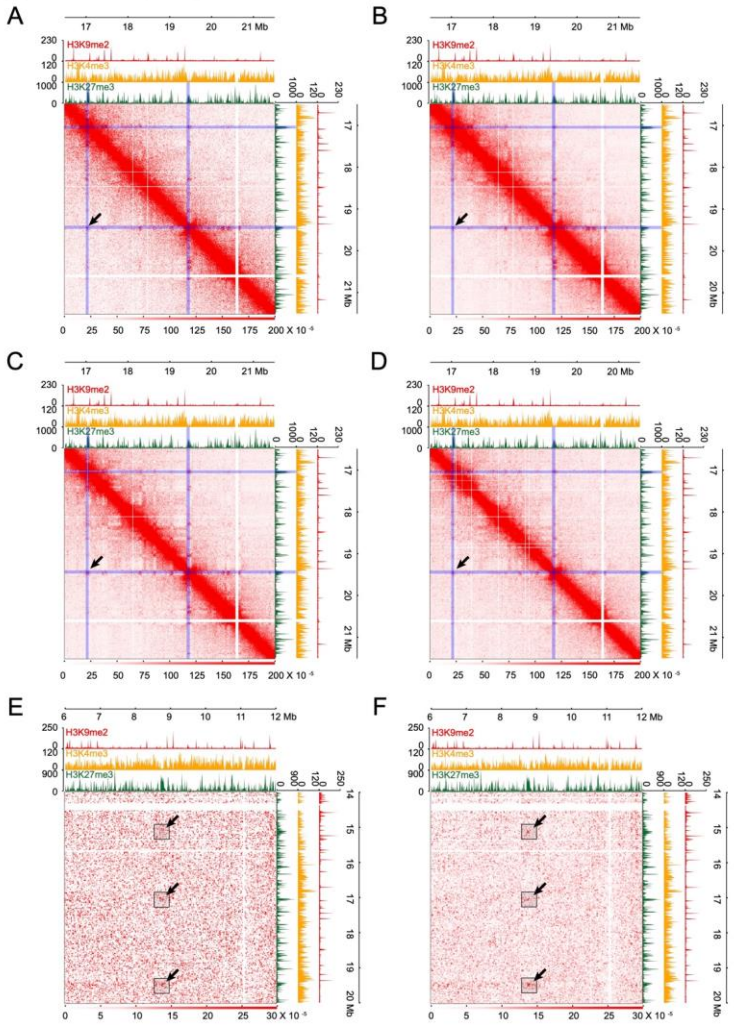
Conserved H3K27me3-associated chromatin looping mediates physical interactions of gene clusters in plants

Linhua Sun[1,2†], Yuxin Cao[2†], Zhu Li[1,3†], Yi Liu[2], Xiaochang Yin[2], Xing Wang Deng[1,2*], Hang He[1,2*], Weiqiang Qian[1,2*]

SUPPORTING INFORMATION

Supplementary Figure 1



Supplementary Figure 1

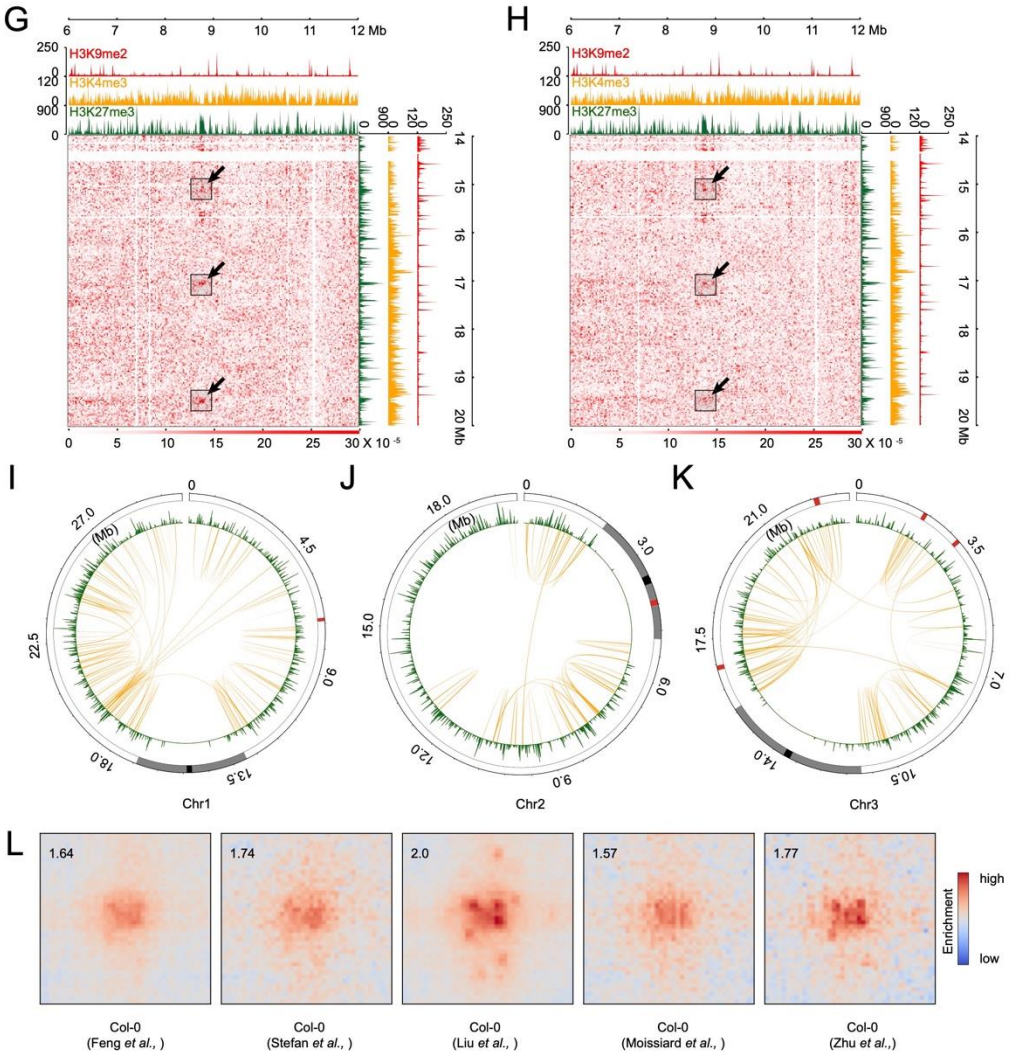


Figure S1. Visible and distinct long-range chromatin loops in two biological replicates and other published Hi-C datasets

(A) Typical *cis* chromatin interactions in a Hi-C map. The black arrow marks the position of a typical chromatin interaction in the Hi-C map. Three histone modifications (H3K9me2, H3K4me3, and H3K27me3) are provided above and to the right of the heatmap. Blue strips mark the relative positions of chromatin interactions to relevant distributions of chromatin modifications. The Hi-C datasets came from the two biological replicates **(A and B)**, aerial parts of Col-0 from Sun *et al*., published Hi-C datasets of Liu *et al*., **(C)** and Feng *et al*. **(D)**. **(E)** Typical *trans* chromatin interactions between chromosomes 4 and 5 in a Hi-C map. Black boxes mark the position of three typical chromatin interactions in the Hi-C map. The Hi-C datasets came from the two biological replicates **(E and F)**, aerial parts of Col-0 from Sun *et al*., published Hi-C datasets of Liu *et al*., **(G)** and Feng *et al*. **(H)** The black arrow marks the positions of typical chromatin interactions in the Hi-C map. Blue strips mark the relative positions of chromatin interactions to relevant distributions of chromatin modifications. **(I-K)** Circos plot shows the long-range chromatin loop (orange links) distribution on *Arabidopsis* chromosomes 1, 2, and 3 identified from Aerial (Control) Hi-C. The green bars show the distribution of associated H3K27me3. *Arabidopsis* cytoband consists of centromeres (black), pericentromeres (gray), KEEs (red), and chromosome arms (white). **(L)** Plots of APA of loops identified from a series of published Hi-C datasets using 10-kb resolution Hi-C with 200-kb padding. The enrichment values are relative to a random background (methods). The central pixel's value is displayed in the top left corner of each heatmap.

Supplementary Figure 2

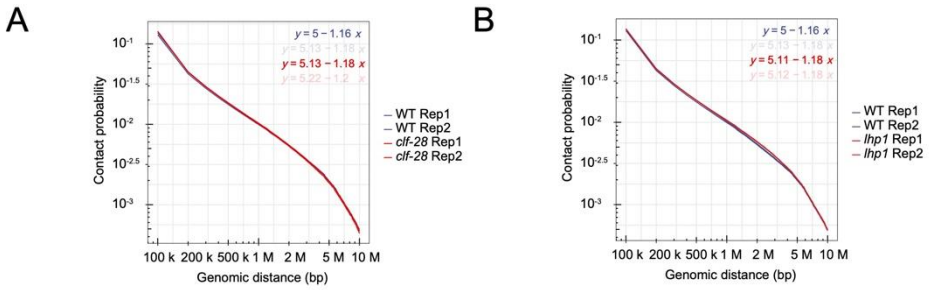


Figure S2. Good reproducibility of *clf-28* and *lhp1* Hi-C datasets

(A-B) Averaged scaling plot of contact probability against increasing genomic distance for all *Arabidopsis* chromosomes in *clf-28* **(A)** and *lhp* **(B)** vs. WT. The genomic bin size is 100 kb.

Supplementary Figure 3

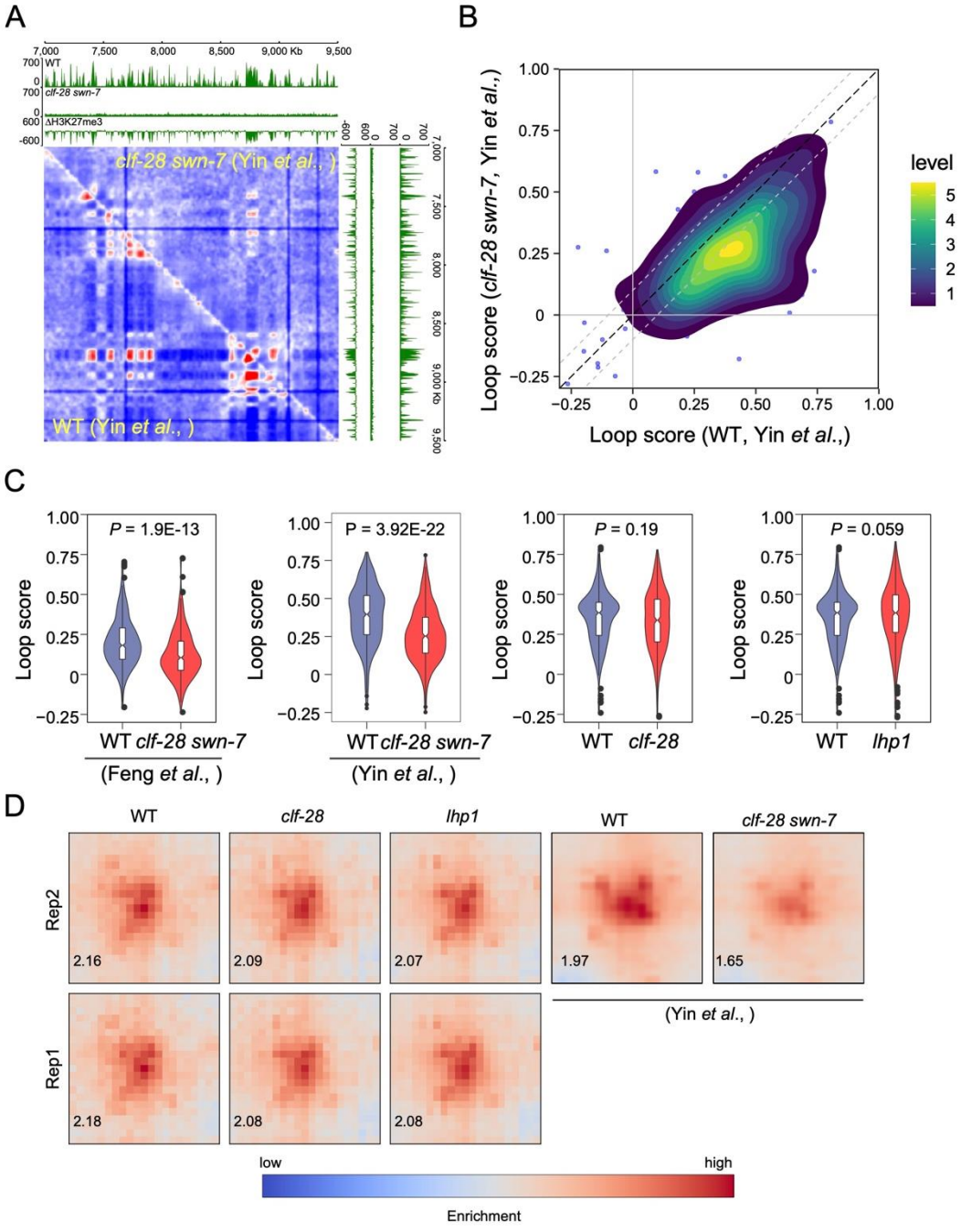


Figure S3. Chromatin interaction patterns of long-range chromatin loops in *clf-28 swn7*, *clf-28*, and *lhp1* Hi-C datasets relative to WT

(A) Asymmetric heatmaps show chromatin interactions mapped in WT and *clf-28 swn-7* from Yin *et al*., . The heatmap shows the observed/expected values of the contact map at 10-kb resolution, and the color bar indicates the range of the values. H3K27me3 levels in the indicated genotypes are provided next to the heatmap body. The Hi-C datasets of WT and *clf-28 swn-7* **(A)** were collected and reanalyzed based on Yin *et al*.,. The H3K27me3 ChIP-Seq datasets (**A**) were collected from Zhou *et al*.,. **(B)** Scatter plots of loop scores in WT vs. *clf-28 swn-7*. **(C)** Violin plots showing the distribution of loop scores of long-range chromatin loops in *clf-28 swn-7* vs. its associated Col-0 from different sources, *clf-28* vs. WT, and *lhp1* vs. WT. In each violin plot, centerlines show the medians; boxes show the 25th and 75th percentiles. A two-sided Mann–Whitney U test was used for each comparison. *P* values are indicated. **(D)** Plots of APA of loops identified from WT, *clf-28*, and *lhp1* Hi-C datasets with two biological replicates using 10-kb resolution Hi-C with padding 100 kb. Plots of APA of loops identified from WT and *clf-28 swn-7* from Yin *et al*., are also appended. The enrichment values are relative to a random background (methods). The central pixel's value is displayed in the top left corner of each heatmap.

Supplementary Figure 4

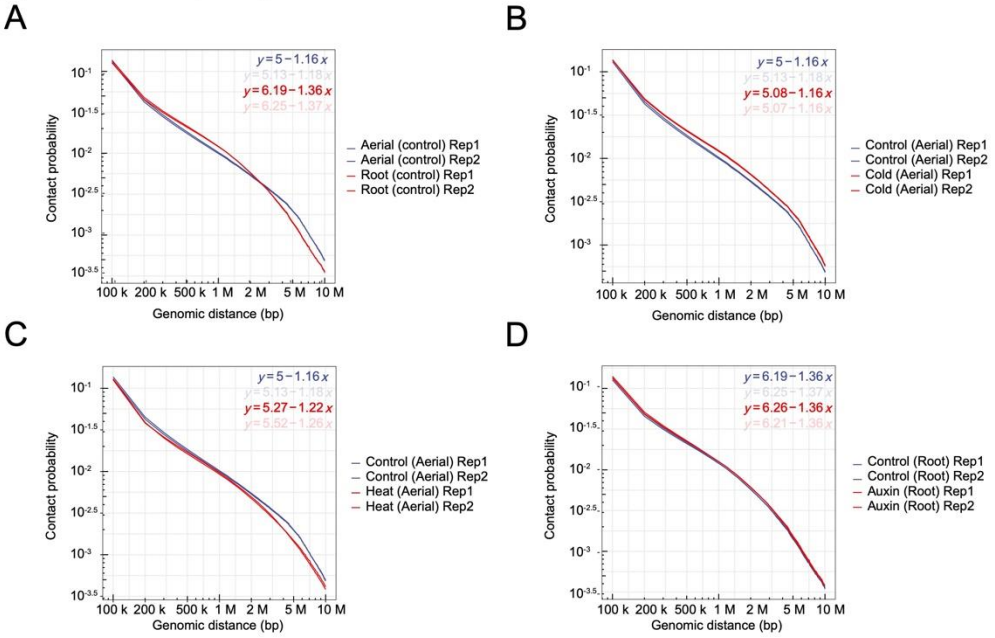
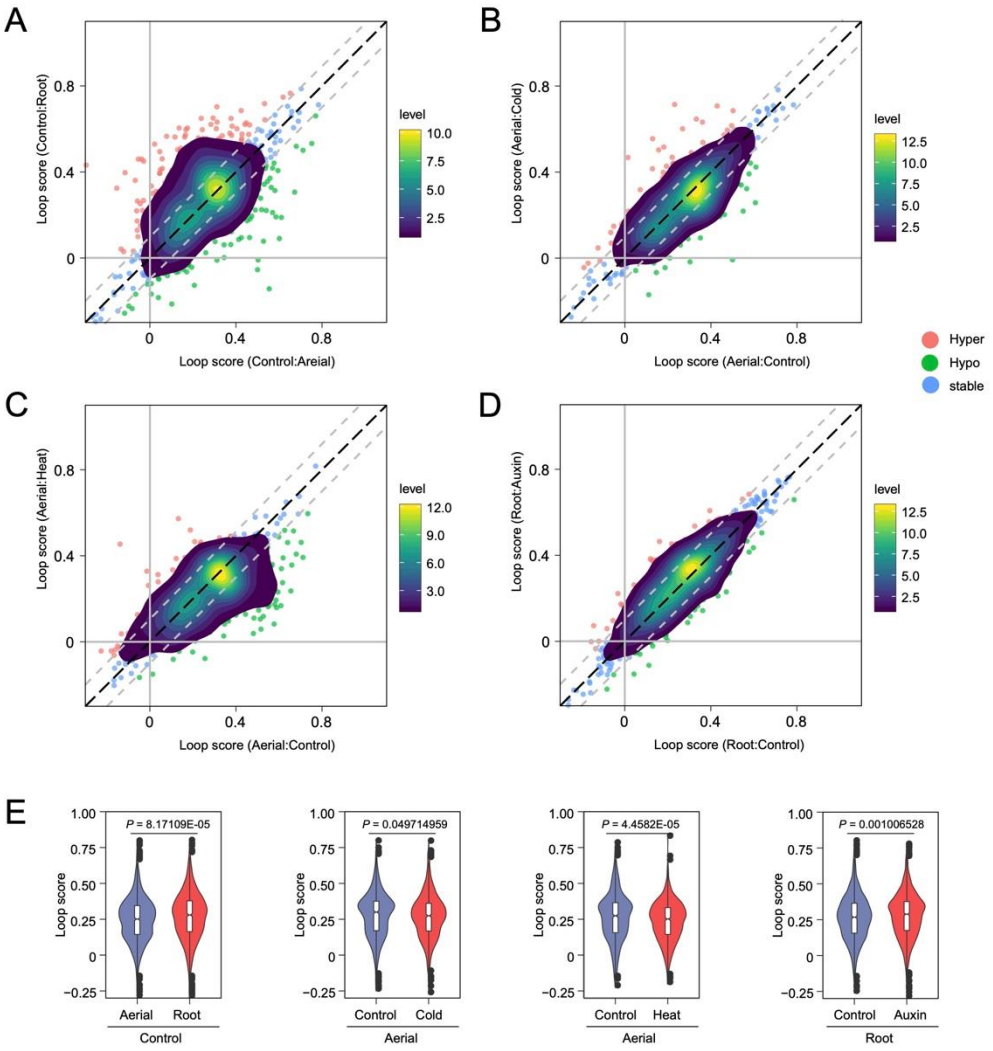


Figure S4. Good reproducibility of Root (Control), Cold (Aerial), Heat (Aerial), and Auxin (Root) Hi-C datasets

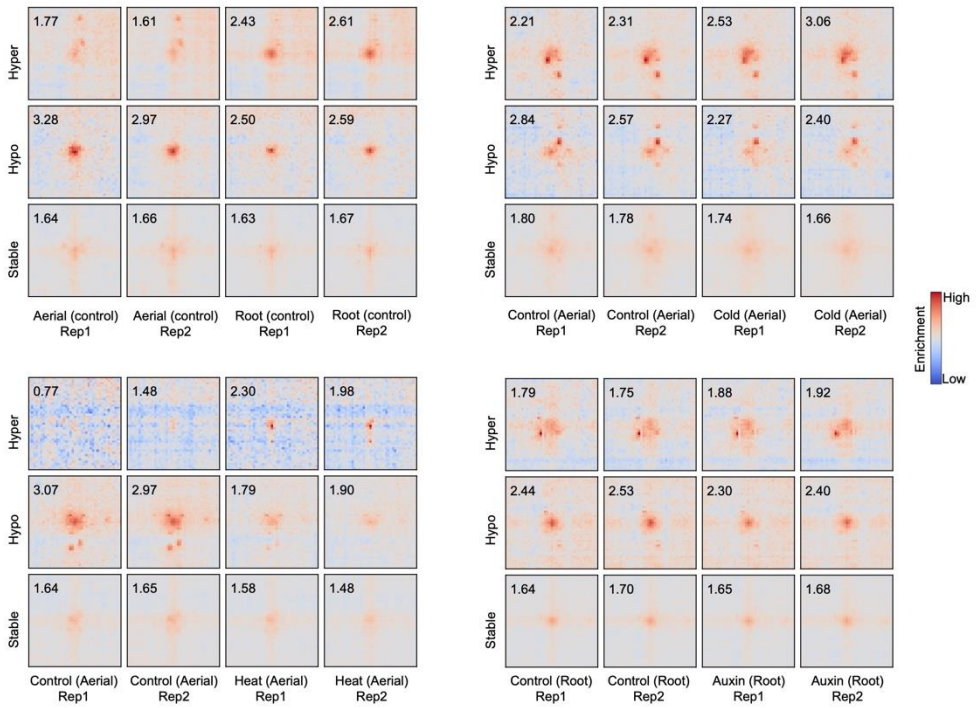
(A-D) Averaged scaling plot of contact probability against increasing genomic distance for all *Arabidopsis* chromosomes in Control: Root vs. Aerial **(A)**, Aerial: Cold vs. Control **(B)**, Aerial: Heat vs. Control **(C)**, and Root: Auxin vs. Root (**D**). The genomic bin size is 100 kb.

Supplementary Figure 5



Supplementary Figure 5

F



Supplementary Figure 5

G

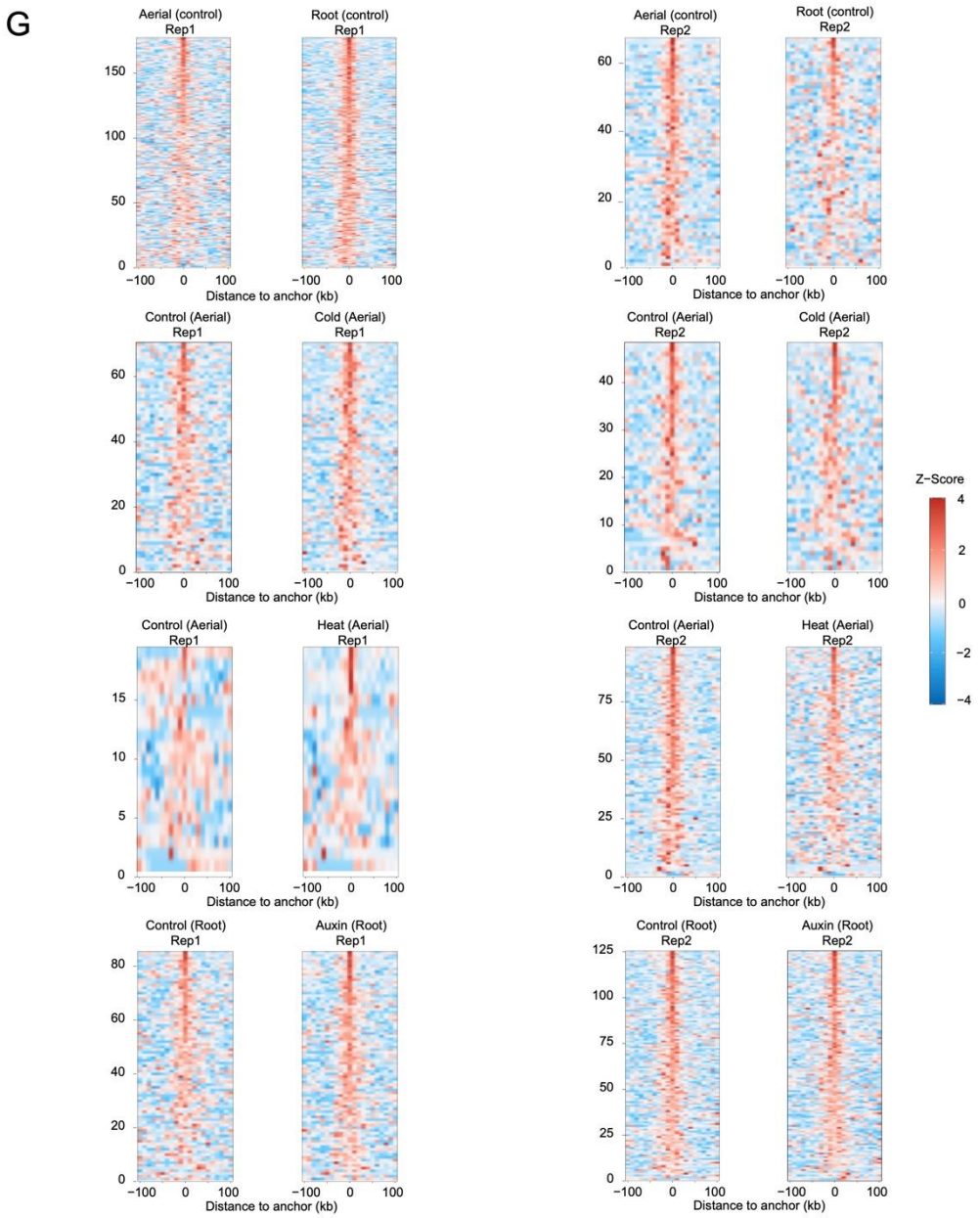


Figure S5. Dynamics of long-range chromatin loops in Control: Root vs. Aerial, Aerial: Cold vs. Control, Aerial: Heat vs. Control, and Root: Auxin vs. Root **(A-D)** Scatter plots of loop scores in Control: Root vs. Aerial **(A)**, Aerial: Cold vs. Control **(B)**, Aerial: Heat vs. Control **(C)**, and Root: Auxin vs. Root **(D)** using loops identified from their merged Hi-C datasets (methods). **(E)** Violin plots showing the distribution of loop scores of long-range chromatin loops in Control: Root vs. Aerial, Aerial: Cold vs. Control, Aerial: Heat vs. Control, and Root: Auxin vs. Root. In each violin plot, centerlines show the medians; boxes show the 25th and 75th percentiles. A two-sided Mann–Whitney U test was used for each comparison. *P* values are indicated. **(F)** Plots of APA of loops identified from Control: Root vs. Aerial, Aerial: Cold vs. Control, Aerial: Heat vs. Control, and Root: Auxin vs. Root Hi-C datasets with two biological replicates using 10-kb resolution Hi-C with padding of 200 kb. The enrichment values are relative to a random background (methods). The central pixel's value is displayed in the top left corner of each heatmap. **(G)** Z score heatmaps of chromatin interactions in hyper- and hypo-loops identified from comparisons of Control: Root vs. Aerial, Aerial: Cold vs. Control, Aerial: Heat vs. Control, and Root: Auxin vs. Root Hi-C datasets. The aggregate tornado plot extracts the scores around and in each loop with padding of 100 kb. Each row serves as a loop in the heatmap.

Supplementary Figure 6

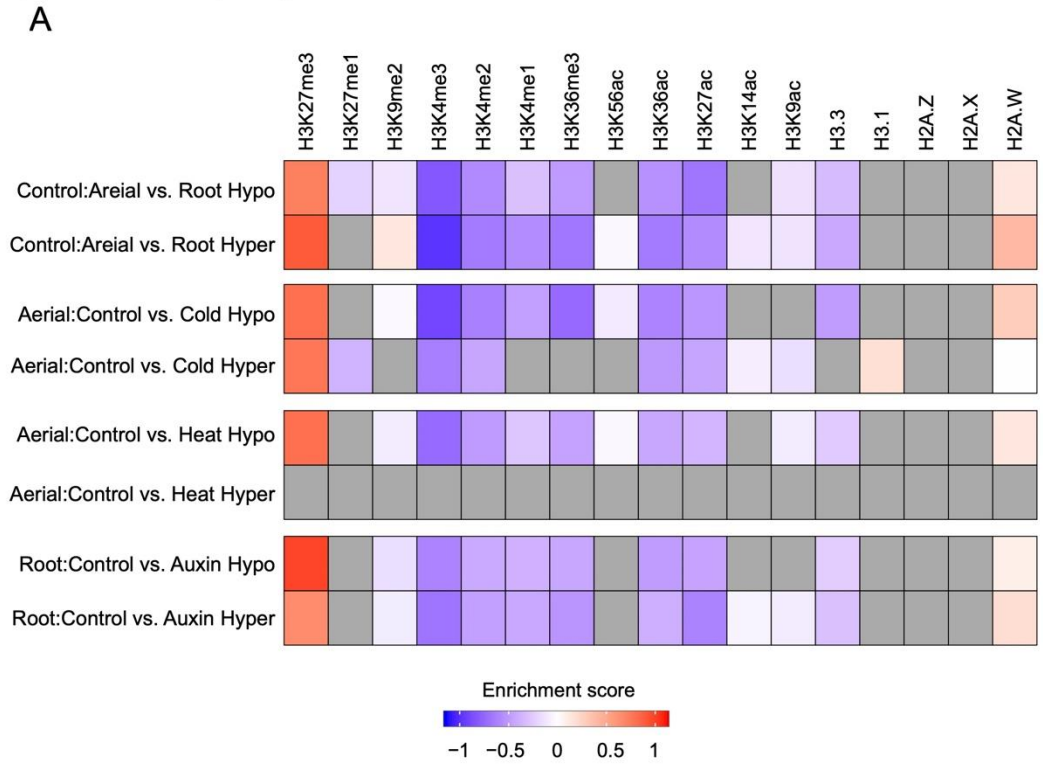


Figure S6. Epigenomic feature enrichment of anchors of hyper- and hypo-loops in Control: Root vs. Aerial, Aerial: Cold vs. Control, Aerial: Heat vs. Control, and Root: Auxin vs. Root

(A) The enrichment (red) and depletion (blue) patterns of epigenetic features in hyper- and hypo-loop anchors compared with their shuffled regions in comparisons of Control: Root vs. Aerial, Aerial: Cold vs. Control, Aerial: Heat vs. Control, and Root: Auxin vs. Root Hi-C datasets. Gray indicates that the epigenetic feature is neither enriched nor depleted (two-sided Mann–Whitney U test, $p > 0.05$).

Supplementary Figure 7

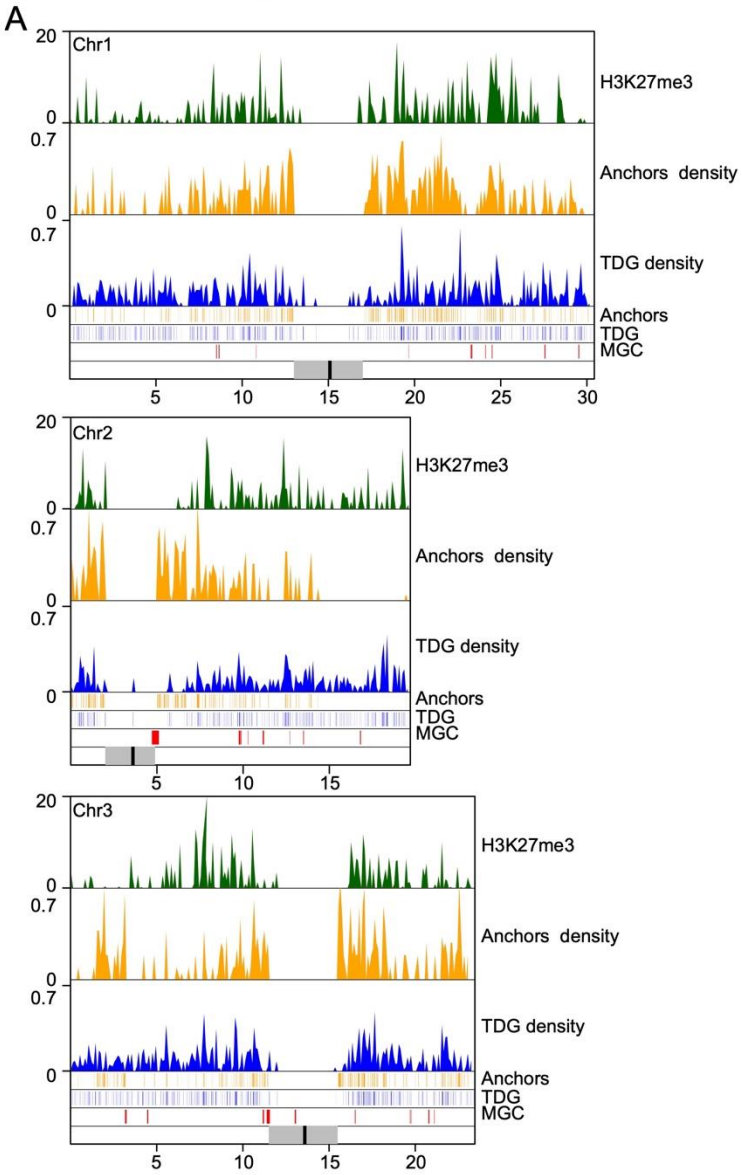
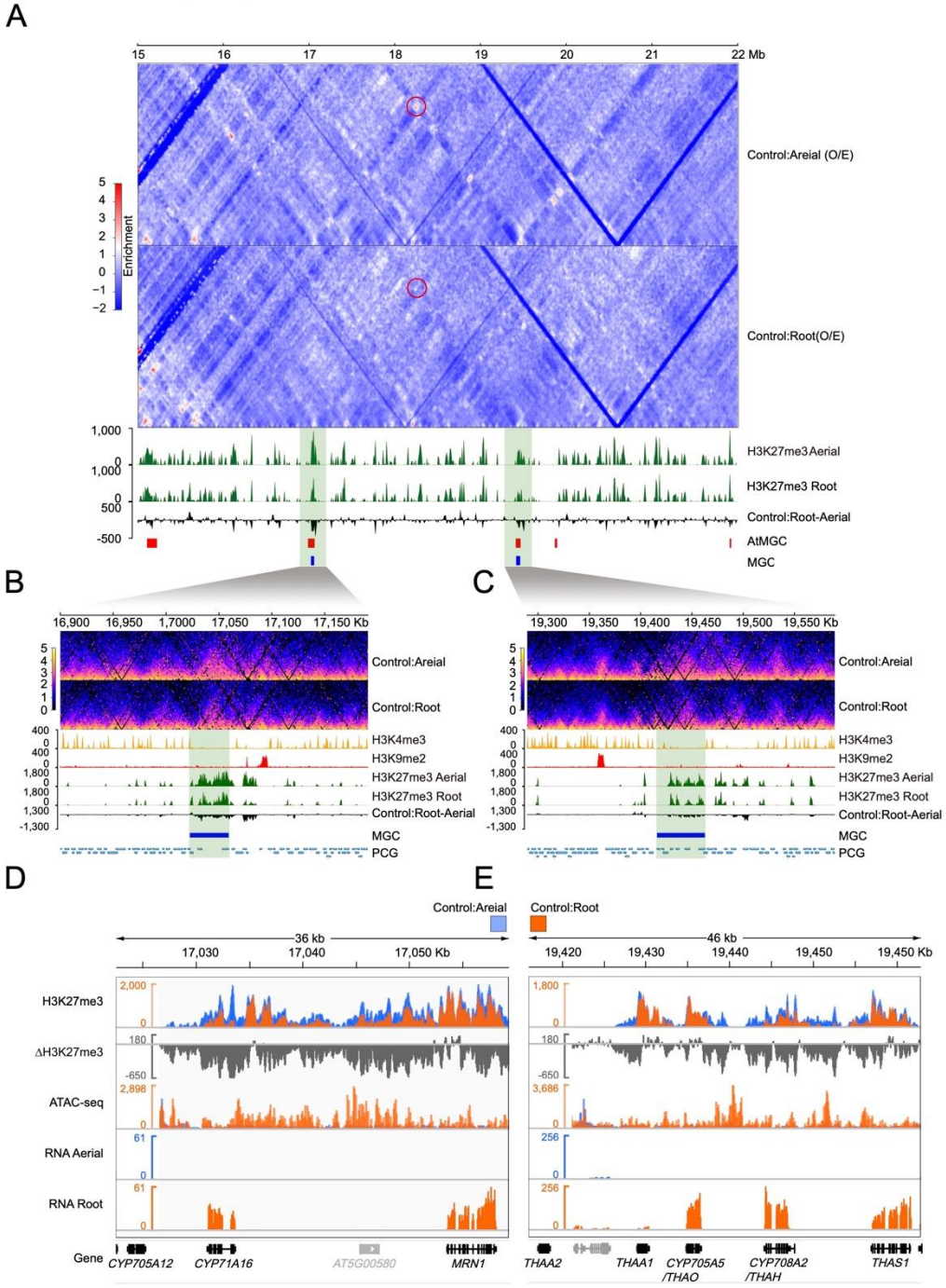


Figure S7. Chromosome-wide distribution of two types of clusters and loop anchors

(A) Chromosome-wide distributions of H3K27me3 modifications, loop anchors, TDGs, and their density on Chr1, Chr2, and Chr3. *Arabidopsis* cytoband consists of centromeres (black), pericentromeres (gray), *KNOT-engaged* elements (KEEs, red), and chromosome arms (white).

Supplementary Figure 8



Supplementary Figure 8

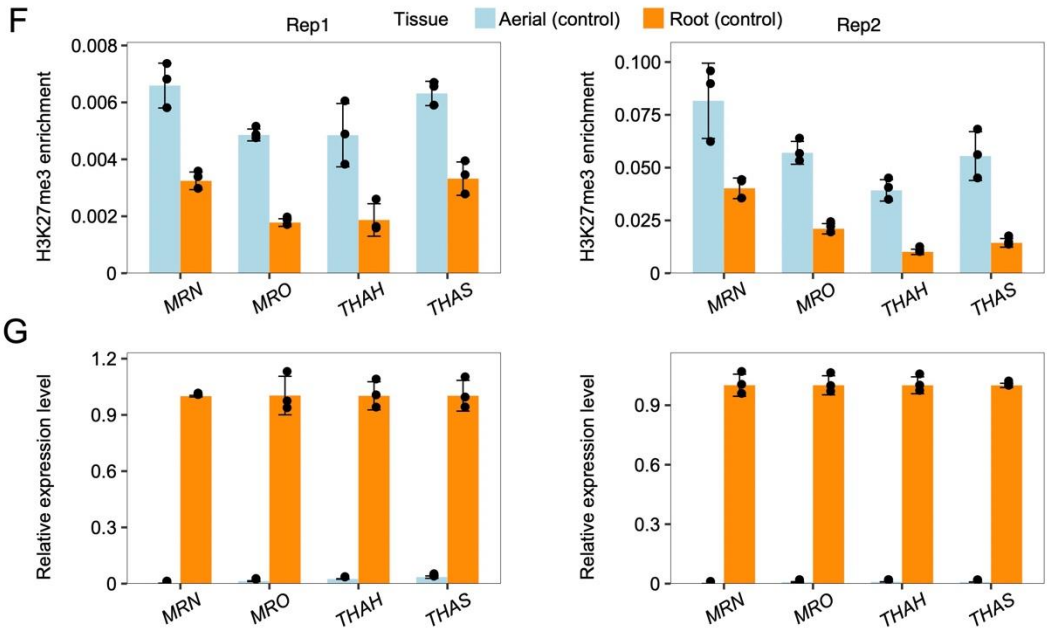


Figure S8. Dynamic regulation of chromatin interactions, H3K27me3 levels, and associated gene expression levels of a typical long-range chromatin loop between two MGCs in Aerial (Control) vs. Root (Control)

(A) The top panels are heatmaps showing the observed/expected values of the contact map at 10-kb resolution in Aerial (Control) vs. Root (Control). Below are distributions of H3K27me3 levels in Aerial (Control), Root (Control), and Root (Control) minus Aerial (Control). MGC locations in red are displayed under the ChIP-seq tracks. More precise locations of MGCs of the marneral (Chr5:17,022,503-17,059,245) and thalianol (Chr5:19,416,021-19,462,706) gene clusters from Huang *et al*. are in blue. Highlighted areas in green indicate the regions to be enlarged in **(B)** and **(C)**. **(B)** Enlarged region from **(A)** contains a marneral gene cluster on Chr5:16,899,999-17,190,874. The top panels are heatmaps showing the raw KR normalized value of the contact map at 1-kb resolution in Aerial (Control) and Root (Control). H3K9me2, H3K4me3, and H3K27me3 (Aerial (Control), Root (Control), and Root (Control) minus Aerial (Control)) are displayed under the Hi-C heatmaps. The highlighted area in green indicates the marneral gene cluster (Chr5:17,022,503-17,059,245). *Arabidopsis* protein-coding genes are listed below the tracks. **(C)** Enlarged region from **(A)** contains a thalianol gene cluster on Chr5:19,289,363-19,589,363. The highlighted area in green indicates the thalianol gene cluster (Chr5:19,416,021-19,462,706). Other information is similar to **(B)**. **(D)** Enlarged

region of the marneral gene cluster on Chr5:17,022,503-17,059,245. H3K27me3 in Root (Control) and Aerial (Control) are overlaid in orange and blue. The H3K27me3 level of Root (Control) minus Aerial (Control) is displayed in the next track. Chromatin accessibility in Root (Control) and Aerial (Control) measured by ATAC-Seq is overlaid in orange and blue in the next track. RNA-Seq data of Root (Control) and Aerial (Control) are attached in the following two tracks. *Arabidopsis* gene annotations are listed at the bottom. **(E)** Enlarged region of the thalianol gene cluster on Chr5:19,416,021-19,462,706. **(F)** Barplot shows the H3K27me3 ChIP–qPCR results of chosen genes in Root (Control) and Aerial (Control). Two biological replicates were performed. **(G)** Barplot shows relative gene expression levels determined by RT–qPCR for chosen genes in Aerial (Control) and Root (Control). Two biological replicates were performed. Error bars show mean ± s.d. Each point shows the value from each technical replicate.

Supplementary Figure 9

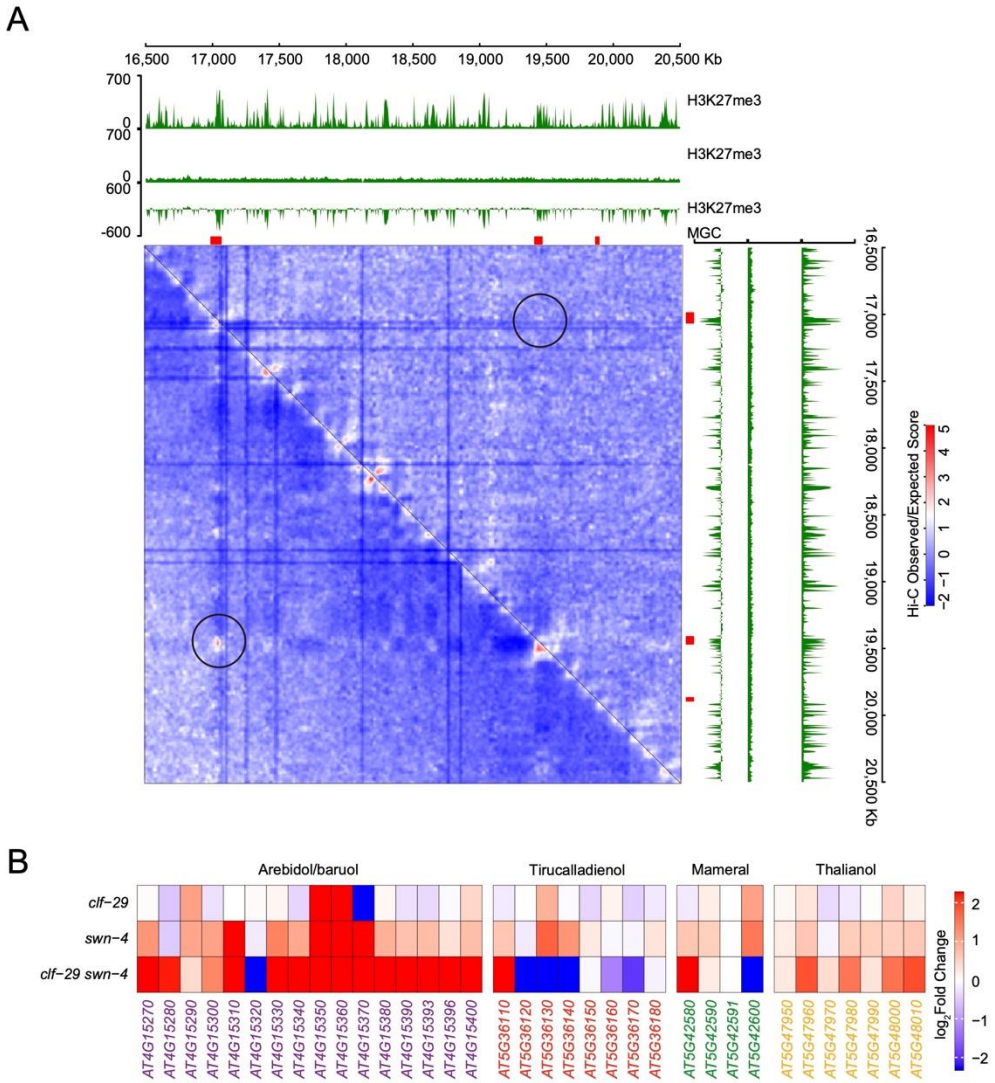


Figure S9. Dynamic regulation of chromatin interactions, H3K27me3 levels, and associated gene expression levels of a typical long-range chromatin loop between two MGCs in clf-28 swn-7 and its associated Col-0

(A) Asymmetric heatmap shows chromatin interactions mapped in Col-0 (left lower corner) and *clf-28 swn-7* (upper right corner). The heatmap shows the observed/expected values of the contact map at 10-kb resolution. H3K27me3 levels in Col-0 and *clf-28 swn-7* are provided next to the heatmap body. The black circles indicate the position of the long-range chromatin loop between the two MGCs. **(B)** Heatmap of log_2 fold change of gene expression level in *clf-29*, *swn-4*, and *clf-29 swn-4* vs. Col-0.

Supplementary Figure 10

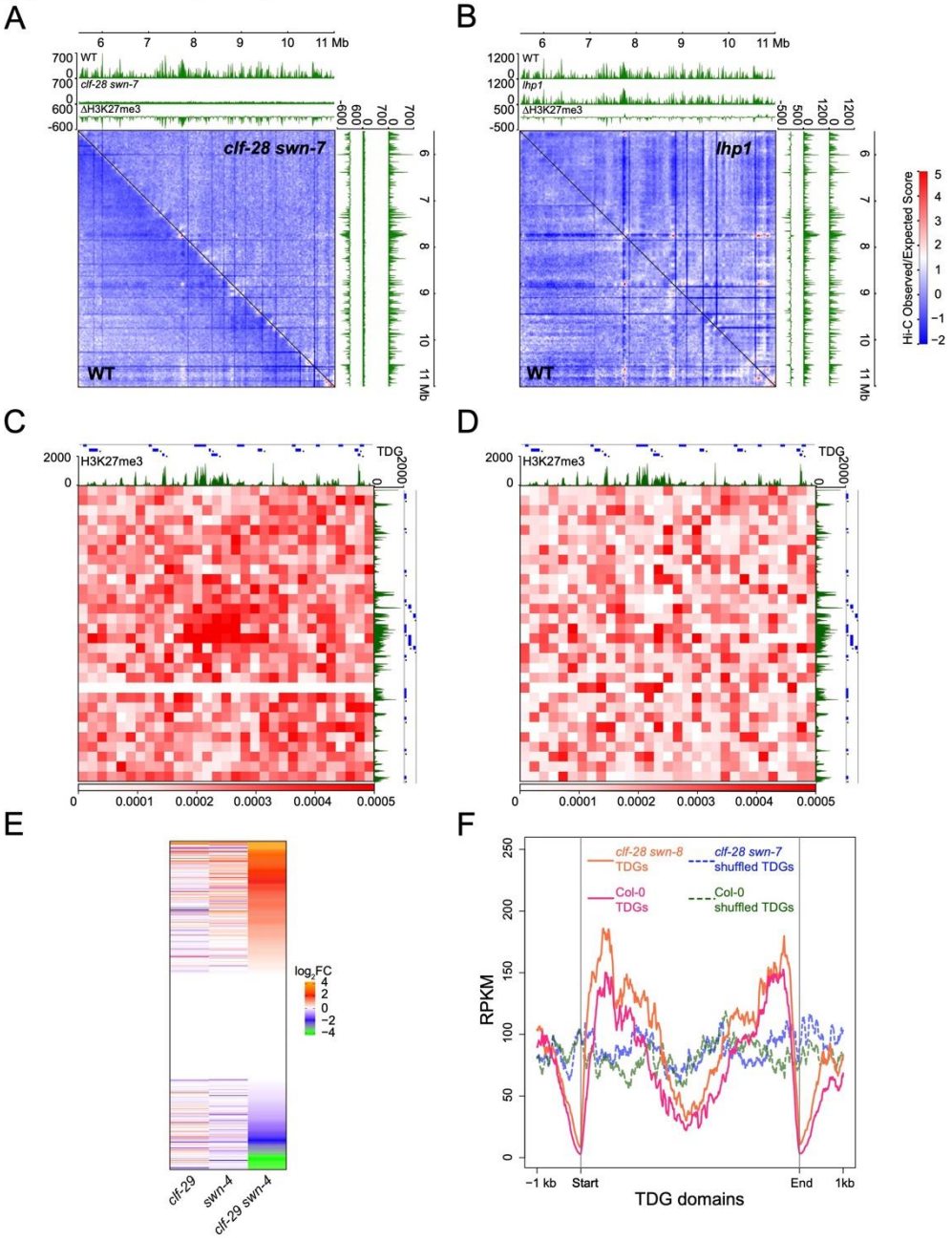


Figure S10. Chromatin interactions and gene expression changes in PcG mutants

(A) Asymmetric heatmap shows chromatin interactions mapped in Col-0 (left lower corner) and *clf-28 swn-7* (upper right corner). The heatmap shows the observed/expected values of the contact map at 10-kb resolution (also for **(B)**). The color bar indicates the range of the values. H3K27me3 levels in Col-0 and *clf-28 swn-7* are provided next to the heatmap body. **(B)** Asymmetric heatmap shows chromatin interactions mapped in aerial parts of Col-0 (left lower corner) and *lhp1* (upper right corner). H3K27me3 levels in Col-0 and *lhp1* are provided next to the heatmap body. **(C)** Far-*cis* symmetrical heatmap for WT using raw normalized Hi-C contact intensity for Col-0 at 20-kb resolution. H3K27me3 from Aerial (Control) are provided on the left and at the top of the heatmap body. MCG and TDG tracks are further provided. **(D)** Similar to (C), the combined heatmap using Hi-C data from *clf-28 swn-7*. **(E)** Heatmap of $\log_2$ fold change of gene expression level in *clf-29*, *swn-4*, and *clf-29 swn-4* vs. Col-0 using all Aerial (Control) loop-targeted PCGs. **(F)** Meta-gene plot for RNA-Seq normalized read coverage for TDGs and shuffled TDGs in Col-0 and *clf-28 swn-7.*

Supplementary Figure 11

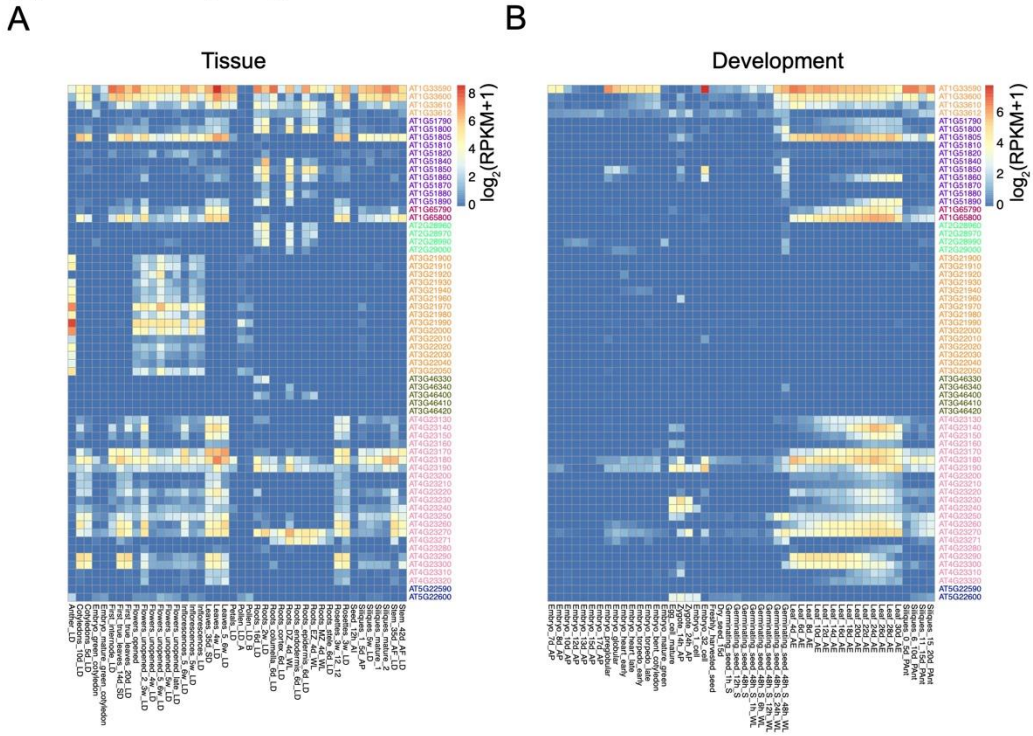


Figure S11. Gene expression changes in tissue- and development-related RNA-Seq datasets

Heatmap showing gene expression levels in $\log_2$(RPKM+1) using tissue **(A)** and developmental **(B)** RNA-Seq datasets for a series of typical selected TDG-associated genes.

Supplementary Figure 12

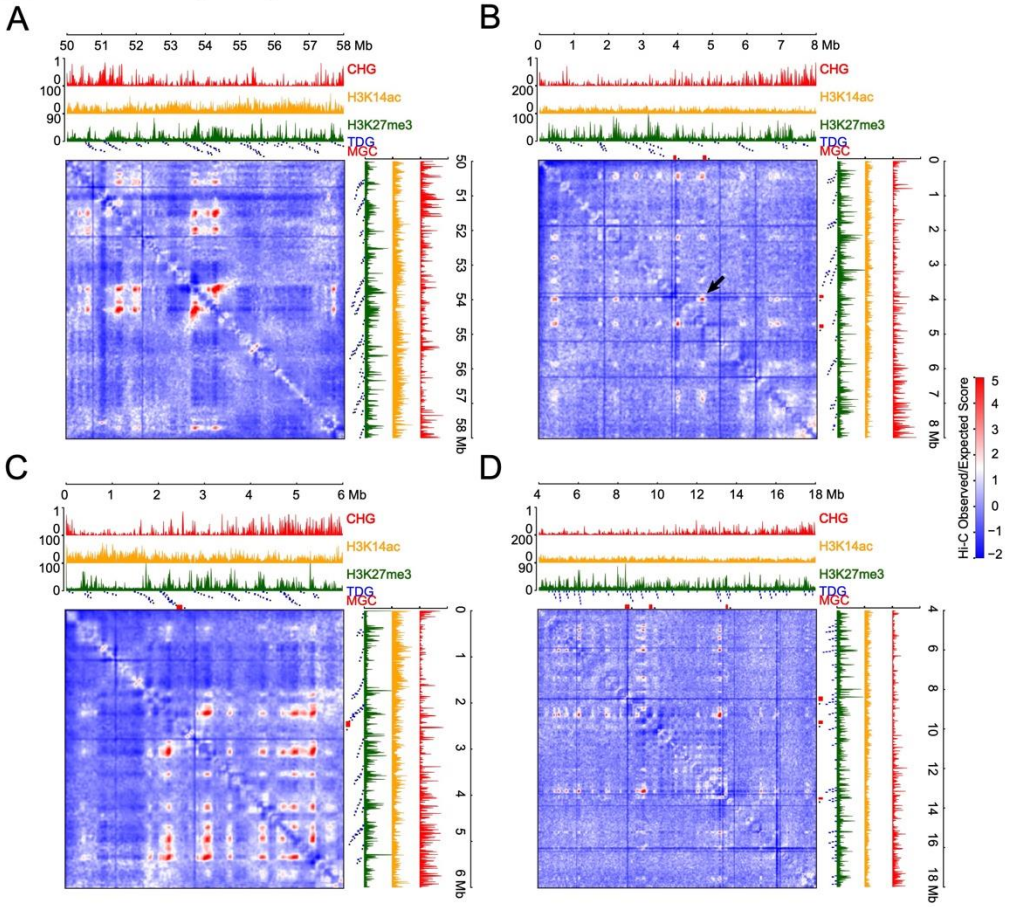


Figure S12. Other typical loci in the H3K27me3-marked chromatin organization and gene clusters in soybeans

(A) Chromatin interaction map and distributions of three typical modifications (CHG, H3K14ac, and H3K27me3), TDGs, and MGCs on *Gm* chromosome 18 (GM18:50,000,000-58,000,000). The heatmap shows the observed/expected values of the contact map at 20-kb resolution. The color bar indicates the range of the values. The redder the color, the stronger the interaction. The bluer the color, the weaker the interaction. TDGs (blue) and MGCs (red) are displayed around the heatmap body. **(B)** Typical loci in GM04:0-8,000,000. The black arrow marks the position of a typical chromatin interaction between two MGCs in the Hi-C map. Left (GM04:3,882,297-3,943,357) and right (GM04:4,733,370-4,812,273). **(C)** Typical loci in GM09:0-6,000,000. **(D)** Typical loci in GM08:4,000,000-18,000,000. All WGBS, ChIP-Seq, and Hi-C data were collected from the same source, Wang, L. *et al*.

Table S1. Long-range cis chromatin loops and anchors identified from Aerial (Control) Hi-C with strict filtering standards.

Table S2. Hi-C quality control for all new generated Hi-C samples.

Table S3. Hyper- and hypo-chromatin loops and their associated loop anchors and genes in 4 different comparisons.

Table S4. Summary of public datasets used in this study.

Table S5. Gene ontologies (GO) results of hyper- and hypo-loops' associated genes in 4 different comparisons.

Table S6. *Arabidopsis*, rice, and soybean TDGs and MGCs and their corresponding shuffled regions used in this study.

Table S7. PCR primers used in this study.