SUVH5, and SUVH6 are histone methyltransferases for H3K9 (Feng and Jacobsen, 2011). The *suvh4 suvh5 suvh6* triple mutant exhibits extensive loss of non-CG methylation (Stroud et al., 2013, 2014) and has been shown to have somewhat decondensed chromocenters (Rajakumara et al., 2011). Consistent with these data, we observed an *atmorc6*-like pericentromeric-interaction pattern in *suvh4 suvh5 suvh6* (Figure S6C) (Stroud et al., 2014). However, unlike the *met1* and *ddm1* mutants, *suvh4 suvh5 suvh6* did not shift the regions interacting with pericentromeric regions (Figure S5), suggesting that *suvh4 suvh5 suvh6* generates a less-severe nuclear organization change than *met1* and *ddm1*. We also observed that all of the IHI interactions still occurred in *suvh4 suvh5 suvh6*, whereas three new IHI loci also appeared, two of which correspond to the same ectopic IHI loci in *met1* and *ddm1* (Figure 7C; Tables S1A and S1C; File S5). ChIP-seq (Stroud et al., 2014) suggests that the majority of IHI loci, including the ones that ectopically interact in *suvh4 suvh5 suvh6*, have lost detectable H3K9me2 signal in *suvh4 suvh5 suvh6* (Figure S7F). Collectively, the findings from *met1*, *ddm1*, and *suvh4 suvh5 suvh6* suggest that, despite the IHIs being enriched in DNA methylation and H3K9me2, the interactions among IHI loci are not dependent on DNA methylation or H3K9me2 marks and that, in fact, additional loci are recruited to IHIs when these marks are reduced.

CMT3 mediates CHG methylation (Stroud et al., 2013), but loss of CMT3 does not lead to chromocenter decondensation (Moissiard et al., 2012). Consistently, we did not observe *atmorc6*-like chromatin-interaction patterns in *cmt3* mutant, and *cmt3* also lacks the dramatic chromocenter-interaction alterations observed in *met1* and *ddm1* (Figures S5 and S6C; File S5). In fact, the difference in Hi-C maps between *cmt3* and WT is quite minimal—not much bigger than the difference between WT replicates (Figures S6A–S6C). This result suggests that CHG methylation alone plays a relatively minor role in regulating chromatin interaction.

Collectively, our investigations of the chromatin-interaction patterns in various epigenetic mutants indicate that loss of DNA methylation and histone H3K9 methylation affects chromatin interaction, leading to losses of interaction among pericentromeric regions and gains in interaction among IHIs. Furthermore, H3K27me3 is also important in regulating interactions within large domains consisting of adjacent H3K27me3-marked genes.

Conclusions

Our Hi-C analyses show that *Arabidopsis* chromosomes interact extensively through their pericentromeric regions, as well as through two domains of the euchromatic arms: one consisting of the proximal half that also interacts with pericentromere-adjacent regions, and one consisting of the distal half that also interacts with telomeres (see model in Figure 1E). A number of IHIs show strong long-range interactions with each other, which are also associated with telomeric regions. In addition, special regions in the euchromatic arms that are either H3K9me2 or H3K27me3 modified form local interactive hot spots. On the other hand, we do not observe strong interactions among highly expressed genes as has been observed in animals. Mutants that affect various repressive epigenetic processes exhibit altered chromosome architectures that are related to the effect of these mutations on heterochromatin condensation or their effects on the maintenance of histone or DNA methylation. These results reveal the complexity of chromatin interactions within the *Arabidopsis* nucleus and will form the basis of future studies on the regulatory mechanisms underlying chromosome folding in plants.

EXPERIMENTAL PROCEDURES

Additional details of experimental and analysis methods can be found in the Supplemental Experimental Procedures.

Hi-C Library Construction and Sequencing

Hi-C libraries compatible with Illumina sequencing were generated as described previously by Moissiard et al. (2012). The libraries were sequenced on HiSeq 2000 DNA sequencers obtaining paired-end 50- or 51-nucleotide reads following manufacturer instructions (Illumina).

Hi-C Data Analysis

The analysis pipeline includes a "dynamic smoothing" algorithm to reduce statistical noise in low-coverage regions but retain high resolution in high-coverage regions. Provision was also made for two aspects of Hi-C signals that can lead to confounding effects. The first phenomenon is that some regions of the genome have much higher or lower counts than average due to issues such as density of HindIII sites, the mapability of 50-mer, the fidelity of the reference genome, and any biases introduced by library amplification or sequencing procedures. The second relates to the fact that interactions on the same chromosome correlate strongly with genomic distance. For instance, interactions between two areas of chromatin are naturally higher if those areas have only a short distance between them (Figure S1A). Raw counts are separated into components consisting of mapability effect, distance effect, and the remaining signal. This procedure allowed for higher sensitivity in detecting subtle interactions occurring both within and between chromosomes.

ACCESSION NUMBERS

Sequencing data were deposited in the NCBI Sequence Read Archive as accession SRP043612.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, one table, and six additional supplemental files and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2014.07.008.

AUTHOR CONTRIBUTIONS

S.F., S.J.C., M.P., and S.E.J. designed the project. S.F. and V.S. performed experiments. S.J.C. conceived, implemented, and ran the analysis pipeline. S.J.C., J.Z., and S.E.J. analyzed the data. S.F., S.J.C., and S.E.J. wrote the manuscript.

ACKNOWLEDGMENTS

We thank M. Akhavan for Illumina sequencing; H. Stroud and G. Moissiard for assistance with plant work; H. Stroud, G. Moissiard, S. Bischof, D. Husmann, C. Hale, and J. Goodrich for seeds; Y. Zhan and J. Dekker for advice on Hi-C library construction; M. Kühne for technical assistance; and I. Schubert, P. Fransz, and members of the S.E.J. laboratory for supportive discussions and comments on the manuscript. Illumina sequencing was performed at the UCLA BSCRC BioSequencing Core Facility. This work was supported by a grant from NIH (GM60398 to S.E.J.). S.F. was a Special Fellow of the Leukemia & Lymphoma Society. S.E.J. is an investigator of the Howard Hughes Medical Institute.

Received: January 21, 2014
Revised: June 2, 2014
Accepted: July 10, 2014
Published: August 14, 2014

REFERENCES

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.

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

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., and Schubert, I. (2007). Interphase chromosome arrangement in *Arabidopsis thaliana* is similar in differentiated and meristematic tissues and shows a transient mirror symmetry after nuclear division. Genetics *176*, 853–863.

Crevillén, P., Sonmez, C., Wu, Z., and Dean, C. (2013). A gene loop containing the floral repressor FLC is disrupted in the early phase of vernalization. EMBO J. *32*, 140–148.

Dekker, J., Marti-Renom, M.A., and Mirny, L.A. (2013). Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. Nat. Rev. Genet. *14*, 390–403.

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.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Duan, Z., Andronescu, M., Schutz, K., McIlwain, S., Kim, Y.J., Lee, C., Shendure, J., Fields, S., Blau, C.A., and Noble, W.S. (2010). A three-dimensional model of the yeast genome. Nature *465*, 363–367.

Feng, S., and Jacobsen, S.E. (2011). Epigenetic modifications in plants: an evolutionary perspective. Curr. Opin. Plant Biol. *14*, 179–186.

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. USA *107*, 8689–8694.

Fransz, P., De Jong, J.H., Lysak, M., Castiglione, M.R., and Schubert, I. (2002). Interphase chromosomes in *Arabidopsis* are organized as well defined chromocenters from which euchromatin loops emanate. Proc. Natl. Acad. Sci. USA *99*, 14584–14589.

Gibcus, J.H., and Dekker, J. (2013). The hierarchy of the 3D genome. Mol. Cell *49*, 773–782.

Grob, S., Schmid, M.W., Luedtke, N.W., Wicker, T., and Grossniklaus, U. (2013). Characterization of chromosomal architecture in *Arabidopsis* by chromosome conformation capture. Genome Biol. *14*, R129.

Harper, L., Golubovskaya, I., and Cande, W.Z. (2004). A bouquet of chromosomes. J. Cell Sci. *117*, 4025–4032.

Heger, P., Marin, B., Bartkuhn, M., Schierenberg, E., and Wiehe, T. (2012). The chromatin insulator CTCF and the emergence of metazoan diversity. Proc. Natl. Acad. Sci. USA *109*, 17507–17512.

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.

Jin, F., Li, Y., Dixon, J.R., Selvaraj, S., Ye, Z., Lee, A.Y., Yen, C.A., Schmitt, A.D., Espinoza, C.A., and Ren, B. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature *503*, 290–294.

Lafos, M., Kroll, P., Hohenstatt, M.L., Thorpe, F.L., Clarenz, O., and Schubert, D. (2011). Dynamic regulation of H3K27 trimethylation during *Arabidopsis* differentiation. PLoS Genet. *7*, e1002040.

Le, T.B., Imakaev, M.V., Mirny, L.A., and Laub, M.T. (2013). High-resolution mapping of the spatial organization of a bacterial chromosome. Science *342*, 731–734.

Li, G., Ruan, X., Auerbach, R.K., Sandhu, K.S., Zheng, M., Wang, P., Poh, H.M., Goh, Y., Lim, J., Zhang, J., et al. (2012). Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell *148*, 84–98.

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.

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.

Moissiard, G., Cokus, S.J., Cary, J., Feng, S., Billi, A.C., Stroud, H., Husmann, D., Zhan, Y., Lajoie, B.R., McCord, R.P., et al. (2012). MORC family ATPases required for heterochromatin condensation and gene silencing. Science *336*, 1448–1451.

Osborne, C.S., Chakalova, L., Brown, K.E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J.A., Lopes, S., Reik, W., and Fraser, P. (2004). Active genes dynamically colocalize to shared sites of ongoing transcription. Nat. Genet. *36*, 1065–1071.

Probst, A.V., Fransz, P.F., Paszkowski, J., and Mittelsten Scheid, O. (2003). Two means of transcriptional reactivation within heterochromatin. Plant J. *33*, 743–749.

Rajakumara, E., Law, J.A., Simanshu, D.K., Voigt, P., Johnson, L.M., Reinberg, D., Patel, D.J., and Jacobsen, S.E. (2011). A dual flip-out mechanism for 5mC recognition by the *Arabidopsis* SUVH5 SRA domain and its impact on DNA methylation and H3K9 dimethylation in vivo. Genes Dev. *25*, 137–152.

Rosa, S., De Lucia, F., Mylne, J.S., Zhu, D., Ohmido, N., Pendle, A., Kato, N., Shaw, P., and Dean, C. (2013). Physical clustering of FLC alleles during Polycomb-mediated epigenetic silencing in vernalization. Genes Dev. *27*, 1845–1850.

Sanyal, A., Lajoie, B.R., Jain, G., and Dekker, J. (2012). The long-range interaction landscape of gene promoters. Nature *489*, 109–113.

Scherthan, H. (2007). Telomere attachment and clustering during meiosis. Cell. Mol. Life Sci. *64*, 117–124.

Schubert, V. (2014). RNA polymerase II forms transcription networks in rye and *Arabidopsis* nuclei and its amount increases with endopolyploidy. Cytogenet. Genome Res. Published online July 18, 2014. http://dx.doi.org/10.1159/000365233.

Schubert, V., Berr, A., and Meister, A. (2012). Interphase chromatin organisation in *Arabidopsis* nuclei: constraints versus randomness. Chromosoma *121*, 369–387.

Schwartz, Y.B., Kahn, T.G., Nix, D.A., Li, X.Y., Bourgon, R., Biggin, M., and Pirrotta, V. (2006). Genome-wide analysis of Polycomb targets in *Drosophila melanogaster*. Nat. Genet. *38*, 700–705.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the *Drosophila* genome. Cell *148*, 458–472.

Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V., and Ren, B. (2012). A map of thecis-regulatory sequences in the mouse genome. Nature *488*, 116–120.

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.

Stroud, H., Hale, C.J., Feng, S., Caro, E., Jacob, Y., Michaels, S.D., and Jacobsen, S.E. (2012). DNA methyltransferases are required to induce heterochromatic re-replication in *Arabidopsis*. PLoS Genet. *8*, e1002808.

Stroud, H., Greenberg, M.V., Feng, S., Bernatavichute, Y.V., and Jacobsen, S.E. (2013). Comprehensive analysis of silencing mutants reveals complex regulation of the *Arabidopsis* methylome. Cell *152*, 352–364.

Stroud, H., Do, T., Du, J., Zhong, X., Feng, S., Johnson, L., Patel, D.J., and Jacobsen, S.E. (2014). Non-CG methylation patterns shape the epigenetic landscape in *Arabidopsis*. Nat. Struct. Mol. Biol. *21*, 64–72.

Tanizawa, H., Iwasaki, O., Tanaka, A., Capizzi, J.R., Wickramasinghe, P., Lee, M., Fu, Z., and Noma, K. (2010). Mapping of long-range associations throughout the fission yeast genome reveals global genome organization linked to transcriptional regulation. Nucleic Acids Res. *38*, 8164–8177.

Tolhuis, B., de Wit, E., Muijrers, I., Teunissen, H., Talhout, W., van Steensel, B., and van Lohuizen, M. (2006). Genome-wide profiling of PRC1 and PRC2 Polycomb chromatin binding in *Drosophila melanogaster*. Nat. Genet. *38*, 694–699.

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.

Zemach, A., McDaniel, I.E., Silva, P., and Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science *328*, 916–919.

Zhang, X., Clarenz, O., Cokus, S., Bernatavichute, Y.V., Pellegrini, M., Goodrich, J., and Jacobsen, S.E. (2007). Whole-genome analysis of histone H3 lysine 27 trimethylation in *Arabidopsis*. PLoS Biol. *5*, e129.

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, Y., McCord, R.P., Ho, Y.J., Lajoie, B.R., Hildebrand, D.G., Simon, A.C., Becker, M.S., Alt, F.W., and Dekker, J. (2012). Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. Cell *148*, 908–921.

Molecular Cell, Volume *55*

Supplemental Information

Genome-wide Hi-C Analyses in Wild-Type and Mutants Reveal High-Resolution Chromatin Interactions in *Arabidopsis*

Suhua Feng, Shawn J. Cokus, Veit Schubert, Jixian Zhai, Matteo Pellegrini, and Steven E. Jacobsen

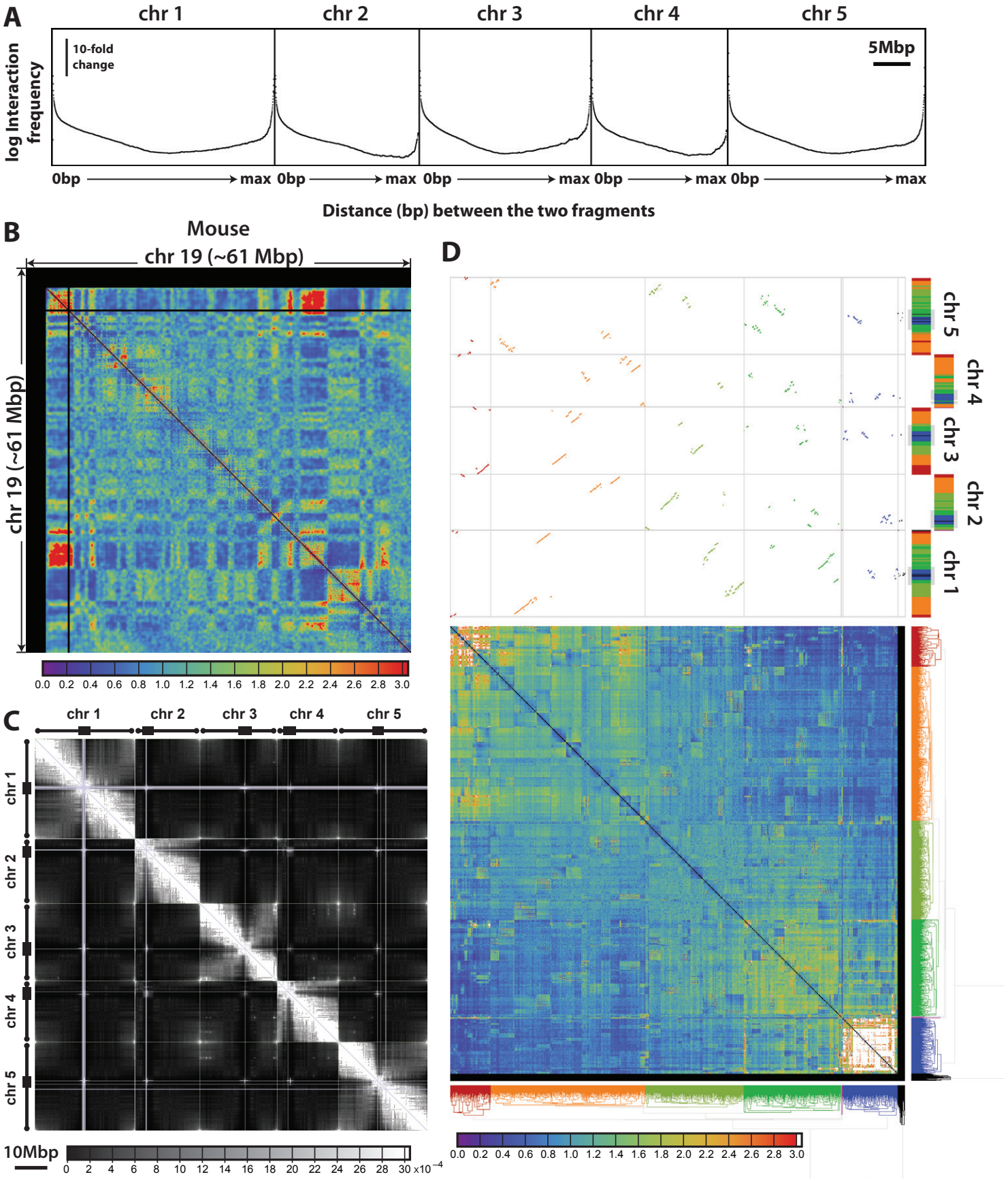
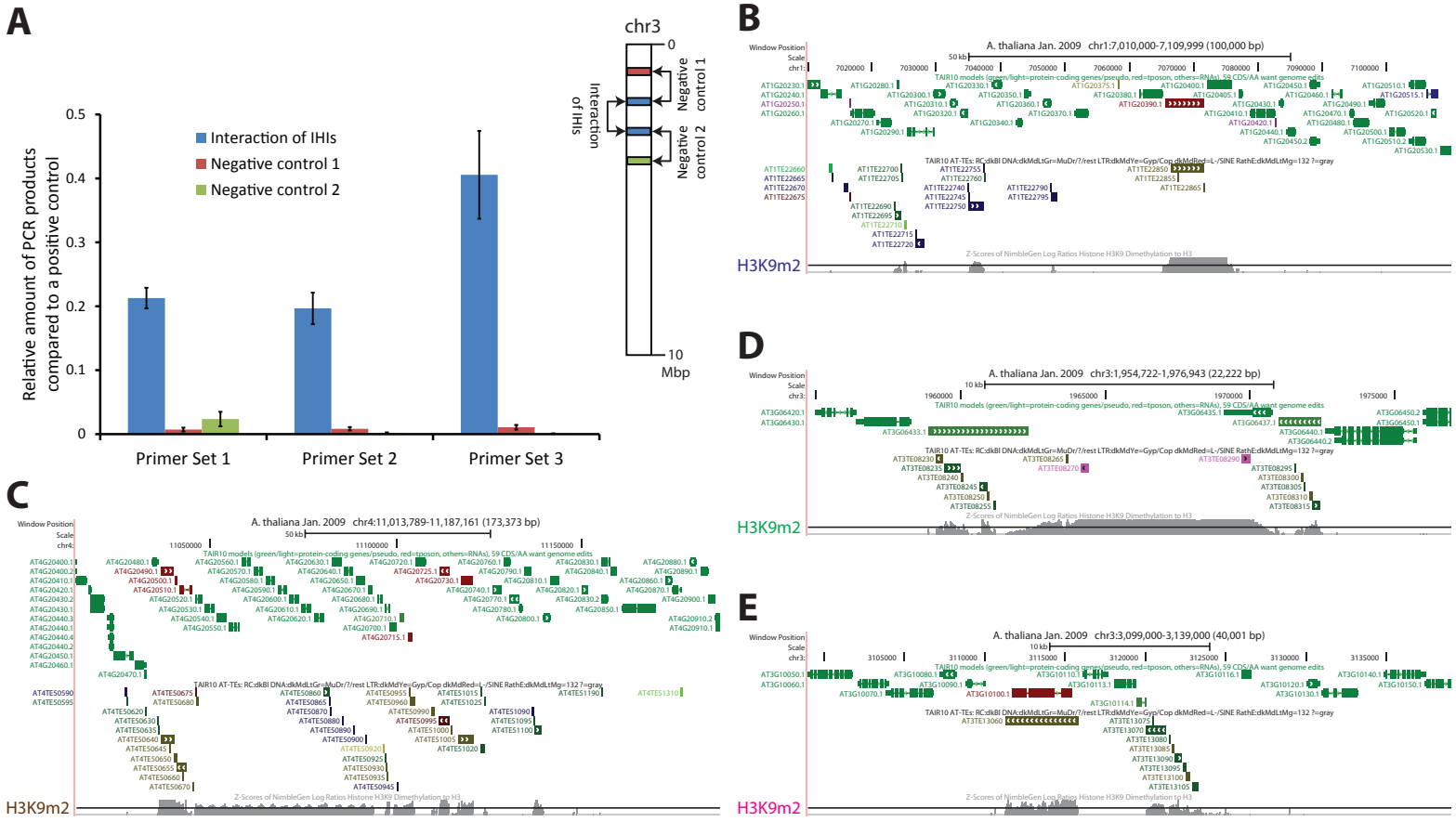

A
chr 1
chr 2
chr 3
chr 4
chr 5
log Interaction frequency
10-fold change
5Mbp
0bp
max
Distance (bp) between the two fragments
B
Mouse
chr 19 (~61 Mbp)
chr 19 (~61 Mbp)
0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4 1.6 1.8 2.0 2.2 2.4 2.6 2.8 3.0
C
chr 1
chr 2
chr 3
chr 4
chr 5
10Mbp
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 x10^-4
D
chr 5
chr 4
chr 3
chr 2
chr 1
0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4 1.6 1.8 2.0 2.2 2.4 2.6 2.8 3.0

Figure S1


A
Relative amount of PCR products compared to a positive control
0.5
0.4
0.3
0.2
0.1
0
Interaction of IHIs
Negative control 1
Negative control 2
Primer Set 1
Primer Set 2
Primer Set 3
chr3
0
10
Mbp
Interaction of IHIs
Negative control 1
Negative control 2
B
A. thaliana Jan. 2009 chr1:7,010,000-7,109,999 (100,000 bp)
H3K9m2
C
A. thaliana Jan. 2009 chr4:11,013,789-11,187,161 (173,373 bp)
H3K9m2
D
A. thaliana Jan. 2009 chr3:1,954,722-1,976,943 (22,222 bp)
H3K9m2
E
A. thaliana Jan. 2009 chr3:3,099,000-3,139,000 (40,001 bp)
H3K9m2

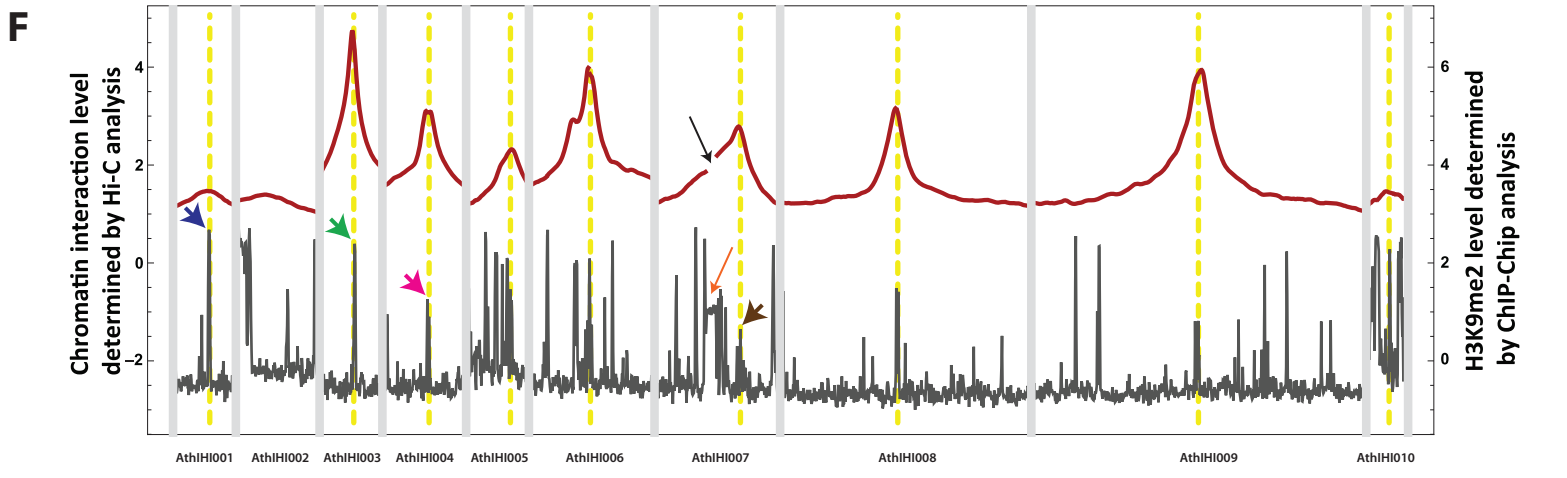

F
Chromatin interaction level determined by Hi-C analysis
H3K9me2 level determined by ChIP-Chip analysis
AthIHI001
AthIHI002
AthIHI003
AthIHI004
AthIHI005
AthIHI006
AthIHI007
AthIHI008
AthIHI009
AthIHI010

Figure S2

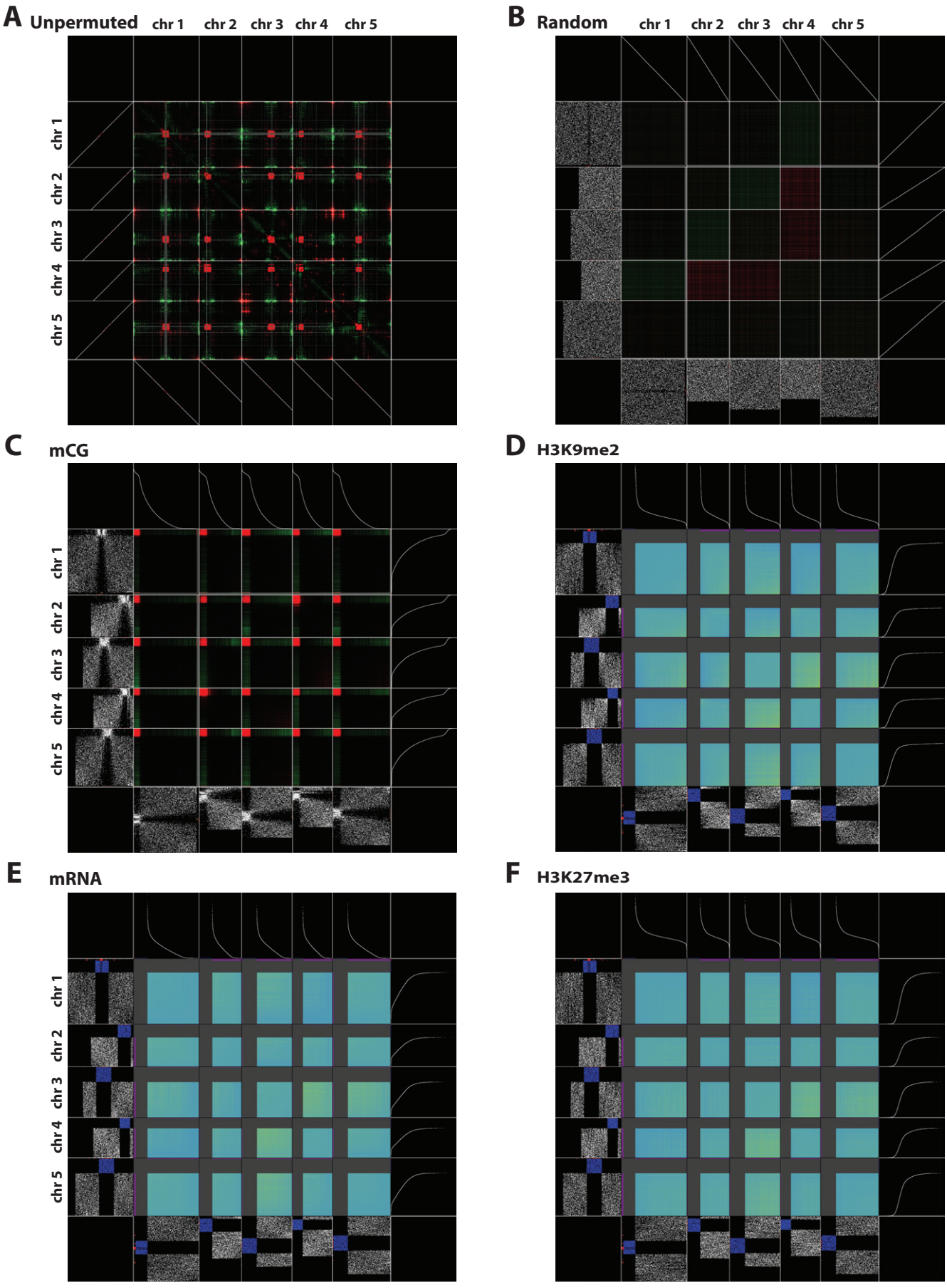

A Unpermuted
chr 1 chr 2 chr 3 chr 4 chr 5
chr 1
chr 2
chr 3
chr 4
chr 5
B Random
chr 1 chr 2 chr 3 chr 4 chr 5
C mCG
chr 1
chr 2
chr 3
chr 4
chr 5
D H3K9me2
E mRNA
chr 1
chr 2
chr 3
chr 4
chr 5
F H3K27me3

Figure S3

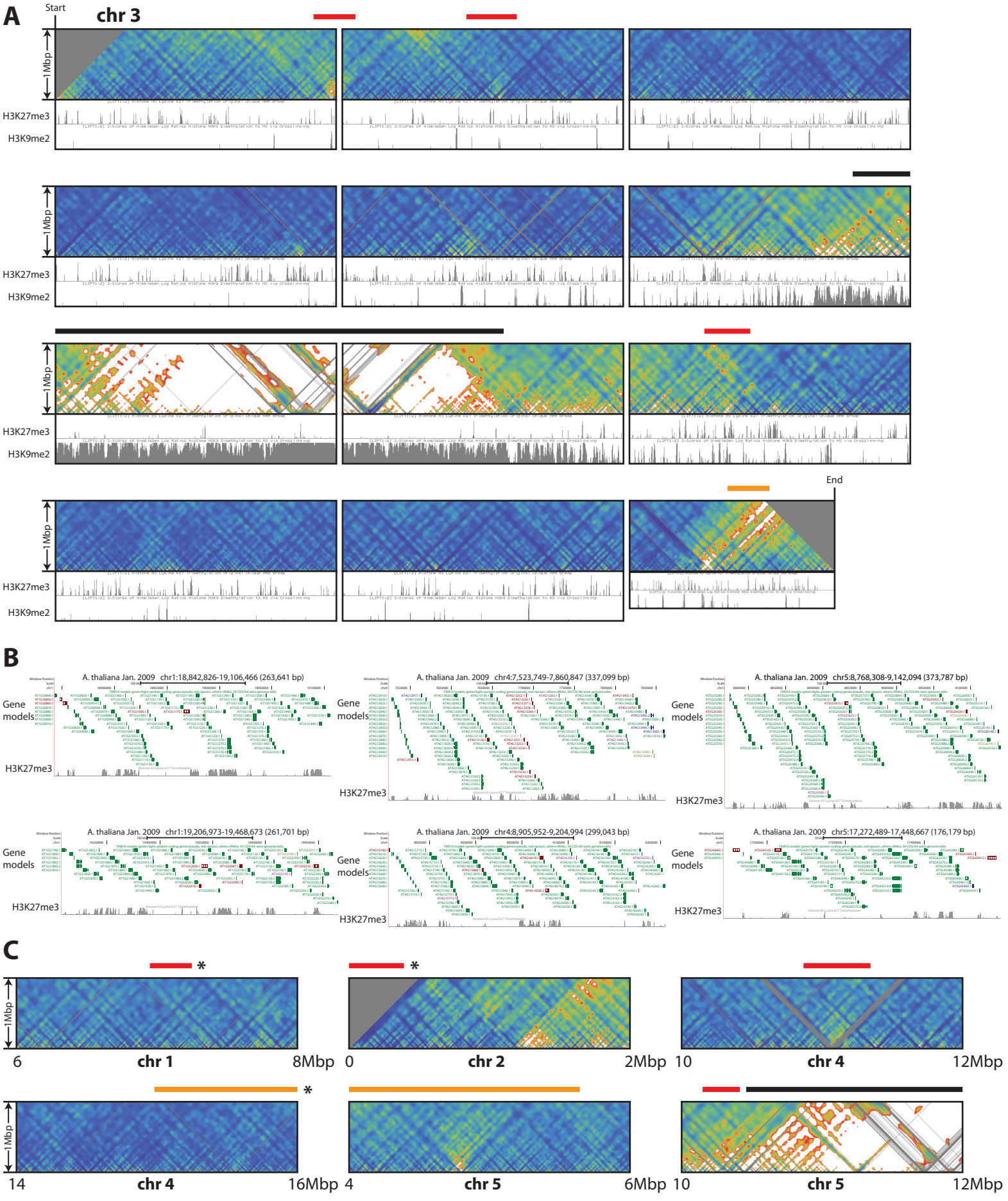

A
Start
chr 3
1Mbp
H3K27me3
H3K9me2
End
B
A. thaliana Jan. 2009 chr1:18,842,826-19,106,466 (263,641 bp)
A. thaliana Jan. 2009 chr4:7,523,749-7,860,847 (337,099 bp)
A. thaliana Jan. 2009 chr5:8,768,308-9,142,094 (373,787 bp)
A. thaliana Jan. 2009 chr1:19,206,973-19,468,673 (261,701 bp)
A. thaliana Jan. 2009 chr4:8,905,952-9,204,994 (299,043 bp)
A. thaliana Jan. 2009 chr5:17,272,489-17,448,667 (176,179 bp)
Gene models
H3K27me3
C
1Mbp
6
chr 1
8Mbp
0
chr 2
2Mbp
10
chr 4
12Mbp
14
chr 4
16Mbp
4
chr 5
6Mbp
10
chr 5
12Mbp

Figure S4

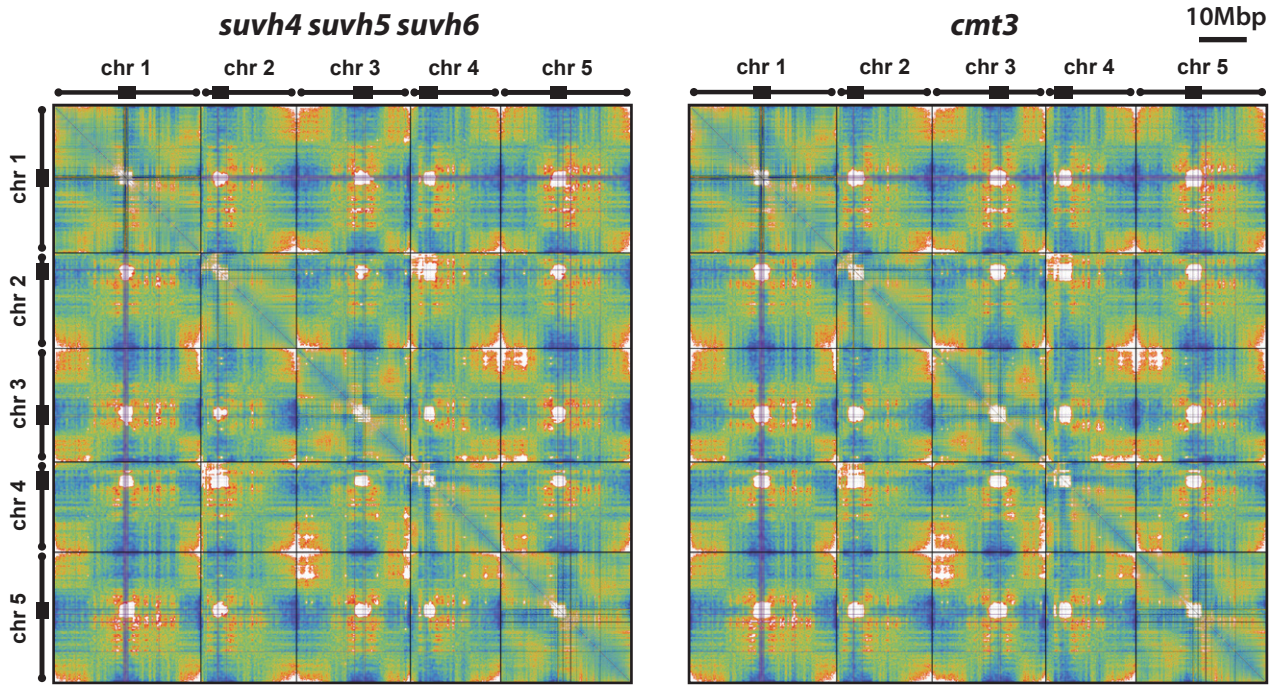


Figure S5

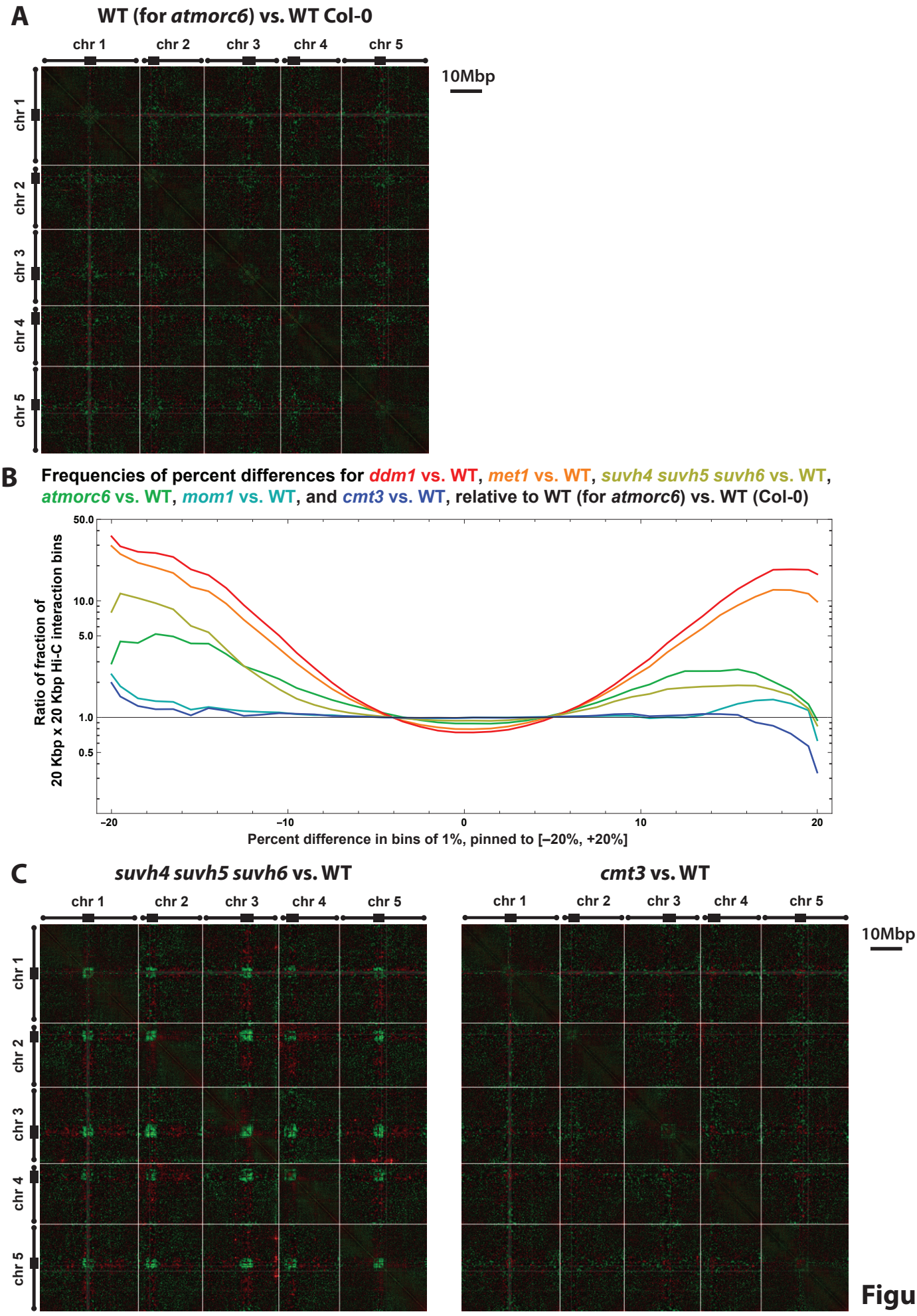

A
WT (for atmorc6) vs. WT Col-0
chr 1
chr 2
chr 3
chr 4
chr 5
10Mbp
B
Frequencies of percent differences for ddm1 vs. WT, met1 vs. WT, suvh4 suvh5 suvh6 vs. WT, atmorc6 vs. WT, mom1 vs. WT, and cmt3 vs. WT, relative to WT (for atmorc6) vs. WT (Col-0)
Ratio of fraction of 20 Kbp x 20 Kbp Hi-C interaction bins
50.0
10.0
5.0
1.0
0.5
−20
−10
0
10
20
Percent difference in bins of 1%, pinned to [−20%, +20%]
C
suvh4 suvh5 suvh6 vs. WT
cmt3 vs. WT
10Mbp

Figure S6

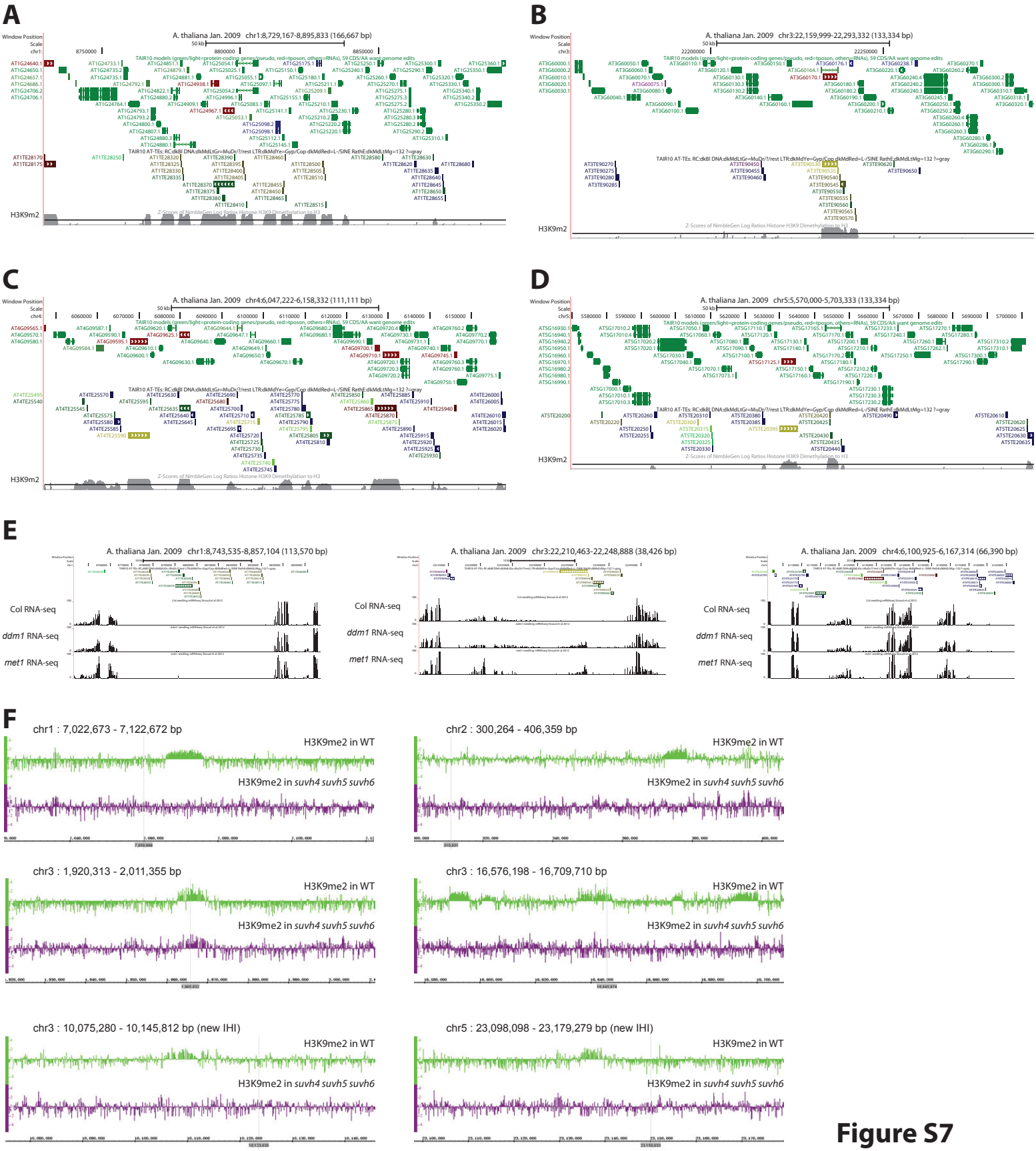

A
A. thaliana Jan. 2009 chr1:8,729,167-8,895,833 (166,667 bp)
H3K9m2
B
A. thaliana Jan. 2009 chr3:22,159,999-22,293,332 (133,334 bp)
H3K9m2
C
A. thaliana Jan. 2009 chr4:6,047,222-6,158,332 (111,111 bp)
H3K9m2
D
A. thaliana Jan. 2009 chr5:5,570,000-5,703,333 (133,334 bp)
H3K9m2
E
A. thaliana Jan. 2009 chr1:8,743,535-8,857,104 (113,570 bp)
A. thaliana Jan. 2009 chr3:22,210,463-22,248,888 (38,426 bp)
A. thaliana Jan. 2009 chr4:6,100,925-6,167,314 (66,390 bp)
Col RNA-seq
ddm1 RNA-seq
met1 RNA-seq
F
chr1 : 7,022,673 - 7,122,672 bp
chr2 : 300,264 - 406,359 bp
chr3 : 1,920,313 - 2,011,355 bp
chr3 : 16,576,198 - 16,709,710 bp
chr3 : 10,075,280 - 10,145,812 bp (new IHI)
chr5 : 23,098,098 - 23,179,279 bp (new IHI)
H3K9me2 in WT
H3K9me2 in suvh4 suvh5 suvh6

Figure S7

SUPPLEMENTAL FIGURE LEGENDS

Figure S1. Additional details of chromatin interaction patterns in Arabidopsis, Related to Figure 1

(A) Log-scale average interaction by genomic distance in wild-type Col-0. The five Arabidopsis chromosomes are shown left to right, and genomic distance increases left to right within each chromosome. Levels are vastly elevated both at short and long genomic distances (with the largest distances in each chromosome arising from the two telomeres on the opposite sides of that chromosome, suggesting that they interact with each other strongly).

(B) Topological domains, when present, can be identified with the processing and visualizations of this study. Mouse chromosome 19 for mESC cells is shown as an example (aligned reads taken from NCBI GEO accession GSE35156 file "GSE35156_GSM862720_J1_mESC_HindIII_ori_HiC.nodup.hic.summary.txt.gz"). Colors blue to red and white are as in Figure 1A. Gray/black indicates areas withheld from analysis due to, e.g., large stretches of "N"s in the mouse reference genome. (Note that the UCSC mm9 mouse reference chr19 sequence begins with 3 Mbp of "N"s.)

(C) Two dimensional interaction map of wild-type Col-0, without removal of distance-related interactivity. Dark (black) to light (white) gray is low to high interaction. Light green lines mark chromosome boundaries and light blue indicates areas withheld from analysis due to, e.g., problematic 50-mer mapping.

(D) Hierarchical clustering analysis of chromatin interaction map of wild-type Col-0 reveals major interactive domains of Arabidopsis chromosomes. Major clusters in the dendrogram are marked by different colors. Correspondingly, the genomic locations of all the 100 Kbp-sized regions from each color-coded cluster are indicated along the five Arabidopsis chromosomes using the same color scheme. Red clusters in general correspond to telomeres and subtelomeric regions, orange clusters in general correspond to centromere distal euchromatin

regions, both light green and green clusters in general correspond to centromere proximal euchromatin regions, and blue clusters in general correspond to pericentromeres. Purple clusters at the beginning of chromosomes 2 and 4 correspond to NORs. Black indicates areas withheld from analysis due to, e.g., problematic 50-mer mapping.

Color bars for panels (B), (C), and (D) are shown at the bottom of the respective panels. See Supplemental Experimental Procedures online for details.

Figure S2. Additional details of Interactive Heterochromatic Islands (IHIs) in Arabidopsis, Related to Figure 2

(A) 3C and quantitative PCR (qPCR) analyses of interaction between IHIs. Three biological replicas of wild-type Arabidopsis were used in the analysis. For each biological replicate, qPCR was performed in duplicate. Data are represented as mean +/– one standard deviation. Blue bars indicate the interaction between the two IHIs shown in Figure 2A. Red bars indicate the interaction between one of the IHIs and a negative control region. Green bars indicate the interaction between the other IHI and another negative control region. Correspondingly, in the diagram of chromosome 3, blue indicates the approximate locations of PCR primers inside the two IHIs, and red and green indicate the approximate locations of PCR primers inside the two control regions. Within all the PCR primer pairs, the linear distance between the two primers is about the same, ~1 Mbp. For positive control of 3C, two primers that are about 5.5 Kbp from each other were used (the two primers are separated by two HindIII restriction sites on the linear chromosome). Sequences of the primers and the combinations used in qPCR are listed in Table S1D.

(B-E) UCSC Genome Browser views of prominent IHIs from wild-type Col-0. Tracks from top to bottom are TAIR10 gene models, TAIR10 transposable elements (TEs), and H3K9me2 ChIP-Chip. The four IHIs chosen are from those shown in Figure 2A/C. Details of IHIs are found in Table S1.

(F) Chromatin interaction levels (*y* axis on the left) and H3K9me2 levels (*y* axis on the right) across the IHIs (*x* axis). The names of IHIs (from AthIHI001 to AthIHI010, as labeled at the bottom of the panel) and the corresponding genomic coordinates of each IHI can be found in Table S1A. Colored arrowheads indicate the H3K9me2 peaks that overlap with the peaks of chromatin interaction and are also shown in panels B to E (blue: panel B; green: panel C; pink: panel D; and brown: panel E). Vertical dotted yellow lines illustrate the strong tendency for tight colocation (within a few tens of Kbp; ~1.6% to ~4% of IHI width for the four largest IHIs) of the highest chromatin interaction levels with sharp spikes in H3K9me2 levels. A black arrow indicates a region withheld from analysis due to, e.g., problematic 50-mer mapping, and the aberrant H3Kme2 level associated with this region (indicated by an orange arrow) probably results from analyzing microarray probes in repetitive DNA. Note that the rightmost IHI shown in the graph (number 10) is located next to the pericentromere of chromosome 5 (bottom right panel in Figure S4C and Table S1), which explains why the overall H3K9me2 level in this IHI is higher than that in other IHIs.

Figure S3. Additional details of genomic features and chromatin interactions, **Related to Figure 3**

(A) Two dimensional interaction map of wild-type Col-0 in the style of Figure 3, but without any permutation of chromosomal positions. Red indicates higher and green lower interactions than averages.

(B) Two dimensional interaction map of wild-type Col-0 in the style of Figure 3, after a random permutation of chromosomal positions.

(C) Two dimensional interaction map of wild-type Col-0 in the style of Figure 3, with chromosomal positions permuted based on intensity of CG methylation.

(D-F) Two dimensional interaction maps of wild-type Col-0 in the style of Figure 3, except showing untransformed Hi-C interaction tendency colored as in Figure S1D, and with

pericentromeric regions (dark blue) eliminated from the analysis (gray bands). Chromosomal positions are permuted based on intensity of H3K9me2 (D), mRNA abundance (E), and H3K27me3 (F).

Color scales in panels (A) to (C) are the same as Figure 3. Color scales in panels (D) to (F) are the same as in Figure S1D.

Figure S4. Additional details of local interactive domains in Arabidopsis, **Related to Figure 4**

(A) Local interaction detail in the style of Figure 4 for the entirety of chromosome 3, in consecutive 2 Mbp-long blocks to a genomic distance of 1 Mbp (except for the last block, which is shorter than 2 Mbp). The H3K27me3 and H3K9me2 tracks shown are from the UCSC Genome Browser.

(B) UCSC Genome Browser views of selected small local interactive domains that overlap predominantly with consecutive H3K27me3-modified regions. Top track is TAIR10 gene models, and bottom track is H3K27me3 ChIP-Chip. The six domains chosen are from those shown in Figure 4.

(C) Local interaction detail in the style of (A) for regions in chromosomes 1, 2, 4, and 5 that contain IHIs.

In (A) and (C), pericentromeres are labeled by black bars on top. Red bars are over the seven IHIs shown in Figure 2A/C. Orange bars label the three IHIs shown in Figure 2D. Most IHIs correspond to local interactive domains (see text for details); the IHIs that do not correspond to local interactive domains are marked by an asterisk to the right of its red or orange bar. Details of IHIs are found in Table S1.

Color scales in panels (A) and (C) are the same as in Figure 4.

Figure S5. Additional details of chromatin interaction patterns in mutants affecting epigenetic processes, **Related to Figure 5**

Two dimensional interaction maps of *suvh4 suvh5 suvh6* and *cmt3*, in the style of Figure 1A. Color scales are the same as in Figure 1A.

See also File S5.

Figure S6. Additional details of comparison of interaction patterns across mutants and wild type, **Related to Figure 6**

(A) Comparison of chromatin interaction maps of the wild type used as control for *atmorc6* vs. wild-type Col-0, in the style of Figure 6.

(B) Comparison of the differences in chromatin interaction observed in mutants vs. wild type and the differences observed in wild type vs. wild type. Colored lines are the ratios of the percent differences of the indicated mutant/wild type pairs over the percent differences of the wild type/wild type pair. A thin black line illustrates a constant ratio of 1.

(C) Comparison of chromatin interaction maps of *suvh4 suvh5 suvh6* and *cmt3* vs. wild type, in the style of Figure 6.

Color scales in panels (A) and (C) are the same as in Figure 6.

See also File S5.

Figure S7. Additional details of dynamics of IHIs in mutants, **Related to Figure 7**

(A-D) UCSC Genome Browser views of four selected newly-recruited IHIs in mutants. Tracks top to bottom are TAIR10 gene models, TAIR10 transposable elements (TEs), and H3K9me2 ChIP-Chip. The four chosen are from those shown in Figure 7.

(E) UCSC Genome Browser views of three selected newly-recruited IHIs in mutants. The top track shows TAIR10 transposable elements (TEs) and the bottom three tracks show RNA-Seq for the indicated genotypes. The three chosen are from those shown in Figure 7. The region on

chromosome 3 shows a slight de-repression of TEs, while the other two regions (on chromosomes 1 and 4) do not show signs of TE de-repression.

(F) Integrated Genome Browser (IGB) views of H3K9me2 ChIP-Seq from wild type Col-0 and *suvh4 suvh5 suvh6* within six regions corresponding to IHIs. The six regions are chosen from those shown in Figures 2 and 7. The four regions in the top and middle rows are IHIs found in wild type, and the two regions in the bottom row are IHIs found in *suvh4 suvh5 suvh6*. Details of IHIs are found in Table S1.

SUPPLEMENTAL TABLE

Table S1. Description of IHIs revealed by Hi-C analysis, Related to Figures 2 and 7

A. Prominent IHIs found in wild-type Col-0:

Genotype	Location	Approximate start position (bp)	Approximate end position (bp)	Interactions presented in…	Forms a local interactive domain in Figure S4?	Label in Figure S2F
Col-0	chr1	6,900,001	7,200,000	Figure 2C	No	AthIHI001
	chr2	1	400,000	Figure 2C	No	AthIHI002
	chr3	1,800,001	2,100,000	Figure 2A, C, and D	Yes	AthIHI003
	chr3	2,900,001	3,300,000	Figure 2A, C, and D	Yes	AthIHI004
	chr3	16,500,001	16,800,000	Figure 2C	Yes	AthIHI005
	chr3	22,300,001	22,900,000	Figure 2D	Yes	AthIHI006
	chr4	10,800,001	11,400,000	Figure 2C	Yes	AthIHI007
	chr4	15,000,001	16,200,000	Figure 2D	No	AthIHI008
	chr5	4,000,001	5,600,000	Figure 2D	Yes	AthIHI009
	chr5	10,200,001	10,400,000	Figure 2C	Yes	AthIHI010

B. Information of BACs used in DNA-FISH:

BACs used to analyze the interaction of the two chr3 IHIs presented in Figure 2A:	Control BACs:
F24P17 (chr3: 1,906,274 – 1,992,295)	MMM17 (chr3: 4,455,128 – 4,536,177)
T22K18 (chr3: 3,047,305 – 3,143,536)	MGL6 (chr3: 5,633,951 – 5,713,409)

C. New and changed IHIs in mutant Arabidopsis:

Genotype	Location	Approximate start position (bp)	Approximate end position (bp)	Interactions presented in…	Color of focus in Figure 7
atmorc6	chr2	4,000,001	4,600,000	Figure 7A	Red
ddm1	chr1	5,000,001	5,200,000	Figure 7B	Red
	chr1	8,500,001	9,000,000	Figure 7B	Red
	chr1	20,200,001	20,500,000	Figure 7B	Red
	chr1	20,900,001	21,400,000	Figure 7B	Red
	chr3	10,000,001	10,200,000	Figure 7B	Red
	chr3	22,100,001	22,300,000	Figure 7B	Red
	chr3	22,700,001	22,800,000	Figure 7B	Red
	chr3	23,000,001	23,200,000	Figure 7B	Red
	chr4	5,800,001	6,300,000	Figure 7B	Red
	chr4	14,900,001	15,100,000	Figure 7B	Red
	chr5	5,600,001	5,800,000	Figure 7B	Red
suvh4 suvh5 suvh6	chr3	10,000,001	10,200,000	Figure 7C	Red
	chr3	22,700,001	22,800,000	Figure 7C	Red
	chr5	23,100,001	23,300,000	Figure 7C	Red

D. PCR primers used in 3C qPCR analysis:

Combination	Names	Sequences	Genomic interval (bp)	Notes
Primer Set 1:	JP11701	5'-TTGTCATTGATGTACTTCACTCTTTTTATC-3'	chr3: 1,973,034 – 1,973,063	Primer for IHI
	JP11707	5'-TAAAGATAATGAGAAATGATGGGAAAGTAG-3'	chr3: 3,130,189 – 3,130,218	Primer for IHI
	JP11712	5'-ATCTATCACCAAAACTCAGAGAAACTAATC-3'	chr3: 1,004,297 – 1,004,326	Control 1, used with JP11701
	JP11717	5'-ATGTTTTTATACTCGTGAACTTGAATTGAG-3'	chr3: 4,016,920 – 4,016,949	Control 2, used with JP11707
Primer Set 2:	JP11696	5'-TACCGTACCCACTTAAAACTATGTTCTG-3'	chr3: 1,957,395 – 1,957,422	Primer for IHI
	JP11703	5'-CTGCCTAGTTCTCAACTTATCTCCTCTTTA-3'	chr3: 3,122,532 –	Primer for IHI

			3,122,561	
	JP11708	5'-AGAGTATGTGGCCTAAGCTCTTTATAACAT-3'	chr3: 998,541 – 998,570	Control 1, used with JP11696
	JP11714	5'-CCATATTACAGCAATGATTATGATTTCAAG-3'	chr3: 4,008,243 – 4,008,272	Control 2, used with JP11703
Primer Set 3:	JP11697	5'-CATAATTGATATCTACGTCCTTGTAAGTCC-3'	chr3: 1,959,079 – 1,959,108	Primer for IHI
	JP11703	5'-CTGCCTAGTTCTCAACTTATCTCCTCTTTA-3'	chr3: 3,122,532 – 3,122,561	Primer for IHI
	JP11710	5'-AGTTAACAAGAAGAAGCAGTAAGATACCTC-3'	chr3: 1,000,437 – 1,000,466	Control 1, used with JP11697
	JP11714	5'-CCATATTACAGCAATGATTATGATTTCAAG-3'	chr3: 4,008,243 – 4,008,272	Control 2, used with JP11703
Positive control:	JP10119	5'-AGTACTTCCCCAGGAGCAACTTTATCACCT-3'	chr1: 20,248,723 – 20,248,752	
	JP10122	5'-GAAAGCAACATAACCTTGCAGTTAGCCGTAG-3'	chr1: 20,254,342 – 20,254,372	

Note: Genomic coordinates are against the TAIR9 Arabidopsis Col-0 assembly.

SUPPLEMENTAL FILES

File S1. Full resolution two dimensional interaction maps for wild-type Col-0 in the style of Figure 1A. Permuted interaction maps for wild-type Col-0 in the style of Figure 3. Complete set of local interaction detail views for wild-type Col-0 in the style of Figure 4. Related to Figures 1, 3, and 4.

File S2. Full resolution two dimensional interaction maps and comparison maps over wild-type control for *clf swn* double mutant in the styles of Figures 1A and 6. Related to Figures 5 and 6.

File S3. Full resolution two dimensional interaction maps and comparison maps over wild-type control for *atmorc6* and *mom1* mutants in the styles of Figures 1A and 6. Related to Figures 5, 6, and 7.

File S4. Full resolution two dimensional interaction maps and comparison maps over wild-type control for *met1* and *ddm1* mutants in the styles of Figures 1A and 6. Related to Figures 5, 6 and 7.

File S5. Full resolution two dimensional interaction maps and comparison maps over wild-type control for *suvh4 suvh5 suvh6* triple and *cmt3* mutants in the styles of Figures 1A and 6. Related to Figures 5, 6 and 7.

File S6. Complete set of local interaction detail views for *clf swn*, *atmorc6*, *mom1*, *met1*, *ddm1*, *suvh4 suvh5 suvh6*, and *cmt3* mutants in the style of Figure 4. Related to Figure 4. Note that H3K27me3 and H3K9me2 tracks shown are from wild type on the UCSC Genome Browser.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Plant Materials

Wild–type Arabidopsis in this study is Columbia-0 (Col-0) accession unless indicated otherwise. The *atmorc6-1*, *ddm1-2*, *met1-3*, *cmt3-11*, *suvh4 suvh5 suvh6* triple, and *clf-28 swn-7* double mutants are as described previously (Lafos et al., 2011; Moissiard et al., 2012; Stroud et al., 2013). The *mom1* EMS mutant (line 337) was identified through a previously described forward genetic screen, and the mutation produces a stop codon at amino acid number 603 in the MOM1 protein (Moissiard et al., 2014; Moissiard et al., 2012). Since *atmorc6* and *mom1* are EMS mutagenesis alleles, we also made Hi-C libraries from the parental lines used for the screen (Moissiard et al., 2012) as wild–type controls. Arabidopsis plants were germinated on soil and grown under continuous light, and tissues were harvested at the same developmental stage (four–week–old rosette leaves) for all genotypes, except that *clf swn* tissues were taken from callus grown in liquid medium due to the growth defects of this double mutant (Lafos et al., 2011). The *clf swn* double mutant was first germinated on plates containing MS medium and then grown in liquid MS medium for one month before the callus–like tissues formed by the double mutant were harvested.

Preparation of Nuclei, Probe Labeling, and Fluorescence *in situ* Hybridization (FISH)

Wild–type Col-0 Arabidopsis were grown on agar plates at 21°C under continuous light for 14 days before seedling tissues (without roots) were harvested. Nuclei were isolated and flow–sorted from these seedlings after formaldehyde fixation using a FACS Aria (BD Biosciences) according to their 2C and 4C ploidy level as described previously (Pecinka et al., 2004). The Arabidopsis BACs used for FISH were obtained from the Arabidopsis Biological Resource Center (Columbus, OH, USA). BAC DNA from positions along chromosomes 3 (Figure 2B and Table S1B) was labelled by nick translation with Alexa488–dUTP, Cy3–dUTP, and Texas Red–dUTP according to previously published protocols (Ward, 2002). FISH was performed as described previously (Schubert et al., 2001). Nuclei and chromosomes were counterstained with DAPI (1 μg/ml) in Vectashield (Vector Laboratories).

Microscopic Evaluation, Image Processing, and Statistics

Analysis of FISH signals was performed with an epifluorescence microscope (Zeiss Axiophot) using a 100×/1.45 Zeiss α plan–fluar objective and a three–chip Sony (DXC-950P) color camera. Images were captured separately for each fluorochrome using appropriate excitation and emission filters. Images

were merged using Adobe Photoshop 6.0 software. Euchromatin associations at the ∼100 Kbp segments labeled by BACs were evaluated as described previously (Schubert et al., 2008). The cohesion frequencies were calculated per homolog. One FISH signal cluster and overlapping signals per homolog were regarded as cohesion, two signal clusters as separated. The frequencies of homologous and heterologous associations and of sister chromatid cohesion at distinct BAC positions were compared by two–sided Fisher's exact test. Note that for both 2C and 4C images in Figure 2B, there are sometimes more signals than expected (e.g., more than one signal — red or green — per homolog). This is because elongated chromatin fibers can lead to split FISH signals, especially when using BAC probes that are ≈100 Kbp long. This effect and the method to appropriately evaluate FISH signals under this circumstance have been described previously (Schubert et al., 2008).

3C and Quantitative PCR Analysis

3C assays were performed in the same way as Hi-C (Moissiard et al., 2012), except the omission of the end filling step after the HindIII restriction digestion step. After the ligation of HindIII fragments, ∼100 ng 3C template DNA was used in PCR analysis. Quantitative real–time PCR was carried out using SYBR Green Supermix (Bio-Rad) in an Mx3000P qPCR system (Stratagene). The PCR conditions were as follows: one cycle of 5 min at 95°C, 40 cycles of 30 sec at 95°C, 30 sec at 55°C, and 1 min at 72°C. PCR primer sequences are listed in Table S1D.

Formation of Raw Hi-C Interaction Matrices

Each of the 10 libraries was sequenced on an Illumina HiSeq 2000 as an entire lane in paired end $50+50$ or $51+51$ cycle mode to obtain ∼175 to 270 million raw spots and ∼162 to 231 million PF1 spots per library. Each end of each spot was independently stringently aligned to the TAIR9 *Arabidopsis* reference genome with Bowtie 0.12.7, only keeping ends with exactly one gapless *zero-mismatch* alignment. (NCBI GEO file `GSE35156_GSM862720_J1_mESC_HindIII_ori_HiC.nodup.hic.summary.txt` from URL ⟨http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE35156⟩ was used for alignments to the UCSC mouse mm9 reference, interpreting coordinates as 0–based aligned minimums quoted against reference plus strands for end–to–end gapless 72–mers. For mouse, only chr18/19 were analyzed, as these total ∼152 Mbp [similar to the entire *Arabidopsis* genome], facilitating comparison.)

Only PF1 spots with both ends aligned and neither end aligning to the chloroplast or mitochondrion were retained. Due to the alignment stringency, chimeric and PhiX spike–in reads rarely align; no special

attempt to identify them was made. Spots were not filtered by strands of ends. A read was considered to cross a HindIII site if and only if the genomic sequence it aligned to contains the literal palindromic DNA sequence `AAGCTT`; spots with either or both ends crossing were rejected. The reference genome was partitioned into nominal "fragments" at each literal occurrence of `AAGCTT` (corresponding to a complete digest by HindIII), placing the nominal cut in the middle of this six–mer. Spots with the two ends aligning to the same fragment were rejected. Libraries have ∼41 to 66 million read pairs (∼23 million for mouse) meeting all requirements, with most of the loss due to pairs with either or both ends not having any gapless end–to–end perfect alignments (unique or not).

Depending on the downstream analysis, the reference genome was partitioned into successive 2,500, 20,000, or 100,000 basepair (50,000 for mouse) intervals ("bins") starting at the beginning of each chromosome. "Raw" whole genome (bin, bin) Hi-C interaction real symmetric entrywise non-negative matrices were formed starting from a zero matrix as follows: given a unit mass to be placed on a fragment pair (f_1, f_2) with f_i composed of basepairs F_i, $i \in 1..2$, the mass is imagined to be uniformly distributed over pairs of basepairs in $F_1 \times F_2$, and as each basepair belongs to exactly one bin, this induces a mass contribution to (bin, bin) pairs (i.e., entries of the matrix). Each surviving aligned read pair (fragment f_1, fragment f_2) effectively contributes +1 mass to (f_1, f_2) and +1 mass to (f_2, f_1). (Computationally, the symmetry of the matrix is typically used to reduce storage needs and is not represented explicitly.)

Blacklisting of Problematic Genomic Regions

As with all high–throughput alignment–based analyses, due to representational inaccuracies in the reference genome, library preparation and sequencing biases, repetitiveness of the genome and limitations of alignment, etc., coverage of certain reference genomic locations (here, bins) is anomalous. Based on raw Hi-C matrices of collections of preliminary experiments, *blacklists* of 779 of 44,306 (∼1.8%), 101 of 5,959 (∼1.7%), and 19 of 1,193 (∼1.6%) *Arabidopsis* bins at resolution 2,500, 20,000, and 100,000 bp, respectively, were composed (130 of 3,043 [∼4.3%] for mouse chr18/19 with 50,000 bp bins).

For each bin and preliminary experiment, the raw Hi-C matrix row for that bin was broken into two parts, interaction to same chromosome vs. to different chromosomes, and each part summarized by the number of non-zero entries and total of entries. Various thresholds on these summaries (collapsing over conditions by sum or maximum) or ratios of these to centered–window sliding medians were used to compose the blacklists. The primary constitutent of the blacklists, unsurprisingly, are genomic intervals

near centromeres proper, at the cores deep inside the large regions of pericentromeric heterochromatin in each chromosome (and, for mouse, note UCSC mm9 reference chr18 and chr19 each have a 3 Mbp stretch of consecutive `N`'s).

Hi-C interaction matrix analyses generally treated the entire row and column of blacklisted bins as missing (i.e., as if the bin was not present). Analyses also generally treated as blacklisted/missing those individual (bin, bin) pairs (i.e., matrix entries) that contain at least one (basepair, basepair) pair with both basepairs belonging to the same fragment.

Dynamic Smoothing of Raw Hi-C Interaction Matrices

While the number of contributing Hi-C read pairs per condition is large (being many millions), hugeness of the space (effectively 2–D whole genome HindIII all fragments to all fragments) they are populating as well as the highly non-uniform distribution (short vs. long genomic distances, same vs. different chromosomes) into the space results in typical Hi-C experiments operating in an undersampled regime relative to the presumed true continuous 2–D density distribution being sampled from. This is exacerbated when high resolution (e.g., small bin size) analyses are attempted (especially as each halving of 1–D bin size tends to quarter the number of counts per raw matrix entry, so that counts rapidly fall as bin sizes decrease). Indeed, a raw Hi-C *Arabidopsis* matrix at 2,500 bp resolution has more than 1.9 billion entries, with most entries (at current sequencing depths) essentially being just discrete "counts" of 0 or 1 (fractional values may arise due to fragments straddling bin boundaries, but this does not change the essence). Difficulties then arise in downstream analyses as Hi-C interaction density estimates from single matrix entries are statistically poor and extremely noisy.

Existing Hi-C analyses have employed constant spatial resolution at the expense of statistical control of individual interaction density estimates. The choice has generally either been a fine bin size to obtain high resolution in high coverage areas, leaving low coverage areas to have poor density estimates, or a coarse bin size to obtain usable density estimates widely across the entire Hi-C interaction matrix, but with reduced spatial resolution. Note that due to the discrete nature of the sampling (digital counting of read pairs), an uncertainty principle applies in the absense of, e.g., detailed *a priori* assumptions on the Hi-C interaction (and we wish to be unbiased here and not make strong assumptions about the distributions): spatial genomic resolution trades against quantitation of interaction density resolution; for any fixed depth of sampling, one of these two can be relatively high, but not both at the same time.

In this work, a different approach was taken, placing a lower bound on statistical quality of density estimates everywhere, but at the expense of constant genomic spatial resolution. Instead, spatial resolution is high in regions of high coverage and lower (by necessity) in regions of low coverage; spatial resolution becomes *dynamic* in response to local density variation. A very similar situation (with comparable statistics) is found in astrophysics: digital cameras (e.g., behind telescopes) produce 2–D images with Poisson (counting) noise per pixel (i.e., matrix entry); there is very high dynamic range variation and a mixture of point and diffuse sources across the field; and there is a general need for accurate photon (interaction) density estimation throughout the field. Thus, a "dynamic smoothing" ("dyna–smoothing") process, eventually realized to be similar to ASMOOTH (Ebeling et al., 2006) used for *Chandra* X–ray images, is being developed with a preliminary version applied to the raw Hi-C interaction matrices here. That many averages and weighted averages of increasing numbers of independent Poissons are increasingly likely to have low relative error to the corresponding average of their true rates is key, with lower bounds on the total "counts" (e.g., 100 or 200 as used here) contributing able to control the relative error with relatively high probability.

A smoothed density estimate at a matrix entry is a weighted average (determined by a smoothing kernel of variable size, such as a Gaussian) of masked matrix entries in a region around the entry. Initially, these regions contain just the entries themselves (i.e., there is no smoothing). Certain entries may have kernel–pooled counts (initially, raw Hi-C interaction matrix values) sufficiently high that it is statistically likely that the weighted average (having Poisson counting ambiguities) is close in relative error to the true density (scaled by the number of counts in the experiment); these entries are done (the density estimate from the average being accepted) and no longer participate or propagate (becoming masked from future iterations). Other entries have lower counts and require more averaging before the statistics of low counts results in an average that is statistically likely to have low relative error to the true average density; kernels are incrementally enlarged and the process repeated until all matrix entries are replaced with density estimates of sufficient probable quality or the smoothing radius (kernel size) is untenably large. In this way, high–count, sharp features are retained (and do not unduly "bleed" to nearby entries), while low–count diffuse regions are smoothed until their density estimates are directly representative: the output of dyna–smoothing gives our spatially sharpest estimate of observed interaction density given the depth of sequencing performed for the desired level of control of error.

To mitigate the computational intensiveness of the many needed algorithm iterations and generally very large (gigabyte–sized) matrices, an efficient implementation was coded in C++ using Intel AVX vector intrinsics and OpenMP–based multithreading, operating on each raw Hi-C interaction submatrix corresponding to one pair of chromosomes at a time. For successive iterations with smoothing radius $R = 0, 1, 2, \ldots$, Gaussian smoothing of a matrix M was approximated as $V(V(V(H(H(H(M,a),b),c),a),b),c)$, where $H(\cdot, r)$ is a centered horizontal (i.e., row) box blur over $\pm r$ entries and $V(\cdot, r)$ is a centered vertical (i.e., column) box blur over $\pm r$ entries, with non-negative integers a, b, $c \in \{\lfloor R/3 \rfloor, \lceil R/3 \rceil\}$, $a+b+c = R$. (This has support of $(2R+1) \times (2R+1)$ bins and approximately corresponds to *Mathematica* kernel `GaussianMatrix[`$\approx 0.726R + 1.179$`]`, which has standard deviation $\approx 0.320R + 0.586$ bins.) The upper limit for radius R was dependent on bin size and high enough to permit the support of the largest kernel to exceed or approach the size of the largest chromosome (or extend ≈5 Mbp for 2,500 bp bins). Faster performance would likely be achieved by moving to a GPGPU–based implementation (assuming large memory GPUs are available), given the natural fit of the simple core loops to GPU–style architectures and the massive bandwidth of such platforms to accelerator–local memory.

The table below gives statistics on the smoothing radius R and approximate equivalent Gaussian standard deviation σ at which entries in this work terminate dynamic smoothing. For *Arabidopsis*, Col0 wild type is presented, which is typical.

Non-blacklisted entries	**Analysis**	**Median σ**	**10$^{\text{th}}$% σ**	**90$^{\text{th}}$% σ**	**$R = 0$**	**$R \leq 1$**
Chromosome to same chromosome	*Ara.* 2,500 bp	∼35 Kbp	∼16 Kbp	∼64 Kbp	∼3%	∼4%
	Ara. 20,000 bp	∼37 Kbp	∼18 Kbp	∼69 Kbp	∼3%	∼24%
	Ara. 100,000 bp	∼91 Kbp	∼59 Kbp	∼91 Kbp	∼32%	∼96%
	Mus 50,000 bp	∼173 Kbp	∼45 Kbp	∼269 Kbp	∼8%	∼12%
Chromosome to different chromosome	*Ara.* 2,500 bp	∼81 Kbp	∼53 Kbp	∼160 Kbp	∼3%	∼3%
	Ara. 20,000 bp	∼89 Kbp	∼57 Kbp	∼165 Kbp	∼3%	∼3%
	Ara. 100,000 bp	∼155 Kbp	∼91 Kbp	∼251 Kbp	∼2%	∼43%
	Mus 50,000 bp	∼717 Kbp	∼541 Kbp	∼861 Kbp	∼9%	∼9%

Empirical confirmation of the efficacy of the dynamic smoothing process and the level of relative error control achieved can be performed as follows: given a binned 2–D probability density as a known truth for interaction, simulate the Poisson sampling process of Hi-C from this density for a total number

of placed read pairs as seen in actual experiments to obtain a raw interaction matrix for which the true density it arises from is known. Dyna–smooth this raw matrix and then examine the relative error of the resultant entries to the entries of the known truth. For a realistic examination, the known truth should be typical for biological Hi-C interactions as experimentally observed; a good choice is the dyna–smoothed result of an actual experiment. For example, suppose *Arabidopsis* Col0 at 20,000 bp resolution is taken as known truth, this having not infrequent ≈4.25 orders of magnitude variation in density across entries. For the upper ≈3.25 orders of magnitude, for a very large fraction of entries, the recovered density closely linearly tracks the true density, and with relative error approximately independent of magnitude and having standard deviation $\approx \pm 15\%$ (or better — as expected, relative errors are even lower for the very highest densities, as for these even unsmoothed observations are already well beyond the level needed to establish the relative error control that lower densities can only achieve with smoothing). For the lowest order of density magnitude, linear tracking is still very good but standard deviation of relative errors gradually rises to $\approx \pm 50\%$ (but this is still a considerable degree of control — note that without dyna–smoothing, relative errors are often extremely large, e.g., in very low density areas where observed raw counts contain the occassional 1 in a sea of zeros).

Modeling of Dyna–Smoothed Raw Hi-C Interaction Matrices

As is clear from existing work as well as preliminary experiments presently, Hi-C interaction matrices as observed are subject to certain strong effects related to the library preparation protocol and limitations of short–read alignments. One such issue ("sequenceability") is bins (rows and columns) have varying numbers of read pairs with one end in the bin due to, e.g., variation in the local density of genome–wide unique 50-mers in interaction with the details of where HindIII fragments lie in the genome and how long the fragments are, together with library preparation details that affect the position and width of the distribution of read starts relative to parent fragments. Another issue is the rapid increase of observed interaction to extremely high frequency as the genomic distance between loci on the same chromosome decreases to zero (which is expected due to each chromosome existing in cells as a linear polymer, so that as genomic distance decreases, 3–D physical distance necessarily decreases, making cross–linking and eventual sequenced Hi-C read pairs more likely).

To tease these effects apart from other chromatin interactions of interest, non-blacklisted entries of submatrices S of an $n \times n$ dyna–smoothed raw Hi-C interaction matrix have their entries modeled as a

multiplicative product of several factors:

$$S(i,j) \;=\; \underbrace{RC(i)}_{\text{sequenceability of row}} \cdot \underbrace{RC(j)}_{\text{sequenceability of column}} \cdot \underbrace{D(|i-j|)}_{\text{effect of genomic distance}} \cdot \underbrace{A(i,j)}_{\text{remaining interaction}}$$

with $RC(i) \in (0,\infty)$, $D(d) \in (0,\infty)$, and $A(i,j) \in (0,\infty)$ for i, $j \in 1..n$ and $d \in 1..(n-1)$, as described next (with $D(0)$ fixed to 1 to avoid degeneracy among model variables). The submatrix of each chromosome to itself is modeled separately, as is all chromosomes to all different chromosomes (treating in this last case entries from a chromosome to itself as temporarily blacklisted/missing and omitting the $D(\cdot)$ factors as there is no natural notion of genomic distance between points on different chromosomes). Models are fitted by taking natual logarithms (ln) of both sides of the equation above (resulting in a linear relationship among ln-scale $S_{\ln}(\cdot,\cdot)$, $RC_{\ln}(\cdot)$, $D_{\ln}(\cdot)$, and $A_{\ln}(\cdot,\cdot)$) and least–squares minimizing the Frobenius norm of $A_{\ln}(\cdot,\cdot)$ as variables $RC_{\ln}(\cdot)$, $D_{\ln}(\cdot)$ vary. (Equations involving blacklisted entries are removed, and any unconstrained ln [additive]–variables fixed to 0, or, equivalently, 1 on the original non-ln [multiplicative] scale. Log–scale also reduces influence of outliers and sensitivity to details of blacklist formation.) Note that per–experiment variation in depth of sequencing (i.e., the total number of read pairs contributing to a raw matrix) is absorbed into the model variables; $A(i,j)$ may be viewed as the ratio of observed interaction relative to the expected interaction given the row–column (sequenceability RC) and diagonal (genomic distance D) effects.

The least squares problems arising are typically very large (e.g., A is 12,172 × 12,172 for *Arabidopsis* chromosome 1 to itself with 2,500 bp bins, hence tens of thousands of variables and more than 100 million equations), but very sparse. Hence, one of the iterative class of least squares solution algorithms that only require access to the model matrix via the action of it and its transpose on the vector of variables was used. LSQR was chosen (Paige et al., 1982a, b; C++ code from ⟨http://www.stanford.edu/group/SOL/software/lsqr/cpp/lsqr++.zip⟩ on 2013–05–30 was taken as a base). The initial approximate solution was taken to be $RC^0_{\ln}(i) :=$ (mean of $S_{\ln}(i,\cdot)$ + mean of $S_{\ln}(\cdot,i)$ − mean of $S_{\ln}(\cdot,\cdot)$)/2 = mean of row i of $S_{\ln}$ minus half mean of $S_{\ln}(\cdot,\cdot)$ for $i \in 1..n$, and $D^0_{\ln}(d) :=$ mean of $(S_{\ln}(i,j) - RC^0_{\ln}(i) - RC^0_{\ln}(j)$ over entries (i,j) such that $|i-j| = d$) for $d \in 1..(n-1)$, restricted to non-blacklisted entries. LSQR parameters were relative solution error tolerance goal 10^{-6}, condition limit 10^{15}, zero relative matrix error, zero damping, and iteration limit $\max(4n, 10)$ (generally not reached, as convergence tolerance was typically met). The cross–chromosome model, lacking the $D(\cdot)$ factors, has $RC^0_{\ln}(\cdot)$ as its simple explicit exact solution.

Construction of Figures

Figures 1ABCD, 2ACD, 4ABC, 5AB, S1B, S4AC, and S5 show $A(\cdot,\cdot)$ in non-ln (multiplicative) scale, all initially at 20,000 bp resolution, except Figures 4ABC and S4AC at 2,500 bp resolution and Figure S1B (mouse) at 50,000 bp resolution, and with Figures 1AB, 5AB, and S5 rendered as pixel bitmaps then shrunk five–fold as images (hence final pixels for these correspond to 100,000 bp). Figure S1D begins with $A(\cdot,\cdot)$ in non-ln scale at 100,000 bp resolution, temporarily replacing (for the purpose of clustering) blacklist values with -1.0 and values above 3.0 with 3.0, and then hierarchically clusters rows via Euclidean distance with average linkage, applying the resulting permutation simultaneously to rows and columns of non-ln scale $A(\cdot,\cdot)$, which the figure exhibits. The y-axis of Figure S1A shows fitted model $D(\cdot)$ in log–scale for 20,000 bp resolution. Base data for percent differences (Figures 6, 7ABC, and S6AC) are $A(\cdot,\cdot)$ on non-ln (multiplicative) scale with 20,000 bp bins. After rendering, Figures 6 and S6AC were five–fold shrunk as pixel–based images, hence the resultant pixels for these correspond to 100,000 bp genomic intervals.

Figure S1C begins with a dyna–smoothed raw Hi-C interaction matrix at 100,000 bp resolution. One hundred iterations of the MLE–based sequenceability modeling of existing work (e.g., Imakaev et al., 2012; Moissiard et al., 2012) were applied, with no modeling of the effect of genomic distance between points on same chromosomes performed. Plotted values are the resulting (bin, bin) contact probabilities in linear scale.

Hi-C interaction maps permuted by "signals" — these being UCSC BED or wiggle (WIG) tracks — as in Figures 3AB and S3ABC, were constructed as follows. Start with $A(\cdot,\cdot)$ for 2,500 bp resolution and take $\log_2$ of every (non-blacklisted) entry. Compute a real signal value associated to each bin, and within each chromosome, simultaneously permute rows and columns so that the signal values decrease (breaking ties arbitrarily, placing bins with no signal value last and blacklisted rows and columns after those): for an unpermuted plot, assign values of a strictly decreasing affine function to successive bins of each chromosome (resulting in the identity permutation, equivalent to no permutation). For a random plot, assign a random real number as signal for each bin (resulting in a uniformly random permutation). For BED signals, the signal value in each bin is the fraction of reference genome basepairs in the bin that belong to at least one interval in the track. For wiggles, the signal value in a bin is a weighted average of the wiggle values for the intervals that have non-empty intersection with the bin, the weights being

proportional to the number of basepairs in the intersection of the bin and the interval; bins disjoint from all intervals have no signal value. Partition the typically permuted matrix into 8×8 submatrices, replacing each submatrix by the average of its entries (hence, each row and column now corresponds to a [generally disconnected] collection of 20,000 genomic basepairs). For the pool of non-blacklisted entries from chromosomes to themselves, convert values to z-scores by subtracting the mean of these values and dividing by their standard deviation, and do the same for the pool of non-blacklisted entries from chromosomes to different chromosomes. Render the result as a pixel bitmap in the colors shown at the bottom of Figure 3, and shrink this image eight–fold, with the result that final pixels are at 160 Kbp resolution. The permuted plots of Figure S3DEF are similar, with these differences: *(i)* entries in a "peri" row or column (i.e., those that intersect the previously–defined pericentromeric regions of Bernatavichute et al., 2008) are effectively treated as blacklisted; *(ii)* the permutation order is slightly different, the blacklisted bins followed by peri bins followed by bins with no signal value being placed before bins by decreasing signal rather than after; and *(iii)* values for plotted colors are original $A(\cdot,\cdot)$ entries in non-ln (multiplicative) scale. Wiggle tracks with widely varying values were first log–transformed before the processing of this paragraph began.

The IHI–to–IHI analysis of Figure S2F begins with $A(\cdot,\cdot)$ at 2,500 bp resolution, restricted to the submatrix given by the (discontinuous) subset of rows and columns from the IHI intervals of Table S1A, with each interval enlarged by 50% of the interval's width on each side (pinned to chromosome boundaries when those are reached). The red curves (summarizing Hi-C interaction of the IHI zones to themselves) give row means of this submatrix (omitting blacklisted entries) for rows in the IHI intervals proper (without enlargement). The H3K9me2 signal (dark gray) is from the appropriate UCSC wiggle track, assigning each signal value to the 2,500 bp Hi-C bin containing the middle of its interval and plotting for each bin the mean of values assigned to it.

Finally, note that the *clf-28 swn-7* double T-DNA mutant has rearranged chromosomes.

SUPPLEMENTAL REFERENCES

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.

Ebeling, H., White, D.A., and Rangarajan, F.V.N. (2006). ASMOOTH: a simple and efficient algorithm for adaptive kernel smoothing of two-dimensional imaging data. Mon Not R Astron Soc *368*, 65-73.

Imakaev, M., Fudenberg, G., McCord, R.P., Naumova, N., Goloborodko, A., Lajoie, B.R., Dekker, J., and Mirny, L.A. (2012). Iterative correction of Hi-C data reveals hallmarks of chromosome organization. Nat Methods *9*, 999-1003.

Lafos, M., Kroll, P., Hohenstatt, M.L., Thorpe, F.L., Clarenz, O., and Schubert, D. (2011). Dynamic regulation of H3K27 trimethylation during Arabidopsis differentiation. PLoS Genet *7*, e1002040.

Moissiard, G., Bischof, S., Husmann, D., Pastor, W.A., Hale, C.J., Yen, L., Stroud, H., Papikian, A., Vashisht, A.A., Wohlschlegel, J.A.*, et al.* (2014). Transcriptional gene silencing by Arabidopsis microrchidia homologues involves the formation of heteromers. Proc Natl Acad Sci U S A *111*, 7474-7479.

Moissiard, G., Cokus, S.J., Cary, J., Feng, S., Billi, A.C., Stroud, H., Husmann, D., Zhan, Y., Lajoie, B.R., McCord, R.P.*, et al.* (2012). MORC family ATPases required for heterochromatin condensation and gene silencing. Science *336*, 1448-1451.

Paige, C.C., and Saunders, M.A. (1982a). Algorithm-583 - Lsqr - Sparse Linear-Equations and Least-Squares Problems. Acm T Math Software *8*, 195-209.

Paige, C.C., and Saunders, M.A. (1982b). Lsqr - an Algorithm for Sparse Linear-Equations and Sparse Least-Squares. Acm T Math Software *8*, 43-71.

Pecinka, A., Schubert, V., Meister, A., Kreth, G., Klatte, M., Lysak, M.A., Fuchs, J., and Schubert, I. (2004). Chromosome territory arrangement and homologous pairing in nuclei of Arabidopsis thaliana are predominantly random except for NOR-bearing chromosomes. Chromosoma *113*, 258-269.

Schubert, I., Fransz, P.F., Fuchs, J., and de Jong, J.H. (2001). Chromosome painting in plants. Methods in cell science : an official journal of the Society for In Vitro Biology *23*, 57-69.

Schubert, V., Kim, Y.M., and Schubert, I. (2008). Arabidopsis sister chromatids often show complete alignment or separation along a 1.2-Mb euchromatic region but no cohesion "hot spots". Chromosoma *117*, 261-266.

Stroud, H., Greenberg, M.V., Feng, S., Bernatavichute, Y.V., and Jacobsen, S.E. (2013). Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. Cell *152*, 352-364.

Ward, P. (2002). FISH probes and labelling techniques. FISH *Beatty B, Squire J (eds).* pp 5–28.