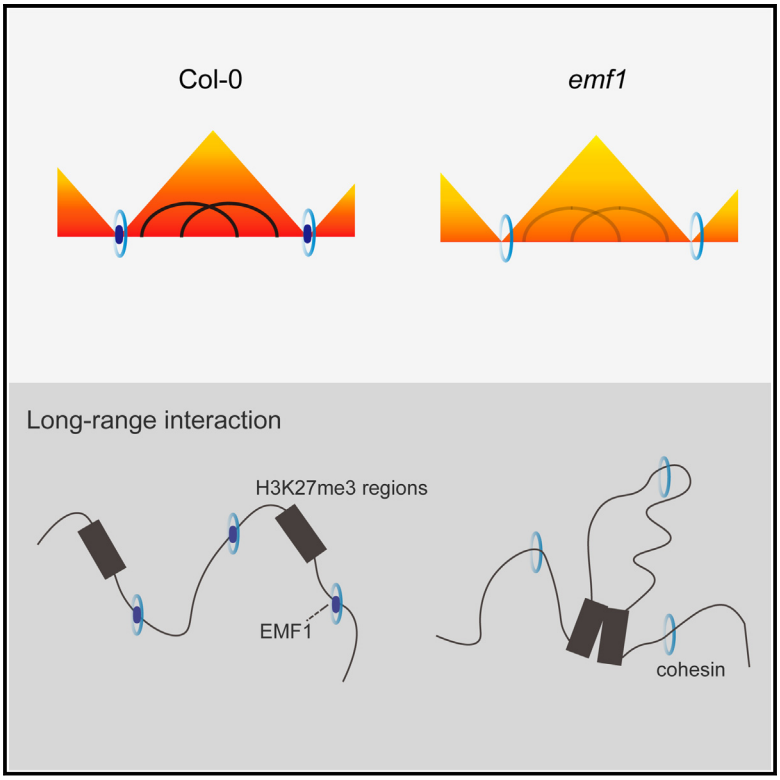
Molecular Cell

Article

EMF1 functions as a 3D chromatin modulator in Arabidopsis

Graphical abstract



Highlights

- EMF1 interacts with SCC3 and colocalizes at the CD boundary in Arabidopsis
- Loss of EMF1 or SCC3 leads to decreased CD boundary strength
- EMF1 maintains gene-resolution interactions and prevents long-range aberrant loops

Authors

Jiayue Shu, Linhua Sun, Dingyue Wang, ..., Xing Wang Deng, Hang He, Yue Zhou

Correspondence

hehang@pku.edu.cn (H.H.), yue_zhou@pku.edu.cn (Y.Z.)

In brief

Stabilizing the three-dimensional (3D) structure of chromatin is important for proper gene expression and normal plant growth. Shu et al. have identified the first 3D chromatin modulator EMF1 in plants that interact with cohesin components to maintain topologically associating domain (TAD) boundary strength. This finding fills a gap in the 3D chromatin structure in plants.

Shu et al., 2024, Molecular Cell *84*, 1–11
December 19, 2024
https://doi.org/10.1016/j.molcel.2024.10.031

CellPress

Please cite this article in press as: Shu et al., EMF1 functions as a 3D chromatin modulator in Arabidopsis, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.10.031

Molecular Cell

CellPress

Article

EMF1 functions as a 3D chromatin modulator in Arabidopsis

Jiayue Shu,[1,7] Linhua Sun,[2,7] Dingyue Wang,[1,7] Xiaochang Yin,[3,7] Minqi Yang,[1,7] Zhijia Yang,[4] Zheng Gao,[1] Yuehui He,[1,5] Myriam Calonje,[6] Jinsheng Lai,[4] Xing Wang Deng,[1,5] Hang He,[1,5,*] and Yue Zhou[1,8,*]

[1]State Key Laboratory of Protein and Plant Gene Research, School of Advanced Agricultural Sciences, Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China

[2]College of Life Sciences, Nanjing Normal University, Nanjing 210023, China

[3]School of Breeding and Multiplication (Sanya Institute of Breeding and Multiplication), Hainan University, Sanya 572025, China

[4]State Key Laboratory of Plant Physiology and Biochemistry & National Maize Improvement Center, Department of Plant Genetics and Breeding, China Agricultural University, Beijing 100183, China

[5]Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agriculture Sciences in Weifang, Weifang 261325, China

[6]Institute of Plant Biochemistry and Photosynthesis (IBVF-CSIC), Avenida Américo Vespucio 49, Seville 41092, Spain

[7]These authors contributed equally

[8]Lead contact

*Correspondence: hehang@pku.edu.cn (H.H.), yue_zhou@pku.edu.cn (Y.Z.)

https://doi.org/10.1016/j.molcel.2024.10.031

SUMMARY

It is well known that genome organizers, like mammalian CCCTC-binding factor (CTCF) or Drosophila architectural proteins CP190 and BEAF-32, contribute to the three-dimensional (3D) organization of the genome and ensure normal gene transcription. However, *bona fide* genome organizers have not been identified in plants. Here, we show that EMBRYONIC FLOWER1 (EMF1) functions as a genome modulator in Arabidopsis. EMF1 interacts with the cohesin component SISTER CHROMATIN COHESION3 (SCC3), and both proteins are enriched at compartment domain (CD) boundaries. Accordingly, *emf1* and *scc3* show a strength decrease at the CD boundary in which these proteins colocalize. EMF1 maintains CD boundary strength, either independently or in cooperation with histone modifications. Moreover, EMF1 is required to maintain gene-resolution interactions and to block long-range aberrant chromatin loops. These data unveil a key role of EMF1 in regulating 3D chromatin structure.

INTRODUCTION

Accurate formation of the intricate three-dimensional (3D) chromatin structure is crucial for the proper functioning of many genomic processes.[1,2] The organization of genomic chromatin within the nucleus varies among species. Chromosome conformation capture followed by high-throughput sequencing (Hi-C) technique was first applied to map mammalian chromatin structure,[3] revealing that the genome can be divided into A/B compartments corresponding to active/inactive chromatin states. A/B compartments frequently interact with other compartments of the same type on a multiple mega-base-pair scale.[3] As sequencing depth increased,[4] topologically associating domains (TADs) were identified as smaller regions that form strong local self-interactions.[5–7] These regions feature highly transcribed genes with binding sites for the DNA-binding protein CCCTC-binding factor (CTCF) at the boundaries. In mammals, TAD formation requires both CTCF and cohesin.[8,9] Cohesin is a ring-like complex that interacts with CTCF to regulate the 3D chromatin structure.[10] During interphase, chromatin loops are extruded by the cohesin complex, which forms between convergent CTCF sites.[11]

Nevertheless, by combining deep-sequencing Hi-C data with the TAD-calling algorithms, CTCF-independent TADs have also been found in vertebrates. They are characterized by the absence of CTCF sites and association with histone modifications,[11–13] which are similar to TAD-like structures or compartment domains (CDs) described in Drosophila.[14] In contrast to mammals, Drosophila exhibits little dCTCF enrichment at the TAD boundary. Their borders tend to be located in gene-rich and highly transcribed open chromatin regions, enriched with active histone post-translational modifications and architectural protein-binding sites[15] such as those of BEAF-32, CP190, or Chromotor.[16] In Drosophila, TAD-like structures strongly correlate with epigenetic domains. These TADs are classified into four types: transcriptionally active TADs associated with active histone modifications such as H3K4me3 or H3K36me3, polycomb-repressed TADs associated with the repressive histone modification H3K27me3, heterochromatin TADs enriched in H3K9me2, and other TADs devoid of known specific epigenetic markers.[17]

Please cite this article in press as: Shu et al., EMF1 functions as a 3D chromatin modulator in Arabidopsis, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.10.031

In plants, there are no homologs of CTCF or CTCF-loop-domain structures. By applying Hi-C techniques in crops such as rice, maize, and pepper, TAD-like structures have been identified as functional structures associated with histone modifications, gene transcription elements, and transposons.[18–20] These TAD-like structures have been classified into four types like in Drosophila.[20] The TEOSINTE BRANCHED 1, CYCLOIDEA, PCF1 (TCP), and BASIC LEUCINE ZIPPER (bZIP) transcription-factor-binding motifs are enriched at TAD boundaries in rice.[18] However, neither TCP1-bound TAD borders nor TCP1-rich TADs show significantly different chromatin genome organization in the Marchantia *tcp* mutant line.[21] Similarly, recent results showed that the Arabidopsis genome is partitioned in different TAD-like structures or CDs enriched in repressive histone marks (H3K27me3 and H3K27me1/H3K9me2) or active histone marks (H3K4me3), which are highly correlated with epigenetic modifications, gene density, and transcription.[22–24] However, insulator proteins functioning as CTCF or BEAF-32, CP190, and Chromotor have not yet been identified in plants.

Polycomb group (PcG) proteins have been involved in the maintenance and regulation of chromatin structure.[14] These proteins associate in two complexes with different histone-modifying activities. Polycomb repressive complex 1 (PRC1), which has H2A monoubiquitin ligase activity, and PRC2, which has H3 lysine 27 trimethyltransferase activity.[25–28] PRC1 and PRC2 have been proven to contribute to the establishment of both local and long-range chromatin interactions.[24,27,28] Here, we reveal that the plant-specific PcG component EMF1 works as an organizer protein in Arabidopsis. Previous studies have shown that EMF1 plays a crucial role in Arabidopsis vegetative development.[29] Accordingly, the *emf1* mutant exhibits an incomplete transition from embryonic to vegetative development; moreover, it skips the vegetative phase and generates floral organs right after germination.[30] Surprisingly, our data show a clear enrichment of EMF1 at CD boundaries and that it interacts with cohesin to regulate and maintain CD structure. Our results in addition unveil that EMF1 is crucial for maintaining strength of gene loops and repressing long-range loops to accurately regulate gene expression.

RESULTS

EMF1 is enriched at CD boundaries to regulate topological structures in Arabidopsis

Although most PcG proteins are conserved between animals and plants, there are some plant-specific proteins associated with these complexes whose roles are still unclear.[25–28] This is the case of EMF1, which has been proposed to play a role in chromatin compaction.[31] To further characterize the EMF1 function, we constructed a transgenic plant with a functional transgene *pEMF1::EMF1-YFP* that can rescue the phenotype of the *emf1* mutant (Figure S1). Then, we explored the genome-wide binding of EMF1-YFP using chromatin immunoprecipitation followed by sequencing (ChIP-seq) (Figure S2).

ChIP-seq results showed 14,170 peaks enriched in EMF1-YFP binding (Figure S2). To confirm the accuracy of our data, we compared ChIP-seq data with previously published ChIP-chip data.[29] We found that ChIP-seq peaks were more numerous and narrower than ChIP-chip peaks, which is consistent with previous observations when comparing these two types of datasets.[32] Moreover, the low correlation between ChIP-chip and ChIP-seq data was also observed in our results. The correlation between ChIP-seq replicates was high (Figure S3A), indicating that our ChIP-seq results have higher reproducibility and resolution than the ChIP-chip results. In any case, we found 3,749 overlapping target genes when comparing ChIP-seq and ChIP-chip (Figure S3C). EMF1 target genes, such as *AG*, *PI*, *SEP1*, and *FLC*,[29] displayed overlapping signals in both datasets (Figure S3B), whereas EMF1 non-target genes, like *FT* and *PHE1*, did not show EMF1 binding in any of them (Figure S3B).

Surprisingly, when we mapped Hi-C matrices with EMF1-YFP ChIP-seq data, we found that 75% of CD boundary regions were enriched with EMF1 binding (Figures S3D and S5; Data S1). We repeated this analysis using ChIP-chip data, which also showed that 52% of CD boundary regions were enriched with EMF1 binding (Figures S3D and S3E). In addition, whereas the binding of the PRC2 components CURLY LEAF (CLF) and SWIGER (SWN)[33] was detected exclusively inside H3K27me3-enriched CDs (Figures S4A and S4B), EMF1 binding was enriched at the boundaries of H3K4me3-, H3K27me3-, and H3K9me2-enriched CDs (Figures 1A–1C), supporting the notion that EMF1 could function independently of PRC2. To explore whether bound EMF1 is crucial for higher-order chromatin organization in Arabidopsis, we analyzed the TAD separation score to show boundary strength in Col-0 (wild type [WT]) and the *emf1* mutant. The results showed that there was a significant decrease in boundary strength in the mutant compared with the WT's three kinds of boundaries for H3K4me3-, H3K27me3-, and H3K9me2-enriched CDs in which EMF1 was found (Figures 1D–1F), supporting the idea that EMF1 plays a crucial role in regulating 3D chromatin structure in Arabidopsis.

In Drosophila, many insulator proteins bind very close to gene promoters, and some of these proteins have been proposed to function as transcriptional regulators.[34,35] Thus, we analyzed the distribution of EMF1-binding sites relative to gene features and found that approximately 75% of these sites were located at promoter regions and 5′ UTR, suggesting that it might be involved in the regulation of gene expression (Figure 1H). Next, we classified EMF1 target genes into two clusters according to the presence or absence of H3K27me3, as some EMF1-binding sites are well correlated with the H3K27me3 signal (Figure S4C). By analyzing the distribution of EMF1 at target genes, we observed that EMF1 peaks tend to be located around the transcription start site (TSS) and transcription terminal site (TTS) in both clusters (Figure 1G). Further, we showed that EMF1 was located at the gene body of H3K27me3 co-target genes. In addition, we found that loss of EMF1 function leads to gene transcription misregulation, but this was not associated with a biased gene expression change, as we found a similar proportion of genes with up- and downregulated expression (Figures 1I and S2I). This feature is distinct from PcG proteins' function, which is mainly to repress gene transcription.

Please cite this article in press as: Shu et al., EMF1 functions as a 3D chromatin modulator in Arabidopsis, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.10.031

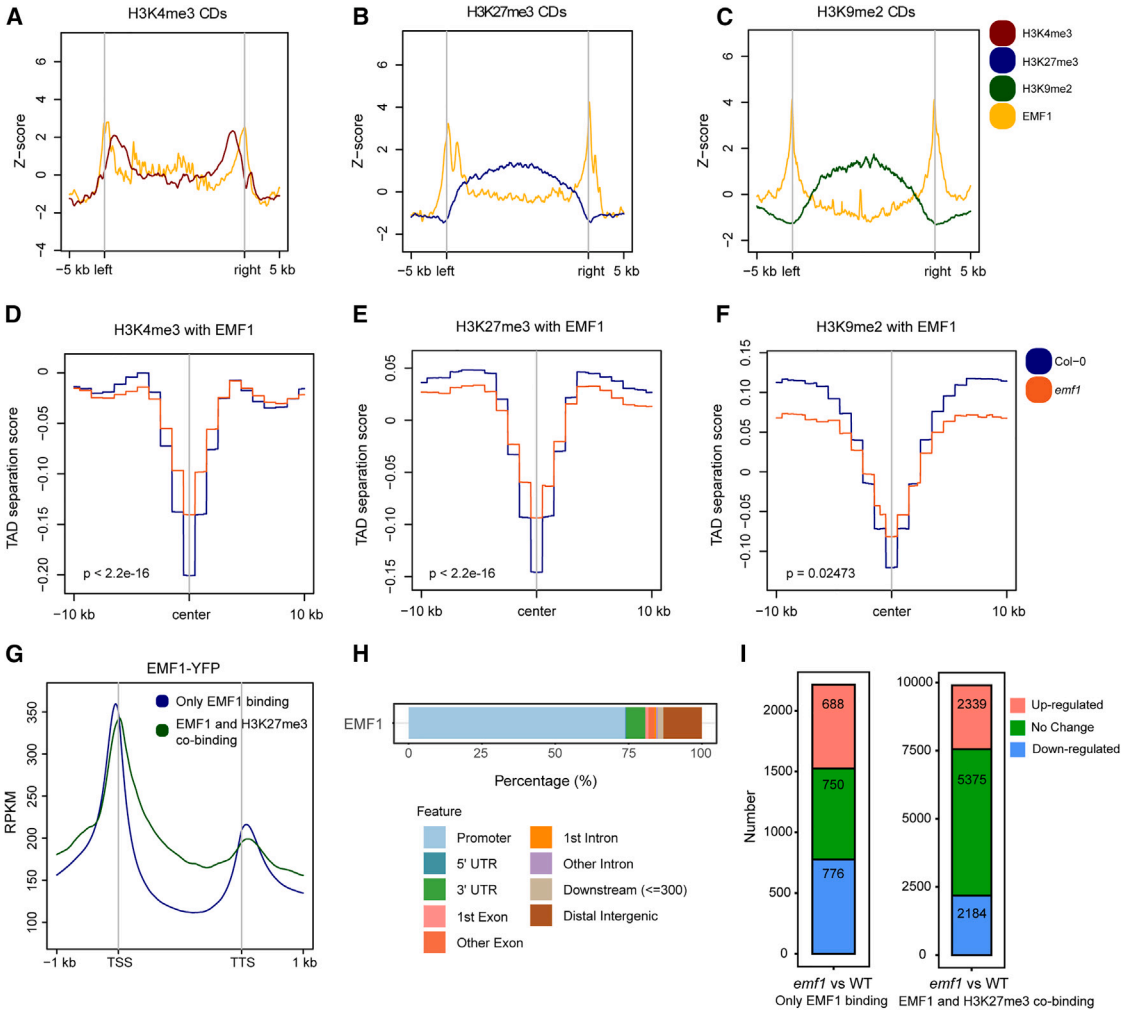


Figure 1. EMF1 is enriched at CD boundaries in Arabidopsis and regulates topological structure

(A–C) Metagene plots indicating the H3K4me3, H3K27me3, H3K9me2, and EMF1 signals at the H3K4me3 (A), H3K27me3 (B), and H3K9me2 (C) compartment domains (CDs), respectively.

(D–F) Metaplots indicating the TAD separation score of the H3K4me3 (D), H3K27me3 (E), and H3K9me2 (F) CDs. The *p* value was calculated by two sample t tests, one tailed.

(G) Distribution of EMF1-binding sites relative to gene feature.

(H) Metagene plots of EMF1 coverage at EMF1 target genes.

(I) The distribution of genes with altered expression in the *emf1* mutant.

EMF1 maintains CD boundary strength, either independently or in cooperation with histone modifications

Like in Drosophila, Arabidopsis chromatin organization is highly correlated with epigenetic modifications. Therefore, we investigated the relationship among EMF1 binding, CD border strength, and the levels of histone modifications. Although the levels of H3K4me3 within H3K4me3-enriched CDs were unchanged in *emf1* compared with WT, CD borders were weaker in the *emf1* mutant than in WT (Figures 2A and S6). This result indicates that the effect of EMF1 on H3K4me3 CDs is independent of the H3K4me3 level and supports the role of EMF1 in regulating CD boundaries.

Interestingly, we found that H3K27me3- and H3K9me2-enriched CDs were grouped in two different clusters, respectively (Figures 2B and 2C), suggesting different EMF1 regulatory functions. For H3K27me3 CDs, we divided them into two clusters according to the EMF1 binding. H3K27me3 CD cluster 1 (black line) represents those CDs in which EMF1 localizes both within CD and at CD boundaries. In the *emf1* mutant, the CD border strength and the levels of H3K27me3 were significantly reduced, which suggests EMF1 and H3K27me3 are synergistically involved in CD border strength regulation (Figures 2B and S6). Differently, H3K27me3 CD cluster 2 (red line) represents those CDs in which EMF1 localizes only at CD boundaries (Figure 2B). In the *emf1* mutant, CD border strength was also reduced, but the change

Please cite this article in press as: Shu et al., EMF1 functions as a 3D chromatin modulator in Arabidopsis, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.10.031

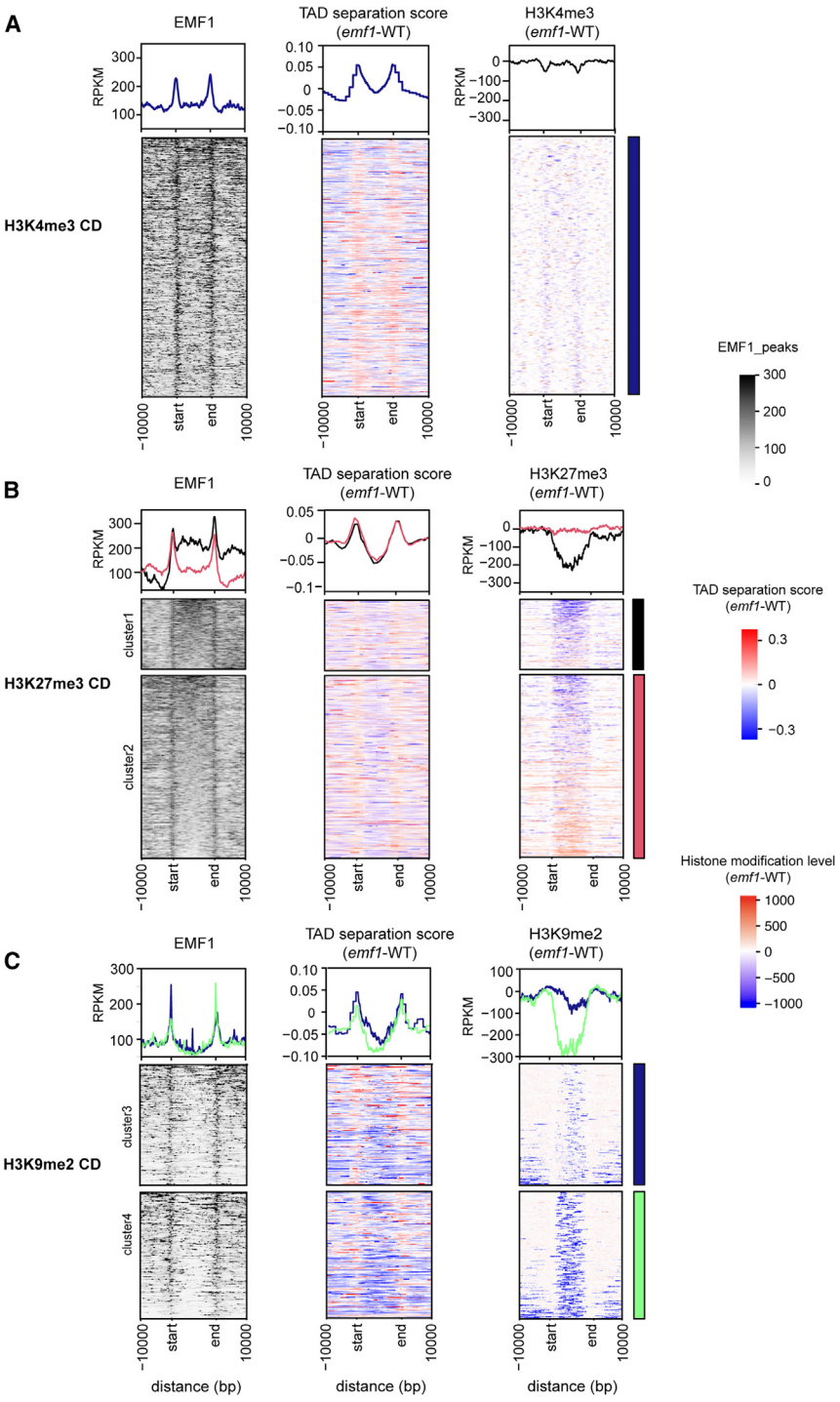


Figure 2. EMF1 maintains CD interactions either independently or in cooperation with histone modifications

Metaplots and corresponding heatmaps indicating the EMF1 binding (left), delta TAD separation score (*emf1*–WT; middle), and delta modification enrichment (*emf1*–WT; right) of the H3K4me3 (A), H3K27me3 (B), and H3K9me2 (C) CDs. H3K27me3 CDs are divided into cluster 1 and cluster 2, according to EMF1 binding (B). H3K9me2 CDs are divided into cluster 3 and cluster 4, according to the H3K9me2 modification level (C).

in H3K27me3 level was indirect and showed no tendency (Figures 2B and S6), which indicates the function of EMF1 at CD border are independent of the H3K27me3 level changes. For H3K9me2 CDs, we also divided them into two clusters according to the change in H3K9me2 level (Figures 2C and S6). The levels of H3K9me2 in the *emf1* mutant were strongly decreased in cluster 4 (green line) and not obvious in cluster 3 (blue line) (Figure 2C). CD border strength was reduced in both clusters 3 and 4 (Figure 2C). Together, these data suggest that EMF1 binding at CD boundaries maintains CD border strength either independently or in cooperation with histone modifications.

EMF1 maintains loops between genes and represses long-range interactions

To further explore the role of EMF1 in 3D chromatin organization, we focused on chromatin loops. A previous study defined >20,000 short-range loops from the euchromatic chromosome arms of Arabidopsis.[23] By measuring the loop strength within EMF1-regulated CDs, we found that gene-loop strength was weakened in the *emf1* mutant compared with WT (Figure 3A). The strength reduction of gene loops with an average length of 6,500 bp is in line with the difference in interaction decay exponents (IDEs) found between WT and the *emf1* mutant (Figure 3B). Comparison of Hi-C interaction matrices between WT and *emf1* revealed that some long-range chromatin interaction or loops were stronger in *emf1* than in WT (Figure 3C). We identified 117 loops with increased strength in the *emf1* mutant than in WT, with an average length of 1,800 kb, which is consistent with IDEs (Figures 3B and 3D; Data S2). The anchor regions of these long-range interactions were enriched in H3K27me3 (Figure S7A). However, we did not find EMF1 binding at the anchor regions of these long-range loops (Figure S7B), indicating that EMF1 does not directly regulate these interactions. The genes annotated at the anchors of *emf1*-specific long-range loops were slightly downregulated (Figure 3E). It could be possible that EMF1 acts as a clip to impose genomic interactions, and the lack of EMF1 allows the establishment of inappropriate long-range H3K27me3 interactions, which subsequently could promote gene downregulation.

Please cite this article in press as: Shu et al., EMF1 functions as a 3D chromatin modulator in Arabidopsis, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.10.031

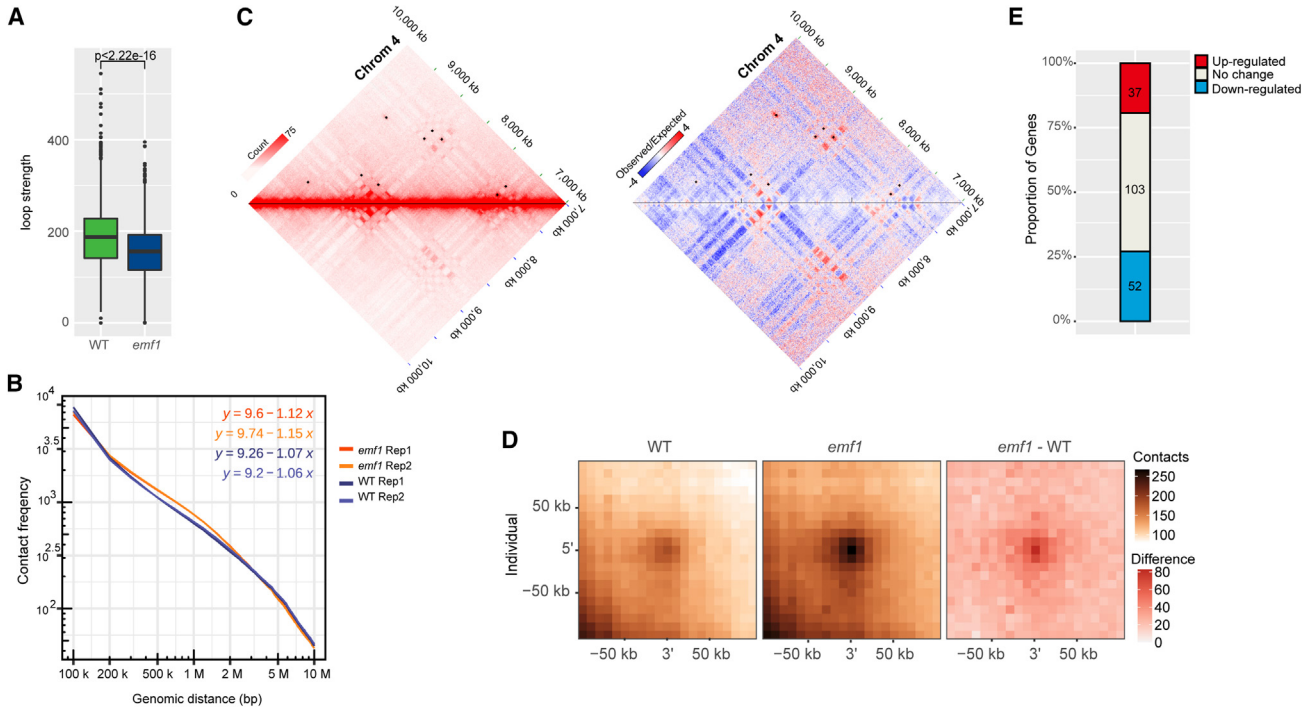


Figure 3. EMF1 maintains loops between genes in CDs and represses H3K27me3 long-range loops by regulating boundaries

(A) The strength of gene loops in wild type (WT) and the *emf1* mutant.

(B) The interaction decay exponents (IDEs) of WT and *emf1* describe the slope of the relative contact probability from 100 kb to 10 Mb decay, with genomic distance using sequencing depth-normalized data.

(C) Representative heatmaps showing the appearance of increased long-range loops (marked in black) in *emf1* mutant (top) compared with WT (bottom).

(D) Aggregate peak analysis (APA) showing the strength of increased long-range loops in individual (left and middle) WT and *emf1* mutant, and the difference between *emf1* mutant and WT (right).

(E) The distribution of genes with altered expression (adjusted p value < 0.05 and $\log_2$ fold-change > 1.5) in the increased long-range loops. Gene counts are marked on the bar.

Collectively, these observations support that EMF1 directly controls CD boundaries and indirectly regulates short- and long-range chromatin loops.

SCC3 and EMF1 interact with each other and co-distribute genome wide

Cohesin is a multi-subunit protein complex that forms a ring-like structure made up of structural maintenance of chromosomes1 (SMC1), SMC3, SCC3, and kleisin.[36,37] In Arabidopsis, there are four kleisin proteins: SYNAPTIC1 (SYN1), SYN2, SYN3, and SYN4.[8,38,39] It has been reported that SCC3 interacts with SYN4.[40] Cohesin is required to hold sister chromatids during cell division and regulates higher-order chromatin structures.[36,37] However, little is known about the function of cohesin in Arabidopsis genome organization. Interestingly, we found that EMF1 interacts with SCC3 by yeast two-hybrid assay (Figure 4A) and further verified the interaction using bimolecular fluorescence complementation (BiFC). The YFP signals were only detected in EMF1-nYFP/cYFP-SCC3 but not in the negative controls (Figure 4B). Moreover, coimmunoprecipitation (coIP) experiments in Col-0 or *EMF1-FLAG*[41] protoplasts in which HA-tagged SCC3 was transiently coexpressed showed that HA-SCC3 copurified with EMF1-FLAG (Figure 4C), which further supported the EMF1-SCC3 interaction *in vivo*.

In addition, we transformed *GST*, *GST-SCC3*, and *MYC-EMF1* into *E. coli* BL21 cells, respectively. Using *in vitro* pull-down assay, GST-fused SCC3—but not GST alone—could pull down EMF1 tagged with MYC from total protein extracts (Figure 4D). Because EMF1 is a plant-specific protein existing in angiosperms (Figure S8A), and yeast two-hybrid results showed that EMF1 and SCC3 from the maize line B73 interacted with each other (Figure S8B), we inferred that the interaction of EMF1-cohesin is conserved in angiosperms.

Next, we constructed *pSCC3*::*SCC3-GFP*/scc3-2$^{-/-}$ and *pSCC3*::*SCC3-FLAG*/scc3-2$^{-/-}$ transgenic lines, finding that both transgenes could rescue the lethal *scc3-2* mutant phenotype (Figures S9A–S9C), and determined the genome-wide occupancy of SCC3 using a green fluorescent protein (GFP) antibody by ChIP-seq (Figure S2). Our ChIP-seq data were similar to previously reported ChIP-seq data obtained with SCC3-specific antibody[40] (Figure S10). We then compared SCC3 and EMF1 occupancy profiles and found a substantial overlap between SCC3- and EMF1-binding sites (Figures 4E and S11A). In addition, the strength of EMF1-SCC3 co-binding peaks is stronger than the EMF1-binding peaks but weaker than

Please cite this article in press as: Shu et al., EMF1 functions as a 3D chromatin modulator in Arabidopsis, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.10.031

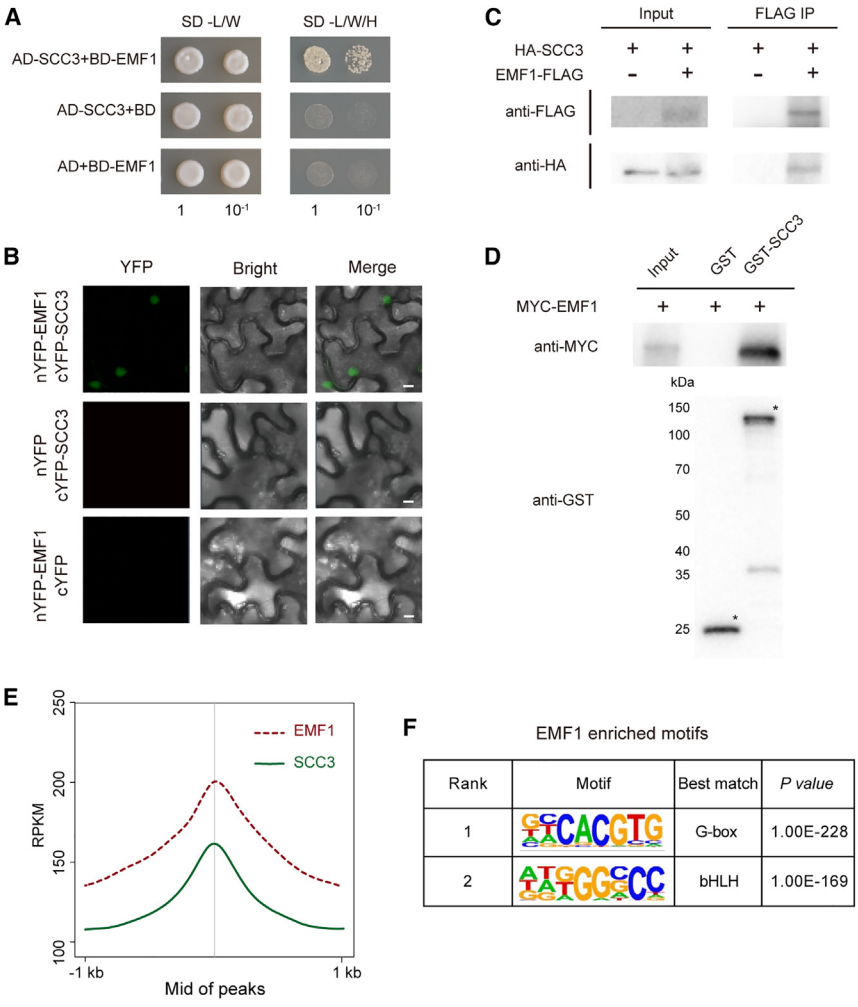


Figure 4. SCC3 and EMF1 interact with each other and co-distribute genome wide

(A) SCC3 interacts with EMF1 in yeast cells.

(B) Bimolecular fluorescence complementation analysis of the physical associations of EMF1 with SCC3 in tobacco cells. Bar, 10 μm.

(C) Coimmunoprecipitation of SCC3 with EMF1.

(D) Pull-down assay for interactions of GST-SCC3 and Myc-EMF1. GST is used as a negative control. We repeated all these experiments twice.

(E) Metaplot showing the ChIP-seq results of EMF1 at SCC3-binding regions (dashed line) and SCC3 at EMF1-binding regions (solid line), respectively.

(F) Motifs enriched at EMF1-binding sites. The *p* values are indicated.

SCC3-binding peaks, which indicates that SCC3 is important for EMF1 binding to chromatin (Figures S11B and S11C). Previous results showed that EMF1 has sequence-unspecific DNA-binding activity[31]; thus, to understand how EMF1 is recruited to target regions, we conducted a motif analysis at EMF1-binding sites, finding that there was a significant enrichment in G-box DNA regulatory elements (Figures 4F and S11E). Interestingly, recent results showed that AtSYN4 could directly bind G-box containing DNA,[40] strongly suggesting that EMF1 could be recruited by cohesin to DNA.

Impaired SCC3 expression affects 3D chromatin structure

In vertebrates, cohesin is essential for loop extrusion and maintenance of promoter-enhancer interactions.[37] Thus, we mapped the genome-wide occupancy of SCC3-GFP along H3K27me3-, H3K9me2-, and H3K4me3-enriched CDs and found that, similar to EMF1, SCC3 binding was enriched at CD boundaries (Figure 5A). In addition, more than 50% of SCC3-binding sites are localized at the promoter region (Figure 5B), and SCC3 tends to be located around TSS and TTS at target genes (Figure 5C) like EMF1. Moreover, we found that 30% of CD boundaries were enriched with both EMF1 and SCC3 (Figure S11D). To verify these results, we performed the same analysis using SYN4-binding-site data,[40] also finding an enrichment of SYN4 at the boundaries of three types of CDs (Figure S12), indicating that EMF1 works in concert with cohesin to regulate the structure of CDs.

Because homozygous *Atscc3-2*$^{-/-}$ mutation is lethal (Figures S9A and S9B),[42] to investigate the function of SCC3 in Arabidopsis, we used RNA interference (RNAi) to obtain two independent RNAi lines: *RNAi-1* and *RNAi-2* (Figure S9D). In comparison with WT, these lines exhibited slower growth and significantly lower *SCC3* expression (Figures S9D–S9F). Both *SCC3 RNAi* lines showed a similar proportion of up- and downregulated genes (Figures 5D and S2I). To further investigate the biological relevance of EMF1 and SCC3 interaction in CD regulation, we used these *SCC3 RNAi* lines to investigate possible changes in CD boundary strength (Figure S5D). We classified CDs into two categories based on the presence or absence of SCC3 and EMF1 binding at their boundaries. Both *RNAi-1* and *RNAi-2* showed reduced boundary strength in the three types of CDs in which SCC3 and EMF1 are present at the boundaries (Figures 5E–5G), indicating that the interaction between SCC3 and EMF1 contributes to maintaining CD boundary strength regardless of the histone modification that they contain.

DISCUSSION

In this work, we combined ChIP-seq, Hi-C, and histone modification data to analyze the role of EMF1 and SCC3 in the formation of 3D chromatin structure in Arabidopsis. Our results support the notion that EMF1 is not only associated with Polycomb silencing but also functions as a 3D chromatin organizer (Figure 6A). The binding of EMF1 at the CD boundaries maintains CD stability and contact frequency. In addition, we found that cohesin is crucial for the maintenance of all kinds of CDs. According to

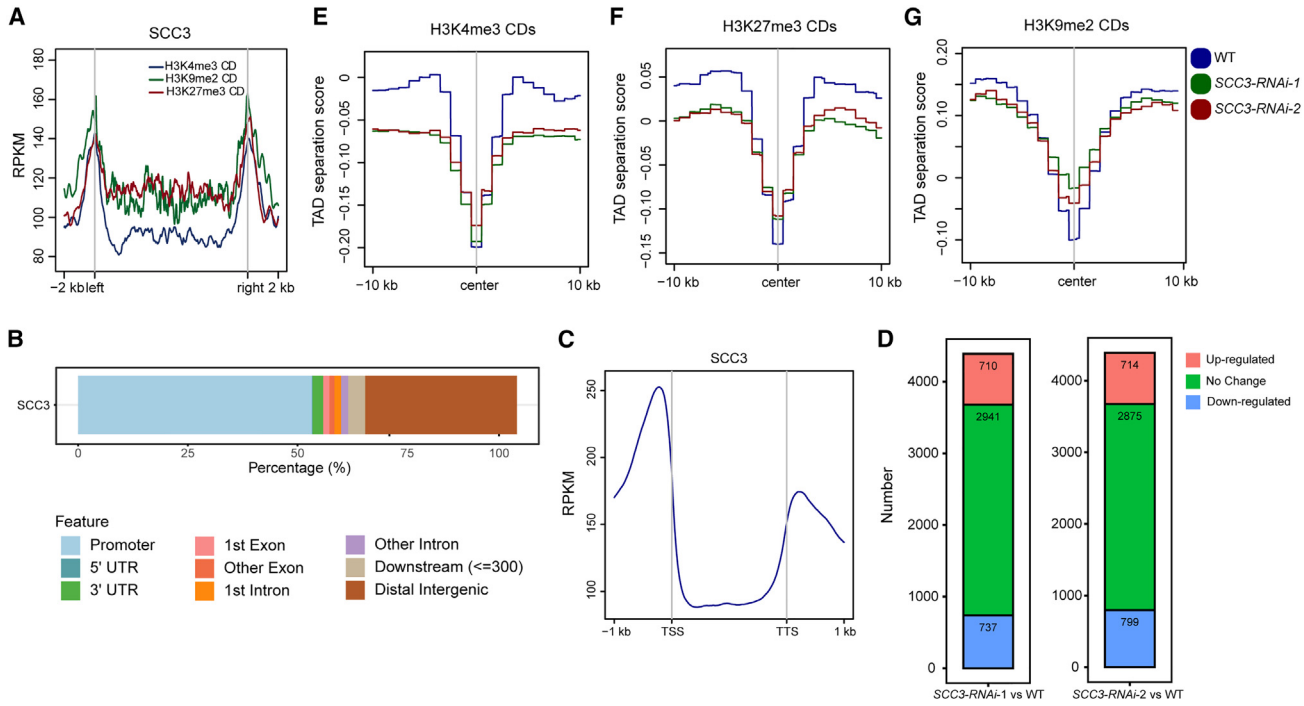


Figure 5. Impaired SCC3 expression affects 3D chromatin structure

(A) Metaplot showing the SCC3 RPKM value within ±2 kb of the H3K27me3 (red), H3K4me3 (blue), and H3K9me2 (green) compartment domain (CD) regions.

(B) Distribution of SCC3-binding sites relative to gene feature.

(C) Metagene plots of SCC3 coverage at SCC3 target genes.

(D) The distribution of genes with altered expression targeting by SCC3 in two *SCC3 RNAi* replications.

(E–G) Metaplots indicating the TAD separation score of H3K4me3 (E), H3K27me3 (F), and H3K9me2 (G) CDs. The *p* value was calculated by two sample t tests, one tailed. (*SCC3-RNAi-1* H3K4me3 CDs: $p = 0.2868$, H3K27me3 CDs: $p = 0.001389$, H3K9me2 CDs: $p = 0.0004484$; *SCC3-RNAi-2* H3K4me3 CDs: $p = 0.004201$, H3K27me3 CDs: $p = 0.0002356$, H3K9me2 CDs: $p = 0.002483$.)

our results, we propose that cohesin could recruit EMF1 to co-regulate higher-order genome organization (Figure 6A). Moreover, EMF1 regulates short- and long-range chromatin loops (Figure 6B).

CTCF- and cohesin-dependent TADs are known to be specific to vertebrates.[11,43–45] In *Drosophila*, CTCF deletion has a limited effect on CD boundaries.[46] TAD-like structure formation has been proposed to arise cooperatively from histone modifications, binding of architectural proteins, and active transcription. In Arabidopsis, recent results also showed that histone modifications are crucial for CD formation and maintenance.[24] This implies that the regulation of chromatin 3D structure is highly similar between Drosophila and Arabidopsis compared with vertebrates. In Drosophila, the definition of organizer, insulator, or architectural proteins is based on the following specificities. First, the boundary-binding specificity is one of the common features of organizer or insulator proteins, such as BEAF-32 and CP190. They function to block chromatin loops. Our data provided enough evidence to show that EMF1 could bind to all three kinds of CDs' boundaries. Moreover, ChIP-seq in our study generates additional, more-precise peaks, whereas the peaks in ChIP-chip are wider and fewer.[32] Nevertheless, the characteristics of bound EMF1 are also proven by ChIP-chip (Figure S3). CD boundary regions were enriched with EMF1, and more than 75% of EMF1 peaks at CD boundaries in ChIP-chip data could be covered in ChIP-seq data (Figure S3). In the *emf1* mutant, long-range interactions are also significantly increased. Second, the strength of boundaries is reduced with the mutation/depletion of organizer or insulator proteins. However, the mutation in different organizers or insulators leads to different changes in 3D chromatin structures. In *Drosophila*, the mutation in *BEAF-32* led to changes in 3D chromatin structure only in BG3 cells but not in S2 cells.[47,48] Our data proved that the mutation in *EMF1* led to a significant strength reduction of boundaries with bound EMF1. Meanwhile, we suggest there may exist other insulators (such as TCP and bZIP[18,21,24]) collaborating with EMF1 to regulate CDs' strength. These insulators may have redundant functions. Architectural protein complexes may exist at the CD boundary to function like Drosophila. Therefore, the deletion of one factor is not sufficient to produce drastic changes in the 3D structure of chromatin.

In Arabidopsis, the *Atscc3-2* mutant has a more severe phenotype than the *emf1* mutant, as loss of SCC3 function is lethal. This is not surprising, as cohesin, in addition to regulating higher-order chromatin structures, plays important roles during cell division. Therefore, the mild phenotype of *SCC3 RNAi* mutants explains the fact that the 3D structural changes of chromatin in these mutants were not as drastic as expected for a loss-of-function mutant. Knockdown of *SCC3* weakens the

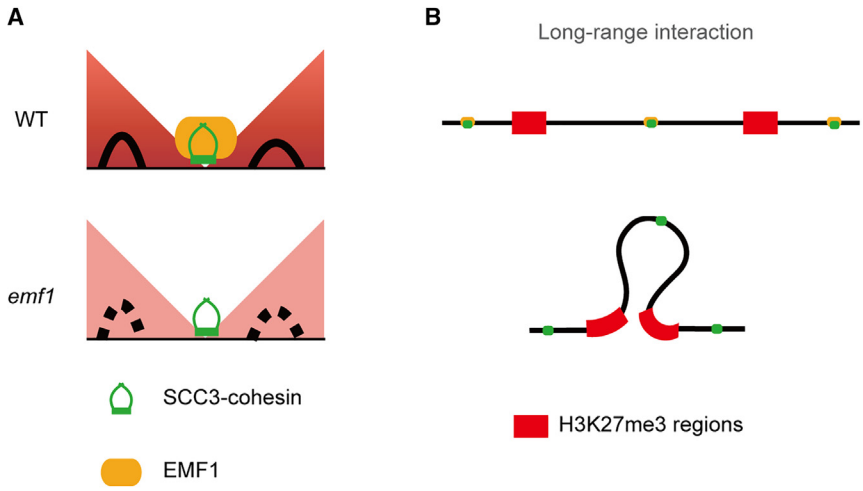


Figure 6. Proposed model of EMF1 3D chromatin structural regulation in *Arabidopsis*

(A) EMF1 is recruited to the boundaries, and cooperates with cohesin to maintain CD boundary strength.

(B) EMF1 maintains gene loops in compartment domains (left) and blocks aberrant long-range loops by regulating boundaries (right).

strength of CD boundaries. Moreover, the same binding motif, the higher binding strength of EMF1 at cohesin-binding sites, and the interaction between EMF1 and cohesin indicate that EMF1 could be recruited to CD borders by cohesin. Nevertheless, we noticed that there are some CD boundaries without the binding of EMF1 or cohesin, also suggesting that there may be other factors participating in shaping the 3D chromatin structure in Arabidopsis. Moreover, there are CD boundaries occupied by EMF1 but not by SCC3, suggesting that EMF1 could be recruited to CD borders by other factors. Besides G-box, there are some other motifs enriched by motif analysis at EMF1-binding site s (Figures S11E and S11F). For example, the basic helix-loop-helix (bHLH) transcription factor family is one of the largest transcription factor gene families in Arabidopsis. We compared ChIP-seq data for some published genes, and the results showed that the EMF1 and PIFs binding sites were highly overlapped, which suggests that EMF1 may interact with PIFs (Figure S11F).

In summary, our study identified EMF1 as the first functional modulator protein, which could contribute to furthering our understanding of the molecular mechanisms of 3D chromatin regulation in plants. Organizers or insulators dependent on 3D chromatin regulation have species specificity. We admitted that the EMF1 function is different from CTCF or its mediated loop domain formation function. Although how EMF1 or even Drosophila insulators organize chromatin 3D structure is elusive, we think the EMF1 is more similar to insulators in Drosophila. A structural organizer such as EMF1 may work as the "glue" to hold different transcription factors and cohesin together and form an architectural protein complex at the CD boundary to regulate 3D chromatin organization.

Limitations of the study

Although our study provides critical evidence for the identification of a higher-order chromatin modulator in Arabidopsis, problems remain. First, the weak molecular phenotype of the *emf1* mutant implies that unknown proteins could act in concert with EMF1 to regulate the 3D structure of chromatin. Architectural protein complexes may exist at the CD boundary to function like Drosophila. However, this conjecture needs to be studied further. Annotating potentially complete complexes remains a challenge for the field. Second, we clearly showed that cohesin is crucial for the maintenance of chromatin organization, but the detailed regulatory processes associated with cohesin need further study. For example, it is unclear how cohesin regulated chromatin loops.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Yue Zhou (yue_zhou@pku.edu.cn).

Materials availability

Constructs and plant seeds generated in this study will be available from the lead contact upon request.

Data and code availability

All sequencing data generated from this study have been deposited into the SRA database (https://www.ncbi.nlm.nih.gov/sra/) under the accession number PRJEB60957.

This paper does not report original code.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

This work was supported by Biological Breeding—National Science and Technology Major Project (2023ZD04073); grant 32370612 (Y.Z.) from the National Natural Science Foundation of China; grant JCTD-2022-06 (Y.Z.) supported by CAS Youth Interdisciplinary Team; PID2019-106664GB-I00 (Myriam Calonje) from the Spanish Ministry of Science and Innovation; 2021T140019, from the China Postdoctoral Science Foundation (X.Y.); and startup funds from the State Key Laboratory for Protein and Plant Gene Research, the School of Advanced Agricultural Sciences, and the Peking-Tsinghua Center for Life Sciences at Peking University (Y.Z.). We thank Dr. Danmeng Zhu for providing the pUCCRNAi and pCAM2300-35s-Ocs vectors. This study was carried out at the Peking University High Performance Computing Platform, and the calculations were performed on CLS-HPC.

AUTHOR CONTRIBUTIONS

J.S., X.Y., and Y.Z. conceived the research. J.S., X.Y., Z.Y., Z.G., and Y.H. conducted the experiments. L.S., D.W., M.Y., X.W.D., and H.H. analyzed the data. J.S., D.W., M.Y., M.C., and Y.Z. wrote the manuscript. All authors read and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Plant materials and growth conditions
- METHOD DETAILS
  - ChIP-seq
  - In situ Hi-C
  - RNA-seq
  - Yeast two-hybrid
  - BiFC
  - Co-IP
  - Pull down
  - Gene resolution loop analysis
  - RNA-seq and ChIP-seq bioinformatics analysis
  - Hi-C sequencing data processing and comparison
  - Identification of CDs from Hi-C
  - Long-range chromatin loop identification, quantitation, and comparison

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.molcel.2024.10.031.

Received: June 17, 2024
Revised: August 30, 2024
Accepted: October 24, 2024
Published: November 19, 2024

REFERENCES

1. Bickmore, W.A. (2013). The spatial organization of the human genome. Annu. Rev. Genomics Hum. Genet. *14*, 67–84. https://doi.org/10.1146/annurev-genom-091212-153515.
2. Zheng, H., and Xie, W. (2019). The role of 3D genome organization in development and cell differentiation. Nat. Rev. Mol. Cell Biol. *20*, 535–550. https://doi.org/10.1038/s41580-019-0132-4.
3. 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. https://doi.org/10.1126/science.1181369.
4. Wetterstrand, K.A., and M.S.. (2022). DNA Sequencing Costs. Data. https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data.
5. 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. https://doi.org/10.1038/nature11082.
6. Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385. https://doi.org/10.1038/nature11049.
7. 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. https://doi.org/10.1016/j.cell.2012.01.010.
8. da Costa-Nunes, J.A., Bhatt, A.M., O'Shea, S., West, C.E., Bray, C.M., Grossniklaus, U., and Dickinson, H.G. (2006). Characterization of the three Arabidopsis thaliana RAD21 cohesins reveals differential responses to ionizing radiation. J. Exp. Bot. *57*, 971–983. https://doi.org/10.1093/jxb/erj083.
9. Vietri Rudan, M.V., Barrington, C., Henderson, S., Ernst, C., Odom, D.T., Tanay, A., and Hadjur, S. (2015). Comparative Hi-C Reveals that CTCF Underlies Evolution of Chromosomal Domain Architecture. Cell Rep. *10*, 1297–1309.
10. Zhu, Z., and Wang, X. (2019). Roles of cohesin in chromosome architecture and gene expression. Semin. Cell Dev. Biol. *90*, 187–193. https://doi.org/10.1016/j.semcdb.2018.08.004.
11. Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., and Aiden, E.L. (2014). A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell *159*, 1665–1680. https://doi.org/10.1016/j.cell.2014.11.021.
12. Bintu, B., Mateo, L.J., Su, J.H., Sinnott-Armstrong, N.A., Parker, M., Kinrot, S., Yamaya, K., Boettiger, A.N., and Zhuang, X.W. (2018). Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. Science *362*, eaau1783. https://doi.org/10.1126/science.aau1783.
13. Szabo, Q., Donjon, A., Jerković, I., Papadopoulos, G.L., Cheutin, T., Bonev, B., Nora, E.P., Bruneau, B.G., Bantignies, F., and Cavalli, G. (2020). Regulation of single-cell genome organization into TADs and chromatin nanodomains. Nat. Genet. *52*, 1151–1157. https://doi.org/10.1038/s41588-020-00716-8.
14. Rowley, M.J., Nichols, M.H., Lyu, X., Ando-Kuri, M., Rivera, I.S.M., Hermetz, K., Wang, P., Ruan, Y., and Corces, V.G. (2017). Evolutionarily Conserved Principles Predict 3D Chromatin Organization. Mol. Cell *67*, 837–852.e7. https://doi.org/10.1016/j.molcel.2017.07.022.
15. Hou, C.H., Li, L., Qin, Z.S., and Corces, V.G. (2012). Gene Density, Transcription, and Insulators Contribute to the Partition of the Drosophila Genome into Physical Domains. Mol. Cell *48*, 471–484. https://doi.org/10.1016/j.molcel.2012.08.031.
16. Wang, Q., Sun, Q., Czajkowsky, D.M., and Shao, Z.F. (2018). Sub-kb Hi-C in D. melanogaster reveals conserved characteristics of TADs between insect and mammalian cells. Nat. Commun. *9*, 188. https://doi.org/10.1038/s41467-017-02526-9.
17. Szabo, Q., Bantignies, F., and Cavalli, G. (2019). Principles of genome folding into topologically associating domains. Sci. Adv. *5*, eaaw1668. https://doi.org/10.1126/sciadv.aaw1668.
18. Liu, C., Cheng, Y.J., Wang, J.W., and Weigel, D. (2017). Prominent topologically associated domains differentiate global chromatin packing in rice from Arabidopsis. Nat. Plants *3*, 742–748. https://doi.org/10.1038/s41477-017-0005-9.
19. Liao, Y., Wang, J.T., Zhu, Z.S., Liu, Y.L., Chen, J.F., Zhou, Y.F., Liu, F., Lei, J.J., Gaut, B.S., Cao, B.H., et al. (2022). The 3D architecture of the pepper genome and its relationship to function and evolution. Nat. Commun. *13*, 3479. https://doi.org/10.1038/s41467-022-31112-x.
20. Dong, P.F., Tu, X.Y., Chu, P.Y., Lü, P.T., Zhu, N., Grierson, D., Du, B.J., Li, P.H., and Zhong, S.L. (2017). 3D Chromatin Architecture of Large Plant Genomes Determined by Local A/B Compartments. Mol. Plant *10*, 1497–1509. https://doi.org/10.1016/j.molp.2017.11.005.
21. Karaaslan, E.S., Wang, N., Faiß, N., Liang, Y., Montgomery, S.A., Laubinger, S., Berendzen, K.W., Berger, F., Breuninger, H., and Liu, C. (2020). Marchantia TCP transcription factor activity correlates with three-dimensional chromatin structure. Nat. Plants *6*, 1250–1261. https://doi.org/10.1038/s41477-020-00766-0.
22. Wang, C., Liu, C., Roqueiro, D., Grimm, D., Schwab, R., Becker, C., Lanz, C., and Weigel, D. (2015). Genome-wide analysis of local chromatin

packing in Arabidopsis thaliana. Genome Res. *25*, 246–256. https://doi.org/10.1101/gr.170332.113.

23. Liu, C., Wang, C.M., Wang, G., Becker, C., Zaidem, M., and Weigel, D. (2016). Genome-wide analysis of chromatin packing in Arabidopsis thaliana at single-gene resolution. Genome Res. *26*, 1057–1068. https://doi.org/10.1101/gr.204032.116.
24. Yin, X., Romero-Campero, F.J., Yang, M., Baile, F., Cao, Y., Shu, J., Luo, L., Wang, D., Sun, S., Yan, P., et al. (2023). Binding by the Polycomb complex component BMI1 and H2A monoubiquitination shape local and long-range interactions in the Arabidopsis genome. Plant Cell *35*, 2484–2503. https://doi.org/10.1093/plcell/koad112.
25. Baile, F., Gómez-Zambrano, Á., and Calonje, M. (2022). Roles of Polycomb complexes in regulating gene expression and chromatin structure in plants. Plant Commun. *3*, 100267. https://doi.org/10.1016/j.xplc.2021.100267.
26. Cheutin, T., and Cavalli, G. (2019). The multiscale effects of polycomb mechanisms on 3D chromatin folding. Crit. Rev. Biochem. Mol. Biol. *54*, 399–417. https://doi.org/10.1080/10409238.2019.1679082.
27. Pachano, T., Crispatzu, G., and Rada-Iglesias, A. (2019). Polycomb proteins as organizers of 3D genome architecture in embryonic stem cells. Brief. Funct. Genomics *18*, 358–366. https://doi.org/10.1093/bfgp/elz022.
28. Huang, Y., Sicar, S., Ramirez-Prado, J.S., Manza-Mianza, D., Antunez-Sanchez, J., Brik-Chaouche, R., Rodriguez-Granados, N.Y., An, J., Bergounioux, C., Mahfouz, M.M., et al. (2021). Polycomb-dependent differential chromatin compartmentalization determines gene coregulation in Arabidopsis. Genome Res. *31*, 1230–1244. https://doi.org/10.1101/gr.273771.120.
29. Kim, S.Y., Lee, J., Eshed-Williams, L., Zilberman, D., and Sung, Z.R. (2012). EMF1 and PRC2 cooperate to repress key regulators of Arabidopsis development. PLoS Genet. *8*, e1002512. https://doi.org/10.1371/journal.pgen.1002512.
30. Yin, X.C., Romero-Campero, F.J., de los Reyes, P., Yan, P., Yang, J., Tian, G.M., Yang, X.Z., Mo, X.R., Zhao, S.S., Calonje, M., and Zhou, Y. (2021). H2AK121ub in Arabidopsis associates with a less accessible chromatin state at transcriptional regulation hotspots. Nat. Commun. *12*, 315. https://doi.org/10.1038/s41467-020-20614-1.
31. Calonje, M., Sanchez, R., Chen, L., and Sung, Z.R. (2008). EMBRYONIC FLOWER1 participates in polycomb group-mediated AG gene silencing in Arabidopsis. Plant Cell *20*, 277–291. https://doi.org/10.1105/tpc.106.049957.
32. Ho, J.W.K., Bishop, E., Karchenko, P.V., Nègre, N., White, K.P., and Park, P.J. (2011). ChIP-chip versus ChIP-seq: lessons for experimental design and data analysis. BMC Genomics *12*, 134. https://doi.org/10.1186/1471-2164-12-134.
33. Shu, J., Chen, C., Thapa, R.K., Bian, S.M., Nguyen, V., Yu, K.F., Yuan, Z.C., Liu, J., Kohalmi, S.E., Li, C.L., and Cui, Y.H. (2019). Genome-wide occupancy of histone H3K27 methyltransferases CURLY LEAF and SWINGER in Arabidopsis seedlings. Plant Direct *3*, e00100. https://doi.org/10.1002/pld3.100.
34. Jiang, N., Emberly, E., Cuvier, O., and Hart, C.M. (2009). Genome-wide mapping of boundary element-associated factor (BEAF) binding sites in Drosophila melanogaster links BEAF to transcription. Mol. Cell. Biol. *29*, 3556–3568. https://doi.org/10.1128/MCB.01748-08.
35. Bartkuhn, M., Straub, T., Herold, M., Herrmann, M., Rathke, C., Saumweber, H., Gilfillan, G.D., Becker, P.B., and Renkawitz, R. (2009). Active promoters and insulators are marked by the centrosomal protein 190. EMBO J. *28*, 877–888. https://doi.org/10.1038/emboj.2009.34.
36. Makrantoni, V., and Marston, A.L. (2018). Cohesin and chromosome segregation. Curr. Biol. *28*, R688–R693. https://doi.org/10.1016/j.cub.2018.05.019.
37. Kim, Y., Shi, Z., Zhang, H., Finkelstein, I.J., and Yu, H. (2019). Human cohesin compacts DNA by loop extrusion. Science *366*, 1345–1349. https://doi.org/10.1126/science.aaz4475.
38. Nasmyth, K., and Haering, C.H. (2009). Cohesin: its roles and mechanisms. Annu. Rev. Genet. *43*, 525–558. https://doi.org/10.1146/annurev-genet-102108-134233.
39. Yuan, L., Yang, X.H., Ellis, J.L., Fisher, N.M., and Makaroff, C.A. (2012). The Arabidopsis SYN3 cohesin protein is important for early meiotic events. Plant J. *71*, 147–160. https://doi.org/10.1111/j.1365-313X.2012.04979.x.
40. Zhang, Y., Ma, M., Liu, M., Sun, A., Zheng, X., Liu, K., Yin, C., Li, C., Jiang, C., Tu, X., and Fang, Y. (2023). Histone H2A monoubiquitination marks are targeted to specific sites by cohesin subunits in Arabidopsis. Nat. Commun. *14*, 1209. https://doi.org/10.1038/s41467-023-36788-3.
41. Li, Z., Fu, X., Wang, Y., Liu, R., and He, Y. (2018). Polycomb-mediated gene silencing by the BAH-EMF1 complex in plants. Nat. Genet. *50*, 1254–1261. https://doi.org/10.1038/s41588-018-0190-0.
42. Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., et al. (2003). Genome-wide insertional mutagenesis of Arabidopsis thaliana. Science *301*, 653–657. https://doi.org/10.1126/science.1086391.
43. Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., Fonseca, N.A., Huber, W., Haering, C.H., Mirny, L., and Spitz, F. (2017). Two independent modes of chromatin organization revealed by cohesin removal. Nature *551*, 51–56. https://doi.org/10.1038/nature24281.
44. Wutz, G., Várnai, C., Nagasaka, K., Cisneros, D.A., Stocsits, R.R., Tang, W., Schoenfelder, S., Jessberger, G., Muhar, M., Hossain, M.J., et al. (2017). Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. EMBO J. *36*, 3573–3599. https://doi.org/10.15252/embj.201798004.
45. Niu, L., Shen, W., Shi, Z., Tan, Y., He, N., Wan, J., Sun, J., Zhang, Y., Huang, Y., Wang, W., et al. (2021). Three-dimensional folding dynamics of the Xenopus tropicalis genome. Nat. Genet. *53*, 1075–1087. https://doi.org/10.1038/s41588-021-00878-z.
46. Kaushal, A., Mohana, G., Dorier, J., Özdemir, I., Omer, A., Cousin, P., Semenova, A., Taschner, M., Dergai, O., Marzetta, F., et al. (2021). CTCF loss has limited effects on global genome architecture in Drosophila despite critical regulatory functions. Nat. Commun. *12*, 1011. https://doi.org/10.1038/s41467-021-21366-2.
47. Ramírez, F., Bhardwaj, V., Arrigoni, L., Lam, K.C., Grüning, B.A., Villaveces, J., Habermann, B., Akhtar, A., and Manke, T. (2018). High-resolution TADs reveal DNA sequences underlying genome organization in flies. Nat. Commun. *9*, 189. https://doi.org/10.1038/s41467-017-02525-w.
48. Chathoth, K.T., Mikheeva, L.A., Crevel, G., Wolfe, J.C., Hunter, I., Beckett-Doyle, S., Cotterill, S., Dai, H.S., Harrison, A., and Zabet, N.R. (2022). The role of insulators and transcription in 3D chromatin organization of flies. Genome Res. *32*, 682–698. https://doi.org/10.1101/gr.275809.121.
49. Yang, T., Wang, D., Tian, G., Sun, L., Yang, M., Yin, X., Xiao, J., Sheng, Y., Zhu, D., He, H., and Zhou, Y. (2022). Chromatin remodeling complexes regulate genome architecture in Arabidopsis. Plant Cell *34*, 2638–2651. https://doi.org/10.1093/plcell/koac117.
50. Moon, Y.H., Chen, L., Pan, R.L., Chang, H.S., Zhu, T., Maffeo, D.M., and Sung, Z.R. (2003). EMF genes maintain vegetative development by repressing the flower program in Arabidopsis. Plant Cell *15*, 681–693. https://doi.org/10.1105/tpc.007831.
51. Chelysheva, L., Diallo, S., Vezon, D., Gendrot, G., Vrielynck, N., Belcram, K., Rocques, N., Márquez-Lema, A., Bhatt, A.M., Horlow, C., et al. (2005). AtREC8 and AtSCC3 are essential to the monopolar orientation of the kinetochores during meiosis. J Cell Sci *118*, 4621–4632. https://doi.org/10.1242/jcs.02583.
52. van der Weide, R.H., van den Brand, T., Haarhuis, J.H.I., Teunissen, H., Rowland, B.D., and de Wit, E. (2021). Hi-C analyses with GENOVA: a case study with cohesin variants. NAR Genom. Bioinform. *3*, lqab040. https://doi.org/10.1093/nargab/lqab040.

53. Zhu, L.J., Gazin, C., Lawson, N.D., Pagès, H., Lin, S.M., Lapointe, D.S., and Green, M.R. (2010). ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. BMC Bioinformatics *11*, 237. https://doi.org/10.1186/1471-2105-11-237.

54. Wu, T.Z., Hu, E.Q., Xu, S.B., Chen, M.J., Guo, P.F., Dai, Z.H., Feng, T.Z., Zhou, L., Tang, W.L., Zhan, L., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation *2*, 100141. https://doi.org/10.1016/j.xinn.2021.100141.

55. Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics *34*, i884–i890. https://doi.org/10.1093/bioinformatics/bty560.

56. Cheng, C.Y., Krishnakumar, V., Chan, A.P., Thibaud-Nissen, F., Schobel, S., and Town, C.D. (2017). Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. Plant J. *89*, 789–804. https://doi.org/10.1111/tpj.13415.

57. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

58. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. https://doi.org/10.1038/nmeth.1923.

59. Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44*, W160–W165. https://doi.org/10.1093/nar/gkw257.

60. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26. https://doi.org/10.1038/nbt.1754.

61. Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.J., Vert, J.P., Heard, E., Dekker, J., and Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol. *16*, 259. https://doi.org/10.1186/s13059-015-0831-x.

62. Wolff, J., Bhardwaj, V., Nothjunge, S., Richard, G., Renschler, G., Gilsbach, R., Manke, T., Backofen, R., Ramírez, F., and Grüning, B.A. (2018). Galaxy HiCExplorer: a web server for reproducible Hi-C data analysis, quality control and visualization. Nucleic Acids Res. *46*, W11–W16. https://doi.org/10.1093/nar/gky504.

63. Flyamer, I.M., Illingworth, R.S., and Bickmore, W.A. (2020). Coolpup.py: versatile pile-up analysis of Hi-C data. Bioinformatics *36*, 2980–2985. https://doi.org/10.1093/bioinformatics/btaa073.

64. Takemata, N., and Bell, S.D. (2021). Multi-scale architecture of archaeal chromosomes. Mol. Cell *81*, 473–487.e6. https://doi.org/10.1016/j.molcel.2020.12.001.

65. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.

66. Durand, N.C., Robinson, J.T., Shamim, M.S., Machol, I., Mesirov, J.P., Lander, E.S., and Aiden, E.L. (2016). Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. Cell Syst. *3*, 99–101. https://doi.org/10.1016/j.cels.2015.07.012.

67. Zhou, Y., Wang, Y.J., Krause, K., Yang, T.T., Dongus, J.A., Zhang, Y.J., and Turck, F. (2018). Telobox motifs recruit CLF/SWN-PRC2 for H3K27me3 deposition via TRB factors in Arabidopsis. Nat. Genet. *50*, 638–644. https://doi.org/10.1038/s41588-018-0109-9.

68. Zhou, Y., Tergemina, E., Cui, H., Förderer, A., Hartwig, B., Velikkakam James, G., Schneeberger, K., and Turck, F. (2017). Ctf4-related protein recruits LHP1-PRC2 to maintain H3K27me3 levels in dividing cells in Arabidopsis thaliana. Proc. Natl. Acad. Sci. USA *114*, 4833–4838. https://doi.org/10.1073/pnas.1620955114.

69. Greenwald, W.W., Li, H., Smith, E.N., Benaglio, P., Nariai, N., and Frazer, K.A. (2017). Pgltools: a genomic arithmetic tool suite for manipulation of Hi-C peak and other chromatin interaction data. BMC Bioinformatics *18*, 207. https://doi.org/10.1186/s12859-017-1621-0.

70. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25. https://doi.org/10.1186/gb-2009-10-3-r25.

71. Sun, L.H., Jing, Y.Q., Liu, X.Y., Li, Q., Xue, Z.H., Cheng, Z.K., Wang, D.W., He, H., and Qian, W.Q. (2020). Heat stress-induced transposon activation correlates with 3D chromatin organization rearrangement in Arabidopsis. Nat. Commun. *11*, 1886. https://doi.org/10.1038/s41467-020-15809-5.

72. Abdennur, N., and Mirny, L.A. (2020). Cooler: scalable storage for Hi-C data and other genomically labeled arrays. Bioinformatics *36*, 311–316. https://doi.org/10.1093/bioinformatics/btz540.

73. Matthey-Doret, C., Baudry, L., Breuer, A., Montagne, R., Guiglielmoni, N., Scolari, V., Jean, E., Campeas, A., Chanut, P.H., Oriol, E., et al. (2020). Computer vision for pattern detection in chromosome contact maps. Nat. Commun. *11*, 5795. https://doi.org/10.1038/s41467-020-19562-7.

Please cite this article in press as: Shu et al., EMF1 functions as a 3D chromatin modulator in Arabidopsis, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.10.031

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-GFP antibody	Abcam	Cat#ab290; RRID:AB_303395
Anti-HA antibody produced in rabbit	Sigma-Aldrich	Cat# H6908; RRID: AB_260070
Monocloneal ANTI-FLAG M2 ANTIBODY	Sigma-Aldrich	Cat# F3165; RRID: AB_259529
ChIPAb+ Trimethyl-Histone H3 (Lys4)	Sigma-Aldrich	Cat# 17-614
Myc-Tag (19C2) mAb	Abmart	Cat# M20002; RRID:AB_2861172
Anti-trimethyl-Histone H3 (Lys27) Antibody	Sigma-Aldrich	Cat# 07-449
Anti-Histone H3 (di methyl K9) antibody	Abcam	Cat# ab1220; RRID: AB_449854
Goat Anti-Mouse IgG (H&L)-HRP	EASYBIO	Cat# BE0102; RRID: AB_2923205
Goat Anti-Rabbit IgG (H&L)-HRP	EASYBIO	Cat# BE0101; RRID: AB_3083002
Bacterial and virus strains		
Trans10 Chemically Competent Cell	TransGen Biotech	Cat# CD101-02
GV3101	Beyotime	Cat# D0392
Chemicals, peptides, and recombinant proteins		
Hygromycin B	Phytotech	Cat# H370
Basta	YuanYe	Cat# S18166
Protease Inhibitor Cocktail	Sigma-Aldrich	Cat# P9599
Silwet-77	YuanYe	Cat# S30920
MG-132	Sigma-Aldrich	Cat# M7449
Acetosyringone	Sigma-Aldrich	Cat# D134406
Formaldehyde solution	Sigma-Aldrich	Cat# F1635
Dynabeads™ MyOne™ Streptavidin C1	ThermoFisher	Cat# 65001
Biotin-14-dCTP	ThermoFisher	Cat# 19518018
Dpn II	NEB	R0543S
T4 ligase	ThermoFisher	Cat# EL0011
Klenow fragment	ThermoFisher	Cat# EP0051
PMSF	Alfa Aesar	Cat# B22146
MES	Amresco	CAT# E169-100G
rProtein A Sepharose Fast Flow	Cytiva	Cat#17127901
VAHTS DNA Clean Beads	Vazyme	Cat# N411-01
ANTI-FLAG M2 magnetic beads	Sigma-Aldrich	Cat# M8823
Gel Extraction Kit	OMEGA	Cat# D2500
Cycle Pure Kit	OMEGA	Cat# D6492
Plasmid Mini Kit I	OMEGA	Cat# D6943
Plant RNA Kit	QMEGA	Cat# R6827-01
Equalbit 1 × dsDNA HS Assay Kit	Vazyme	Cat#EQ121-01
VAHTS Universal V8 RNA-seq Library Prep Kit for MGI	Vazyme	Cat# NDM605-01
VAHTS Universal Pro DNA Library Prep Kit for MGI	Vazyme	Cat# NDM608-01
Deposited data		
Raw data	This paper	PRJEB60957
Hi-C in Col-0	Yin et al., [24]	PRJEB52473

(*Continued on next page*)

Please cite this article in press as: Shu et al., EMF1 functions as a 3D chromatin modulator in Arabidopsis, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.10.031

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
ChIP-seq of H3K4me3 and RNA-seq in Col-0	Yang et al.[49]	PRJNA780072
ChIP-seq of H3K27me3 and RNA-seq in Col-0 and *emf1*	Yin et al.[30]	PRJNA649545
Experimental models: Organisms/strains		
Arabidopsis emf1	Moon et al.[50]	N/A
Arabidopsis scc3-2	Chelysheva et al.[51]	N/A
Arabidopsis ProEMF1::EMF1-YFP/emf1	This paper	N/A
Arabidopsis ProSCC3::SCC3-GFP/scc3-2	This paper	N/A
Arabidopsis ProEMF1::EMF1-FLAG	Li et al.[41]	N/A
Oligonucleotides		
Primers are listed in Data S3	This paper	N/A
Recombinant DNA		
pER8-ProEMF1::EMF1-YFP	This paper	N/A
pCAMBIA1300-ProSCC3::SCC3-GFP	This paper	N/A
pGADT7-SCC3	This paper	N/A
pGADT7-ZmSCC3	This paper	N/A
pGBKT7-EMF1	This paper	N/A
pGBKT7-ZmEMF1	This paper	N/A
pDONR421-SCC3	This paper	N/A
pDONR421-EMF1	This paper	N/A
pCL112-nYFP-EMF1	This paper	N/A
pCL113-cYFP-SCC3	This paper	N/A
p35S-3HA-SCC3	This paper	N/A
pDEST-5Myc-EMF1	This paper	N/A
pDEST-GST-SCC3	This paper	N/A
pDEST-GST	This paper	N/A
Software and algorithms		
GENOVA	van der Weide et al.[52]	N/A
ChIPpeakAnno	Zhu et al.[53]	N/A
clusterProfiler	Wu et al.[54]	N/A
ggpubr	https://github.com/kassambara/ggpubr	N/A
Trim Galore	https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/	N/A
Fastp	Chen et al.[55]	N/A
STAR	Cheng et al.[56]	N/A
SAMtools	Li et al.[57]	N/A
Bowtie	Langmead et al.[58]	N/A
Picard	http://broadinstitute.github.io/picard/	N/A
deepTools	Ramirez et al.[59]	N/A
Integrative Genomics Viewer	Robinson et al.[60]	N/A
HiC-Pro	Servant et al.[61]	N/A
HiCExplorer	Wolff et al.[62]	N/A
HiCdatR	https://github.com/MWSchmid/HiCdat	N/A
pyGenomeTracks	https://github.com/deeptools/pyGenomeTracks/	N/A
seqplots	http://przemol.github.io/seqplots/	N/A
coolpup.py	Flyamer et al.[63]	N/A
hic2cool	https://github.com/4dn-dcic/hic2cool	N/A

(*Continued on next page*)

Please cite this article in press as: Shu et al., EMF1 functions as a 3D chromatin modulator in Arabidopsis, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.10.031

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chromosight	Takemata et al.[64]	N/A
Bedtools	Quinlan et al.[65]	N/A
JuiceBox	Durand et al.[66]	N/A
Other		
PVDF-Western Blotting Membranes	Roche Diagnostics	03010040001

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Plant materials and growth conditions

The wild-type plants used in this study include the Columbia ecotype (Col) and Wassilewskija (Ws) of Arabidopsis thaliana. Several mutants and transgenic lines used in this study were described previously: *emf1-2*,[50] *scc3-2*[42] and *ProEMF1::EMF1-FLAG*.[41] Generation of plant materials was described in the method details.

The seeds were surface-sterilized by soaking in 10% (v/v) NaClO for around 10 min and washing in sterile distilled water for at least four times. After surface-sterilization, the seeds were plated on MS medium (2.22 g/L MS, 1% (w/v) sucrose, 0.7% (w/v) agar, pH 5.8), and cold-treated at 4 C for 2 to 3 days in the dark. Arabidopsis seedlings were grown under long-day conditions (16 h light/8 h dark) at 22°C on MS for 10 days. Adult Arabidopsis plants were grown under a 16 h/8 h day/night photoperiod at 22 C.

METHOD DETAILS

ChIP-seq

For EMF1, a 8,367-bp genomic DNA fragment containing 4,579 bp upstream of the *EMF1* start codon and lacking the stop codon was PCR-amplified using specific primers (Data S3) and recombined into the PXCG-YFP vector. For SCC3, a 5,822-bp genomic DNA fragment containing 449 bp upstream of the *SCC3* start codon and lacking the stop codon was PCR-amplified using specific primers (Data S3) and recombined into the 1300-GFP vector. ChIP-seq experiments were performed as previously described[67] using 10-day-old seedlings. Whole seedlings were fixed in 1% formaldehyde, ground in liquid N_2 and resuspended with 30 mL NIB(50 mM HEPES, pH 8.0, 5% sucrose, 5 mM $MgCl_2$, 5 mM NaCl, 40% (v/v) Glycerol, 0.25% Triton X-100, 0.1 mM Phenylmethylsulfonyl Fluoride (PMSF), 0.1% (v/v) 2-mercaptoethanol (2-ME)). The nuclei were isolated by filtering the slurry with double-layered miracloth. Chromatin was sheared into 200- to 500-bp fragments by sonication in TE-SDS buffer(10 mM Tris-HCl, pH 7.4, 1 mM EDTA, 0.25% SDS). Anti-H3K4me3 (Millipore, 17-614), anti-H3K27me3 (Millipore, 07-449), anti-H3K9me2 (Abcam, ab1220), and anti-GFP (Abcam, ab290) antibodies were used for chromatin immunoprecipitation in IP buffer(80 mM Tris-HCl, pH7.4, 230 mM NaCl, 1.7% NP40, 0.17% DOC). The chromatin was captured with rProtein A Sepharose Fast Flow for antibody recovery. After washing the Protein-A beads five times and decrosslinking at 65°C overnight, DNA was extracted with phenol-chloroform followed by ethanol precipitation. For ChIP-seq, two immunoprecipitations from independent biological replicates were processed for next-generation sequencing library preparation. All libraries were prepared with end repair, A-tailing, and ligation of Illumina-compatible adapters using the VAHTS® Universal Pro DNA Library Prep Kit for Illumina® (Vazyme, ND608). The ligated product was amplified with 14 cycles of PCR. DNA fragments of 200-600 bp in size were selected using VAHTS™ DNA Clean Beads (Vazyme, N411). Libraries were sequenced on the Illumina NovaSeq PE150 platform by generating 2 × 150-bp paired-end reads.

In situ Hi-C

In situ Hi-C was performed using 10-day-old seedlings. Tissue fixation, nuclei isolation, chromatin digestion, ligation, and DNA purification were performed as previously described.[22] Whole seedlings were fixed in 1% formaldehyde, ground in liquid N_2 and resuspended with 30 mL NIB(20 mM HEPES, pH 8.0, 250 mM sucrose, 1 mM Magnesium Chloride (MgCl2), 5 mM Potassium Chloride (KCl), 40% (v/v) Glycerol, 0.25% Triton X-100, 0.1 mM Phenylmethylsulfonyl Fluoride (PMSF), 0.1% (v/v) 2-mercaptoethanol (2-ME)). The nuclei were isolated by filtering the slurry with double-layered miracloth. A typical in situ Hi-C library preparation requires 10^7-10^8 nuclei. The pellet was resuspended with 10 x RE(1 M NaCl, 500 mM Tris-HCl, pH 7.9, 100 mM $MgCl_2$, 10 mM DTT) and 150 μL 0.5% SDS successively. The DpnII (New England Biolabs, R0543S) was used as the restriction enzyme for chromatin digestion at 37°C overnight. On the next day, digested DNA was blunt-ended by filling nucleotides with Klenow Fragment(40U) during which biotin-14-dCTP (Invitrogen) was incorporated. Next, the DNA ligation was carried out at 16°C 4 hr and 25°C 1 hr, with each tube having 50 Weiss units of T4 DNA ligase. The crosslinks were reversed by 65°C-overnight incubation. DNA was extracted with phenol-chloroform followed by ethanol precipitation and sheared into 200- to 600-bp fragments by sonication in Tris-HCl pH 8.0. DNA fragments of 200-600 bp in size were selected using VAHTS™ DNA Clean Beads (Vazyme, N411). The total volume was added to 100 μL. The sample was mixed with 60 μL VAHTS™ DNA Clean Beads and incubated. The supernatant was collected and incubated with 40 μl VAHTS™ DNA Clean Beads again. After cleaning the beads with 80% alcohol, the DNA was eluted with Tris- HCl (pH 8.0).

DNA libraries were constructed and amplified using the VAHTS® Universal Pro DNA Library Prep Kit for Illumina® (Vazyme, ND608). Dynabeads™ MyOne™ Streptavidin C1 beads (Invitrogen, 65001) were used to enrich an aliquot of proximally ligated DNA. Two independent biological replicates were processed for next-generation sequencing library preparation. Libraries were purified using VAHTS™ DNA Clean Beads (Vazyme, N411). Libraries were sequenced on the Illumina NovaSeq PE150 platform by generating 2 × 150-bp paired-end reads.

RNA-seq

Total RNA was extracted using the E.Z.N.A.® Plant RNA Kit (Omega Bio-Tek, R6827-01) from 10-day-old whole seedlings. Three independent biological replicates were processed for next-generation sequencing library preparation. Libraries were prepared according to the manufacturer's instructions for the VAHTS™ Total RNA-seq (H/M/R) Library Prep Kit for Illumina® (Vazyme, NR605). Libraries were sequenced on the Illumina NovaSeq PE150 platform by generating 2 × 150-bp paired-end reads.

Yeast two-hybrid

Coding sequences for SCC3, ZmSCC3, ZmEMF1 and EMF1 or different domains of SCC3 were PCR-amplified using specific primers (Data S3) and cloned into the yeast two-hybrid vectors, pGADT7 and pGBKT7. Yeast two-hybrid assays were performed according to the Matchmaker™ GAL4 Two-Hybrid System manual (Clontech). SD/-Leu -Trp and SD/-Leu -Trp -His plates were used for selection.

BiFC

Coding sequences for SCC3 and EMF1 were PCR-amplified using specific primers (Data S3) and recombined into the split-YFP binary vectors, p112-sYFPn and p113-sYFPc. BiFC assays were conducted as previously described.[68] *Agrobacterium tumefaciens* strains carrying plasmids were grown overnight at 28 °C in a selective medium. The strains were collected by centrifugation and resuspended in an infiltration medium (1 mM MgCl2, 150 μM acetosyringone, and 1 mM MES-KOH, pH 5.6). Different experiment combinations were co-infiltrated *N. benthamiana* using a 1 mL needleless syringe. The relevant proteins were transiently expressed in tobacco leaves. YFP fluorescence was imaged two days after infiltration and visualized using a ZEISS LSM 800 confocal laser-scanning microscope.

Co-IP

Coding sequences for target proteins SCC3 were fused to HA and EMF1 were fused to FLAG and inserted into the pAM vector. Co-IP assays were performed as previously described.[41] In brief, the leaves were shredded, vacuumed for 15 min and incubated for 3-4 hr in 30 ml of working solution (0.4 g Cellulose R10, 0.09 g Macerozyme R10, 15 mL 0.8 M Mannitol, 0.6 mL 1 M KCl, 0.6 mL 1M MES, pH 5.7, 0.3 mL 1 M $CaCl_2$, 0.03 g BSA). The protoplasts were isolated by filtering the slurry with double-layered miracloth. After washing with W5 solution(154 mM NaCl, 125 mM $CaCl_2$, 5 mM KCl, 2 mM MES pH 5.7) twice, the pellet was resuspended in 2 mL MMG(0.4 M Mannitol, 15 mM $MgCl_2$, 4 μM MES pH 5.7), added 80 μg plasmids and 2.2 mL PEG solution(40% PEG, 0.2 M Mannitol, 0.1 M $CaCl_2$) incubating for 12 min. A typical Co-IP preparation requires 10^6 protoplasts. After resuspending the pellet with 10 mL W5 solution, the relevant proteins were expressed in Col-0 or *EMF1-FLAG* protoplasts overnight. The protoplasts were harvested and lysed in lysis buffer (50 mM Tris-HCl, pH 7.5; 150 mM KCl; 2 mM EDTA; 0.1% Triton X-100; 10% glycerol; and 5 mM DTT) with freshly added Protease Inhibitor Cocktail (Sigma-Aldrich, P9599) and MG-132 (Sigma-Aldrich, M7449). Total protein was extracted and incubated for 2 hr using Anti-FLAG M2 Magnetic Beads (Sigma-Aldrich, M8823) from Col-0 or *EMF1-FLAG* protoplasts expressing SCC3-HA. The beads were washed three times with lysis buffer, diluted in 2× SDS loading buffer, boiled for 5 min, subjected to SDS-PAGE, and then transferred to PVDF membrane. Membranes were probed with anti-HA (Sigma-Aldrich, H6908) and anti-FLAG (Sigma-Aldrich, F3165) antibodies.

Pull down

Coding sequences for target proteins SCC3 were fused to GST and EMF1 were fused to Myc and inserted into the *pDEST* vector. *pDEST* was used as negative control for GST expression. Each construct was transformed into *E. coli* BL21 cells and grown in resistant medium at 37°C until the OD600 reached 0.6-0.8. Cultures were induced with 0.3 mM isopropyl β-D-1 thiogalactopyranoside(IPTG) overnight at 16°C. Harvested cells were resuspended in 1x PBST buffer (137 mM NaCl, 10 mM Na_2HPO_4, 2.7 mM KCl, 2 mM KH_2PO_4, 0.1% Tween-20) and sonicated. PMSF was added to the solution to inhibit protein degradation. The disrupted cells were centrifuged and the supernatant containing GST or GST-SCC3 protein was incubated with Glutathione Sepharose 4B (GE Healthcare) beads for 2 hours at 4°C. The beads were washed three times with 1x PBST and transferred to the supernatant containing Myc-EMF1 protein for incubating. The pull-down fractions were analyzed by Western blot using antibodies (Abmart, M20002S) against MYC.

Gene resolution loop analysis

More than 20,000 chromatin loops from the euchromatic chromosome arms have been previously identified.[23] For analytical purposes, these loops were shifted to 1-kb resolution. Pgltools, particularly the *intersect* and *intersect1D* operations, were used to confirm their relationships with EMF1 binding sites and CDs,[69] and GENOVA was used to calculate loop strength.[52] The bioconductor

Please cite this article in press as: Shu et al., EMF1 functions as a 3D chromatin modulator in Arabidopsis, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.10.031

R packages ChIPpeakAnno[53] and clusterProfiler[54] were used to annotate loop anchors. Statistically significant differences in loop strength were assessed using the nonparametric Mann–Whitney–Wilcoxon test in the R package ggpubr.

RNA-seq and ChIP-seq bioinformatics analysis

Trim Galore (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) with flags -pairs and -length 70 (minimum read length after trimming) was used for removal of adapter sequences and quality control of reads. The clean reads were then aligned to the Arabidopsis genome (TAIR10, https://www.arabidopsis.org) using STAR default parameters under the guidance of Araport11 (https://www.araport.org) annotation.[56] Next, the reads were sorted, indexed, and compressed using SAMtools,[57] and those with multiple alignments were discarded due to ambiguity of their true locations.

Fastp[55] was used to remove adaptor sequences and low-quality reads. Alignment of the Arabidopsis clean reads to TAIR10 (http://www.arabidopsis.org/) using Bowtie[70] (v1.2.1.1) resulted in two mismatches. PCR duplicates were marked by MarkDuplicates in Picard (http://broadinstitute.github.io/picard/). Unique mapped reads were retained in SAMtools[57] for downstream analysis. The Bigwig file was generated with the following options: binSize 10, normalizeUsing RPKM, and corresponding effective genome size analysis was performed in deepTools[59] through bamCoverage. For the Arabidopsis samples, the blacklist regions were selected from our previously published study[30] and removed with the options –blackListFileName in bamCoverage and bamCompare in deepTools.[59] Typical loci were plotted using the Integrative Genomics Viewer (IGV) software.[60] To compare the reproducibility between ChIP-seq and ChIP-chip data of EMF1, ChIP-seq data were mapped to the TAIR 8 genome. multiBigwigSummary and plotCorrelation tools in deepTools were used. Binsize was set at 1000, which is also the resolution of Hi-C data. The average score is based on the 1 kilobase size bins, which consecutively cover the entire genome. The signal intensities were presented by RPKM in ChIP-seq and $\log_2$(test/input-DNA) ratios in ChIP-chip.

For the violin plot, first, we employed featureCounts with the parameter -p to tally meta-features such as genes, considering fragments instead of reads. Subsequently, we utilized DESeq2 for the differential expression analysis of these genes. We compared the gene expression levels across various groups, applying stringent criteria: a adjusted P-value less than 0.05 and a fold change greater than 2. Volcano plots are constructed using log2FoldChange and adjusted P-value.

Hi-C sequencing data processing and comparison

The raw Hi-C sequencing reads were filtered by fastp,[55] the adapters were trimmed, and low-quality reads were discarded. Next, the output data were mapped, read pair-filtering was performed, and the ligation validity was controlled using HiC-Pro.[61] First, Bowtie2[58] with the end-to-end method was used to align single clean mates from the read pairs to the TAIR10 reference genome (https://www.arabidopsis.org) in order to avoid linear constraints between read pairs. Additionally, a small number of chimeric read pairs, including those reads spanning junction sites, were saved using the two-step method in HiC-Pro, which eliminates singletons, multiple mappings, and low-quality mapped reads. According to the reference genome sequence and restriction enzymes, only well-mapped reads were retained and assigned to fragments. Next, we constructed our downstream Hi-C contact map by selecting read pairs consisting of two different restriction fragments, while dangling-end and self-circulation pairs were discarded as invalid products. After the removal of read pairs from PCR duplicates, we calculated Hi-C resolution based on the established definition,[11] and generated Hi-C contact maps with fixed-sized bins at different resolutions. We transformed the HiC-Pro contact matrixes and bins into different formats for specific analysis. Hi-C contact matrixes at different resolutions were loaded into HiCExplorer[62] (https://github.com/deeptools/HiCExplorer/) in h5 format using hicTransform with fixed bins, and cooler files were generated through hicTransform.

To compare the different Hi-C samples and biological replicates globally, we used the RCP function in GENOVA[52] to measure the distance-dependent decay of the interaction frequency, and ggplot2 (http://ggplot2.tidyverse.org) was used for decay plot visualization. The interaction decay index (IDE) is defined as the linear-fitting slope of the average probability of chromatin interaction with increasing genomic distance, which can be used as a measure to explain the differences between Hi-C samples in our previously published study.[71] The HiCdatR tool was applied to calculate the compartment score of each replicate to further validate the reproducibility (https://github.com/MWSchmid/HiCdat).

Identification of CDs from Hi-C

All Hi-C datasets were subsampled to the same sequencing depth. Using hicCorrectMatrix in HiCExplorer,[62] the raw matrix was further transformed into a normalized Hi-C contact matrix using the Knight and Ruiz (KR) balancing method. The obs_ exp_ Lieberman method was then used to generate the Hi-C Observed/Expected matrixes through hicTransform. Typical loci were generated by pyGenomeTracks (https://github.com/deeptools/pyGenomeTracks/) based on our Hi-C data, RNA-Seq data, and annotations.

The degree of segregation between the left and right parts of each Hi-C genome bin was defined by the TAD separation score. We used the hicFindTADs tool in HiCExplorer to calculate the TAD separation fraction with the following options: –minDepth 3000 –maxDepth 30000 –step 1000 –thresholdComparison 0.05 –delta 0.01 –correctForMultipleTesting fdr at 1 kb resolution using Hi-C. The metagene plot and associated heatmap of the TAD separation scores were generated using seqplots (http://przemol.github.io/seqplots/). Next, local rescaled pileup plots normalized by local backgrounds were generated by coolpup.py[14] (https://github.com/Phlya/coolpuppy/) and GENOVA.[52] The hicInterIntraTAD tool in HiCExplorer was applied to quantified the contacts within CDs.

Please cite this article in press as: Shu et al., EMF1 functions as a 3D chromatin modulator in Arabidopsis, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.10.031

Long-range chromatin loop identification, quantitation, and comparison

We identified unique visible long-range *cis* chromatin interactions (loops) from the genome-wide Hi-C map. hic2cool (https://github.com/4dn-dcic/hic2cool) was employed to generate 5-, 10-, and 20-kb resolution Hi-C maps in cooler format. These maps were further normalized to the equilibrium function in cooler.[72] For WT data, maximizing recognition accuracy was our primary goal; therefore, we adopted a higher threshold. Based on the Pearson correlation coefficient,[73] the quantization module Chromosight was used to calculate a metric called the loop score, which is defined as the similarity between the detected patterns and the loop template. This metric has also been used as a measure of loop strength in other studies[64] in addition to the present study. To identify the most stringent loops, we retained only those with loop scores >0.45, >0.4, and >0.35 at a 20-kb, 10-kb, or 5-kb resolution, respectively. Based on our practical experience, we observed that the patterns around the pericentromeric and centromeric regions were unique relative to the loops from the chromosome arm, and these regions had relatively poor localization quality; therefore, we used pairToBed from bedtools[65] to remove all loops detected in the heterochromatin region. Next, all remaining loops identified at the three different resolutions were merged into a union set using hicMergeLoops in HiCExplorer[62] with the option –lowestResolution 20000. We also used the hic format in JuiceBox[66] to manually check the identified chromatin loops directly. Then, coolpup.py was used to perform aggregate peak analysis (APA) in order to evaluate the genome-wide loop in *cis*, which was similar to our previous studies.[71] We also performed APAs for a series of Hi-C maps. To examine the dynamics of each loop, we used an aggregate tornado-plot, which extracts each loop to a certain distance and displays it as a heatmap in GENOVA.[52]