nature portfolio

Peer Review File

Epigenomic and 3D genomic mapping reveals developmental dynamics and subgenomic asymmetry of transcriptional regulatory architecture in allotetraploid cotton

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

This is a review of the manuscript “Epigenomic and 3D genomic mapping reveals developmental dynamics and subgenomic asymmetry of transcriptional regulatory architecture in allotetraploid cotton” submitted by Wang and colleagues to Nature Communications.

The cotton genome provides a special window for studying the transcriptional regulation of two subgenomes and duplicated genes. The author used cotton to study how epigenetics or chromatin structure affects gene expression. This question is very interesting and important. The author utilized 12 major issues/developmental stages of cotton (G. hirsutum J668), one of the main cultivated varieties, to conduct various omics sequencing, including Hi-C, ChIP-Seq (only K4me3 and K27ac), ATAC-Seq, and RNA-Seq. It can be considered as a small-scale data resource or atlas. The relevant data may have a certain value for researchers in the fields of cotton biology and comparative epigenomics. However, there may be significant flaws in experimental design and interpretation.

The main issues are as follows:

1. The data quality of Hi-C is concerning. The resolution of the data is not high (for example, as stated by the author: ultra-high resolution (3 kb) frm L98, etc). As is well known, a resolution of 3 kb is not considered high for Hi-C data. More importantly, Hi-C data should present a cis ratio of at least 50% (as a QC), but I carefully checked the author's appendix table and found that the cis ratio is below 50% for some tissues. These tissues do not seem to have obvious specificity unless strong evidence such as immune staining of active and inactive chromatin is sufficient to prove the specificity of their nucleus, otherwise, the quality of Hi-C data is of concern. The proportion of cis contacts varies greatly among different tissues, with many tissues being around 50% (anther 53%, hypocotyl 53%, two fiber 53%, and stigma 48%, calculated from Supplementary Table 11). Moreover, the author did not provide information on various biological replicates and related quality controls, only the merged information was provided. I can understand that there may indeed be a low proportion of cis contacts for some special cell types. However, the related work provides strong imaging evidence to prove the authenticity of their data. See the work of this Nature (Extended Data Fig. 4e, of https://www.nature.com/articles/s41586-019-1275-3).

2. The data quality of ATAC-Seq is concerning. In general, ATAC-Seq should have a high peak in the TSS region. Unfortunately, I did not observe this phenomenon. I found it difficult to believe that such data is sufficient to draw those complex conclusions. ATAC-Seq relies on effectively extracting nuclei from different tissues, and from this heatmap (Figure S9), it can be seen that the data quality of ATAC-Seq varies greatly among different tissues. ATAC-Seq from some tissues have significantly higher enrichment of ATAC-Seq signals in the gene body (Figure S9 and Figure 2C, 2H, 2J, etc). The number of peaks varies a lot between their replicates based on Supplementary Table 7. Overall, the data quality of ATAC-Seq presented in this work is unstable and has a serious impact on downstream complex analysis.

3. Lack of DNA methylation and H3K27me3 in experimental design:

Since the author wishes to study gene expression regulation from the perspectives of epigenome and three-dimensional genome, why not include information on DNA methylation and H3K27me3, which have been considered key factors in regulating related processes, such as the imprinting phenomenon in human or paramutation?

4. The definition, analysis, and interpretation of chromatin loops.

What is a loop? How to define it as one? The author has identified a large number of loops (1,478,067 in total, ~ 50,000 for each tissue from Supplementary Table 16) with the 3-kb resolution Hi-C maps, but I have doubts about the authenticity of these loops and they are not very reliable (like Figure 5H, 6A-D). It cannot be proven to be accurate, especially from the APA diagram based on Figure 6A. These figures are not sufficient to determine that these are reliable loops. Similarly, a landmark study on 3D genomics in 2014 Rao et al., drew a 1kb resolution Hi-C and identified only ~10000 loops, all of which were clearly visible point-like structures on the map (see also Figure 6 in that study, http://dx.doi.org/10.1016/j.cell.2014.11.021). I am quite concerned about the loop section in this cotton study. Describing them as “loops” will add to confusion in the field and is a misnomer. The interaction pattern seems to be more consistent with compartments. Evidence for this is seen by their checkerboard-like pattern. Additionally, they appear as rectangular interactions, instead of punctate loops in these figures, and even in the APA plots. These patterns in plants have been described by many including by Dong et al., 2017 Mol Plant and Rowley et al 2017 Mol Cell.

5. There is a certain degree of separation between the author's figures and the author's results/conclusions. The presentation of data in the figure does not support the conclusion of the article. It seems that they are talking about different things.

Reviewer #2 (Remarks to the Author):

In this study, Huang et al. generated a comprehensive multi-omics dataset comprising Hi-C, ChIP-Seq, ATAC-Seq, and RNA-seq data from 12 main tissues/developmental stages of allotetraploid cotton. They investigated the dynamics of genome organization and gene expression across tissues and highlighted the subgenomic asymmetry in transcriptional regulation. The researchers constructed an extensive atlas of functional genome elements, some of which were implicated in fiber quality. Overall, this study provided a high-quality omics dataset for establishing the ENCODE project in cotton and conducted a thorough analysis of subgenome-coordinated regulation of gene expression dynamics. While the manuscript is well-organized, I have several suggestions for enhancing its presentation.

The ratio of Differentially Expressed Homoeologous Genes (DEHGs) between different subgenomes appears to be higher than within the same subgenome, as mentioned in lines 136-139. It remains unclear whether this observation is linked to subgenomic divergence. Please provide a detailed explanation of the relationship between them.

The authors categorized biased genes into three groups based on subgenomic divergence across various tissues. Among these, 4069 homoeologous genes displayed opposing subgenome dominance in different tissues, as outlined from lines 154 to 160. These genes seem to lean towards subfunctionalization. It is important to ascertain whether these genes are tissue-specific or contribute to environmental adaptation.

The authors employed the TAU algorithm to identify a set of Tissue-specific Promoter Genes (TPGs) in lines 180-182 and noted that these genes exhibit a higher bias ratio compared to random genes in lines 190-191. While the TAU algorithm is used to identify tissue-specific genes (TSG), they are referred to as TPGs in this study. Clarification is needed regarding the absence of distinct Gene Ontology terms from TPGs in the fiber_10DPA sample. Additionally, Figure 1F compares TPGs with random genes, where the latter are described as non-TPGs in line 191. For clarity, it is advisable to use consistent terminology for ease of comprehension.

In lines 225-226, the authors mentioned a strong signal strength for H3K27ac and H3K4me3 near TPGs. How was normalization handled across different ChIP-Seq samples in this context?

The authors analyzed Transcription Factor Binding Sites (TFBS) in the promoter regions of TPGs and identified certain TFs that may preferentially bind to these regions. How did the authors address the potential impact of varying TPG numbers across different tissues on their findings? Is the number presented in Figure 2F a percentage?

Further detailed analysis is warranted on how functional elements in noncoding genomic regions regulate TPGs, as mentioned in lines 261-262.

The authors discovered that the Dt subgenome exhibited a higher ratio in both dynamic A/B compartments and sub-compartments compared to the At subgenome in lines 293-298. It would be valuable to determine if the overlap region between dynamic A/B compartments and sub-compartments aligns with this finding.

The manuscript stated the identification of a range of 7,196 to 7,855 TAD-like domains across tissues in line 323. However, in Figure 4A, the vertical axis scale ranged from 0 to 50,000 in the box

plot. It is unclear whether this number represents the sum across 12 tissues, and this should be explicitly stated, possibly in the figure legend.

The statement "Approximately 85% of the boundaries were conserved among tissues," in lines 333-334 prompts consideration of whether the approximately 15% tissue-specific boundaries are enriched with TPGs and whether they exhibit a preference for either the A or B compartments.

In Figure 4F, the terms "active," "stable," and "deactivate" require clarification. Additionally, what does the orange line in Figure 4I signify?

The authors noted that the intra-A compartment primarily displayed short-range interactions in line 410. It would be beneficial to explore any significant differences in loop numbers within the A/B compartment and their impact on the results.

Supplementary Figure 23's biological significance needs to be elucidated further.

The authors utilized Figure 6E to depict multi-level 3D spatial structural differences among organizations. It would be valuable to explain the relationship between large-scale spatial structure reorganization and fine spatial structure rewiring.

The term "stigmas" appeared twice in lines 650 and 656. Please ensure consistency in language usage.

References for the TrimGalore tool need to be included in line 698.

"PATERs" should be corrected to "PATREs" in line 723.

The designation of "OCRs" or "ORCs" should be standardized throughout the text, starting from line 235 onwards. Please make the necessary revisions.

Reviewer reports:

Reviewer #1:

The cotton genome provides a special window for studying the transcriptional regulation of two subgenomes and duplicated genes. The author used cotton to study how epigenetics or chromatin structure affects gene expression. This question is very interesting and important. The author utilized 12 major issues/developmental stages of cotton (*G. hirsutum* J668), one of the main cultivated varieties, to conduct various omics sequencing, including Hi-C, ChIP-Seq (only K4me3 and K27ac), ATAC-Seq, and RNA-Seq. It can be considered as a small-scale data resource or atlas. The relevant data may have a certain value for researchers in the fields of cotton biology and comparative epigenomics. However, there may be significant flaws in experimental design and interpretation.

Response: Thank you very much for your comments and suggestions that help us craft a better manuscript. We have revised this manuscript carefully and made a point-to-point response to each of your suggestions below. We wish our responses would highlight the challenges of 3D genome study in plants.

The main issues are as follows:

1. The data quality of Hi-C is concerning. The resolution of the data is not high (for example, as stated by the author: ultra-high resolution (3 kb) frm L98, etc). As is well known, a resolution of 3 kb is not considered high for Hi-C data. More importantly, Hi-C data should present a cis ratio of at least 50% (as a QC), but I carefully checked the author's appendix table and found that the cis ratio is below 50% for some tissues. These tissues do not seem to have obvious specificity unless strong evidence such as immune staining of active and inactive chromatin is sufficient to prove the specificity of their nucleus, otherwise, the quality of Hi-C data is of concern. The proportion of cis contacts varies greatly among different tissues, with many tissues being around 50% (anther 53%, hypocotyl 53%, two fiber 53%, and stigma 48%, calculated from Supplementary Table 11). Moreover, the author did not provide information on various biological replicates and related quality controls, only the merged information was provided. I can

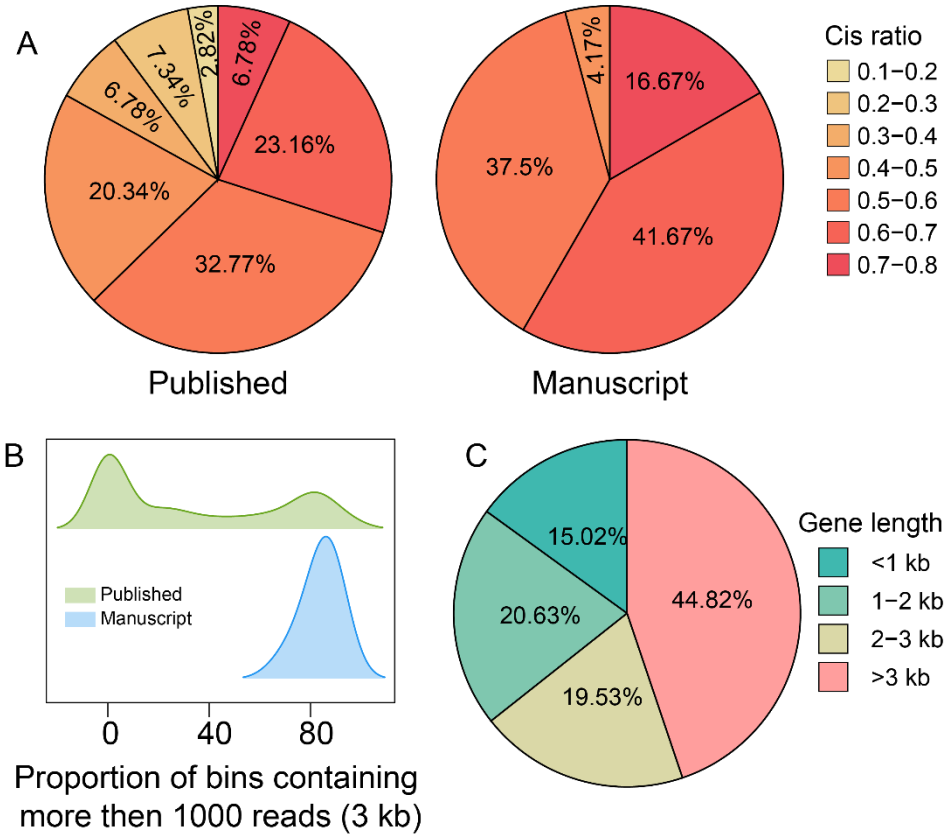
understand that there may indeed be a low proportion of cis contacts for some special cell types. However, the related work provides strong imaging evidence to prove the authenticity of their data. See the work of this Nature (Extended Data Fig. 4e, of https://www.nature.com/articles/s41586-019-1275-3).

Response: We greatly appreciate your valuable comments and insightful suggestions**.** Indeed, considering the three-dimensional genomic datasets from mammals, 3 kb is not considered ultra-high resolution. Hence, the description in the main text has been modified, in lines 97 and 262. Based on your suggestions, we have reconstructed the Hi-C libraries for the anther, stigma and hypocotyl and provided the complete information of two biological replicates for 12 tissues in **Supplementary Table 11**. The cis ratio of replicates from the two fiber samples exceeded 50%, so we did not reconstruct their libraries. We evaluated the consistency of the replicates using GenomeDISCO, HiCRep, HiC-Spector, and QuASAR-Rep, and found that all tissues exhibited high consistency (**Supplementary Figure 13 C**). Moreover, we updated all Hi-C related results in the manuscript, including Figures 3-7, Supplementary Figures 13-27, Supplementary Tables 10-18, and all statistical results related Hi-C in lines 28, 281, 284, 286, 307-308, 319, 329, 342, 367-368, 381, 385-386, 398, 402, 425, 443, 445, 452, 456-459, 463, 473-474, 476, 479, 486, 496, 501-502, 523. In general, updating the Hi-C libraries slightly changed some statistical results, but did not affect the overall conclusion.

Compared to vertebrates, 3D genome research in plants faces widespread challenges due to factors such as recalcitrant cell walls and heterogeneous cell types in tissues, which limited the efficiency and quality of nucleus extraction. We collected and evaluated the quality of all published Hi-C libraries in plants. In evaluating the cis ratio of Hi-C libraries between all published studies in plants, we found that 64.44% of the samples (merged replicate libraries) presented a cis ratio beyond 50% **(Re comments Figure 1A)**. This result indicates that our libraries have generally achieved a satisfactory quality in plants.

Moreover, only 17.61% of the published Hi-C libraries in plants have attained a resolution of 3 kb **(Re comments Figure 1B)**. Currently, a 3 kb resolution of Hi-C

library may not be considered high for animal studies, while achieving the same resolution in plant studies is not easy. Comparing with the published Hi-C libraries in plants, the resolution of datasets generated in this manuscript was satisfactory. Moreover, in *Gossypium hirsutum*, 44.82% of genes exhibited a length greater than 3 kb between the transcription start site (TSS) and transcription end site (TES) **(Re comments Figure 1C)**. The three-dimensional genomic datasets in this study are suited at the gene level, which is the primary focus of this study.



Re comments Figure 1.

A. The pie chart illustrates a comparison of the cis ratios of Hi-C libraries in previously published studies and our manuscript. **B**. The density plot presents a comparison of the resolution of Hi-C libraries. **C**. The pie chart illustrates the distribution of various gene lengths.

2. The data quality of ATAC-Seq is concerning. In general, ATAC-Seq should have a high peak in the TSS region. Unfortunately, I did not observe this phenomenon. I found it difficult to believe that such data is sufficient to draw those complex conclusions.

ATAC-Seq relies on effectively extracting nuclei from different tissues, and from this heatmap (Figure S9), it can be seen that the data quality of ATAC-Seq varies greatly among different tissues. ATAC-Seq from some tissues have significantly higher enrichment of ATAC-Seq signals in the gene body (Figure S9 and Figure 2C, 2H, 2J, etc). The number of peaks varies a lot between their replicates based on Supplementary Table 7. Overall, the data quality of ATAC-Seq presented in this work is unstable and has a serious impact on downstream complex analysis.

Response: Thank you for your valuable suggestions. ATAC-Seq exhibits instability in plants due to the cell wall, which hinders the isolation of pure nuclei. While the signals from our ATAC-Seq library exhibited notable enrichment at the Transcription Start Site (TSS) location **(Supplementary Figure 9C)**, the presence of mixed nuclei could influence the signal intensity of tissues within the TSS location and gene body. In addition, to maximize the accuracy of identifying peak sites in ATAC-Seq data, we implemented a five-step quality filtering process. Initially, we applied a strict parameter setting (-q 20, --stringency 3, --length 20) with Trim_galore to filter out low-quality reads from the ATAC libraries. Next, a stringent threshold of -p 0.005 was set using MACS2 to identify reliable peak sites. Subsequently, false-positive peaks in blacklist regions were removed. Additionally, peak sites consistently present in both replicates with an IDR $\leq$ 0.05 were retained. Furthermore, some reliable peak sites were recalled (Details in methods). Final, the number of peaks sites from ATAC libraries across multiple tissues showed stability. In downstream analyses, these stable peaks can be considered. Moreover, this study primarily focused on investigating the correlation between the presence or absence of open chromatin regions (OCRs) and gene expression, rather than considering the complex dynamics of signal intensity at the OCRs across various tissues. Subsequent analyses focused on the three-dimensions regulatory relationships between genes and CREs. Furthermore, the identification of CREs employed all OCRs from multiple tissues without considering the differences in peak sites across these tissues. In summary, the stable and reliable peak loci are appropriate to support the analyses in this study.

3. Lack of DNA methylation and H3K27me3 in experimental design:

Since the author wishes to study gene expression regulation from the perspectives of epigenome and three-dimensional genome, why not include information on DNA methylation and H3K27me3, which have been considered key factors in regulating related processes, such as the imprinting phenomenon in human or paramutation?

Response: We appreciated very much your valuable suggestions. Absolutely, DNA methylation and H3K27me3 play critical roles in the regulation of a wide array of biological processes. Both are involved in gene silencing and may serve as markers of silencer. In contrast to the fungal infestation experiment where gene activation and repression status changes are crucial to consider between the control and treatment groups, our study focuses solely on identifying active modification sites across various tissue samples. The primary goal of this study is to establish an initial cotton ENCODE. The reason why we only selected H3K4me3 and H3K27ac is that the former is well-known at promoters for transcription regulation, and the latter is well-known as an enhancer marker in animals and humans. As you know, the regulation of related processes involves a multitude of key factors, such as H3K4me1 and H3K4me2 linked to gene transcription, as well as H3K9me2 and H3K9me3, which are enriched in heterochromatin compartments and associated with gene silencing. In the future, we plan to generate more omics-data to enrich the ENCODE project in cotton. The current study focuses on chromatin interactions and active modifications.

4. The definition, analysis, and interpretation of chromatin loops.

What is a loop? How to define it as one? The author has identified a large number of loops (1,478,067 in total, ~ 50,000 for each tissue from Supplementary Table 16) with the 3-kb resolution Hi-C maps, but I have doubts about the authenticity of these loops and they are not very reliable (like Figure 5H, 6A-D). It cannot be proven to be accurate, especially from the APA diagram based on Figure 6A. These figures are not sufficient to determine that these are reliable loops. Similarly, a landmark study on 3D genomics in 2014 Rao et al., drew a 1kb resolution Hi-C and identified only ~10000 loops, all of which were clearly visible point-like structures on the map (see also Figure 6 in that

study, http://dx.doi.org/10.1016/j.cell.2014.11.021). I am quite concerned about the loop section in this cotton study. Describing them as “loops” will add to confusion in the field and is a misnomer. The interaction pattern seems to be more consistent with compartments. Evidence for this is seen by their checkerboard-like pattern. Additionally, they appear as rectangular interactions, instead of punctate loops in these figures, and even in the APA plots. These patterns in plants have been described by many including by Dong et al., 2017 Mol Plant and Rowley et al 2017 Mol Cell.

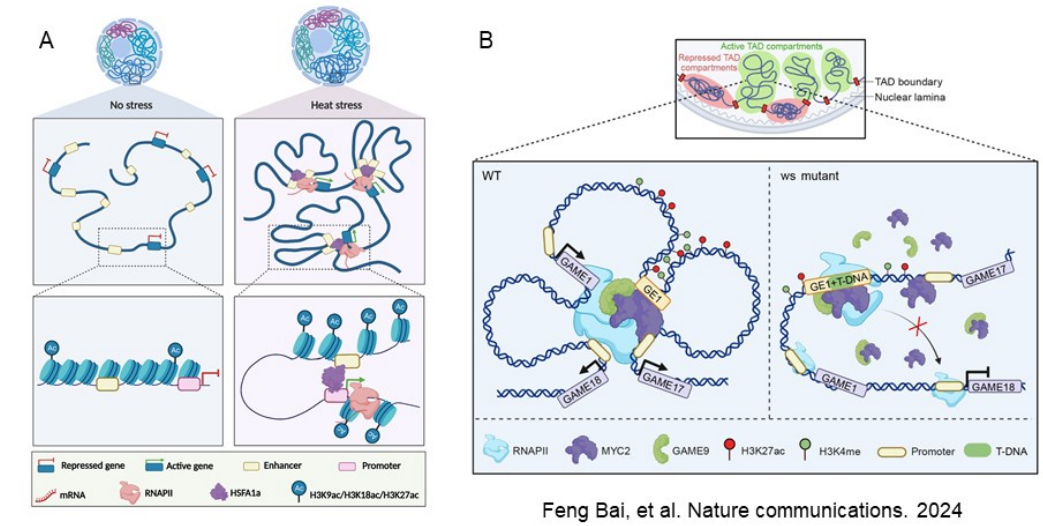
Response: Thank you very much for your constructive feedback and valuable suggestions. Based on your suggestions, we have conducted an extensive review of the background literature related to the three-dimensional genome in plants. The first enhancer-promoter loop in plants was identified in maize in 2009 by chromosome conformation capture (3C) technology[1]. CTCF and the structural maintenance of chromosomes (SMC) proteins, such as cohesin and condensin, are fundamental components in the formation of loop structures in animal and human studies. While animal CTCF homologs are not identified in plants[2]. Recent studies have uncovered that transcription factors could be pivotal in preserving the stability of chromatin loops in plants[3, 4, 5]. For example, the heat shock factor (HSF) transcription factor HSFA1a can promote the formation of promoter-enhancer loops in tomato in response to heat stress **(Re comments Figure 2A)**[3]. Moreover, the GAME enhancer 1 (GE1) can recruit the transcription factor MYC2 and GAME9 complex to from chromatin loops for regulating the expression of GAME cluster in tomato **(Re comments Figure 2B)**[4] . According to overall statistics, about 4–30% of transcription factor binding sites are located at both anchors of loops formed between distal open chromatin regions (dOCRs) and genes in maize[5]. Additionally, some studies also revealed that noncoding transcription of genes and the SWI/SNF chromatin-remodeling complex play crucial roles in the formation of chromatin loops in plants[6, 7]. Although the mechanisms of loop formation differ between mammals and plants, the aforementioned studies do confirm the existence of chromatin loop structures in plants. To avoid conceptual confusion, we decided to keep using the term ‘chromatin loops’ to follow previous studies in plants.

With the advancement of experimental and sequencing technologies, chromatin loop

structures have been discovered in diverse plant species by using Hi-C, ChIA-PET, HiChIP and Capture-C[4, 8, 9, 10, 11, 12, 13], including wheat[8] (293,044, shoot, HOMER, Hi-C), soybean[11] (32,181, leaf, FitHiC, Hi-C) and tomato (3597, fruits at 7 DPA, HICCUPS, Hi-C)[4]. As you know, loops are fine three-dimensional genomic structures whose number is significantly influenced by resolution, tissue/organ type and species. To eliminate the impact of differences in matrix resolution and identification parameters applied in previous studies, we standardized the parameters of software (FitHiC, HiCExplorer and HiCCUPS) and matrix resolution (10 kb) to identify chromatin loops from Hi-C libraries of plants, the results indicate a significant correlation between loop numbers and the software choice **(Re comments Figure 3A)**. To minimize software bias in the results and enhance the accuracy of identifying loop structures, we employed tools Fit-Hi-C[14], Mustache[15], Chromosight[16] and Peakachu[17], selecting only the highest-ranking 25% of loops based on the confidence levels and the overlapped loops identified by four tools for further analysis. Details can be found Methods section.

We agree with you that the article (Rao et al., 2014, Cell) displays clearly visible point-like structures on heatmaps. We had evaluated HICCUPS, the loops identification tools mentioned in the article (Rao et al., 2014, Cell), but it does not appear suitable for cotton **(Re comments Figure 3B)**. Differences exist in the number of identified loops between our study and the article you cited (Rao et al., 2014, Cell), which may result from discrepancies in the loop identification tools utilized.

One problem that should not be overlooked is that, unlike animal cell lines with a highly homogenized cell type, plant tissues contain multiple cell types. As a result, the extracted nuclei may exist in a multicellular mixed state, potentially affecting the visualization of loop interactions. Indeed, the APA figure exhibit the loops anchor is not obvious in the manuscript, we have updated these figures **(Figures 5H, 6A-D)**.

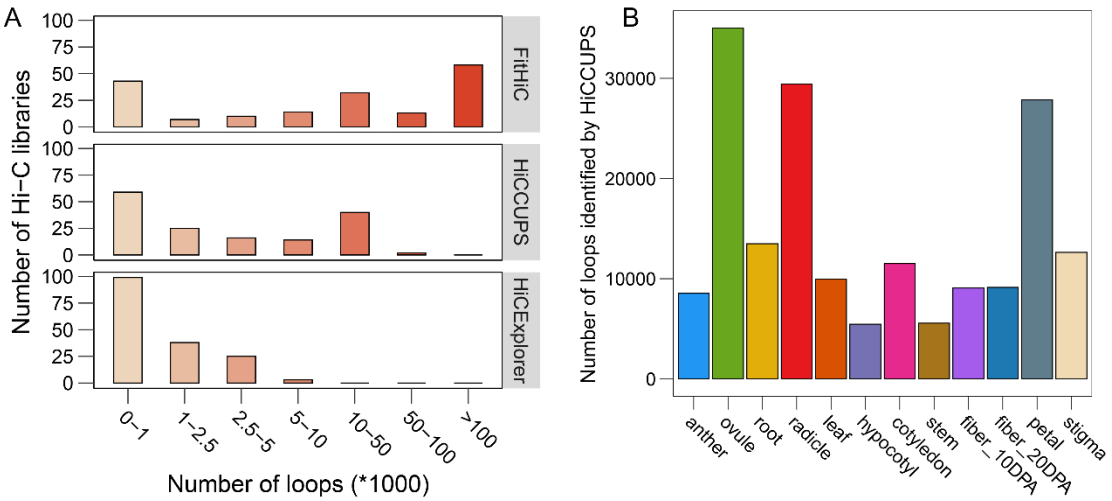


Ying Huang, et al. Nature communications. 2023

Feng Bai, et al. Nature communications. 2024

Re comments Figure 2

A. The pattern diagram illustrates the formation of chromatin loops of tomato under the heat stress. B. The pattern diagram demonstrates that the transcription factor facilitates the formation of chromatin loops in plants.



Re comments Figure 3

A. The bar plot displays the number of Hi-C libraries with different loops identified by FitHiC, HiCCUPS and HiCExplorer. **B**. The bar plot shows the number of loops identified by HiCCUPS.

5. There is a certain degree of separation between the author's figures and the author's results/conclusions. The presentation of data in the figure does not support the

conclusion of the article. It seems that they are talking about different things.

Response: Thanks for your comments. We meticulously reviewed each figure and the related conclusion, providing a more detailed description of the figures and revising conclusions to enhance their readability and facilitate clearer comprehension for the reader. The enhancements to the conclusion can be found on lines 120-121, 123-125, 129-130, 135-138, 156-161, 185, 262, 276, 340, 344-345, 349-353, 400-405, 445-446, 450. The more details of figures are provided in lines 1156-1159, 1163-1166, 1174, 1183-1184, 11890-1191, 1194-1195, 1200, 1202, 1204-1205, 1220, 1223-1226, 1248-1249, 1255-1256, 1263-1265, 1272, 1282-1283, 1299-1301, 1304-1305.

Reviewer #2:

In this study, Huang et al. generated a comprehensive multi-omics dataset comprising Hi-C, ChIP-Seq, ATAC-Seq, and RNA-seq data from 12 main tissues/developmental stages of allotetraploid cotton. They investigated the dynamics of genome organization and gene expression across tissues and highlighted the subgenomic asymmetry in transcriptional regulation. The researchers constructed an extensive atlas of functional genome elements, some of which were implicated in fiber quality. Overall, this study provided a high-quality omics dataset for establishing the ENCODE project in cotton and conducted a thorough analysis of subgenome-coordinated regulation of gene expression dynamics. While the manuscript is well-organized, I have several suggestions for enhancing its presentation.

Response: Thank you very much for your comments and suggestions. We have revised the manuscript carefully. Please see our point-to-point response below.

1. The ratio of Differentially Expressed Homoeologous Genes (DEHGs) between different subgenomes appears to be higher than within the same subgenome, as mentioned in lines 136-139. It remains unclear whether this observation is linked to subgenomic divergence. Please provide a detailed explanation of the relationship

between them.

Response: Thanks for your valuable comments. Homologous genes are two copies of a gene that result in similar functions and expression levels. The number of differentially expressed homologous genes within the same subgroup should be similar to those within different subgroups across tissues. However, in our manuscript, we observed a dissimilarity, implying subgenome divergence. We revised the statements describing the conclusions to enhance the readers' understanding, in lines 129-130, 135-139.

2. The authors categorized biased genes into three groups based on subgenomic divergence across various tissues. Among these, 4069 homoeologous genes displayed opposing subgenome dominance in different tissues, as outlined from lines 154 to 160. These genes seem to lean towards subfunctionalization. It is important to ascertain whether these genes are tissue-specific or contribute to environmental adaptation.

Response: Thank you for your suggestion. We focus on the homoeologous genes with contrary bias. As predicted, noticeable enrichment of TPGs is witnessed on these genes **(Supplementary Figure 5C)**. This finding indicates that the two copies of these genes not only show biased expression towards different sub-genomes in various tissues but also demonstrate a higher degree of tissue specificity. In other words, the divergent tissue specificity of homoeologous gene pairs in different tissues reflects a trend towards subfunctionalization. However, additional experiments are required for further validation of this trend in further studies. Upon further analysis the functions of contrary genes, we observed that the Gene Ontology (GO) terms associated with these genes include response to external stimulus (eg: light, biotic and abiotic, et al). This implies a potential correlation between contrary genes and environmental adaptation **(Supplementary Figure 5D)**. We have added the aforementioned two conclusions in lines 156-161, and updated **Supplementary Figure 5**.

3. The authors employed the TAU algorithm to identify a set of Tissue-specific Promoter Genes (TPGs) in lines 180-182 and noted that these genes exhibit a higher

bias ratio compared to random genes in lines 190-191. While the TAU algorithm is used to identify tissue-specific genes (TSG), they are referred to as TPGs in this study. Clarification is needed regarding the absence of distinct Gene Ontology terms from TPGs in the fiber_10DPA sample. Additionally, Figure 1F compares TPGs with random genes, where the latter are described as non-TPGs in line 191. For clarity, it is advisable to use consistent terminology for ease of comprehension.

Response: Thank you for your suggestion. We meticulously examined the TPGs identified by the TAU algorithm, and found that not all genes are specifically expressed in a single tissue. This phenomenon was also observed in human and mouse. Therefore, we suggest that the terminology "preferential expression" better characterizes these genes. Moreover, the term 'tissue-preferential gene' has been employed in previous study[18]. Indeed, we were unable to find suitable GO terms from the TPGs within fibers at 10 DPA. Given the existence of fiber_20DPA in tissues, our speculation suggests that the inefficacy might be attributed to the TAU algorithm's inadequacy in identifying period-specific genes, which may lead to the insufficient identification of specific genes of fiber_10DPA. To ensure clarity for the reader, we modified "non-TPGs" to "random" in line 185, and provided a detailed description of random sampling in line 1174.

4. In lines 225-226, the authors mentioned a strong signal strength for H3K27ac and H3K4me3 near TPGs. How was normalization handled across different ChIP-Seq samples in this context?

Response: Your valuable comments and insightful suggestions are greatly appreciated. In this section, we compare the H3K27ac/H3K4me3 signals between TPGs of different tissues within a tissue, so data normalization is not needed. To make it clear, we revised the figure legend of **Figure 2D**, in lines 1194-1195. In this manuscript, we utilized the BPM normalization method in deepTools to mitigate the impact of sequencing depth on data signal intensity, and the detail can be found in the Methods, in line 704.

5. The authors analyzed Transcription Factor Binding Sites (TFBS) in the promoter regions of TPGs and identified certain TFs that may preferentially bind to these regions.

How did the authors address the potential impact of varying TPG numbers across different tissues on their findings? Is the number presented in Figure 2F a percentage?

Response: Thanks for your valuable comments. To circumvent the influence of the number of TPGs, the number of annotated TPGs was divided by the total number of TPGs in each respective tissue. We have provided a more detailed description in the figure legend of Figure 2F, in line 1200.

6. Further detailed analysis is warranted on how functional elements in noncoding genomic regions regulate TPGs, as mentioned in lines 261-262.

Response: Thank you for your suggestion. Our statistical analysis revealed that an average of 42.33% of TPGs are linked by tissue-specific loops **(Supplementary Figure 21C)** and a total of 625 TPGs are regulated by putative functional elements through chromatin loops in individual tissues **(Supplementary Figure 21D)**. These results imply a potential relationship between the organization of tissue-specific intricate domains and the function. We have added this analysis and conclusion, in lines 402-405.

7. The authors discovered that the Dt subgenome exhibited a higher ratio in both dynamic A/B compartments and sub-compartments compared to the At subgenome in lines 293-298. It would be valuable to determine if the overlap region between dynamic A/B compartments and sub-compartments aligns with this finding.

Response: Thanks for your valuable suggestions. We conducted an analysis of the consistency between the A/B compartment and sub-compartments in 12 tissues. The results revealed that, on average, 85.50% of chromatin regions exhibit the same compartment status **(Supplementary Figure 14B)**. Then, upon comparing the compartment status of these regions across 12 tissues, we discovered that 35.80% of the regions exhibited difference between the A/B compartment and sub-compartment status in some tissues **(Supplementary Figure 14C)**. Interestingly, regions with different compartment status demonstrated greater dynamics across tissues, particularly in the Dt subgenome **(Supplementary Figure 14C,D)**. Additionally, 94.92% of

dynamic A/B compartment regions overlapped with regions exhibiting different compartment status **(Supplementary Tables 12,13)**. These findings suggest that the divergence regions within both the A/B compartment and the sub-compartment are more susceptible to experiencing compartment alterations across tissues. We added the analysis of the overlap of A/B compartment and sub-compartments lines 290-299, and updated the **Supplementary Figure 14**.

8. The manuscript stated the identification of a range of 7,196 to 7,855 TAD-like domains across tissues in line 323. However, in Figure 4A, the vertical axis scale ranged from 0 to 50,000 in the box plot. It is unclear whether this number represents the sum across 12 tissues, and this should be explicitly stated, possibly in the figure legend.

Response: Thank you for your suggestion. Indeed, the number represents the sum of boundaries of TADs. We have supplemented the detailed characterization using boxplots in **Figure 4a**, in lines 1248-1249.

9. The statement "Approximately 85% of the boundaries were conserved among tissues," in lines 333-334 prompts consideration of whether the approximately 15% tissue-specific boundaries are enriched with TPGs and whether they exhibit a preference for either the A or B compartments.

Response: Thank you for your suggestion. Through an analysis of TPG enrichment and A/B compartment status, we observed a significant enrichment of the TPGs and a lower ratio of A compartment status near tissue-specific boundaries **(Supplementary Figure 17B,C)**. These results further demonstrate the tissue specificity of the spatial chromatin structure. We have added this result in lines 350-353.

10. In Figure 4F, the terms "active," "stable," and "deactivate" require clarification. Additionally, what does the orange line in Figure 4I signify?

Response: Thanks for your question. We have provided the details for "active," "stable," and "deactivate" in Figure 4F, in line 1263-1265. An orange line represents a gene located in isolated TAD. We added a detailed description of these genes in line

1272.

11. The authors noted that the intra-A compartment primarily displayed short-range interactions in line 410. It would be beneficial to explore any significant differences in loop numbers within the A/B compartment and their impact on the results.

Response: Thank you for your suggestion. The proportions of length difference of loop interactions in intra-A, intra-B and inter-AB compartment illustrate that short-range interactions are primarily distributed in intra-A compartment (**Figure 5F and Supplementary Figure 22C**). According to your suggestion, we calculated the ratio of loops of different length in intra-A compartment and intra-B compartment, respectively. The result also shows that the regions within the intra-A compartment are more enriched in short-range interactions compared with the intra-B compartment regions **(Supplementary Figure 22D)**. These showed that more short-range interactions are enriched within the A compartment. In a word, the difference in the number of loops between A and B compartment does not affect the result. We have supplemented the Supplementary Figure 22D, in line 410.

12. Supplementary Figure 23's biological significance needs to be elucidated further.

Response: We identified 59,842 genes connected by imbalance loops. By analyzing the GO terms of these genes, we did not find any enrichment of biological processes. We hypothesize that the reason why these genes do not show biological process enrichment is because they constitute 85.25% of the total genes, thus spanning nearly all biological processes.

13. The authors utilized Figure 6E to depict multi-level 3D spatial structural differences among organizations. It would be valuable to explain the relationship between large-scale spatial structure reorganization and fine spatial structure rewiring.

Response: Thank you for your suggestion. At different resolutions, the spatial organization of chromatin exhibits distinct characteristics, such as A/B compartments, sub-compartments, TADs, and loops. Previous research has indicated that the formation

of TADs is associated with the compartmentalization and the extrusion of loops in mammals[19]. The TAD rearrangement is closely related to enrichment/digestion of CTCF binding, the loops rewiring is associated with cohension[19]. However, the orthologous protein of CTCF has never been found in plants, the current speculation is that transcription factors may maintain the stability of TAD structures in plants. This question cannot be addressed in this study. The nature of the relationship between large-scale spatial structure reorganization and fine spatial structure rewiring requires in-depth exploration and more extensive research in plants.

14. The term "stigmas" appeared twice in lines 650 and 656. Please ensure consistency in language usage.

Response: Thanks for your suggestion. We have modified the term 'stigmas' to 'radicle' in line 647.

15. References for the TrimGalore tool need to be included in line 698.

Response: Thanks for your suggestion. We have added the reference of TrimGalore in line 689.

16. "PATERs" should be corrected to "PATREs" in line 723.

Response: Thanks for your suggestion. We have corrected the ‘PATERs’ to ‘PATREs’ in line 713.

17. The designation of "OCRs" or "ORCs" should be standardized throughout the text, starting from line 235 onwards. Please make the necessary revisions.

Response: Thanks for your comments. The term 'ORCs' has been rectified to 'OCRs' in lines 227, 233, 239, 241, 246, and 252.

References:

1. Louwers M, Bader R, Haring M, van Driel R, de Laat W, Stam M. Tissue- and expression level-specific chromatin looping at maize b1 epialleles. *Plant Cell* **21**, 832-842 (2009).

2. Domb K, Wang N, Hummel G, Liu C. Spatial Features and Functional Implications of Plant 3D Genome Organization. *Annu Rev Plant Biol* **73**, 173-220 (2022).

3. Huang Y*, et al.* HSFA1a modulates plant heat stress responses and alters the 3D chromatin organization of enhancer-promoter interactions. *Nat Commun* **14**, 469 (2023).

4. Bai F*, et al.* A distal enhancer guides the negative selection of toxic glycoalkaloids during tomato domestication. *Nat Commun* **15**, 2894 (2024).

5. Sun Y*, et al.* 3D genome architecture coordinates trans and cis regulation of differentially expressed ear and tassel genes in maize. *Genome Biol* **21**, 143 (2020).

6. Ariel F*, et al.* Noncoding transcription by alternative RNA polymerases dynamically regulates an auxin-driven chromatin loop. *Mol Cell* **55**, 383-396 (2014).

7. Jégu T*, et al.* The BAF60 subunit of the SWI/SNF chromatin-remodeling complex directly controls the formation of a gene loop at FLOWERING LOCUS C in Arabidopsis. *The Plant Cell* **26**, 538-551 (2014).

8. Concia L*, et al.* Wheat chromatin architecture is organized in genome territories and transcription factories. *Genome Biol* **21**, 104 (2020).

9. Liu C, Wang C, Wang G, Becker C, Zaidem M, Weigel D. Genome-wide analysis of chromatin packing in Arabidopsis thaliana at single-gene resolution. *Genome Res* **26**, 1057-1068 (2016).

10. Peng Y*, et al.* Chromatin interaction maps reveal genetic regulation for quantitative traits in maize. *Nat Commun* **10**, 2632 (2019).

11. Wang L, Jia G, Jiang X, Cao S, Chen ZJ, Song Q. Altered chromatin architecture and gene expression during polyploidization and domestication of soybean. *The Plant Cell* **33**, 1430-1446 (2021).

12. Zhao L*, et al.* Chromatin loops associated with active genes and heterochromatin shape rice genome architecture for transcriptional regulation. *Nat Commun* **10**, 3640 (2019).

13. Huang X*, et al.* Multi-omics mapping of chromatin interaction resolves the fine hierarchy of 3D genome in allotetraploid cotton. *Plant Biotechnol J* **20**, 1639-1641 (2022).

14. Ay F, Bailey TL, Noble WS. Statistical confidence estimation for Hi-C data reveals

regulatory chromatin contacts. *Genome Res* **24**, 999-1011 (2014).

15. Roayaei Ardakany A, Gezer HT, Lonardi S, Ay F. Mustache: multi-scale detection of chromatin loops from Hi-C and Micro-C maps using scale-space representation. *Genome Biol* **21**, 256 (2020).

16. Matthey-Doret C*, et al.* Computer vision for pattern detection in chromosome contact maps. *Nat Commun* **11**, 5795 (2020).

17. Salameh TJ*, et al.* A supervised learning framework for chromatin loop detection in genome-wide contact maps. *Nat Commun* **11**, 3428 (2020).

18. Cervantes S, Vuosku J, Pyhajarvi T. Atlas of tissue-specific and tissue-preferential gene expression in ecologically and economically significant conifer Pinus sylvestris. *PeerJ* **9**, e11781 (2021).

19. Beagan JA, Phillips-Cremins JE. On the existence and functionality of topologically associating domains. *Nat Genet* **52**, 8-16 (2020).

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

Thanks for the authors that most concerns were well addressed, especially for that they have greatly improved the data quality by adding more Hi-C sequencing and ATAC-seq data quality filter processing. I only keep my concerns about the loop identification. And also the author did not respond to my questions about the relationships between loops and compartment. The authors compared the effectiveness of different software in identifying loops. However, they just extensively used the results based on loops analysis in the article, with no directly evidence or other supports. I suggest the author to provide direct evidence for that these loops objectively exist, such as graphs (alignments on regions, e.g. tornado-plots) other than APA.

Reviewer #2 (Remarks to the Author):

The authors have adequately addressed all of my previous concerns. The revisions have improved the manuscript, and I now believe it is suitable for publication in Nature Communications.

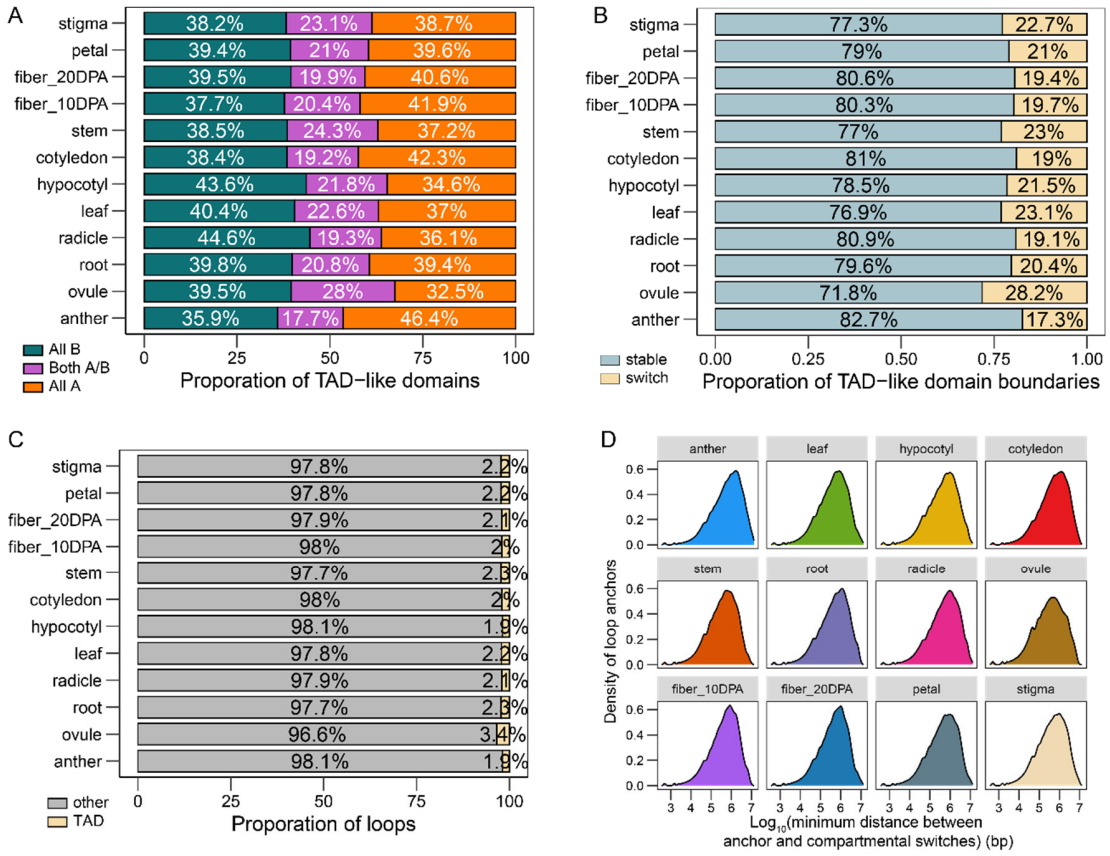
Reviewer reports:

Reviewer #1:

1. Thanks for the authors that most concerns were well addressed, especially for that they have greatly improved the data quality by adding more Hi-C sequencing and ATAC-seq data quality filter processing. I only keep my concerns about the loop identification. And also the author did not respond to my questions about the relationships between loops and compartment. The authors compared the effectiveness of different software in identifying loops. However, they just extensively used the results based on loops analysis in the article, with no directly evidence or other supports. I suggest the author to provide direct evidence for that these loops objectively exist, such as graphs (alignments on regions, e.g. tornado-plots) other than APA.

Response: Thank you very much for your valuable comments. The mechanisms underlying the formation of plant loops remain a topic of ongoing debate, such as compartmentalization and transcription factor binding, among others[1–3]. Further in-depth research may be necessary to reveal the underlying mechanisms, which is beyond the focus of this manuscript. In our manuscript, we performed a simple analysis of the distribution of loops with different lengths in A and B compartments (**Figure 5F, Supplementary Figure 22C and 22D**). In line with your suggestion, we carefully read the articles by Dong et al. (2017, Molecular Plant) and Rowley et al. (2017, Molecular Cell), again. Following the ideas and analysis strategies proposed in these articles, we have deeply explored the relationship between A/B compartment, TAD-like domains and loops. We found that ~21% of TAD-like domains contain more than one compartment switching (**Re comments Figure 1A**), and ~14% of TAD-like domain boundaries overlap with A/B compartment switching regions **(Re comments Figure 1B)**. The A/B compartment is identified in the Hi-C matrix with a resolution of 100 kb. Upon further analysis, we found that only ~2% of loops are located within the TAD-like domains, and boundaries are related to compartment status switching **(Re comments Figure 1C)**. Furthermore, we analyzed the minimum distance between loop anchors and the sites of A/B compartment switching. The results indicate that most loop anchors are situated approximately 1 Mb or more from the compartment switching **(Re comments Figure 1D)**. The above results suggest that compartmentalization may facilitate the segregation of TAD-like domains to some degree. However, it appears to exert a negligible influence on loop formation in cotton. Herein, we propose an immature conjecture that the correlation between A/B compartments and loops observed in the Dong et al. (2017, Molecular Plant) could be due to the

low resolution (10 kb) at which the loops were identified, and in this case a single anchor of loops may be considered a mini-TAD.

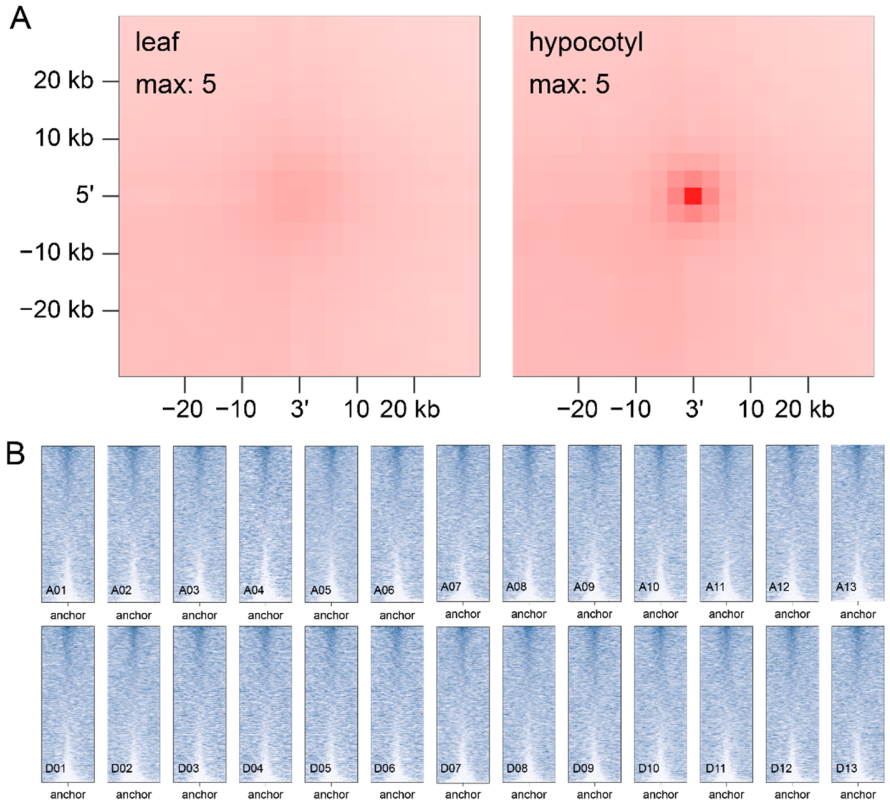


Re comments Figure 1.

A. The bar plot shows the proportion of TAD-like domains with different A/B compartment statues. 'All A' or 'All B' refer to TAD-like domains exhibiting only A or B compartment status, respectively. 'Both A/B' refers to the TAD-like domains contains A/B compartment switching. **B.** The bar plot illustrates the proportion of TAD-like domain boundaries. 'Stable' refers to the boundaries (approximately 200 kb) that do not overlap with A/B compartment switching regions. 'Switch' implies a situation contrary to this. **C.** The bar plot shows the proportion of loops. 'TAD' means the loops located in TAD-like domains related to A/B compartment switching. 'other' means the other loops. **D.** The density plot shows the minimum distance between anchor of loops and sites of A/B compartment switching.

To further confirm the authenticity of chromatin structures identified in this manuscript, we produced genome-wide average features of Hi-C dataset across 12 tissues using the aggregation tools of GENOVA[4]. From the Aggregate TAD Analysis (ATA) and Aggregate Peak Analysis (APA) plots, we clearly observed TAD-like domain boundaries and peak enrichment of loops, respectively **(Supplementary Figure 15A and Supplementary Figure 18A)**. The main text modifications occur in lines 319-321, 324, 326, 386. In addition, we analyzed leaf APA using the loops identified from hypocotyl. The results did not show obvious highlighted pixels in the center (**Re comments Figure 2A**). This result reflects the tissue specificity and non-randomness (authenticity) of the loops identified in this manuscript. **We should note that the heatmaps in Figures 5H and 6A-D actually do not represent the widely known APA plots, but rather depict the interaction frequency within a specific region.**

Moreover, we also calculated the coverage of mapped reads around anchor of loops related to genes and presented the results using tornado-plots[5]. We observed an enrichment of mapping reads of Hi-C at the loop anchor positions **(Re comments Figure 2B)**. Unlike the CTCF-binding loops studied in humans, the loops analyzed here are non-specific loops. Additionally, the heterogeneity of plant tissues may also introduce noise into the results. In human studies, significant CTCF enrichment can be observed at loop anchor points by tornado-plot[6]. However, no homologous proteins have been found in plants. To further demonstrate that the loops identified in this manuscript are non-random (authenticity), we analyzed the enrichment of active histone modification sites at loop anchors. Compared to random anchors, loop anchors show a significant enrichment of H3K4me3 and H3K27ac signals (**Supplementary Figure 18B**). Collectively, these results show that the chromatin loops identified in this study are reliable. The main text modifications occur in lines 384-387, 390, 393. We also supplemented the references Liu, Z et al (2017, Molecular Cell) and van der Weide, R. H. et al (2021, NAR Genomics and Bioinformatics) in the revised manuscript.



Re comments Figure 2.

A. The left APA plot was performed in leaf using chromatin loops identified in hypocotyl, and the right APA plot was performed in hypocotyl using chromatin loops identified in this tissue. The redder the color, the higher the frequency of interaction. **B.** The tornado-plots display the coverage of mapping reads from Hi-C data of hypocotyl at the anchors of loops related genes.

References

1. Dong, P. *et al.* 3D Chromatin Architecture of Large Plant Genomes Determined by Local A/B Compartments. *Molecular Plant* **10**, 1497–1509 (2017).
2. Huang, Y. *et al.* HSFA1a modulates plant heat stress responses and alters the 3D chromatin organization of enhancer-promoter interactions. *Nat Commun* **14**, 469 (2023).

3. Ariel, F. *et al.* Noncoding Transcription by Alternative RNA Polymerases Dynamically Regulates an Auxin-Driven Chromatin Loop. *Molecular Cell* **55**, 383–396 (2014).
4. van der Weide, R. H. *et al.* Hi-C analyses with GENOVA: a case study with cohesin variants. *NAR Genomics and Bioinformatics* **3**, lqab040 (2021).
5. Liu, Z. & Kraus, W. L. Catalytic-Independent Functions of PARP-1 Determine Sox2 Pioneer Activity at Intractable Genomic Loci. *Molecular Cell* **65**, 589-603.e9 (2017).
6. Yang, B. *et al.* CTCF controls three-dimensional enhancer network underlying the inflammatory response of bone marrow-derived dendritic cells. *Nat Commun* **14**, 1277 (2023).

REVIEWERS' COMMENTS

Reviewer #1 (Remarks to the Author):

Thank for the responses that they addressed my concern well. I have no further questions.