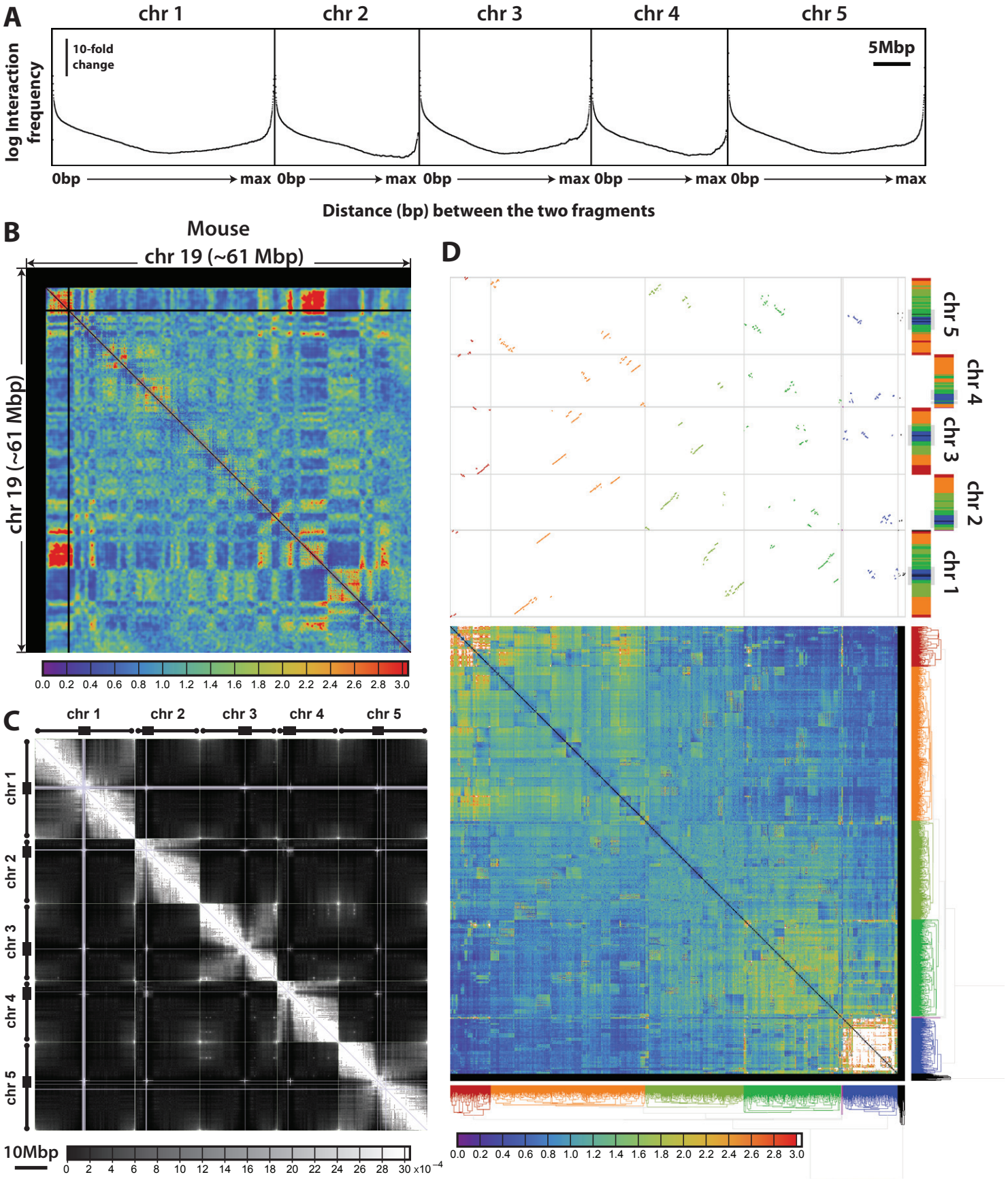
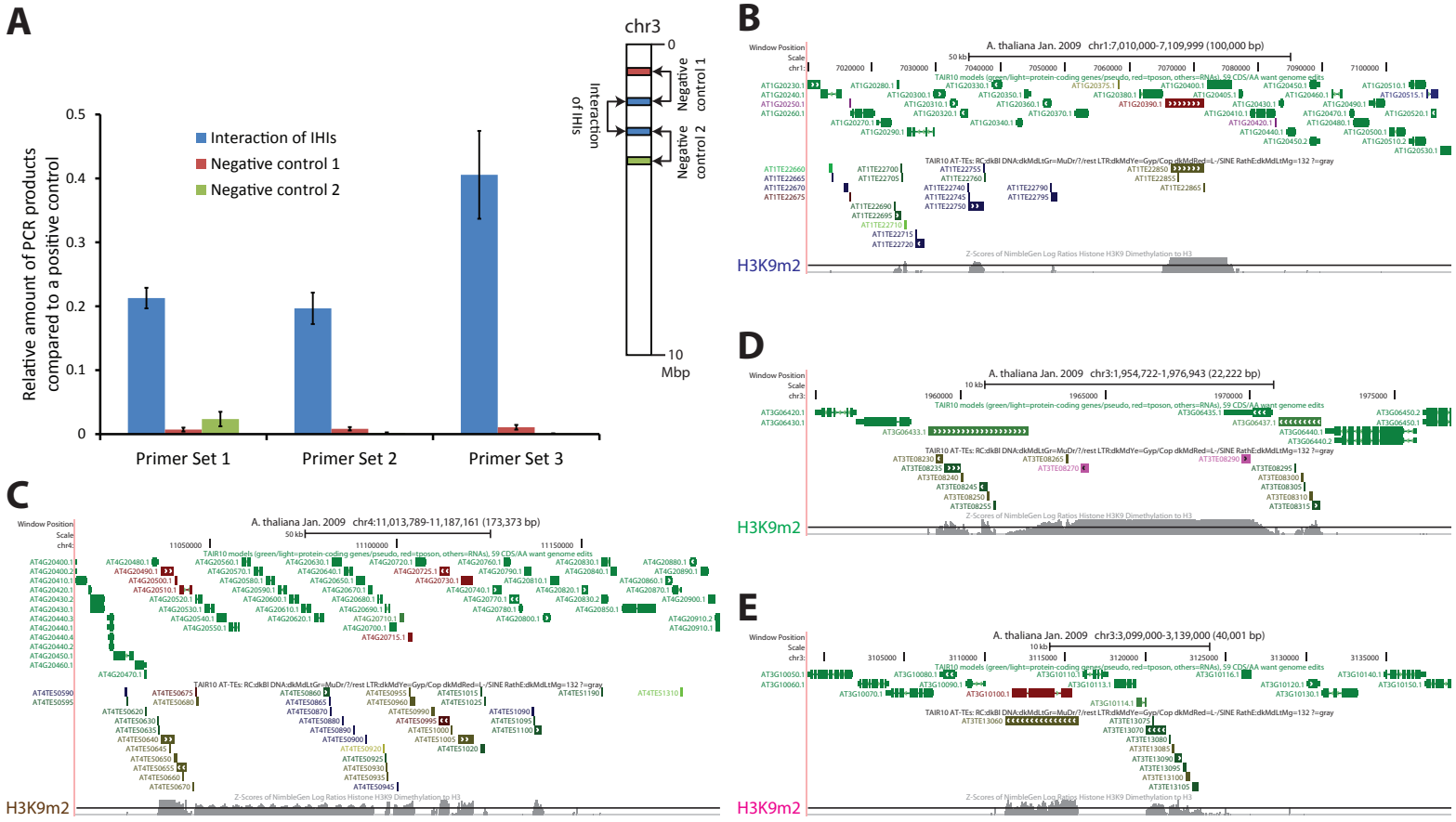

A
chr 1
chr 2
chr 3
chr 4
chr 5
log Interaction frequency
10-fold change
5Mbp
0bp
max
Distance (bp) between the two fragments
B
Mouse
chr 19 (~61 Mbp)
chr 19 (~61 Mbp)
0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4 1.6 1.8 2.0 2.2 2.4 2.6 2.8 3.0
C
chr 1
chr 2
chr 3
chr 4
chr 5
10Mbp
0 2 4 6 8 10 12 14 16 18 20 22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

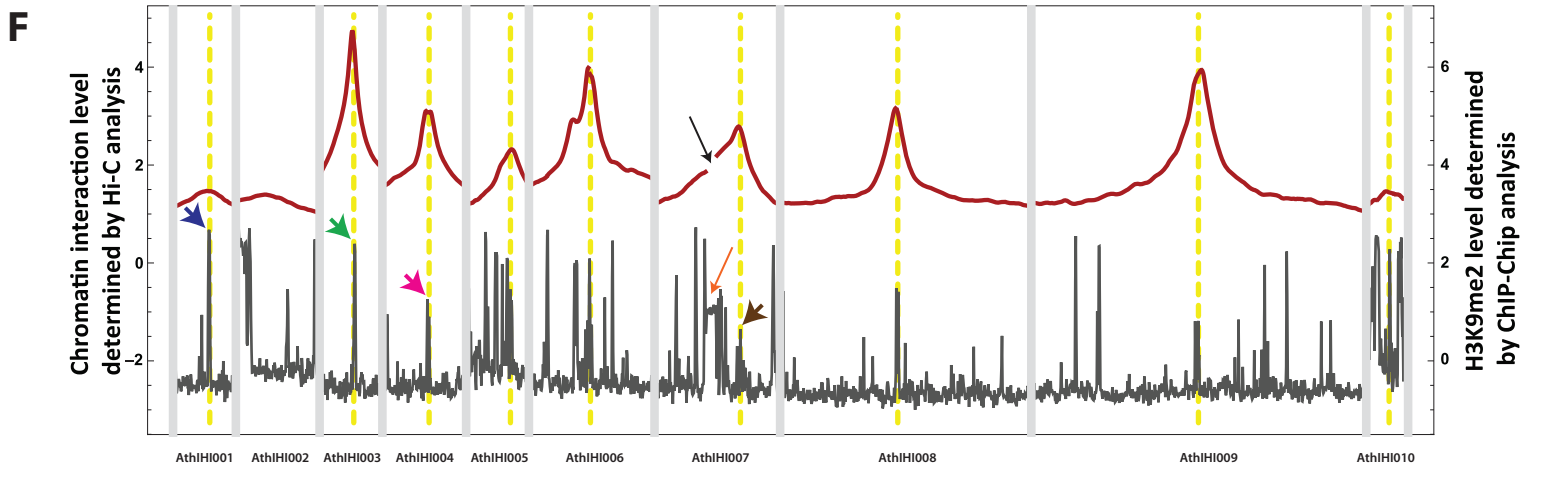

F
Chromatin interaction level determined by Hi-C analysis
H3K9me2 level determined by ChIP-Chip analysis
AthIHI001
AthIHI002
AthIHI003
AthIHI004
AthIHI005
AthIHI006
AthIHI007
AthIHI008
AthIHI009
AthIHI010

Figure S2

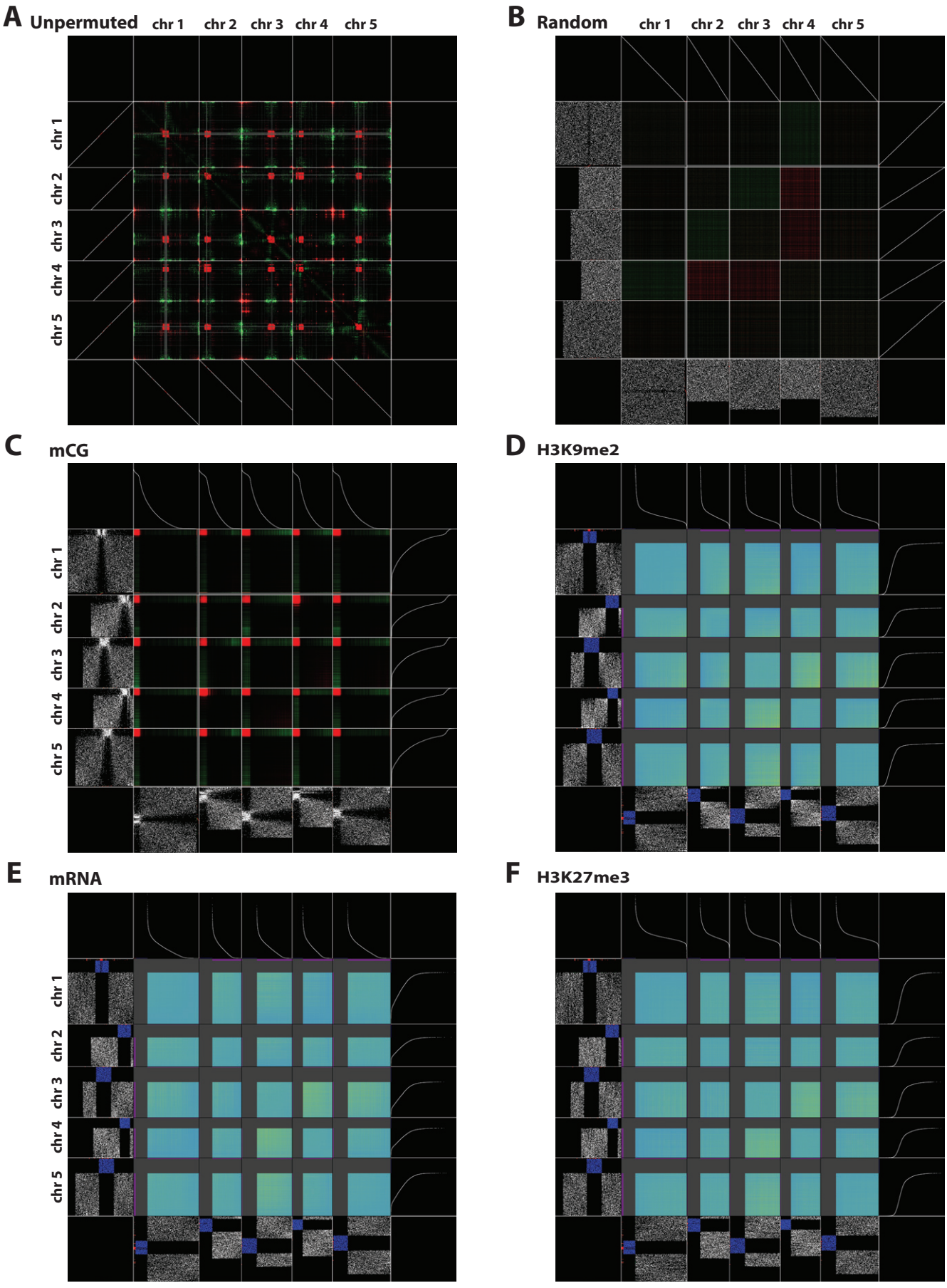

A Unpermuted
chr 1 chr 2 chr 3 chr 4 chr 5
chr 1
chr 2
chr 3
chr 4
chr 5
B Random
chr 1 chr 2 chr 3 chr 4 chr 5
C mCG
chr 1
chr 2
chr 3
chr 4
chr 5
D H3K9me2
E mRNA
chr 1
chr 2
chr 3
chr 4
chr 5
F H3K27me3

Figure S3

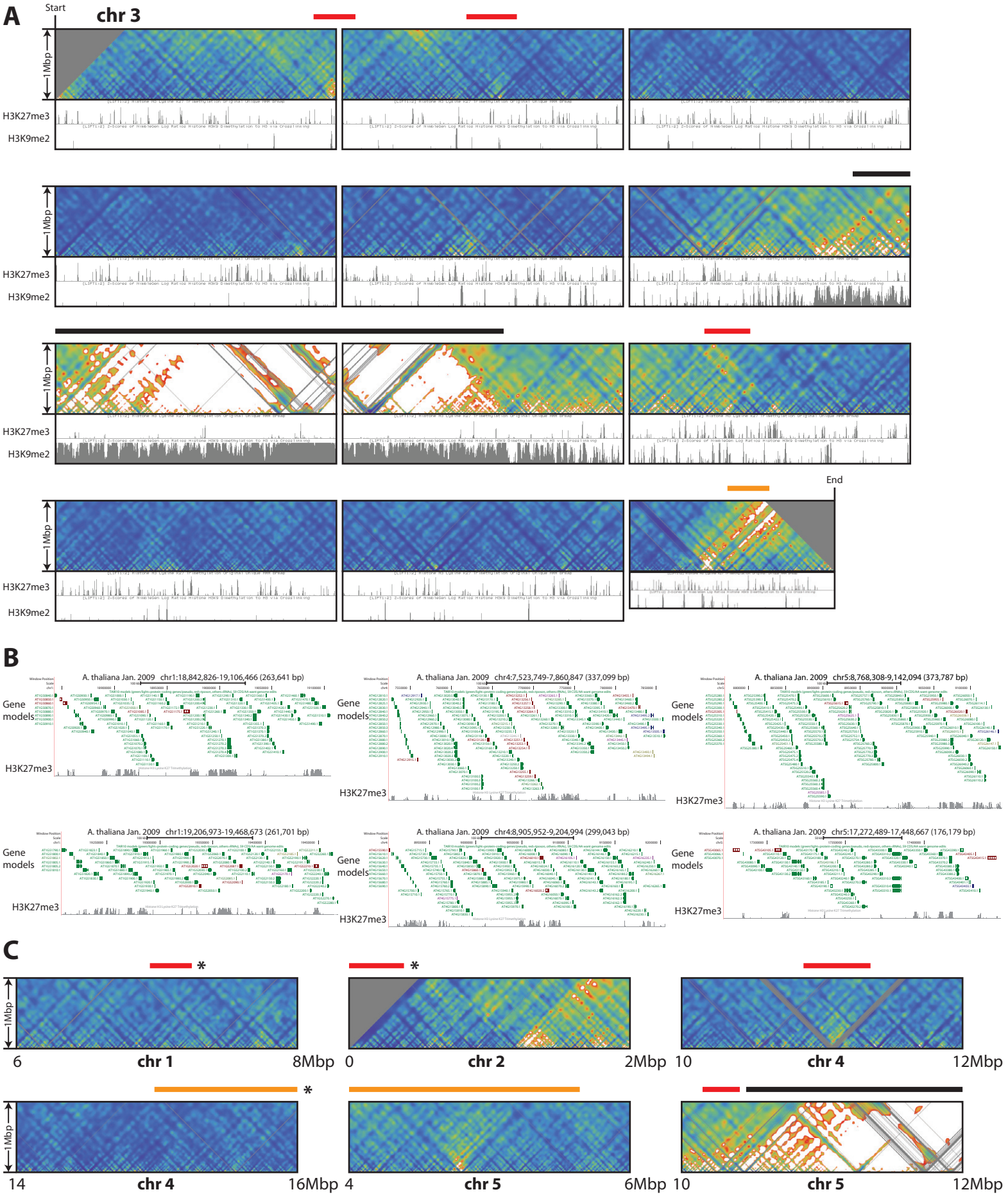

A
Start
chr 3
1Mbp
H3K27me3
H3K9me2
End
B
A. thaliana Jan. 2009 chr1:18,842,826-19,106,466 (263,641 bp)
A. thaliana Jan. 2009 chr4:7,523,749-7,860,847 (337,099 bp)
A. thaliana Jan. 2009 chr5:8,768,308-9,142,094 (373,787 bp)
A. thaliana Jan. 2009 chr1:19,206,973-19,468,673 (261,701 bp)
A. thaliana Jan. 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

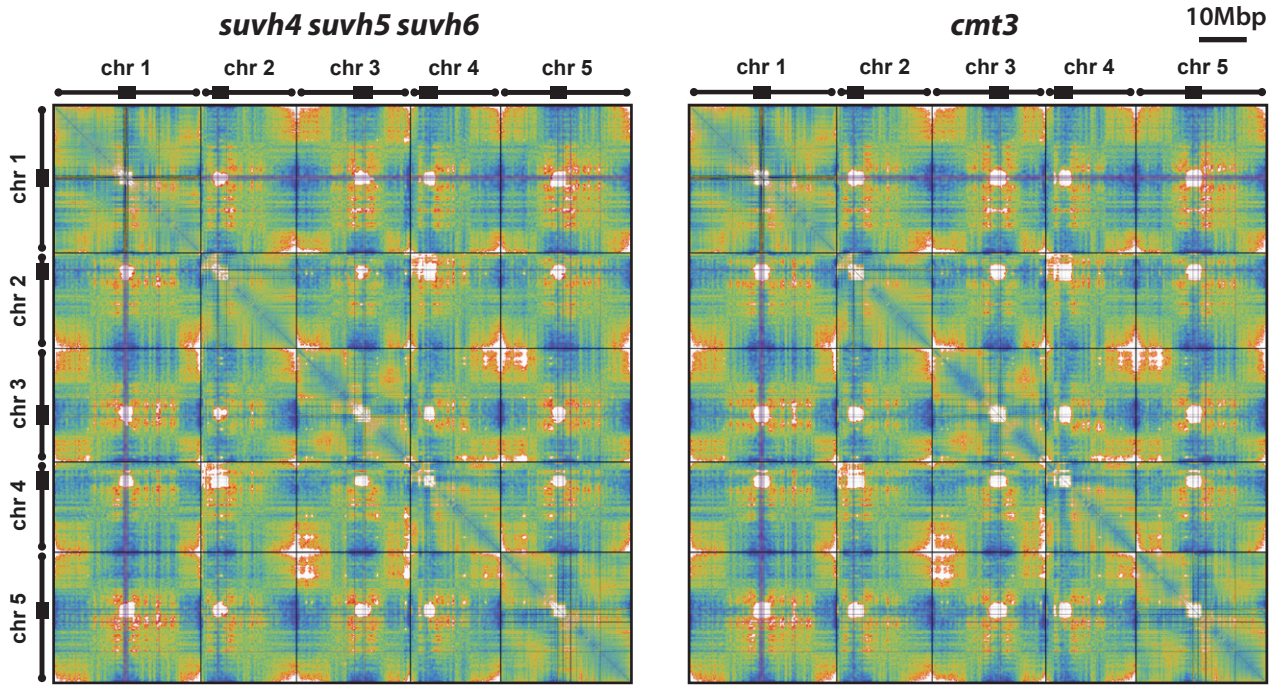


Figure S5

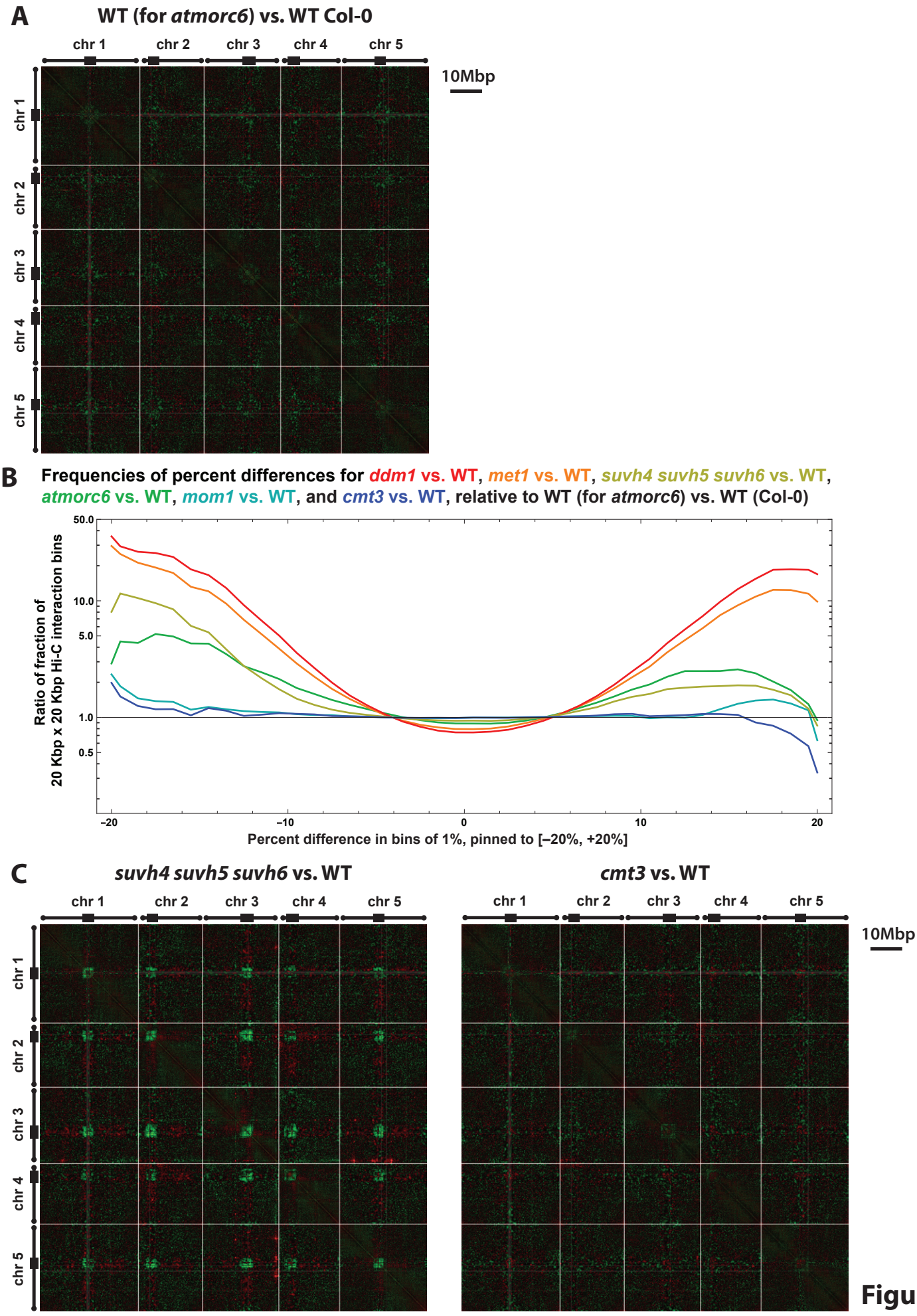

A
WT (for atmorc6) vs. WT Col-0
chr 1
chr 2
chr 3
chr 4
chr 5
10Mbp
B
Frequencies of percent differences for ddm1 vs. WT, met1 vs. WT, suvh4 suvh5 suvh6 vs. WT, atmorc6 vs. WT, mom1 vs. WT, and cmt3 vs. WT, relative to WT (for atmorc6) vs. WT (Col-0)
Ratio of fraction of 20 Kbp x 20 Kbp Hi-C interaction 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

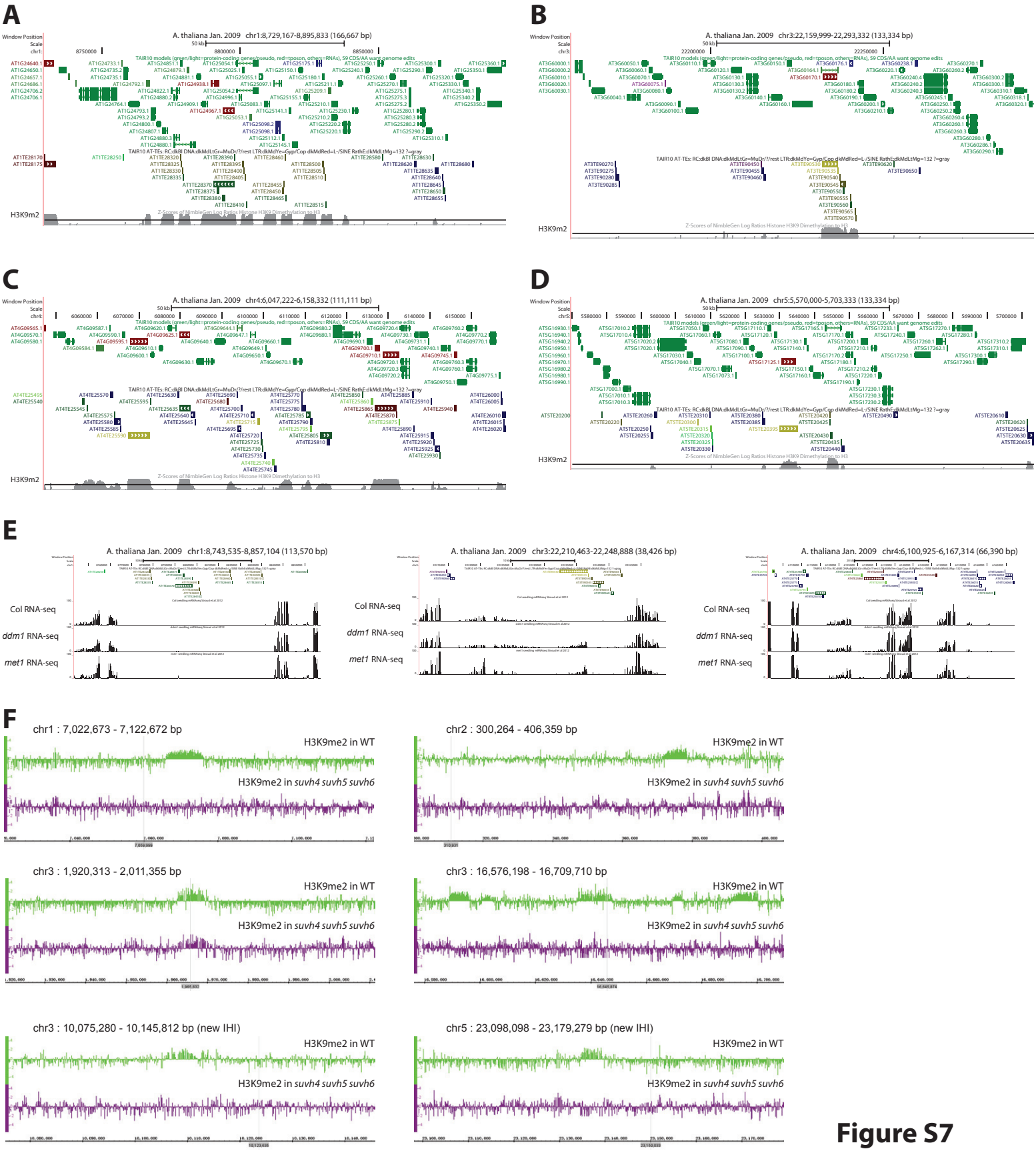

A
A. thaliana Jan. 2009 chr1:8,729,167-8,895,833 (166,667 bp)
H3K9m2
B
A. thaliana Jan. 2009 chr3:22,159,999-22,293,332 (133,334 bp)
H3K9m2
C
A. thaliana Jan. 2009 chr4:6,047,222-6,158,332 (111,111 bp)
H3K9m2
D
A. thaliana Jan. 2009 chr5:5,570,000-5,703,333 (133,334 bp)
H3K9m2
E
A. thaliana Jan. 2009 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 Data 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 Data 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 DATASETS

Data 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.

Data 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.

Data 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.

Data 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.

Data 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.

Data 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.