nature communications

Article

https://doi.org/10.1038/s41467-024-55309-4

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

Received: 24 December 2023

Accepted: 6 December 2024

A list of authors and their affiliations appears at the end of the paper

Although epigenetic modification has long been recognized as a vital force influencing gene regulation in plants, the dynamics of chromatin structure implicated in the intertwined transcriptional regulation of duplicated genes in polyploids have yet to be understood. Here, we document the dynamic organization of chromatin structure in two subgenomes of allotetraploid cotton (*Gossypium hirsutum*) by generating 3D genomic, epigenomic and transcriptomic datasets from 12 major tissues/developmental stages covering the life cycle. We systematically identify a subset of genes that are closely associated with specific tissue functions. Interestingly, these genes exhibit not only higher tissue specificity but also a more pronounced homoeologous bias. We comprehensively elucidate the intricate process of subgenomic collaboration and divergence across various tissues. A comparison among subgenomes in the 12 tissues reveals widespread differences in the reorganization of 3D genome structures, with the Dt subgenome exhibiting a higher extent of dynamic chromatin status than the At subgenome. Moreover, we construct a comprehensive atlas of putative functional genome elements and discover that 37 *cis*-regulatory elements (CREs) have selection signals acquired during domestication and improvement. These data and analyses are publicly available to the research community through a web portal. In summary, this study provides abundant resources and depicts the regulatory architecture of the genome, which thereby facilitates the understanding of biological processes and guides cotton breeding.

Polyploidization, which is a type of whole genome duplication (WGD), entails the merging of two or more genomes into a single nucleus, changing not only the distribution and interactions of intranuclear chromatin but also influencing the local microenvironment, such as in terms of histone modification and chromatin accessibility[1–7]. As a significant driving force in the evolutionary process, plant polyploidy, including its impacts on genetic diversity and environmental adaptation, among other aspects, has been widely studied[8–15]. During the process of polyploidization and domestication, significant changes in the chromatin spatial architecture of soybean subgenomes have been

e-mail: mjwang@mail.hzau.edu.cn

observed[16]. In wheat, epigenetic modifications exhibit divergence among subgenomes at the genome-wide level, influencing the dynamic balance of homoeologous gene expression during development and stress response processes[14,17]. Tetraploid rice exhibits more DNA hypomethylation than diploid rice, which enhances its salt tolerance[18]. An analysis of the effect of whole-genome duplication on the phyllosphere microbiome using synthetic autotetraploid *Arabidopsis* revealed that polyploidization improves immunity[19]. At present, an increasing number of studies have focused on coordination among subgenomes in regulating diverse biological processes in allopolyploid plants[14,17,20–24].

Single or multiple genomes fold in an organized manner within a nucleus to form spatial chromatin structural units, which include chromosome territories (CTs) that depict the spatial positioning of chromosomes, A/B compartment that distinguish active and inactive chromatin status, topologically associating domains (TADs) where chromatin is tightly folded inside, and fine chromatin folding structures called loops[25]. A recent study illustrated that chromatin status encompasses multiple status (A1, A2, A3, A3, B1, B2, B3 and B4), which are defined as refined sub-compartments[26,27]. The multilevel 3D genome structure plays a key role in various tissues and different developmental periods[28,29]. The comparison of spatial genome architecture across multiple tissues in rice, millet, and maize has revealed the local A/B compartment that exhibits tissue-specific dynamics[30]. The TAD boundary regions are enriched in active genes, and the reorganization of these boundaries is related to multiple life processes, such as development, differentiation and stress response[31,32]. Loops mediate the close spatial distances between distal regulatory elements and genes[33,34]. Dynamic loops regulate the expression of tissue-specific genes in maize and the response to heat stress in tomato[34,35]. The arrangement of the hierarchy of 3D genome architectures effectively influences various biological processes by coordinating long-range interactions between CREs and genes[25,34,36,37].

CREs are noncoding DNA sequences with functions that ensure the precise spatiotemporal expression pattern of genes by dynamically changing their own status and regulatory relationships with genes[37–40]. The advent of multiple assays utilizing high-throughput sequencing techniques has greatly enhanced the process of identifying CREs[37,39]. Chromatin immunoprecipitation assay based on DNA sequencing (ChIP-Seq) is widely employed for identifying histone modification and transcription factor-binding sites[41,42]. Assay for transposase-accessible chromatin using sequencing (ATAC-Seq) can be utilized to identify open chromatin regions (OCRs) or the locations of nucleosomes in the whole genome[43,44]. In animals, the identification of CREs has become routine due to the generation of comprehensive OCRs and epigenome maps[45–48]. Recently, major advancements have been made in the identification and functional analysis of CREs in plants[40,49–54]. Extensive differences in OCRs have been observed when comparing chloroplasts and meristematic stem cells in *Arabidopsis*[55]. Plant OCRs exhibit dynamic changes during development and in response to biotic and abiotic stress[17,37,56,57]. For instance, *Arabidopsis* OCR levels increase under heat or cold stress[58]. Abundant CREs have been identified in maize, and changes in regulatory networks of CREs influence the expression of ear- and tassel-specific genes[35,52]. Moreover, the variation in CREs results in the presence or absence of transcription factor-binding sites (TFBSs), which in turn influences the response to cold and heat stress[59]. Thus, in-depth exploration of CREs significantly enhances our comprehension of the functional genome and unlocks its immense potential for crop improvement[37,39,59,60].

The allotetraploid cotton *Gossypium hirsutum* offers considerable economic and social benefits by serving as a source for textile fiber, oil and protein[61–65]. Although cotton genomes have been extensively studied, referable functional component mapping has yet to be conducted[66,67]. Here, we generated in situ Hi-C chromatin contact maps at 3 kb resolution, ChIP-Seq for two active histone modifications (H3K4me3 and H3K27ac), ATAC-Seq and RNA-Seq datasets from twelve tissues spanning the major developmental programs. Utilizing these multi-omic datasets, we uncovered the developmental dynamics and subgenomic divergence in the 3D genome organization, epigenome modification and gene expression. We observed abundant dynamic rearrangement in the hierarchical 3D genome architecture that exhibited both tissue specificity and subgenome disparities. Comparing the epigenome modification profiles of the two subgenomes revealed that these differences also play a role in subgenome divergence. In summary, this study employs established comprehensive epigenomic and three-dimensional genomic maps to systematically elucidate the cooperative interactions and regulatory divergence of subgenomes throughout the entire developmental stage of allopolyploid cotton. Additionally, it provides a high-quality data resource with great potential for application in cotton breeding.

Results

Transcriptomic atlas across multiple *G. hirsutum* tissues

To clearly describe the dynamics of coordination and divergence between the two subgenomes of allotetraploid cotton, we constructed a comprehensive *G. hirsutum* transcriptome atlas using samples from 12 tissues or developmental stages, including radicle, root, stem, leaf, hypocotyl, cotyledon, ovule, anther, petal, stigma, and developmental fibers at 10 days post anthesis (DPA) and 20 DPA (Fig. 1A, Supplementary Table 1, Supplementary Data 1 and Supplementary Fig. 1). The overview of this atlas revealed pronounced tissue variations and dynamic divergence of subgenomes across the 12 tissues (Fig. 1A and Supplementary Data 1, 2). A total of 28,169 genes exhibiting higher expression levels were detected (transcripts per million, TPM ≥ 0.5) in all tissues and showed functional enrichment in basic gene ontology (GO) terms (Supplementary Fig. 2A–C). Based on their inherent characteristics, we defined these genes as core genes. Furthermore, approximately 10% of single-copy core homoeologous genes may exhibit potential subgenomic divergence in fundamental biological processes (Supplementary Fig. 2D). Subsequently, we tabulated the differentially expressed genes (DEGs) between the any two tissues (Supplementary Fig. 3). We analyzed the differentially expressed homoeologous genes (DEHGs) between the same and between different subgenomes in different tissues. This analysis revealed that the degree of gene divergence between two subgenomes is greater than that observed between different tissue types (AA DEHGs: 17.94%–55.05%, DD DEHGs: 18.46%–55.23%, AD DEHGs: 35.46%–58.69%, DA DEHGs: 35.63%–58.92%) (Supplementary Fig. 3). To comprehensively investigate the DEHGs, we classified the homoeologous gene pairs with differential expression into eight patterns (At-Dt: D–D, U–U, D–N, N–D, U–N, N–U, D–U and U–D) based on their expression types (Fig. 1B). Interestingly, approximately half of homoeologous gene pairs exhibited directionally inconsistent expression patterns between tissues (D–N, N–D, U–N, N–U, D–U and U–D), suggesting a widespread presence of inconsistent subgenomic divergence across different tissues (Fig. 1B and Supplementary Fig. 4). These results reinforce our speculation that each copy of homoeologous genes may make variable contribution to the developmental programs of different tissues.

To further investigate the divergence of subgenomes, we conducted statistical analyses and observed that 69.04% of homoeologous genes exhibited a dynamic direction of expression bias toward the At or Dt subgenome across tissues (Fig. 1A, C and Supplementary Data 2). We classified 17,851 homoeologous gene pairs with biased expression into three patterns: constant bias (389, 2.18%), in which homoeologous genes exhibited bias in all the tissues and only one biased direction (biased At or biased Dt); semi-constant bias (13,393, 75.03%), in which homoeologous genes exhibited bias in some tissues and only one biased direction; and contrary bias (4069, 22.79%), in which homoeologous genes exhibited different bias directions (biased At and

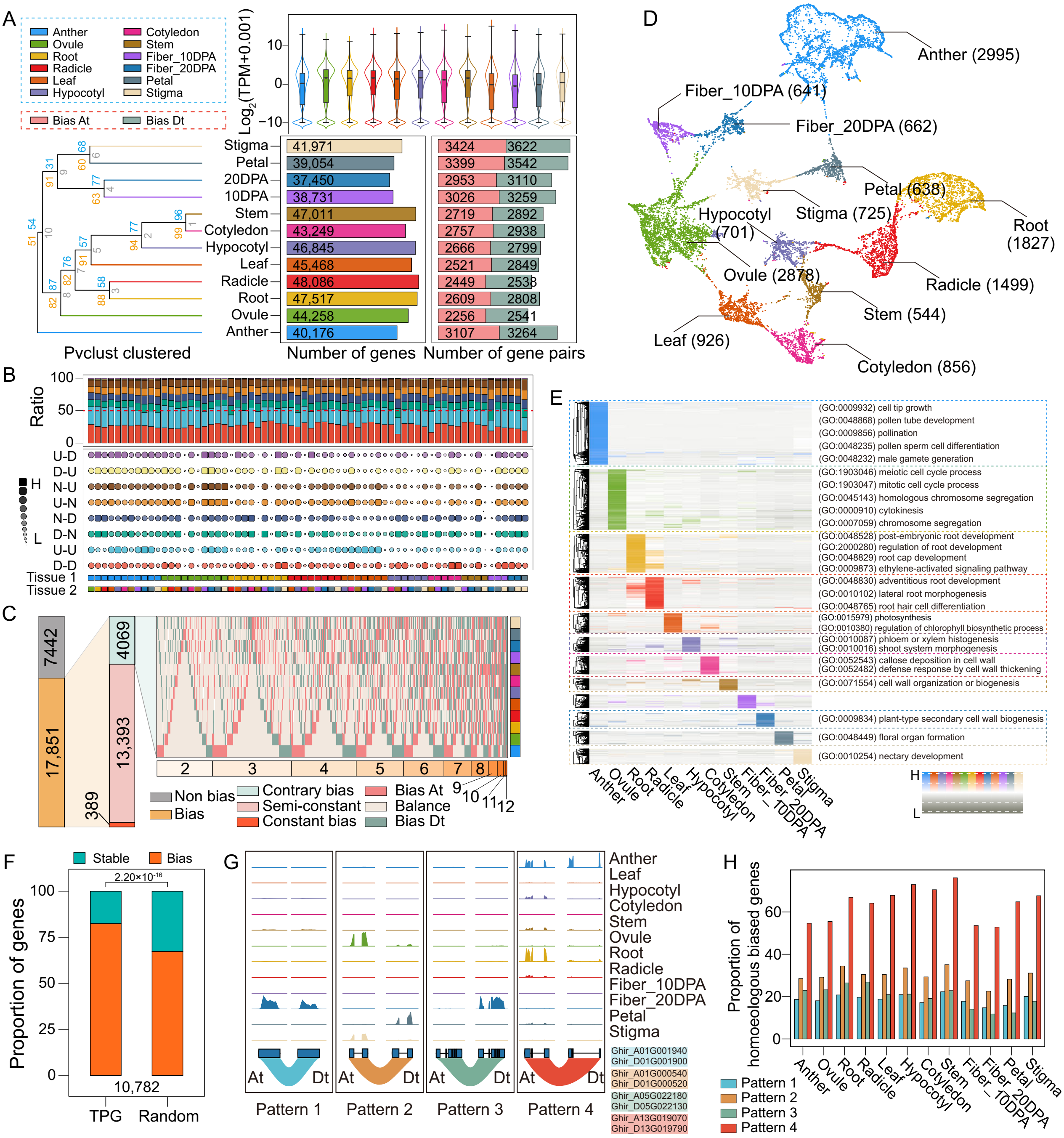

Fig. 1 | Dominance switching of homoeologous gene expression and tissue-specific genes in multiple tissues of cotton. A Overview of transcriptomes of 12 cotton tissues, focusing on the clustering of tissues (tree plot), the expression levels of genes (box plot), the number of expressed genes (bar plot on left) and homoeologous gene pairs with biased expression (bar plot on right). Center line, median; box limits, first and third quartiles; whiskers, 1.5 × interquartile range. **B** The point plot displaying the number of eight different patterns of DEHGs (At-Dt) among tissues. "D", downregulated expression; "U", upregulated expression. The tissue colors align with those in (**A**). **C** The bar plot depicting the count of homoeologous genes with no bias, dynamic bias, and stable bias in 12 tissues. The heatmap showing the details of biased homoeologous genes with contrasting bias patterns across 12 tissues. **D** The point plot displaying the clustering of TPGs using UMAP arithmetic. Each point on the plot represents a gene, which is colored based on the tissue where it exhibits the highest expression level. **E** The heatmap shows the clustering of the *z*-score normalized expression levels of TPGs, while the annotation provides information on the corresponding Gene Ontology (GO) terms associated with these genes. **F** The bar plot illustrates the comparison of biased proportions between homoeologous TPGs and random genes. The random genes were extracted from the homoeologous subset excluding TPGs through over a thousand random samplings. Two-sided Pearson's chi-squared test. **G** The four expression patterns of homoeologous TPGs. Pattern 1: Both copies of homoeologous gene pairs show tissue preferential expression in the same tissues; Pattern 2: Both copies of homoeologous gene pairs show tissue preferential expression in different tissues; Pattern 3: One copy of homoeologous gene pairs shows tissue preferential expression and the other copy is silenced; Pattern 4: One copy of homoeologous gene pairs shows tissue preferential expression and the other copy shows expression in multiple tissues. The tracks are colored by tissues. **H** The bar plot showing the proportion of four expression patterns of homoeologous TPGs with biased expression in each tissue. Source data are provided as a Source Data file.

biased Dt) in the 12 tissues (Fig. 1C). Our findings showed that approximately 20% of the contrary bias genes exhibited a high-frequency biased transition (≥7) (Fig. 1C). Further analysis revealed that 4247 homoeologous gene pairs exhibited tissue-specific biased expression (Supplementary Fig. 5A). In addition, 83.78% of the contrary bias genes overlapped with the absolute reverse homoeologous DEHGs (D–U and U–D) (Supplementary Fig. 5B). Analysis of contrary biased genes showed that the two homoeologous copies exhibited not only differential expression levels in various tissues, but also pointed to a higher degree of tissue specificity (Supplementary Fig. 5C), suggesting that these genes may exhibit potential subgenome subfunctionalization. Moreover, we observed that GO terms of these genes showcased response to external stimuli (Supplementary Fig. 5D). These results provide detailed illustrations of the differential expression of homoeologous genes among the different tissues and the widespread dynamic bias of homoeologous genes across the 12 tissues. The disparity of homoeologous genes ranges from tissue-specific bias to a semi-constant bias, followed by a constant bias, and eventually to a contrary bias, which indirectly reflects the direction of subgenome divergence.

Diverse biological processes are controlled by distinct sets of genes, some of which are tissue-preferential genes (TPGs) that may have tissue-specific functional implications[68,69]. We conducted a statistical analysis to determine the number of genes exhibiting differential expression across all different tissues. Interestingly, we observed that a significantly higher number of genes exhibited differential expression (Supplementary Fig. 6A and Supplementary Data 1). To investigate the underlying reasons, our subsequent analysis revealed that the majority of these genes exhibited high levels of expression in a specific tissue (Supplementary Fig. 6B, C). To comprehensively identify this type of genes, we calculated the TAU value for each gene using the TAU algorithm to evaluate their tissue specificity and identified a total of 14,892 TPGs (TAU ≥ 0.85)[70] (Supplementary Data 3 and Supplementary Fig. 7A–C). By studying the properties of such genes, we found that they can be clustered well using the uniform manifold approximation and projection (UMAP) method[71] based on their expression levels (Fig. 1D). Additionally, these genes showed strong co-expression correlations in each tissue (Supplementary Fig. 7D). Further in-depth analysis revealed that the enriched GO terms for the TPGs were closely associated with the biological processes of each corresponding tissue (Fig. 1E). The above results indicate that we identified a set of high-quality TPGs. Moreover, we found that homoeologous TPG pairs exhibit a greater propensity towards single subgenome dominance expression compared to random (Fig. 1F and Supplementary Fig. 7E). Furthermore, we classified the homoeologous TPGs into four patterns according to the biased status in multiple tissues (Fig. 1G). We found that the TPGs with pattern 4 exhibited preference for biased expression in multiple tissues (Fig. 1H). Homoeologous TPGs exhibit a stronger degree of dominant expression in the single subgenome, especially in pattern 4, implying that subgenome divergence may stem from tissue-specific functions.

Annotation of candidate functional elements

Gene region annotation is a routine operation, while the identification of noncoding regulatory elements that control gene expression levels is still at the early stage in plants[39]. To mitigate the detection of false positive peaks, we inferred 445 blacklist regions in the cotton genome using the GreenScreen tool[72] (Supplementary Data 4 and Supplementary Fig. 8A). We inferred active histone modification sites and open chromatin regions (OCRs) in the cotton genome using ChIP-Seq (H3K4me3 and H3K27ac) and ATAC-Seq datasets from multiple tissues (Supplementary Tables 2, 3 and Supplementary Fig. 8B–D). These datasets exhibited obvious enrichment signals in the regions near the transcription start sites (TSS) (Supplementary Fig. 9A–C). We identified a total of 455,194 H3K4me3 and 292,160 H3K27ac modification peaks and 222,978 OCRs from all the tissues (Supplementary Tables 2 and 3). By annotating these peaks, we observed that an average of 82.90% (H3K4me3) and 76.13% (H3K27ac) of peaks identified from ChIP-Seq data were annotated within promoter regions (2 kb upstream of TSS), while more peaks (55.20%) identified from ATAC-Seq data were located in distal intergenic regions (Supplementary Fig. 9D–F).

Previous studies have indicated an association between alterations in gene expression and specific types of local histone modifications[73,74]. We discovered that a total of 16,840 nonredundant genes were annotated by both ChIP-Seq (H3K27ac and H3K4me3) and ATAC-Seq peaks (Fig. 2A). Their expression levels in each tissue were significantly higher, and the proportion of core genes within this gene set was notable, which indicates that the core genes had more stable epigenomic modifications (Supplementary Fig. 10). The comparison of peaks among tissues revealed that the presence or absence of histone modification peaks within promoter regions was closely associated with the upregulation or downregulation of genes (Fig. 2B, C and Supplementary Fig. 11A, B). Additionally, our analysis detected a noticeably strong signal strength for H3K27ac and H3K4me3 near TPGs (Fig. 2D and Supplementary Fig. 11C, D). These results unveiled the underlying causes of tissue-specific expression resulting from the differences in chromatin status with histone modifications between tissues.

Chromatin accessibility has the potential to influence gene expression (Supplementary Fig. 11E). We have thoroughly explored the tissue specificity of OCRs that provide the binding probability for transcription factors[75]. We predicted binding sites for 656 motifs from the JASPAR database and analyzed the motif enrichment in OCRs from multiple tissues by using the HINT tool[76] (Fig. 2E and Supplementary Data 5). The results showed that certain motifs within OCRs were enriched in specific tissues (Fig. 2E). To further explore the function of TFs in tissue-specific regulation, we analyzed the TFBSs located in the promoter regions of the TPGs. The results revealed that some TFs may preferentially bind the promoter regions of TPGs (Supplementary Data 6). For instance, FLC (Flowering locus C), which is associated with flowering, exhibits more preferential binding in the promoter regions of TPGs associated with flower-related tissues (Fig. 2F). The absence of OCRs leads to a difference in the binding site of the FLC motif between the hypocotyl and stigma that influences the expression level of Ghir_A03G019940 (Fig. 2G). The above results demonstrate that OCRs were dynamically present across tissues and exhibited tissue-specific patterns.

Currently, there is limited research on the effects of the dynamic epigenome on homoeologous genes[14,17]. Our analysis revealed that the differences in histone modifications and OCRs also influence the biased expression of homoeologous genes (Fig. 2H and Supplementary Fig. 12A, B). We also found that not all of the homoeologous genes with biased expression exhibited differences in OCRs in promoter regions (Supplementary Fig. 12C). To determine the reason, we analyzed the Levenshtein distance, which describes the degree of difference between two targets, and revealed a positive correlation between the difference in TFBSs and the difference in the expression level of homoeologous gene pairs (Fig. 2I). This result suggests that biased gene expression can be influenced not only by the absence of active histone modification and OCRs but also by differences in TFBSs located in the promoter regions (Fig. 2H, J and Supplementary Fig. 12C).

By integrating ATAC-Seq and ChIP-Seq data, we constructed a map of predicted ATAC-Seq-supported tissue regulatory elements (PATREs)[77], which were perfectly clustered using UMAP algorithms[71], based on the H3K27ac, H3K4me3 and ATAC signal strength in OCRs (Fig. 2K). Further analysis revealed that the map contained 13,992 putative enhancers, 10,132 promoters, and 5859 other elements (Supplementary Data 7). Both candidate enhancers and promoters were located within the peak regions of the ATAC-Seq data and had

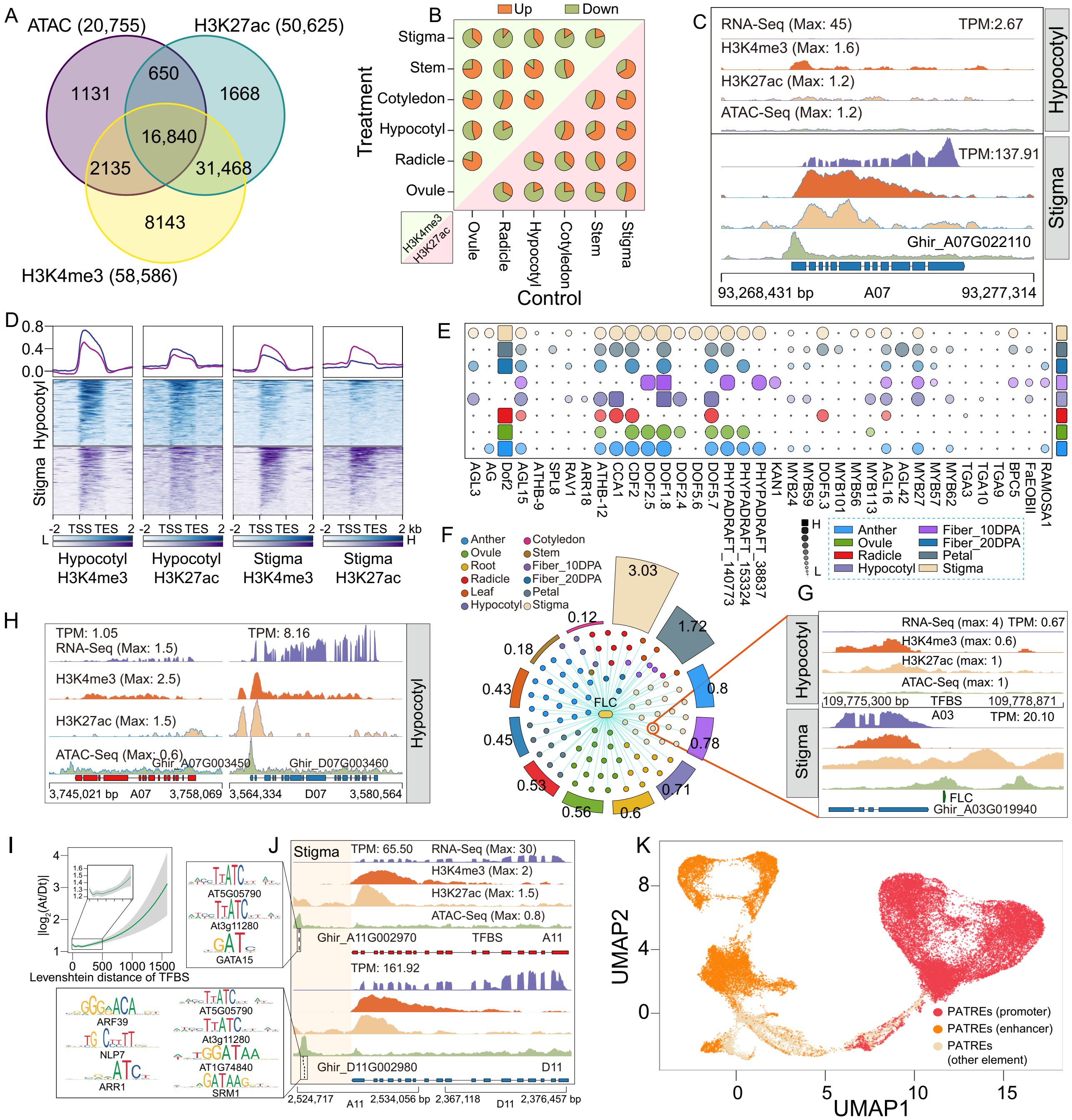

Fig. 2 | Dynamics of epigenomes in multiple tissues, with implications for gene transcriptional regulation. A Venn diagram illustrating the overlap of genes annotated by peaks inferred from ChIP-Seq (H3K4me3 and H3K27ac) and ATAC-Seq datasets. **B** The pie chart displaying the ratio of DEGs with the presence or absence of H3K4me3 and H3K27ac peaks in promoter regions of genes. **C** Genome browser screenshot displays gene expression, H3K4me3, H3K27ac and ATAC-Seq signals in the 2 kb region surrounding Ghir_A07G022110. **D** Heatmap showing the signal strength of histone modifications surrounding TPGs from different tissues. **E** The point plot demonstrates the enrichment of different motifs with tissues, where the size of each point corresponds to the number of TFBSs. **F** The network plot shows the TPGs with FLC motif annotation in stigma. Each point in the network plot represents an annotated TPG, and the size of the circle bars represents the proportion of annotated TPGs relative to the total TPGs in each tissue. **G** The genomic region tracks highlight the region located 2 kb upstream of the TSS of Ghir_A03G019940 in both stigma and hypocotyl tissues. **H** Genome region track plot displays the presence and absence of the peaks of ChIP-Seq (H3K27ac and H3K4me3) and ATAC-Seq within promoter regions of homoeologous biased gene pairs (Ghir_A07G003450/Ghir_D07G003460). **I** Line plot showing the relationship between the Levenshtein distance of TFBSs in homoeologous gene pairs and the absolute value of the $\log_2(\mathrm{TPM} + 0.01)$ ratio for homoeologous gene pairs. The LOESS method, incorporated in ggplot2 form R (version 4.2.2), is utilized to fit the curve, while the shaded region signifies an interval at a 95% confidence level. **J** The genome tracks displaying differences in TF motif binding within the open chromatin region of homoeologous genes, as well as a close-up view of the TFBSs motifs. **K** The UMAP plot displaying the clustering of PATREs. Source data are provided as a Source Data file.

high signals of active histone modifications. This dataset provides global annotation of functional elements in noncoding genomic regions.

Tissue dynamics of the A/B compartment and sub-compartments

Notably, 3D multi-layered structures can serve as bridges to facilitate interactions between regulatory elements and genes[78]. The spatial genome architecture for individual tissues from a few plants has been elucidated in previous studies[25,40]. We generated highly reproducible in situ Hi-C libraries from 12 tissues, achieving a resolution of 3 kb (Supplementary Fig. 13 and Supplementary Data 8). The A/B compartment and sub-compartments were identified at 100 kb and 20 kb, respectively (Supplementary Data 9, 10). By comparing the proportions of the A/B compartment between the At and Dt subgenomes, we observed that their ratios were not always consistent across all the 12 tissues (Fig. 3A). To investigate whether this difference might influence the biased expression of homoeologous genes, our subsequent analysis revealed that a higher number of homoeologous genes were located in the A compartment regions in the At subgenome than in the Dt subgenome across the 12 tissues (Fig. 3B). However, previous transcriptome analyses indicated that a higher number of genes exhibited bias toward the Dt subgenome in all the tissues (Fig. 1A). These results illustrate that the correlation between the biased expression of homoeologous genes and their A/B compartment state was relatively weak. Notably, the noncoding regions of the Dt subgenome exhibited higher signals of active A compartment status (Supplementary Fig. 14A), which might slightly influence the biased expression of homoeologous genes towards the Dt subgenome.

A global analysis revealed widespread switching of A/B compartment and sub-compartments among the 12 tissues (Fig. 3C and Supplementary Data 9, 10). The comparison of the A/B compartment among all the tissues revealed the presence of 424.00 Mb regions exhibiting dynamic switching of the A/B compartment status (Fig. 3D and Supplementary Data 9). Interestingly, when considering the two subgenomes separately, a larger dynamic compartment was observed in the Dt subgenome than in the At subgenome (23.29% versus 15.30%) (Fig. 3D and Supplementary Data 9). Similar to the A/B compartment, the switching regions of sub-compartments were larger in the Dt subgenome than in the At subgenome (81.78% versus 76.90%) (Fig. 3E and Supplementary Data 10). The Dt subgenome exhibited a greater degree of chromatin status dynamics across the 12 tissues.

Compared with the A/B compartment, the sub-compartments displayed more intricate and complex dynamics across the 12 tissues (Fig. 3C, E). We found that an average of 85.50% of chromatin regions maintained consistent A/B compartment and sub-compartment status (Supplementary Fig. 14B). In-depth analysis revealed that 35.60% of regions in different tissues displayed divergence for A/B compartment and sub-compartments, particularly in the Dt subgenome (Supplementary Fig. 14C, D). Interestingly, 94.92% of dynamic A/B compartment regions across tissues overlapped with regions exhibiting different chromatin status of A/B compartment and sub-compartments (Supplementary Data 9, 10). These findings suggest that the divergence regions between the A/B compartment and the sub-compartment are more susceptible to experiencing compartment status alterations across tissues. A subsequent analysis demonstrated that genes located in the dynamic sub-compartments displayed a significantly higher coefficient of variation (CV) than those located in the stable sub-compartments, but this difference was not observed in the A/B compartment (Fig. 3F). This result suggests that gene activity is more strongly correlated with the sub-compartment status than with the A/B compartment. We found that genes located in sub-compartments with higher signals of A status showed elevated expression levels (Fig. 3G). In our investigation of the tissue-specific activity of the sub-compartments, our statistical analysis revealed that 73.24% of the TPGs exhibited multiple sub-compartment statuses, with approximately 86.11% having the highest or second highest active sub-compartment status in the corresponding tissue (Fig. 3H, I). These results indicate that the sub-compartments were more strongly associated with gene expression than the A/B compartment and exhibited tissue-specific patterns in particular regions.

The organization of the TAD-like domains in different tissues

TADs with high-frequency self-interaction within their interior regions act as modules implicated in transcriptional regulation[79]. The dynamic change in TAD-like domain boundaries influences various cellular processes, including cell differentiation, tissue development and immune response, among others[32]. Due to the absence of CTCF homologous proteins in plants, such structures are defined as TAD-like domains[80–82]. We identified a range of 7248 to 7855 TAD-like domains which exhibited clear domain features across tissues (20 kb) using the HiTAD tool[83] (Supplementary Data 11 and Supplementary Fig. 15A, B). The average size of these domains was approximately 300 kb, and the size of these domains from the At subgenome was larger than those from the Dt subgenome in the 12 tissues (Supplementary Data 11 and Supplementary Fig. 15C). The boundary regions exhibited significantly stronger signals of active histone modifications (H3K27ac and H3K4me3) than the interior regions of the TAD-like domains (Supplementary Fig. 15D, E).

Previous studies demonstrated that the disruption of TADs could disrupt the normal expression of genes[84]. By analyzing the conservation of the TAD-like boundaries between all tissues, we found that approximately 85% of the boundaries were conserved among tissues, while 55.94% of the boundaries were completely conserved in all the tissues with very low insulation scores (Fig. 4A). Further analysis of the dynamic boundaries showed that many DEGs shifted from boundary positions to interior positions, or vice versa (Fig. 4B and Supplementary Fig. 16A). When comparing the TAD-like domains identified in both subgenomes, we observed that approximately 50% of the homoeologous gene pairs with biased expression were located at the boundaries or interior regions of these domains (Supplementary Fig. 16B). Half of these gene pairs were located at boundaries in one subgenome, and they exhibited biased expression toward the corresponding subgenome (Fig. 4C and Supplementary Fig. 16B). The results suggest that the dynamic boundaries have an impact on the differential expression of genes, while the rearrangement of the TAD-like domains between the two subgenomes may contribute to biased expression of homoeologous genes.

The TAD structure potentially serves as a functional module[85]. To further investigate the distinctive 3D genome architectures in 12 tissues, we identified 2,566 tissue-specific TAD-like domains with high activity (TS-TADs) (Fig. 4D and Supplementary Data 11). These TS-TADs contained genes with higher TAU values, serving as a measure to evaluate the tissue specificity of spatial chromatin structure (Supplementary Fig. 17A). Upon further analysis of the genes contained within the TS-TADs, we observed differences in the expression levels of genes among the different tissues (Fig. 4E). Specifically, the genes located in the TS-TADs of anthers exhibited significantly higher expression levels and TAU values (Fig. 4E and Supplementary Fig. 17A), which implied greater specialization within the structural domains of certain tissues. Moreover, we observed a significant enrichment of the TPGs and a lower ratio of A compartment status near tissue-specific boundaries (Supplementary Fig. 17B, C). These results demonstrate the tissue specificity of the spatial chromatin structures. Further exploration of the chromatin status activity revealed that the TS-TADs in the radicle exhibited a more active sub-compartment status than the corresponding chromatin regions in other tissues (Fig. 4F). These findings suggest that the TS-TADs contained genes with high TAU values and high sub-compartment activity, implying their potential as structural units implicated in tissue-specific biological/evolutionary processes.

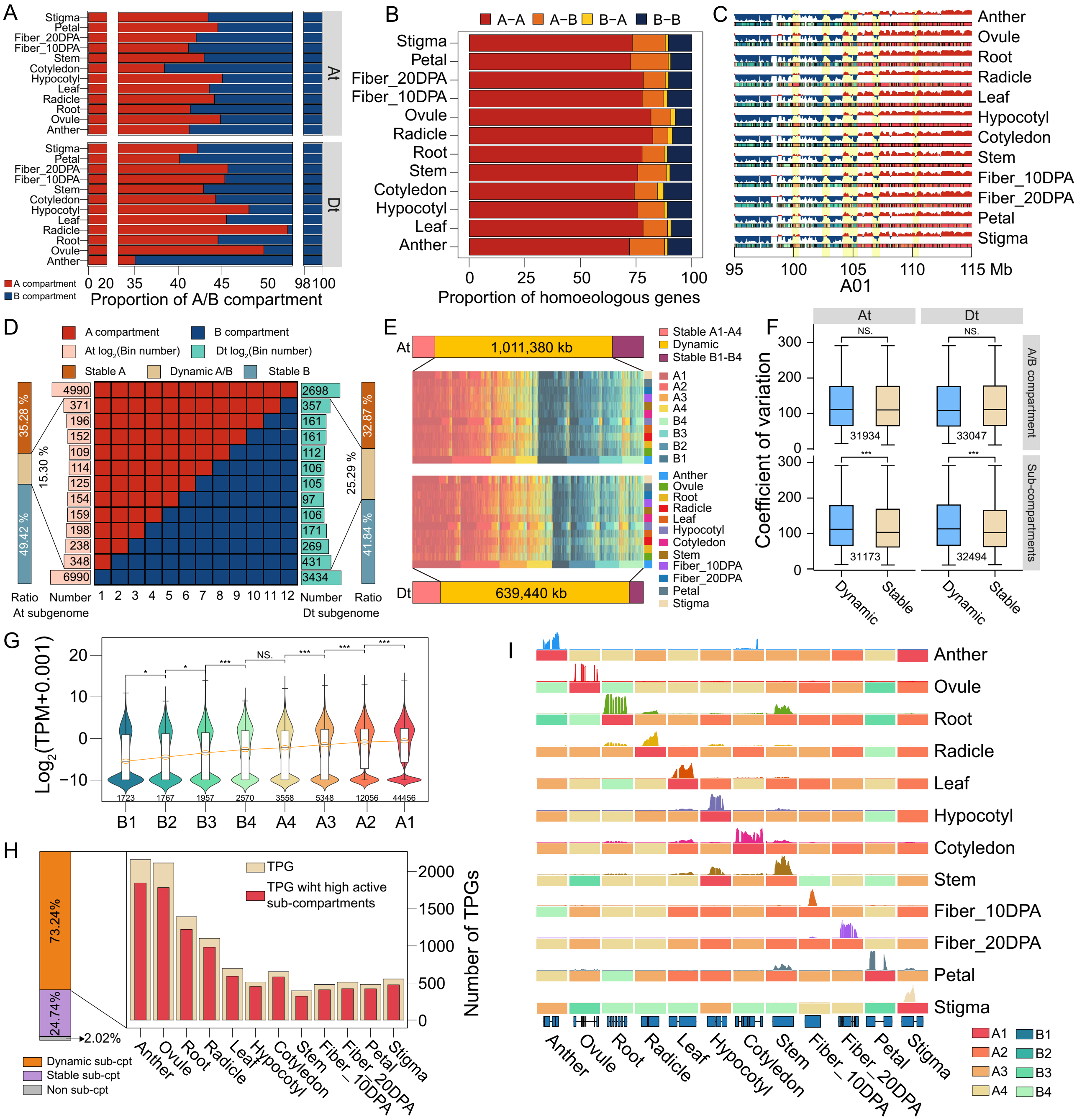

Fig. 3 | The switching of the A/B compartment and sub-compartment status between multiple tissues. **A** The bar plot showing the proportion of the A/B compartment in both subgenomes across the 12 tissues. **B** The bar plot showing the proportion of homoeologous gene pairs (At-Dt) based on their A/B compartment status in two subgenomes. **C** The genomic region exemplifies the dynamic and stable regions of the A/B compartment and sub-compartment status in 12 tissues. Yellow bars highlight distinct dynamic regions. **D** An overview plot displaying the statistical results of the stability and dynamics of the A/B compartment in the At and Dt subgenomes across the 12 tissues. The vertical bar plot depicts the ratio of absolute stable A/B compartment and dynamic A/B compartment across 12 tissues. The horizontal bar plot provides details of stable and dynamic A/B compartment in 12 tissues. **E** An elaborate overview plot showing the stability and dynamics sub-compartments across the 12 tissues. The heatmap shows the details of the dynamic sub-compartments in 12 tissues. **F** The box plot showing the CV of TPM for genes located in the stable and dynamic A/B compartment and sub-compartments. **G** The violin plot showing the expression levels of genes contained within different sub-compartments in fiber at 20 DPA. **H** The left hand bar plot displaying the proportion of stable and dynamic sub-compartment regions across tissues, and the righthand bar plot shows the number of TPGs located in the dynamic sub-compartments. The TPGs with highly active sub-compartments refer to those with first or second active sub-compartment status in the corresponding tissue. **I** The genomic region plot showing sub-compartment status and tracks of RNA-Seq data for TPGs in multiple tissues. Data from **F** and **G** were analyzed by two-sided Wilcoxon signed-rank test. NS, $p > 5{\times}10^{-2}$; *, $1 \times 10^{-3} \le p < 5 \times 10^{-2}$; **, $2.2 \times 10^{-16} \le p < 1 \times 10^{-3}$; ***, $p < 2.2 \times 10^{-16}$. Center line, median; box limits, first and third quartiles; whiskers, 1.5 × interquartile range. Source data are provided as a Source Data file.

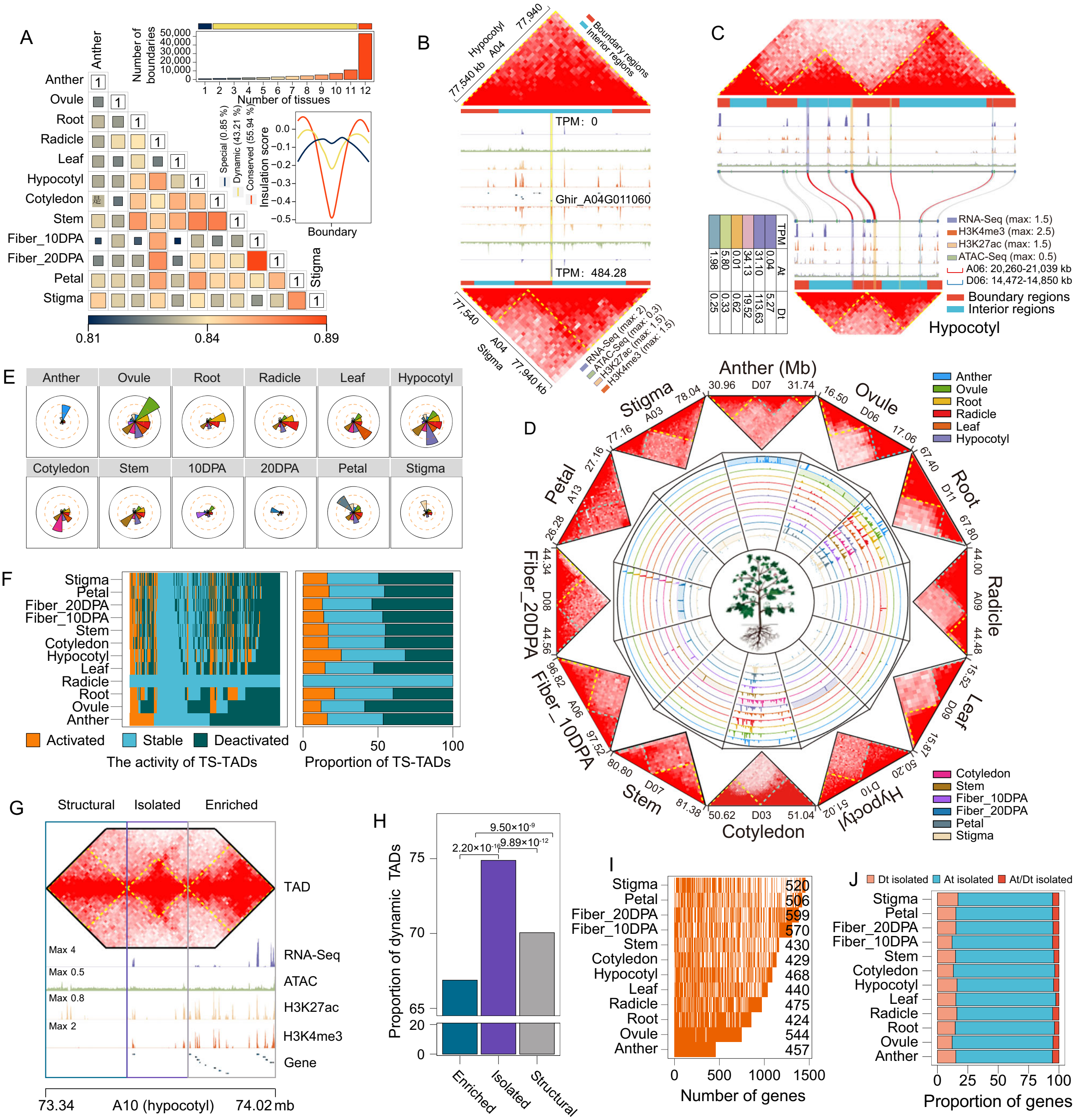

Fig. 4 | Dynamic and specific TAD-like domains across tissues. A Overview of dynamic TAD-like domain boundaries observed in 12 tissues. The heatmap showing the similarity of TAD-like domain boundaries between tissues, the line plot displaying the insulation scores of three types of TAD-like domain boundaries (conserved, dynamic and specific) and the bar plot showcasing the total number of special (blue), dynamic (yellow) and conserved (orange) TAD-like domain boundaries. **B** The Hi-C interaction heatmap and genome region tracks illustrate the rearrangement of the TAD-like domains, which alters the spatial position of Ghir_A04G011040 (interior vs. boundary) and impacts its expression level. **C** The Hi-C interaction heatmap and genome region tracks illustrate the rearrangement of the TAD-like domains in the subgenomes. The colors of table were consistent with bar highlighting in the genome tracks. **D** Overview of the TS-TADs from each tissue marked with blue dotted lines. The genomic region tracks represent the RNA-Seq signal. **E** The circular bar plot showing the median expression level of genes located in TS-TADs. **F** Dynamic activation of the TS-TAD of the radicle among multiple tissues. The heatmap shows the changes in TS-TAD activation based on sub-compartments, while the bar plot provides an overview of the total count. Activated/Deactivated means the TADs contains more/fewer higher sub-compartments regions. Stable means no change of sub-compartment status between tissues. **G** The heatmap showing three types of TADs: enriched TAD containing more than one gene within a TAD (represented by blue box), isolated TAD containing a single gene in a TAD (represented by purple box) and structural TAD that do not contain any genes (represented by gray box). **H** The bar plot illustrating the proportion of dynamic TADs for three types of TADs. Two-sided Pearson's chi-squared test. **I** The heatmap displaying the dynamics of genes (orange line) located in isolated TADs across multiple tissues. **J** The bar plot showing the proportion of homoeologous gene pairs located in isolated TADs. At/Dt isolated indicates that both homoeologous copies are located in isolated TADs, and At isolated and Dt isolated indicate that only one homoeologous copy is located in an isolated TAD. Source data are provided as a Source Data file.

To further explore the relationship between genes and TAD-like domains, we classified the domains into three types based on the number of genes they contained: enriched TAD (containing more than one gene), isolated TAD (containing only one gene) and structural TAD (not containing a gene) (Fig. 4G and Supplementary Data 11). We speculated that structural TADs might play a role in maintaining the stability of the 3D genome architecture. In comparison to enriched TADs, isolated TADs display a simpler and more directly regulated relationship between genes and functional elements[31,86]. A thorough analysis of the isolated TADs revealed that they were more prone to rearrangements among tissues (Fig. 4H). Subsequently, we identified 1632 nonredundant isolated genes within isolated TADs (Supplementary Data 12). Among them, 89.28% were located in the dynamic TAD-like domains (Fig. 4I). Additionally, we analyzed the homoeologous genes for isolated genes in each tissue and found that there were more isolated genes in the At subgenome than in the Dt subgenome (Fig. 4J). The underlying reasons for the high tissue dynamics and subgenomic difference of isolated TADs warrant further investigation.

Regulatory effects of chromatin loops on gene expression

The chromatin loop, a finer chromatin 3D architecture that allows genes to approach regulatory elements that are distant in the linear genome, frequently undergoes dynamic rewiring both in animals and plants[33,78,87]. The regulatory networks of loops are intricate, and even slight changes can have far-reaching effects[34,88,89]. To provide an overview of the rewiring of the 3D genome organization in the 12 tissues of *G. hirsutum*, we identified a total of 1,609,540 nonredundant chromatin loops (3 kb) in the 12 tissues by combining a few tools, including Fit-Hi-C, Mustache, Chromosight and Peakachu[90–95] (Supplementary Fig. 18A and Supplementary Table 4). In addition to enrichment of loop peaks, we observed notable peaks of H3K4me3 and H3K27ac signals around the loop anchors compared to random bins (Supplementary Fig. 18A, B). Through a statistical analysis of the fine 3D genome alterations, we revealed that the majority of loops are uniquely present in one tissue, with only 249,354 (15.49%) of loops being identified repeatedly in multiple tissues (Fig. 5A). Focusing on genes, we observed that 65.16% of the genes were linked with multiple loops (Supplementary Fig. 18C). Moreover, our further analysis revealed that the gene expression level did not always exhibit a positive correlation with the number of loops (Fig. 5B and Supplementary Figs. 18D, 19A). Anchors with an excessive number of loops not only had fewer expressed genes but also contained fewer peaks of active histone modifications and OCRs (Supplementary Figs. 19B and 20). This observation implies that excessive extrusion by loops in a local chromatin region is likely to be accompanied by a reduction in local microenvironmental activity.

The rewiring of loops also exhibits shifts between long-range and short-range interactions[96]. However, the relationship between loop distance and gene expression level remains elusive[96,97]. We classified the loops into three types (G–G loop, G–N loop and N–N loop) based on whether the loop anchor contained a gene or not (Supplementary Fig. 21A). Subsequently, our analysis revealed that 37.29% of the loops contained genes (G–G loop and G–N loop) that exhibited a higher conservation ratio among tissues compared with N–N loops, which suggests that tissue-specific loops are more prevalent in the non-genic regions (Fig. 5C and Supplementary Fig. 21B). Interestingly, an average of 42.33% of TPGs are linked by tissue-specific loops and 9.74% of these are associated with putative functional elements in noncoding regions, implying a potential relationship between the organization of tissue-specific intricate domains and the functions (Supplementary Fig. 21C, D). We also found a higher proportion of these loops involving short-range interactions (Fig. 5D and Supplementary Fig. 22A). Subsequently, the results revealed that genes linked to excessively long loops demonstrated decreased levels of expression (Fig. 5E and Supplementary Fig. 22B). We also found that the intra-A compartment predominantly exhibited short-range interactions (Fig. 5F and Supplementary Fig. 22C, D). The above result indicates that compared to long-range interactions, short-range interactions entailed more G–G loops and tended to be distributed in the active A compartment.

The mechanism of loop regulation is still worth exploring in greater detail. By thoroughly investigating the G–G loops, we uncovered a fascinating phenomenon in which approximately 75% of gene pairs located at both ends of the G–G loops exhibited an imbalanced expression pattern, with TPMs differing by more than two-fold (Fig. 5G, H). Exploring the underlying reasons for this phenomenon revealed that compared with random anchors, these anchor regions exhibited notably higher signal strength in active histone modification (H3K27ac and H3K4me3) and open chromatin, and these signals did not exhibit a significant difference between anchors with imbalanced G–G loops (Fig. 5H and Supplementary Fig. 23). We hypothesize that the sequences of genes with an active status could serve as candidate regulatory elements for other genes by spatial folding.

Long-range loops can bring enhancers close to a promoter directly[98]. We discovered that 9134 (90.15%) of candidate enhancers overlapped with the N anchors, which were defined as potential distal CREs (Fig. 5I, J). Utilizing the Hi-C matrix, we discovered that in addition to one-step regulation (G–G loops and G–CRE loops), the two-step regulation of the loops (G–N–G loops and G–N–CRE loops) also influences the expression level of genes (Fig. 5K and Supplementary Fig. 24). Limited by the fundamental principles of in situ Hi-C, multiplex chromatin interaction methods, such as Pore-C, ChIA-Drop, and SPRITE, could be more suitable for further study of multiple-step regulation[99–101]. Ultimately, we integrated multiple loop-related factors and analyzed their effects on the biased expression of homoeologous genes. Our findings show that the number of the loops is the primary factor influencing biased expression of homoeologous genes, followed by the expression level of genes and CREs (Fig. 5L).

Overview of subgenome divergence and tissue heterogeneity

Genome allopolyploidization allows different subgenomes to be coresident in one nucleus, resulting in the formation of specific and conserved CREs for homoeologous genes[102,103]. Several studies have illustrated that CREs can influence subgenome divergence[6,14]. However, the regulatory mechanisms underlying this process are still worth exploring. By conducting a BLAST analysis on CRE sequences from the At/Dt subgenomes, we inferred that 31.63% of the CREs were homoeologous (Supplementary Data 13). By analyzing the direct spatial interactions between homoeologous CREs and homoeologous genes, we discovered that an average of 6.19% of homoeologous biased genes (type1) exhibited loss of the loops and absence of CRE sequences in a specific subgenome (Fig. 6A and Supplementary Fig. 25A). We subsequently analyzed four regulatory patterns for homoeologous genes and functional elements, based on sequence conservation and the diversity of loop interactions. Our observations indicated that the absence of the loops (type2) and interaction difference of loops (type3) serves as significant influencing factors in the biased expression of homoeologous (Fig. 6B, C and Supplementary Fig. 25A). Additionally, a total of 55,625 loops were identified as homoeologous loops, in which two anchors contained homoeologous sequences between the two subgenomes (Fig. 6D and Supplementary Fig. 25B).

Widespread rearrangement of multilevel 3D genome architectures was observed among the various tissue. Our observations reveal that 18.97% of the A/B compartment and 78.62% of the sub-compartments exhibit dynamics across 12 different tissues (Fig. 6E). Furthermore, there was an average difference of 10.36% in TAD-like domains and 92.02% of loops associated genes across different tissues. Additionally, we found a strong consistency in the large-scale 3D genome architectures of the A/B compartment, sub-compartments and TAD-like domains between fiber_10DPA and fiber_20DPA. A

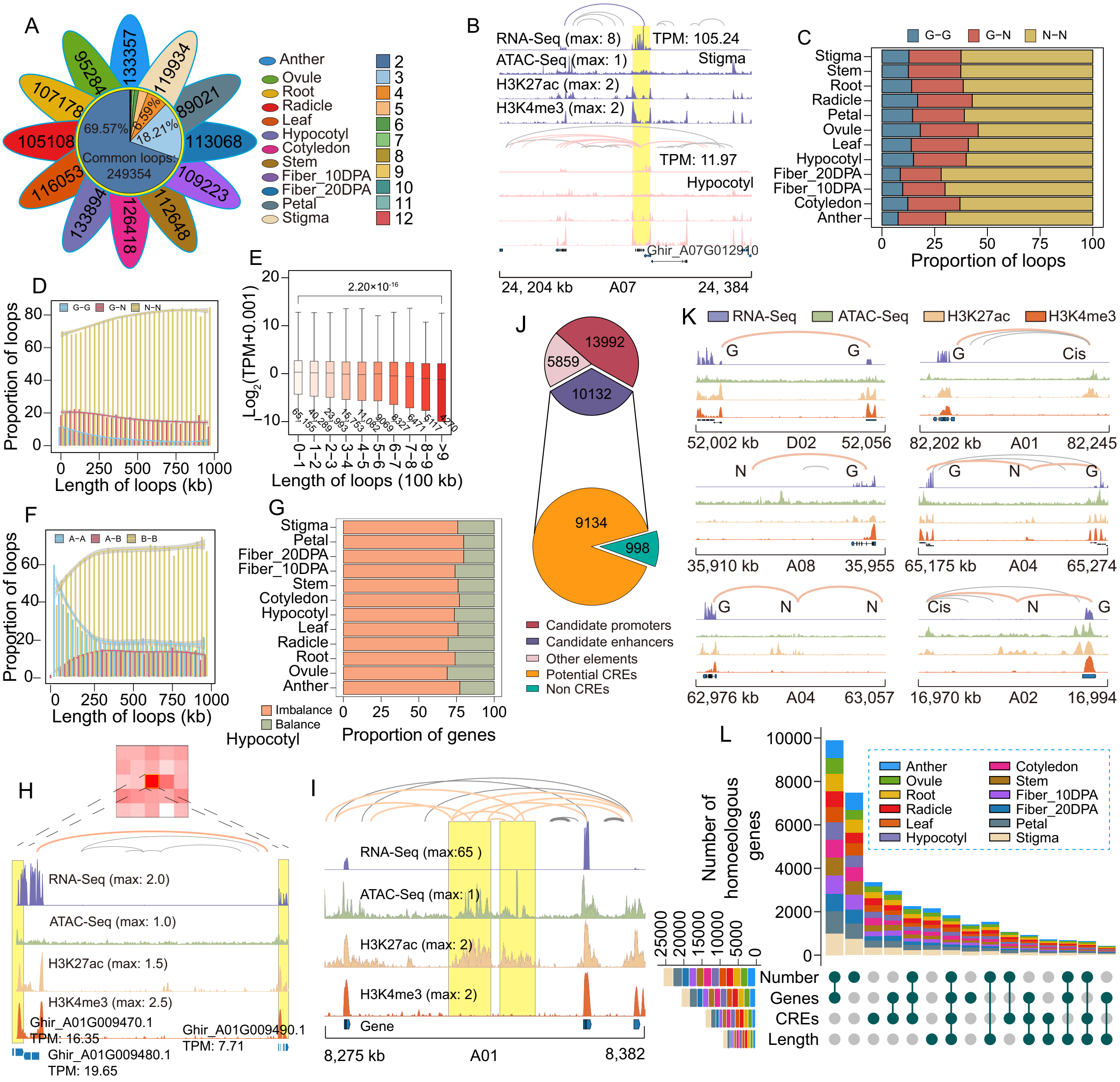

Fig. 5 | Chromatin loops and gene expression. A Petal diagram showing the number of tissue-specific loops and the pie plot showing proportion of common loops among multiple tissues. **B** An example showing the influence of excess loops for gene expression. Yellow bar denotes excess loops. **C** Bar plot showing the proportion of three types of the loops (G–G loop, G–N loop and N–N loop). G represents gene-inclusive loop anchors and N, gene-exclusive. **D** The bar plot showing the relationship between the length of the loops and the proportion of different types of loops. **E** The box plot showing the relationship between the length of the loops and the expression level of genes located in loop anchors. Two-sided Wilcoxon signed-rank test. Center line, median; box limits, first and third quartiles; whiskers, 1.5 × interquartile range. **F** The bar plot showing the relationship between the length of loops and the ratio of loops types. A–A refers to the genes located in the different loop anchors with A compartment status, whereas B–B represents genes in both anchors of loops with B compartment status. A–B refer to genes in different loops ends with different compartment statuses. **G** Bar plot showing the proportion of gene pairs linked by G–G loops. "Imbalance" indicates the gene pairs with an expression level difference of more than twofold, while "Balance" represents the opposite. **H** An example demonstrating an expression imbalance of the gene pair connected by chromatin loops. **I** The genome region showing the potential CREs, highlighted by yellow bars. **J** The pie plot displaying the ratio of candidate functional elements and potential CREs, which are candidate enhancers overlapping with loops. **K** The genome browser screenshot showing different types of loops. The orange lines represent examples of loop types. **L** The UpSet plot displaying the number of homoeologous genes with expression bias associated with chromatin loops. Data from **D** and **F** were analyzed by LOESS method, incorporated in ggplot2 form R (version 4.2.2), is utilized to fit the curve, while the shaded region signifies an interval at a 95% confidence level. Source data are provided as a Source Data file.

thorough analysis of the large-scale 3D genome architectures revealed that they influenced 5.42% (A/B compartment), 36.18% (sub-compartments) and 25.24% (TAD-like domains) of the DEGs (Fig. 6E). Additionally, multiple factors associated with loops, including the number and length of loops, CREs and the expression level of genes by looped, also influenced the differential expression of genes (Fig. 6E).

In this study, we generated multi-omics datasets from 12 tissues/developmental stages of *G. hirsutum* (Fig. 6F). Using these datasets, we constructed a cotton Encyclopedia of DNA Elements (ENCODE) database (http://cotten.hzau.edu.cn) that offers two crucial functions: visualization of multi-omics datasets in chromatin regions and searching according to chromatin region and gene ID (Supplementary

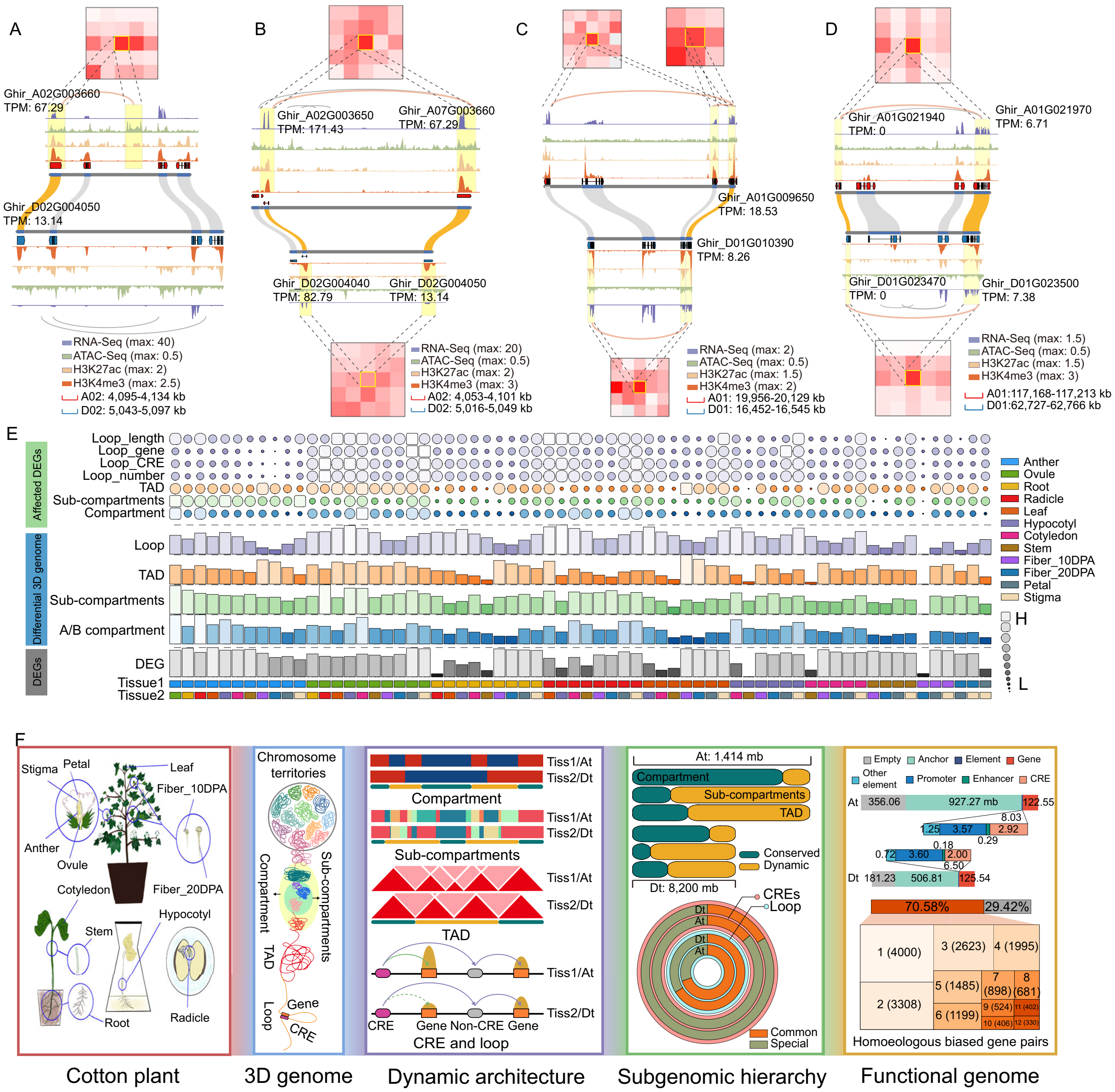

Fig. 6 | Overview of subgenome divergence and tissue difference. A An example in which deletion of the CRE sequence and loop interaction is implicated in the expression bias of homoeologous genes. **B** An example in which the loop absence between homoeologous genes is implicated in expression bias. **C** An example illustrating the changes in loop interactions between homoeologous genes is implicated in expression bias. **D** The conserved loops between homoeologous genes with no expression bias. **E** The overview plot showing the differences in multi-level 3D genome organization among tissues and the DEGs influenced by them. The bar plot displays the number of different architectures. The point plot showing the number of DEGs influenced by various structures. **F** A proposed model showing an overview of dynamic hierarchies of 3D genome organization in two subgenomes, illustrating its potential influence on the expression level of genes and homoeologous expression bias. The left panel presents the sampling process for 12 different tissues/developmental stages. The 3D genome section demonstrates multilevel 3D genome organization, with its dynamic models displayed in the dynamic architecture framework. The subgenomic hierarchy shows the stable and dynamic ratios of compartment, sub-compartment, TAD and the ratio of common and special CREs and loops. The functional genome section provides a detailed ratio of genome components within two subgenomes and the number of shared and unique homoeologous genes with biased expression in 12 tissues. The cartoon elements in the figure were designed with the aid of Adobe Illustrator and Procreate. Source data are provided as a Source Data file.

Fig. 26). In addition, our analysis revealed widespread dynamic regions of multilevel architecture among the 12 tissues that occupied 15.27%/25.22% (compartment), 78.51%/82.63% (sub-compartments) and 68.73%/65.83% (TAD-like domains) of the At/Dt subgenome, respectively (Fig. 6F and Supplementary Data 9, 11). We also found that 86.25%/83.43% of CREs and 30.32%/33.78% of the loops in the At/Dt subgenome were shared among several tissues. Ultimately, through the integration of the aforementioned results, we discovered that the potential functional regions, which encompass the final spatial structural and functional sequences, accounted for 74.82% and 77.90% of the At and Dt subgenomic length and 70.58% of homoeologous genes with biased expression in the 12 tissues (Fig. 6F).

Characterization of fiber quality-related CREs

To investigate the association between CREs and fiber quality-related traits, we compared the genomic regions of published quantitative trait loci based on copy number variation (CNV-QTL) with CREs[104]. We found that 18 CREs were located within 14 CNV-QTLs associated with agronomic traits, including fiber length (FL), fiber strength (FS), fiber elongation rate (FE), flowering date (FD), lint percent (LP), lint index (LI) and fiber maturity (MAT), of which 8 were identified as fiber-related CREs that were associated with changes in fiber traits (Fig. 7A). Of these, 7 were positively correlated and 1 was negatively correlated with phenotype. Cis6160 and Cis6164 were identified as pleiotropic CREs, both of which were associated with LI and LP, two important indicators for evaluating the yield of cotton fiber (Fig. 7A, B). Interestingly, these elements are located adjacent to each other (at a distance of 102 kb). The phenotypic values increased as the copy number of Cis6160 increased, while Cis6164 showed the opposite trend. For 7 CREs that positively regulate the formation of phenotypes, 5 CREs presented only 0 or 1 copy in >70% of accessions, which suggested great potential for improvement (Fig. 7C).

To explore more CREs related to fiber quality, we compared published quantitative trait loci based on single nucleotide polymorphisms (SNP-QTLs) with the CREs[104]. We found that 329 CREs were located within 52 SNP-QTLs associated with traits including FL, FS, FE, fiber length uniformity (FU) and fiber micronaire (FM), with only 29 identified as fiber-related CREs inducing the change in phenotype values (Fig. 7D). On the basis of the analysis of the genes, the SNP combinations within the CREs were identified as haplotypes of CREs. The haplotypes corresponding to accessions with higher phenotype values were identified as favorable haplotypes, the haplotypes corresponding to <18 (2% of 890) accessions were identified as haplotypes with insufficient accessions, and the other haplotypes were identified as unfavorable haplotypes. We found that 18 fiber-related CREs were present as favorable haplotypes in <70% of improved accessions, which indicated that accessions can be improved by editing CREs from those of unfavorable haplotypes to those of favorable haplotypes (Fig. 7D, E). The phenotype values increased as favorable haplotypes accumulated, confirming the feasibility of enhancing fiber quality by inserting favorable CREs (Fig. 7F and Supplementary Fig. 27).

The target genes of fiber-related CREs can be predicted by Hi-C loops, and published expression quantitative trait loci (eQTLs) can assist in screening more accurate target genes[105]. For example, a FU-related *cis*-element (Cis7599) regulated Ghir_D08G021330 (*APX1*) in developing fibers at 20 DPA and was validated with both Hi-C loops and eQTLs (Fig. 7G). *APX1* has been found to participate in the lignin (one component of fiber) biosynthesis pathway in *Arabidopsis*[106]. The SNPs within Cis7599 were associated with the expression of *APX1*, which suggested that Cis7599 may regulate fiber development by regulating the expression of key genes (Fig. 7G, H). In total, we identified 37 fiber-related CREs and 31 candidate genes that can serve as resources for cotton fiber improvement (Fig. 7I and Supplementary Data 14).

Discussion

Following allopolyploidy, the subgenomes of plants have undergone gradual divergent evolution as an adaptive response to variable environments[107]. Subgenome divergence has been observed in various plant species[11,108]. A comparison of cultivated and wild cotton has revealed an expression bias toward the D subgenome in fiber development during domestication[109]. An in-depth study of the reasons revealed that the global difference in *cis/trans* regulatory elements was associated with the expression bias of homoeologous genes[110]. A comprehensive and in-depth study of subgenomes contributes to understanding adaptive evolutionary processes in polyploid plants. However, there is currently a lack of systematic research on the collaboration and divergence of subgenomes across various stages of development in polyploid plants. In this study, we integrated various omics datasets generated from 12 tissues of allotetraploid cotton to comprehensively elucidate the dynamic regulatory role of the two subgenomes from a spatial genome and epigenome perspective.

Extensive research has been conducted on multiple hierarchical spatial structures, including compartments, sub-compartments, topologically associating domains (TADs), and loops[25,111]. The reorganization of the multi-level 3D genome architecture can impact the interaction between CREs and genes, leading to changes in the regulatory network of genes[112–114]. Studies on the spatial architecture differences between subgenomes have contributed to understanding subgenome divergence[115]. However, studies on the regulatory divergence of the 3D genome between subgenomes are uncommon[116,117]. In this study, we compared and analyzed the multi-level 3D genome structures of subgenomes in multiple tissues. We discovered widespread differences in the 3D genome organization between subgenomes across various tissues. The compartment and sub-compartment status in the Dt subgenome exhibited more dynamics than that in the At subgenome. The sub-compartment status was strongly associated with the TPGs, contributing to our understanding of the roles played by dynamic 3D genome structures in specific biological processes. Moreover, we identified a subset of TS-TADs and isolated TADs and revealed the region of TS-TADs with both high 1D (gene expression) and 3D (sub-compartment status) activity in specific tissues. The isolated TADs demonstrated high dynamics among tissues and displayed a preference within the At subgenome. The investigation of the factors contributing to the formation of stable isolated TADs during the evolutionary process is necessary and valuable. To gain a comprehensive understanding of the complex and intricate interactions of loops, it is necessary to thoroughly investigate the functional regulation of multi-step loops between CREs and genes using multiple chromatin interaction capture techniques.

Changes in chromatin accessibility and epigenome modifications are typically accompanied by alterations in the expression levels of regulated genes[17,118,119]. A previous analysis of chromatin accessibility, histone modifications, and sequence conservation in 13 plant species revealed the prevalence of distal CREs in plants[120]. The natural variation in *cis*-elements influences the expression bias (Biased At subgenome) of PRE1, which acts as a positive regulator of fiber elongation in allotetraploid cotton[121]. Analysis of the landscape of accessible chromatin regions (ACRs) during wheat grain development has revealed that the TF bonding sites exhibit subgenome specificity[122]. The sequences of active transcriptional regulatory elements display accessible chromatin, TF binding, low DNA methylation, and histone modifications such as acetylation or methylation of specific lysine residues in histone H3[39]. In this study, we constructed a comprehensive landscape of the dynamic regulation of active functional elements in allotetraploid cotton across 12 different tissues, revealing that the absence or presence of active histone modification and an open chromatin state can impact the biased expression of homoeologous genes. Additionally, variations in motif sequences and loop interactions between two subgenomes can also affect homoeologous expression bias. Similar to the finding that sequence variations in CREs have the potential to lead to phenotypic changes[123,124], we also found that a portion of the SNPs associated with fiber quality were situated within the identified CRE regions.

Understanding the differences in cellular and tissue functions requires studying the regulation of specifically expressed genes, particularly when investigating tissue-specific diseases prevalent in both animals and plants, as well as tissue specificity during developmental stages[69,125–127]. The polyploidization process changes various aspects, including tissue morphology, cell types, biological functions, and more[107,128,129]. Many studies have demonstrated that polyploid organisms exhibit excellent performance in tissue repair and regeneration, which may be attributed to the functional redundancy of homoeologous genes within the subgenomes[130]. Moreover, homoeologous gene pairs exhibit specialized and biased expression in different

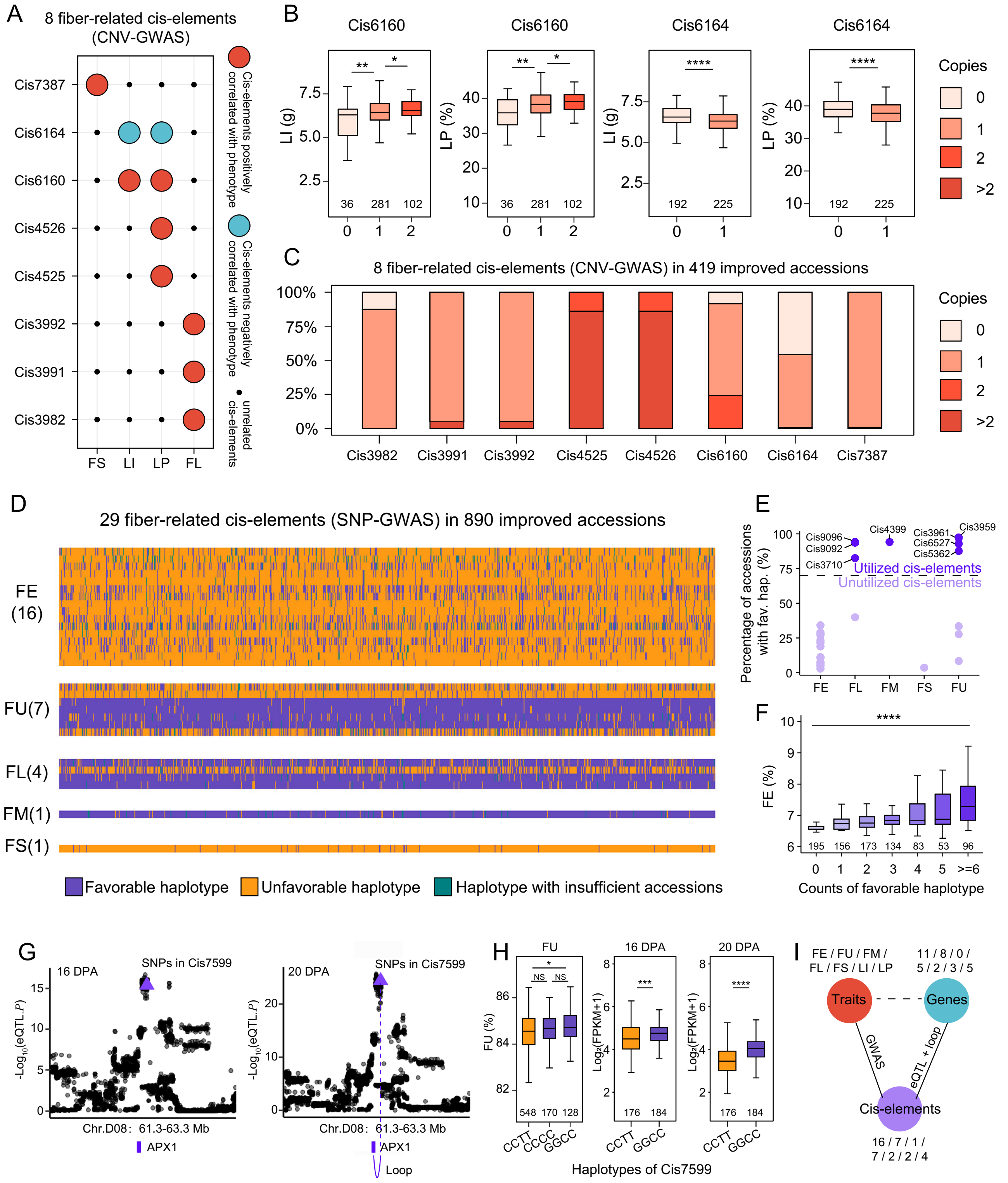

Fig. 7 | Fiber-related CREs. **A** Eight fiber-related CREs identified with CNV-GWAS. Red points represent the CREs positively correlated with phenotype. Green points refer to the CREs negatively correlated with phenotype. Black points indicate the CREs unrelated to phenotype. **B** Box plots of 2 pleiotropic CREs. *X*-axis represents the counts of CRE copies, while *Y*-axis represents the phenotype value. Two-sided Wilcoxon signed-rank test, $^{****}p < 0.0001$; $^{**}p < 0.01$; $^{*}p < 0.05$. **C** The different distributions of 8 fiber-related CREs in improved accessions. The colors of the bar indicate cotton accessions with different CRE copies, as in (**B**). **D** The haplotype distributions of 29 fiber-related CREs in 890 improved accessions. These CREs were identified with SNP-GWAS. Insufficient accessions, count of accessions <18 (2% of 890). **E** The utilized CREs and unutilized CREs. **F** The phenotype value of accessions with different counts of favorable haplotypes. Two-sided Wilcoxon signed-rank test, $^{***}p < 0.0001$. **G** Manhattan plot of the *APX1* gene in fibers at 16 DPA and 20 DPA. *X*-axis represents the location of each SNP and *Y*-axis represents the *p* value of eQTL analysis. Black points represent normal SNPs, while purple triangles indicate SNPs in the fiber-related CRE Cis7599. The *p* values were calculated by fastQTL, which applies the "qvalue" packages of R (version 4.2.2). **H** The phenotype value and gene expression of accessions with different haplotypes. Two-sided Wilcoxon signed-rank test, $^{***}p < 0.0001$; $^{**}p < 0.001$; $^{*}p < 0.05$; NS, $p > 0.05$. **I** A pattern diagram showing the counts of fiber-related CREs and genes regulated by these CREs. In (**B**, **F**, **H**), Center line, median; box limits, first and third quartiles; whiskers, 1.5 × interquartile range. Source data are provided as a Source Data file.

tissues, suggesting rapid subfunctionalization of subgenomes[131]. The primary aim of the Genotype–Tissue Expression (GTEx) project is to improve our understanding of how genetic variants impact gene regulation in various tissues and complex traits[132]. However, this project has had limited prior implementation within plant studies. In this study, we have investigated the subgenome divergence specific to tissues by integrating multiple omics datasets. We utilized the TAU method, an outstanding approach for identifying TPGs[133]. The comparison of TAU curves among human, mouse and cotton demonstrates that it is applicable in plants. The surrounding regions of these TPGs exhibit a strong signal of active histone modification and belong to the high A sub-compartments, indicating variations in the dynamics of the local gene microenvironment among different tissues. Additionally, we observed that homoeologous TPGs display a pronounced biased expression trend, with pattern 4 TPGs exhibiting significantly higher biased expression, which further supports the subfunctionalization of subgenomes.

The primary purpose of the Encyclopedia of DNA Elements (ENCODE) project is to annotate all potential functional elements across entire genomes[134]. Currently, only a few ENCODE projects have been initiated for plant species, with a focus on aggregating published datasets[135–138]. In this study, we initiated the cotton ENCODE project, in which potential functional elements in the entire genome were annotated by utilizing multi-omics datasets generated from twelve cotton tissues. This study provides mapping of 3D genome structure and spatial chromatin interaction linking genes and identifies candidate functional elements. This also provides a comprehensive understanding of the dynamics of subgenomic coordination and divergence, thereby fostering advancements in cotton breeding.

Methods

Collection of cotton tissue samples

Samples of twelve different tissues were collected from *G. hirsutum* J668, which was cultivated in the field of Huazhong Agricultural University and consistently managed throughout 2022. Young leaves at the apex of vigorously growing plants were plucked The petals, stigmas and anthers were collected from the flowers on the day after blooming. The ovules, fiber_10DPA and fiber_20DPA were stripped from cotton bolls on the day of flowering (0 day post anthesis, 0 DPA), 10 DPA, and 20 DPA, respectively. All the aforementioned tissues were obtained from field-grown cotton plants. The cotyledons, stems, and roots were obtained from cotton plants grown in pots for approximately 14 days. The hypocotyls were collected from cotton plants cultured in a sterile environment for approximately 7 days. The radicles were collected from cotton plants sprouting in petri dishes for approximately 3 days. All tissues were rapidly immersed in liquid nitrogen after collection and subsequently stored in a −80 °C refrigerator for future use.

RNA-Seq and gene expression analysis

We take about 50–100 mg of cotton tissue (fiber tissue requires 100–300 mg) for RNA seq experiments, and use the RNAprep Pure polysaccharide polyphenol plant total RNA extraction kit (TIANGEN; DP441) to collect total RNA from the tissue. The RNA-Seq libraries were constructed using the DNBSEQ platform with DNBSEQ-T7 module. These libraries were mapped to the reference genome of *G. hirsutum* utilizing HISAT2 (version 2.2.1) with default settings[139,140]. The mapping results were sorted, filtered (-q 25) and converted from SAM to BAM format using SAMtools (version 1.9)[141]. The expression levels of genes were calculated using StringTie (version 2.1.4) with parameter settings (--fr -e -G)[142] and quantified based on TPM.

Comprehensive transcriptome analysis

The differential expression of genes ($|\log_2(\text{fold change})| \geq 1$ and $p\text{adj} \leq 0.01$) among tissues was calculated using DESeq2 (version 1.38.3) with default parameters[143]. The homoeologous genes were defined repeatedly in subgenomes comparisons using the BLASTP function from DIAMOND (version 0.9.27) with default settings[144]. The biased expression of homoeologous genes was needed to meet the condition that $|\log_2(\text{FC})| \geq 1$, $p\text{adj} \leq 0.01$, $|\log_2(\text{TPM-At/TPM-Dt})| \geq 1$ and $\text{TPM\_(At/Dt)} > 0.5$. The TAU values of genes were calculated using TAU algorithms, which have demonstrated excellent performance in previous studies[70,133]. We identified the TPGs with TAU ≥ 0.85:

$$\text{TAU}(\tau) = \frac{\sum_{i=1}^{N}(1-\hat{x}_i)}{N-1};\ \hat{x}_i = \frac{x_i}{\max_{1\leq i\leq N}(xi)} \tag{1}$$

where N represents the number of tissues, and x_i represents the log(TPM) value for each tissue. To mitigate noise resulting from lowly expressed genes, we replaced values where log(TPM) ≤ 0 with 0.0001.

The genes were clustered using UMAP algorithms based on the transcripts per kilobase million (TPM) values that were z score normalized[71]. The WGCNA of genes was built using WGCNA tools (version 1.71) with default settings[145–147]. Gene Ontology (GO) enrichment analysis was performed using a two-sided Fisher's exact test, and GO terms with a false discovery rate (FDR) of less than 0.05 were selected.

ChIP-Seq and ATAC-Seq data analysis

The tissue powders (~1.5 g) were fixed in NI buffer containing 1% (v/v) formalde-hyde for 10 min. Glycine terminated the fixation process. Purify the nuclei using the miracloth and 20 µm nylon filter membrane. The purified nucleus was subjected using a lysis buffer containing 1% SDS to lysis. Fragment the DNA to approximately 500 base pairs by ultrasound. Mix dynabeads protein A beads (ThermoFisher;10001D), 5 µl antibody (1ug/ul) and DNA fragment, then incubate for 15–20 h. Subsequently, wash and purify the DNA through both aqueous and organic phases. The purified DNA fragments were amplified by PCR and finally sequenced using the DNBSEQ-T7 sequencing mode of the DNBSEQ sequencing platform. Two antibodies, H3K27ac (Abcam, ab4729) and H3K4me3 (Abcam, ab8580), were used in this study.

Cotton tissues (~1 g) were ground into a powder in liquid nitrogen and incubated with lysis buffer at 4 °C for 10 min. The mixture was filtered through a 40 µm strainer, and the cell suspension was layered onto a 60% Percoll solution, then centrifuged at 1800 × *g* for 20 min. The intermediate phase was extracted, diluted with RSB buffer, and centrifuged at 1000 × *g* for 10 min to collect nuclei, which were washed with RSB buffer. Approximately 50,000 nuclei underwent tagmentation in the transposition solution. Tn5-transposed DNA was purified using AMPure DNA magnetic beads and PCR amplification. The final library was sequenced on the PE150 system.

The input datasets of ChIP-Seq from the different tissues can be utilized to infer the blacklist regions in the *G. hirsutum* genome by employing Greenscreen (version 1.0.0) with the default parameter[72]. TrimGalore (version 0.6.6)[148] was used to filter the low-quality reads and remover the adapter with parameter settings (-q 20 –phred33 –stringency 3 –length 20 –paired –gzip –trim-n). The clean data were mapped to the reference genome of *G. hirsutum* using Bowtie2 (Version 2.4.4)[149]. SAMtools (version 1.9) was used to sort the mapping results and convert the data format (BAM to SAM)[141]. The ChIP-Seq and ATAC-Seq peaks were called using MACS2 (version 2.2.7.1) with parameter settings (ChIP-Seq: -B -q 0.01, ATAC-Seq: -B –SPMR –extsize 38 –shift -15 –call-summits -p 0.005)[150]. The IDR (Irreproducible Discovery Rate, version 2.0.3) was employed with specific parameter settings (--input-file-type narrowPeak) to infer the high-confidence ChIP-Seq peaks (IDR ≤ 0.05)[151]. The final set of peaks from ATAC-Seq consisted of peaks with IDR ≤ 0.05 and the recalled peaks. In brief, the recalled peaks are those that are repeatedly identified in multiple tissues, exhibit overlap between duplicate samples, and intersect with peaks having an IDR ≤ 0.05. Any peaks that overlapped with the identified blacklist regions were removed. The ChIP-Seq and ATAC-Seq peaks were annotated by using ChIPseeker (version 1.34.1) with default settings[152]. The signal distribution and normalization (BPM) of ChIP-Seq and ATAC-Seq were analyzed using deepTools (version 3.5.0)[153].

Inferring footprints from ATAC-Seq datasets

The footprints were identified using HINT (Hmm-based Identification of Transcription factor footprints) with default settings[76]. The DNA motifs used for analysis were obtained from the JASPAR Plantae database[154]. The motif enrichment was analyzed using HINT, and the regions of the motif were annotated using ChIPSeeker (version 1.3.41)[152]. The genomic region plot was generated using pyGenomeTracks[155].

Identification of PATREs and candidate functional elements

The method of identifying the PATREs was modified from identifying the PADREs (predicted ATAC-Seq-supported developmental regulatory elements)[77]. First, the ATAC-Seq peak summit is extended for 1050 bp in each direction and split into 21 bins. Subsequently, the signal strength of ChIP-Seq (H3K4me3 and H3K27ac) and ATAC-Seq for each bin was calculated. Finally, the PATRE region dimensionality ranged from 252 to 2 using the UMAP algorithm. We added tags for the PATREs based on whether they were located in the promoter region and whether they exhibited a ChIP-Seq signal. The ChIP-Seq and ATAC-Seq signals were calculated using deepTools (version 3.5.0)[153]. The candidate functional elements were merged from the ATAC-Seq peak regions of all the tissues that overlapped with PATREs.

Evaluation of the Hi-C library

The nuclear fixation and purification steps are consistent with ChIP-seq. The purified nuclear was centrifuged and gently resuspended in DpnII buffer containing 0.16% SDS, then incubated at 62 °C for 7–10 min. After digestion with the DpnII restriction enzyme, biotin labeling, ligation, and reverse cross-linking. One to 1.5 µg of DNA was used for further processing, including removing end markers, ultrasonic fragmentation, end repair, and fragment selection. DNA fragments were purified using Dynabeads MyOne Streptavidin T1 (Invitrogen; 65601) magnetic beads, and libraries were constructed using the VAHTS Universal DNA Library Construction Kit (vazyme; NDM607). The libraries were sequenced on the DNBSEQ platform with DNBSEQ-T7 module.

The HiC-Pro pipeline (version 2.11.1) was utilized for constructing the chromosome interaction matrix with 2, 3, 4, 5, 10, 20, 40, 50, 100, and 200 kb resolution[156]. Two methods are employed to determine the appropriate resolution for the Hi-C matrix. First, the proportion of bins containing more than 1000 reads (more than 80% indicates qualification)[26]. Second, the resolution for each Hi-C library was predicted based on the valid read pairs using HiCRes (version 1.0.0) with default settings[157]. The consistency of replicate libraries was assessed using the 3Dchromatin_ReplicateQC pipeline (version 1.0.1), which includes the GenomeDISCO, HiCRep, HiC-Spector, and QuASAR-Rep tools[158].

Identification of the A/B compartment and sub-compartments

To increase the efficiency, we split the 3D genome interaction matrix into single chromosome interactions. We identified the compartment region by using cworld (version 1.0.1) at 100 kb resolution with default parameters[159]. The status of sub-compartments can be inferred by using CALDER (version 0.7) with default parameters, which will provide the status of sub-compartments at 10 kb[27]. The CALDER defined eight sub-compartments statuses (A.1.1, A1.2, A2.1, A2.2, B1.1, B1.2, B2.1, B2.2) ranging from activity to inhibition. These sub-compartment statuses have been renamed A1, A2, A3, A4, B1, B2, B3, and B4.

Analysis of the TAD-like domains

The TAD-like domains were identified using HiTAD (version 1.0.0) with default parameters at 20 kb resolution[83]. In this process, the toCool function from HiCpeaks (version 0.3.5) was used to convert the matrix format generated from the HiC-Pro pipeline into the cool format[160]. The Aggregate TAD Analysis plot was produced using GENOVA tools (version 0.94) with default parameters[95]. The final TAD-like domains with the lowest hierarchy levels were used for further analysis. Tissue-specific highly active TAD-like domains (TS-TADs) contain TPGs and exhibit significantly higher expression levels of genes in special tissues. In this process, we initially chose TAD-like domains containing TPGs. Then, we compared the gene expression levels within these TAD-like domains across different tissues. Finally, we identified the TS-TADs that showed statistical significance in gene expression values using a *t*-test analysis with a p value ≤ 0.05. The TAD-like domain boundary region is the portion that starts from the boundary site and occupies 40% of the entire TAD-like domain[161]. The definition of a conserved TAD-like domain boundary is very strict. The boundary sites of the TAD-like domains were extended by 40 kb on both the left and right sides. These extended regions were considered conserved and exhibited overlap between tissues.

Identification of loops

The final loops dataset consists of the loops that are repeatedly identified and those with the top 25% confidence using Fit-Hi-C (version 2.0.8), Mustache (version 1.0.1), Chromosight (version 1.6.3) and Peakachu (version 1.2.0) at 3 kb resolution[90,92–94]. The loops with high confidence (FDR ≤ 0.001 and contactCount ≥10) were identified using Fit-Hi-C. The loops were identified using Mustache with parameter settings (p value ≤ 0.01). The loops were identified using Chromosight with parameter settings (--min-dist 6000 --max-dist 3000000 --iterations 5 --Pearson 0.1). The Peakachu training dataset comprises 3000 loops, the vast majority of which are double-recognized by Fit-HiC, Mustache and Chromsight. The loops were identified by Peakachu with the parameter settings (-r 3000 -t 0.85). The Aggregate Peak Analysis plot was produced using GENOVA tools (version 0.94) with default parameters[95].

Inferring of homoeologous CREs

The potential distal CREs were characterized as candidate enhancer regions overlapping with loop anchors. The At-subgenomic homoeologous CRE regions in the Dt subgenome were identified using BLASTX (version 2.9.0) with the parameter settings (-evalue 1e−5)[162]. The homoeologous CREs needed to exhibit an overlap of over 50% with the homoeologous regions identified in the At subgenome.

Identification of fiber-related CREs

The published fiber-related CNV-QTLs[104] were collected. The CREs that had >50% of length overlapped with CNV-QTLs were retained for further analysis. As the CNVs were longer than CREs (average length: CNVs, ~800 bp; CREs, ~400 bp), the copy counts of the CREs were assessed by the overlapping CNVs. For example, the genotype of CNV was 2, indicating 2 copies of relative CREs. Only if the phenotype values were significantly different (Two-sided Wilcoxon signed-rank test, $p < 0.05$) between accessions with different CRE copy counts, the CREs were identified as fiber-related CREs.

The published fiber-related SNP-QTLs[104] were collected. The CREs that >50% of length overlapped with SNP-QTLs were retained for further analysis. As SNPs were shorter than CREs (average length: SNPs, 1 bp; CREs, ~400 bp), the haplotype of the CREs was assessed by the overlapping SNP combinations. Only if phenotype values were significantly different (LSD.test in R, $p < 0.05$) between accessions with different haplotypes of CREs, the CREs were identified as fiber-related CREs. The haplotypes corresponding to accessions with higher phenotype values were identified as favorable haplotypes, the haplotypes corresponding to <20 accessions were identified as haplotypes with insufficient accessions, and the other haplotypes were identified as unfavorable haplotypes.

Statistics and reproducibility

Two-sided Wilcoxon signed-rank test was used to conduct comparisons between different groups with continuous variables. Two-sided Pearson's chi-squared test was employed to compare the differences in response rates between groups. The Pearson correlation coefficient was used to evaluate the reproducibility of RNA-Seq, ChIP-Seq and ATAC-Seq datasets. All statistical analyses were done using standard R packages (R, 4.2.2). The reproducibility of Hi-C dataset was accessed using the 3Dchromatin_ReplicateQC pipeline (version 1.0.1).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All the raw sequencing data generated during the current study are available in the NCBI BioProject database under accession number PRJNA1041939 (ATAC-Seq: SAMN38304764, Hi-C: SAMN38304760, H3K27ac: SAMN38304763, H3K4me3: SAMN38304762, RNA-Seq: SAMN38304761). The processed Hi-C matrices and epigenome maps can be visualized in the cotton ENCODE database (http://cotten.hzau.edu.cn). Source data are provided with this paper.

References

1. Liu, B. & Wendel, J. F. Epigenetic phenomena and the evolution of plant allopolyploids. *Mol. Phylogenetics Evol.* **29**, 365–379 (2003).
2. Soltis, P. S., Marchant, D. B., Van de Peer, Y. & Soltis, D. E. Polyploidy and genome evolution in plants. *Curr. Opin. Genet. Dev.* **35**, 119–125 (2015).
3. Song, Q., Zhang, T., Stelly, D. M. & Chen, Z. J. Epigenomic and functional analyses reveal roles of epialleles in the loss of photoperiod sensitivity during domestication of allotetraploid cottons. *Genome Biol.* **18**, 99 (2017).
4. Alix, K., Gérard, P. R., Schwarzacher, T. & Heslop-Harrison, J. S. (Pat). Polyploidy and interspecific hybridization: partners for adaptation, speciation and evolution in plants. *Ann. Bot.* **120**, 183–194 (2017).
5. Wang, M. et al. Evolutionary dynamics of 3D genome architecture following polyploidization in cotton. *Nat. Plants* **4**, 90–97 (2018).
6. Han, J. et al. Genome-wide chromatin accessibility analysis unveils open chromatin convergent evolution during polyploidization in cotton. *Proc. Natl Acad. Sci. USA* **119**, e2209743119 (2022).
7. Li, X. et al. Genomic rearrangements and evolutionary changes in 3D chromatin topologies in the cotton tribe (Gossypieae). *BMC Biol.* **21**, 56 (2023).
8. Gaut, B. S., Le Thierry d'Ennequin, M., Peek, A. S. & Sawkins, M. C. Maize as a model for the evolution of plant nuclear genomes. *Proc. Natl Acad. Sci. USA* **97**, 7008–7015 (2000).
9. Soltis, P. S. & Soltis, D. E. The role of genetic and genomic attributes in the success of polyploids. *Proc. Natl Acad. Sci. USA* **97**, 7051–7057 (2000).
10. Feldman, M. & Levy, A. A. Allopolyploidy—a shaping force in the evolution of wheat genomes. *Cytogenet. Genome Res.* **109**, 250–258 (2005).
11. Zhao, M., Zhang, B., Lisch, D. & Ma, J. Patterns and consequences of subgenome differentiation provide insights into the nature of paleopolyploidy in plants. *Plant Cell* **29**, 2974–2994 (2017).
12. Zhang, L. et al. The ancient wave of polyploidization events in flowering plants and their facilitated adaptation to environmental stress. *Plant Cell Environ.* **43**, 2847–2856 (2020).
13. Román-Palacios, C., Molina-Henao, Y. F. & Barker, M. S. Polyploids increase overall diversity despite higher turnover than diploids in the Brassicaceae. *Proc. R. Soc. B Biol. Sci.* **287**, 20200962 (2020).
14. Wang, M. et al. An atlas of wheat epigenetic regulatory elements reveals subgenome divergence in the regulation of development and stress responses. *Plant Cell* **33**, 865–881 (2021).
15. Zhang, R.-G. et al. Subgenome-aware analyses suggest a reticulate allopolyploidization origin in three Papaver genomes. *Nat. Commun.* **14**, 2204 (2023).
16. Wang, L. et al. Altered chromatin architecture and gene expression during polyploidization and domestication of soybean. *Plant Cell* **33**, 1430–1446 (2021).
17. Zhao, L. et al. Dynamic chromatin regulatory programs during embryogenesis of hexaploid wheat. *Genome Biol.* **24**, 7 (2023).
18. Wang, L. et al. DNA hypomethylation in tetraploid rice potentiates stress-responsive gene expression for salt tolerance. *Proc. Natl Acad. Sci. USA* **118**, e2023981118 (2021).
19. Mehlferber, E. C. et al. Polyploidy and microbiome associations mediate similar responses to pathogens in Arabidopsis. *Curr. Biol.* **32**, 2719–2729.e5 (2022).
20. Yang, C. et al. Evolution of physiological responses to salt stress in hexaploid wheat. *Proc. Natl Acad. Sci. USA* **111**, 11882–11887 (2014).
21. Li, Z. et al. The bread wheat epigenomic map reveals distinct chromatin architectural and evolutionary features of functional genetic elements. *Genome Biol.* **20**, 139 (2019).
22. Jia, J. et al. Homology-mediated inter-chromosomal interactions in hexaploid wheat lead to specific subgenome territories following polyploidization and introgression. *Genome Biol.* **22**, 26 (2021).
23. Ma, H., Lin, J., Mei, F., Mao, H. & Li, Q. Q. Differential alternative polyadenylation of homoeologous genes of allohexaploid wheat ABD subgenomes during drought stress response. *Plant J.* **114**, 499–518 (2023).
24. Li, M. et al. DNA methylation role in subgenome expression dominance of Juglans regia and its wild relative J. mandshurica. *Plant Physiol.* **193**, 1313–1329 (2023).
25. Domb, K., Wang, N., Hummel, G. & Liu, C. Spatial features and functional implications of plant 3D genome organization. *Annu. Rev. Plant Biol.* **73**, 173–200 (2022).
26. Rao, S. S. P. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
27. Liu, Y. et al. Systematic inference and comparison of multi-scale chromatin sub-compartments connects spatial organization to cell phenotypes. *Nat. Commun.* **12**, 2439 (2021).
28. Parada, L. A., McQueen, P. G. & Misteli, T. Tissue-specific spatial organization of genomes. *Genome Biol.* **5**, R44 (2004).
29. Chen, Y. et al. Genome-wide high-throughput chromosome conformation capture analysis reveals hierarchical chromatin interactions during early somatic embryogenesis. *Plant Physiol.* **193**, 555–577 (2023).
30. Dong, P. et al. Tissue-specific Hi-C analyses of rice, foxtail millet and maize suggest non-canonical function of plant chromatin domains. *J. Integr. Plant Biol.* **62**, 201–217 (2020).
31. Beagan, J. A. & Phillips-Cremins, J. E. On the existence and functionality of topologically associating domains. *Nat. Genet.* **52**, 8–16 (2020).
32. Tena, J. J. & Santos-Pereira, J. M. Topologically associating domains and regulatory landscapes in development, evolution and disease. *Front. Cell Dev. Biol.* **9**, 702787 (2021).
33. Karpinska, M. A. & Oudelaar, A. M. The role of loop extrusion in enhancer-mediated gene activation. *Curr. Opin. Genet. Dev.* **79**, 102022 (2023).
34. 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).

35. 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).
36. Oudelaar, A. M. & Higgs, D. R. The relationship between genome structure and function. *Nat. Rev. Genet.* **22**, 154–168 (2021).
37. Marand, A. P., Eveland, A. L., Kaufmann, K. & Springer, N. M. cis-Regulatory elements in plant development, adaptation, and evolution. *Annu. Rev. Plant Biol.* **74**, 111–137 (2023).
38. Andersson, R. & Sandelin, A. Determinants of enhancer and promoter activities of regulatory elements. *Nat. Rev. Genet.* **21**, 71–87 (2020).
39. Schmitz, R. J., Grotewold, E. & Stam, M. Cis-regulatory sequences in plants: their importance, discovery, and future challenges. *Plant Cell* **34**, 718–741 (2022).
40. Deng, L. et al. 3D organization of regulatory elements for transcriptional regulation in Arabidopsis. *Genome Biol.* **24**, 181 (2023).
41. Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**, 1497–1502 (2007).
42. Haring, M. et al. Chromatin immunoprecipitation: optimization, quantitative analysis and data normalization. *Plant Methods* **3**, 11 (2007).
43. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
44. Lu, Z., Hofmeister, B. T., Vollmers, C., DuBois, R. M. & Schmitz, R. J. Combining ATAC-seq with nuclei sorting for discovery of cis-regulatory regions in plant genomes. *Nucleic Acids Res.* **45**, e41 (2017).
45. Yang, H. et al. A map of cis-regulatory elements and 3D genome structures in zebrafish. *Nature* **588**, 337–343 (2020).
46. Son, K. H. et al. Integrative mapping of the dog epigenome: reference annotation for comparative intertissue and cross-species studies. *Sci. Adv.* **9**, eade3399 (2023).
47. Li, Y. E. et al. An atlas of gene regulatory elements in adult mouse cerebrum. *Nature* **598**, 129–136 (2021).
48. Pan, Z. et al. An atlas of regulatory elements in chicken: a resource for chicken genetics and genomics. *Sci. Adv.* **9**, eade1204 (2023).
49. Li, E. et al. Long-range interactions between proximal and distal regulatory regions in maize. *Nat. Commun.* **10**, 2633 (2019).
50. Peng, Y. et al. Chromatin interaction maps reveal genetic regulation for quantitative traits in maize. *Nat. Commun.* **10**, 2632 (2019).
51. Zhao, L. et al. Chromatin loops associated with active genes and heterochromatin shape rice genome architecture for transcriptional regulation. *Nat. Commun.* **10**, 3640 (2019).
52. Ricci, W. A. et al. Widespread long-range cis-regulatory elements in the maize genome. *Nat. Plants* **5**, 1237–1249 (2019).
53. Concia, L. et al. Wheat chromatin architecture is organized in genome territories and transcription factories. *Genome Biol.* **21**, 104 (2020).
54. Mao, H. et al. Variation in cis-regulation of a NAC transcription factor contributes to drought tolerance in wheat. *Mol. Plant* **15**, 276–292 (2022).
55. Sijacic, P., Bajic, M., McKinney, E. C., Meagher, R. B. & Deal, R. B. Changes in chromatin accessibility between Arabidopsis stem cells and mesophyll cells illuminate cell type-specific transcription factor networks. *Plant J.* **94**, 215–231 (2018).
56. Bhadouriya, S. L., Mehrotra, S., Basantani, M. K., Loake, G. J. & Mehrotra, R. Role of chromatin architecture in plant stress responses: an update. *Front. Plant Sci.* **11**, 603380 (2021).
57. Pandey, G., Sharma, N., Sahu, P. P. & Prasad, M. Chromatin-based epigenetic regulation of plant abiotic stress response. *Curr. Genomics* **17**, 490–498 (2016).
58. Raxwal, V. K. et al. Abiotic stress-mediated modulation of the chromatin landscape in Arabidopsis thaliana. *J. Exp. Bot.* **71**, 5280–5293 (2020).
59. Zhou, P. et al. Prediction of conserved and variable heat and cold stress response in maize using cis-regulatory information. *Plant Cell* **34**, 514–534 (2022).
60. Ding, Y. et al. Targeting cis-regulatory elements for rice grain quality improvement. *Front. Plant Sci.* **12**, 705834 (2021).
61. Du, S. et al. High pressure assist-alkali pretreatment of cotton stalk and physiochemical characterization of biomass. *Bioresour. Technol.* **148**, 494–500 (2013).
62. Chbib, H., Samy, I. & Azouz, M. The economic potential of using cotton stalks to produce alternative wooden materials. In *2019 Novel Intelligent and Leading Emerging Sciences Conference (NILES 2019)* 1, 25–28 (2019).
63. Kumar, M. et al. Cottonseed: a sustainable contributor to global protein requirements. *Trends Food Sci. Technol.* **111**, 100–113 (2021).
64. Wu, M., Pei, W., Wedegaertner, T., Zhang, J. & Yu, J. Genetics, breeding and genetic engineering to improve cottonseed oil and protein: a review. *Front. Plant Sci.* **13**, 864850 (2022).
65. Viot, C. R. & Wendel, J. F. Evolution of the cotton genus, gossypium, and its domestication in the Americas. *Crit. Rev. Plant Sci.* **42**, 1–33 (2023).
66. Li, F. et al. Genome sequence of cultivated Upland cotton (Gossypium hirsutum TM-1) provides insights into genome evolution. *Nat. Biotechnol.* **33**, 524–530 (2015).
67. Wang, M. et al. Asymmetric subgenome selection and cis-regulatory divergence during cotton domestication. *Nat. Genet.* **49**, 579–587 (2017).
68. Wang, K. et al. Multi-strategic RNA-seq analysis reveals a high-resolution transcriptional landscape in cotton. *Nat. Commun.* **10**, 4714 (2019).
69. Sonawane, A. R. et al. Understanding tissue-specific gene regulation. *Cell Rep.* **21**, 1077–1088 (2017).
70. Yanai, I. et al. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* **21**, 650–659 (2005).
71. McInnes, L., Healy, J. & Melville, J. UMAP: uniform manifold approximation and projection for dimension reduction. *arXiv* https://doi.org/10.48550/arXiv.1802.03426 (2018).
72. Klasfeld, S., Roulé, T. & Wagner, D. Greenscreen: a simple method to remove artifactual signals and enrich for true peaks in genomic datasets including ChIP-seq data. *Plant Cell* **34**, 4795–4815 (2022).
73. Calo, E. & Wysocka, J. Modification of enhancer chromatin: what, how, and why? *Mol. Cell* **49**, 825–837 (2013).
74. Stillman, B. Histone modifications: insights into their influence on gene expression. *Cell* **175**, 6–9 (2018).
75. Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* **20**, 207–220 (2019).
76. Li, Z. et al. Identification of transcription factor binding sites using ATAC-seq. *Genome Biol.* **20**, 45 (2019).
77. Baranasic, D. et al. Multiomic atlas with functional stratification and developmental dynamics of zebrafish cis-regulatory elements. *Nat. Genet.* **54**, 1037–1050 (2022).
78. Pei, L., Li, G., Lindsey, K., Zhang, X. & Wang, M. Plant 3D genomics: the exploration and application of chromatin organization. *N. Phytol.* **230**, 1772–1786 (2021).
79. Dixon, J. R., Gorkin, D. U. & Ren, B. Chromatin domains: the unit of chromosome organization. *Mol. Cell* **62**, 668–680 (2016).
80. Dong, P., Tu, X., Liang, Z., Kang, B.-H. & Zhong, S. Plant and animal chromatin three-dimensional organization: similar structures but different functions. *J. Exp. Bot.* **71**, 5119–5128 (2020).

81. Tang, Z. et al. CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell* **163**, 1611–1627 (2015).
82. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
83. Wang, X.-T., Cui, W. & Peng, C. HiTAD: detecting the structural and functional hierarchies of topologically associating domains from chromatin interactions. *Nucleic Acids Res.* **45**, e163 (2017).
84. Lupiáñez, D. G. et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012–1025 (2015).
85. Szabo, Q., Bantignies, F. & Cavalli, G. Principles of genome folding into topologically associating domains. *Sci. Adv.* **5**, eaaw1668 (2019).
86. Hsieh, T.-H. S. et al. Resolving the 3D landscape of transcription-linked mammalian chromatin folding. *Mol. Cell* **78**, 539–553.e8 (2020).
87. Rowley, M. J. & Corces, V. G. Organizational principles of 3D genome architecture. *Nat. Rev. Genet.* **19**, 789–800 (2018).
88. Banigan, E. J. et al. Transcription shapes 3D chromatin organization by interacting with loop extrusion. *Proc. Natl Acad. Sci. USA* **120**, e2210480120 (2023).
89. Lohia, R., Fox, N. & Gillis, J. A global high-density chromatin interaction network reveals functional long-range and trans-chromosomal relationships. *Genome Biol.* **23**, 238 (2022).
90. Ay, F., Bailey, T. L. & Noble, W. S. Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts. *Genome Res.* **24**, 999–1011 (2014).
91. Kaul, A., Bhattacharyya, S. & Ay, F. Identifying statistically significant chromatin contacts from Hi-C data with FitHiC2. *Nat. Protoc.* **15**, 991–1012 (2020).
92. Salameh, T. J. et al. A supervised learning framework for chromatin loop detection in genome-wide contact maps. *Nat. Commun.* **11**, 3428 (2020).
93. Roayaei Ardakany, A., Gezer, H. T., 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).
94. Matthey-Doret, C. et al. Computer vision for pattern detection in chromosome contact maps. *Nat. Commun.* **11**, 5795 (2020).
95. van der Weide, R. H. et al. Hi-C analyses with GENOVA: a case study with cohesin variants. *NAR Genomics Bioinforma.* **3**, lqab040 (2021).
96. Beagan, J. A. et al. Three-dimensional genome restructuring across timescales of activity-induced neuronal gene expression. *Nat. Neurosci.* **23**, 707–717 (2020).
97. Dean, A. In the loop: long range chromatin interactions and gene regulation. *Brief. Funct. Genomics* **10**, 3–10 (2011).
98. Furlong, E. E. M. & Levine, M. Developmental enhancers and chromosome topology. *Science* **361**, 1341–1345 (2018).
99. Quinodoz, S. A. et al. Higher-order inter-chromosomal hubs shape 3D genome organization in the nucleus. *Cell* **174**, 744–757.e24 (2018).
100. Deshpande, A. S. et al. Identifying synergistic high-order 3D chromatin conformations from genome-scale nanopore concatemer sequencing. *Nat. Biotechnol.* **40**, 1488–1499 (2022).
101. Zheng, M. et al. Multiplex chromatin interactions with single-molecule precision. *Nature* **566**, 558–562 (2019).
102. Shi, X. et al. Cis- and trans-regulatory divergence between progenitor species determines gene-expression novelty in Arabidopsis allopolyploids. *Nat. Commun.* **3**, 950 (2012).
103. Hu, G. & Wendel, J. F. Cis–trans controls and regulatory novelty accompanying allopolyploidization. *N. Phytol.* **221**, 1691–1700 (2019).
104. Li, J. et al. Cotton pan-genome retrieves the lost sequences and genes during domestication and selection. *Genome Biol.* **22**, 119 (2021).
105. You, J. et al. Regulatory controls of duplicated gene expression during fiber development in allotetraploid cotton. *Nat. Genet.* **55**, 1987–1997 (2023).
106. Barros, J. et al. 4-Coumarate 3-hydroxylase in the lignin biosynthesis pathway is a cytosolic ascorbate peroxidase. *Nat. Commun.* **10**, 1994 (2019).
107. Van de Peer, Y., Ashman, T.-L., Soltis, P. S. & Soltis, D. E. Polyploidy: an evolutionary and ecological force in stressful times. *Plant Cell* **33**, 11–26 (2021).
108. Wang, Z. et al. Musa balbisiana genome reveals subgenome evolution and functional divergence. *Nat. Plants* **5**, 810–821 (2019).
109. Hovav, R. et al. Partitioned expression of duplicated genes during development and evolution of a single cell in a polyploid plant. *Proc. Natl Acad. Sci. USA* **105**, 6191–6195 (2008).
110. Yoo, M.-J., Szadkowski, E. & Wendel, J. F. Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity* **110**, 171–180 (2013).
111. Zheng, H. & Xie, W. The role of 3D genome organization in development and cell differentiation. *Nat. Rev. Mol. Cell Biol.* **20**, 535–550 (2019).
112. Li, T., Jia, L., Cao, Y., Chen, Q. & Li, C. OCEAN-C: mapping hubs of open chromatin interactions across the genome reveals gene regulatory networks. *Genome Biol.* **19**, 1–14 (2018).
113. Kumar, S., Kaur, S., Seem, K., Kumar, S. & Mohapatra, T. Understanding 3D genome organization and its effect on transcriptional gene regulation under environmental stress in plant: a chromatin perspective. *Front. Cell Dev. Biol.* **9**, 774719 (2021).
114. Batut, P. J. et al. Genome organization controls transcriptional dynamics during development. *Science* **375**, 566–570 (2022).
115. Pei, L. et al. Dynamic 3D genome architecture of cotton fiber reveals subgenome-coordinated chromatin topology for 4-staged single-cell differentiation. *Genome Biol.* **23**, 45 (2022).
116. Liao, Y. et al. The 3D architecture of the pepper genome and its relationship to function and evolution. *Nat. Commun.* **13**, 3479 (2022).
117. Ouyang, W., Xiong, D., Li, G. & Li, X. Unraveling the 3D genome architecture in plants: present and future. *Mol. Plant* **13**, 1676–1693 (2020).
118. Chen, Q., Zhang, J. & Li, G. Dynamic epigenetic modifications in plant sugar signal transduction. *Trends Plant Sci.* **27**, 379–390 (2022).
119. Liu, Z. & Kraus, W. L. Catalytic-independent functions of PARP-1 determine Sox2 pioneer activity at intractable genomic loci. *Mol. Cell* **65**, 589–603.e9 (2017).
120. Lu, Z. et al. The prevalence, evolution and chromatin signatures of plant regulatory elements. *Nat. Plants* **5**, 1250–1259 (2019).
121. Zhao, B. et al. Core cis-element variation confers subgenome-biased expression of a transcription factor that functions in cotton fiber elongation. *N. Phytol.* **218**, 1061–1075 (2018).
122. Pei, H. et al. Chromatin accessibility landscapes revealed the subgenome-divergent regulation networks during wheat grain development. *aBIOTECH* **4**, 8–19 (2023).
123. Cherry, T. J. et al. Mapping the cis-regulatory architecture of the human retina reveals noncoding genetic variation in disease. *Proc. Natl Acad. Sci. USA* **117**, 9001–9012 (2020).
124. Yocca, A. E. & Edger, P. P. Current status and future perspectives on the evolution of cis-regulatory elements in plants. *Curr. Opin. Plant Biol.* **65**, 102139 (2022).
125. Yamamoto, R. et al. Tissue-specific impacts of aging and genetics on gene expression patterns in humans. *Nat. Commun.* **13**, 5803 (2022).

126. Basu, M., Wang, K., Ruppin, E. & Hannenhalli, S. Predicting tissue-specific gene expression from whole blood transcriptome. *Sci. Adv.* **7**, eabd6991 (2021).
127. Haidoulis, J. F. & Nicholson, P. Tissue-specific transcriptome responses to Fusarium head blight and Fusarium root rot. *Front. Plant Sci.* **13**, 1025161 (2022).
128. Baker, R. L., Yarkhunova, Y., Vidal, K., Ewers, B. E. & Weinig, C. Polyploidy and the relationship between leaf structure and function: implications for correlated evolution of anatomy, morphology, and physiology in Brassica. *BMC Plant Biol.* **17**, 3 (2017).
129. Anatskaya, O. V. & Vinogradov, A. E. Polyploidy as a fundamental phenomenon in evolution, development, adaptation and diseases. *Int. J. Mol. Sci.* **23**, 3542 (2022).
130. Bailey, E. C., Kobielski, S., Park, J. & Losick, V. P. Polyploidy in tissue repair and regeneration. *Cold Spring Harb. Perspect. Biol.* **13**, a040881 (2021).
131. Adams, K. L., Cronn, R., Percifield, R. & Wendel, J. F. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proc. Natl Acad. Sci. USA* **100**, 4649–4654 (2003).
132. The Gtex Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
133. Kryuchkova-Mostacci, N. & Robinson-Rechavi, M. A benchmark of gene expression tissue-specificity metrics. *Brief. Bioinforma.* **18**, 205–214 (2017).
134. ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306**, 636–640 (2004).
135. Wang, Z. et al. AraENCODE: a comprehensive epigenomic database of Arabidopsis thaliana. *Mol. Plant* **16**, 1113–1116 (2023).
136. Xie, L. et al. RiceENCODE: a comprehensive epigenomic database as a rice Encyclopedia of DNA Elements. *Mol. Plant* **14**, 1604–1606 (2021).
137. Lü, P. et al. Genome encode analyses reveal the basis of convergent evolution of fleshy fruit ripening. *Nat. Plants* **4**, 784–791 (2018).
138. Lane, A. K. et al. pENCODE: A plant encyclopedia of DNA elements. *Annu. Rev. Genet.* **48**, 49–70 (2014).
139. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
140. Wang, M. et al. Reference genome sequences of two cultivated allotetraploid cottons, Gossypium hirsutum and Gossypium barbadense. *Nat. Genet.* **51**, 224–229 (2019).
141. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021).
142. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
143. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
144. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
145. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinforma.* **9**, 559 (2008).
146. Li, Y. et al. Gene Networks in Plant Biology: Approaches in reconstruction and analysis. *Trends Plant Sci.* **20**, 664–675 (2015).
147. Ko, D. K. et al. Network-based approaches for understanding gene regulation and function in plants. *Plant J.* **104**, 302–317 (2020).
148. Krueger, F. et al. FelixKrueger/TrimGalore: v0.6.10—add default decompression path. Zenodo https://doi.org/10.5281/zenodo.7598955 (2023).
149. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
150. Zhang, Y. et al. Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
151. Li, Q., Brown, J. B., Huang, H. & Bickel, P. J. Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat.* **5**, 1752–1779 (2011).
152. Yu, G., Wang, L.-G. & He, Q.-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015).
153. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
154. Castro-Mondragon, J. A. et al. JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **50**, D165–D173 (2022).
155. Lopez-Delisle, L. et al. pyGenomeTracks: reproducible plots for multivariate genomic datasets. *Bioinformatics* **37**, 422–423 (2021).
156. Servant, N. et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).
157. Marchal, C., Singh, N., Corso-Díaz, X. & Swaroop, A. HiCRes: a computational method to estimate and predict the genomic resolution of Hi-C libraries. *Nucleic Acids Res.* **50**, e35 (2022).
158. Yardımcı, G. G. et al. Measuring the reproducibility and quality of Hi-C data. *Genome Biol.* **20**, 57 (2019).
159. Crane, E. et al. Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523**, 240–244 (2015).
160. Open2C et al. Cooltools: enabling high-resolution Hi-C analysis in Python. *PLoS Comput. Biol.* **20**, e1012067 (2024).
161. Gong, Y. et al. Stratification of TAD boundaries reveals preferential insulation of super-enhancers by strong boundaries. *Nat. Commun.* **9**, 542 (2018).
162. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinforma.* **10**, 421 (2009).

Acknowledgements

This study was supported by the National Natural Science Foundation of China (W2411020, 32170645, 31830062), and the National Key Research and Development Program of China (2021YFF1000900, 2021YFF1000100). This study was also supported by the Fundamental Research Funds for the Central Universities (X2662024ZKPY002) and the Tianchi Talent Introduction Plan of Xinjiang Uyghur Autonomous Region. We thank the high-performance computing platform at the National Key Laboratory of Crop Genetic Improvement at Huazhong Agricultural University.

Author contributions

M.W. designed the experiments and managed the project. X.H., Y.W. and S.Z. conducted the cultivation of cotton plants. Y.W., S.Z., L.P. and X.H. collected the cotton tissue samples. L.P., Y.W. and S.Z. constructed the RNA-Seq libraries. Y.W. and L.P. constructed the in situ Hi-C libraries. S.Z., Y.W. and L.P. constructed the ChIP-Seq libraries. S.Z. and Y.W. performed the experiment of ATAC-Seq. X.H. analyzed the Hi-C, ChIP-Seq, ATAC-Seq and RNA-Seq data. J.Y. and X.H. analyzed the CRE-related fiber quality. J.L., Y.L., X.Z. and L.Z. contributed to the project discussion. X.H. wrote the manuscript draft, and M.W. revised it.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-55309-4.

Correspondence and requests for materials should be addressed to Maojun Wang.

Peer review information *Nature Communications* thanks Hang He and Jin-Quan Huang for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

---

Xianhui Huang[1,3], **Yuejin Wang**[1,3], **Sainan Zhang**[1,3], **Liuling Pei**[1], **Jiaqi You**[1], **Yuexuan Long**[1], **Jianying Li**[1], **Xianlong Zhang**[1], **Longfu Zhu**[1] & **Maojun Wang**[1,2] ✉

[1]National Key Laboratory of Crop Genetic Improvement, Hubei Hongshan Laboratory, Huazhong Agricultural University, Wuhan 430070, China. [2]College of Life Science, Shihezi University, Shihezi 832003, China. [3]These authors contributed equally: Xianhui Huang, Yuejin Wang, Sainan Zhang.
✉e-mail: mjwang@mail.hzau.edu.cn