Dynamics of duplicated gene regulatory networks governing cotton fiber development following polyploidy

Xianpeng Xiong[1,*], De Zhu[1,*], Corrinne E. Grover[2], Jonathan F. Wendel[2], Xiongfeng Ma[3,4,#], Guanjing Hu[1,3,4,#]

[1] Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China 518120

[2] Department of Ecology, Evolution and Organismal Biology, Iowa State University, Ames, IA 50011

[3] State Key Laboratory of Cotton Bio-breeding and Integrated Utilization, Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang, China 455000

[4] Western Research Institute, Chinese Academy of Agricultural Sciences, Changji, 831100, China

Short title: Fiber development gene regulatory networks

* These two authors contributed equally.

Corresponding Author(s): Guanjing Hu (huguanjing@caas.cn) and Xiongfeng Ma (maxiongfeng@caas.cn)

Abstract

Cotton fiber development entails complex genome-wide gene regulatory networks (GRN) that remain mostly unexplored. Here we present integrative analyses of fiber GRNs using public RNA-seq datasets, integrated with multi-omics genomic, transcriptomic, and cistromic data. We detail the fiber co-expression dynamics and regulatory connections, validating findings with external datasets and transcription factor (TF) binding site data. We elucidate previously uncharacterized TFs that regulate genes involved in fiber-related functions and cellulose synthesis, and identify the regulatory role of two homoeologous G2-like transcription factors on fiber length. Analysis of duplicated gene expression and network relationships in allopolyploid cotton, which has two co-resident genomes (A, D), revealed novel aspects of asymmetric subgenomic developmental contributions. Whereas D-based homoeolog pairs drive higher overall gene expression from the D subgenome, TFs from the A subgenome play a preferential regulatory role in the fiber gene regulatory network. Following allopolyploid formation, it appears that the trans-regulatory roles of TFs diversified more rapidly between homoeologs than did the cis-regulatory elements of their target genes. Our approach underscores the utility of network analysis for detection of master regulators and provides fresh perspectives on fiber development and polyploid functional genomics, through the lens of co-expression and GRN dynamics.

Keywords: Upland cotton, fiber development, gene regulatory network (GRN), transcription factors, *GhMYS1*

Introduction

Cotton ranks among the world's most important agricultural plants, supplying most of our natural textile fibers. The remarkable cotton "fibers", which are extensively elongated and naturally twisting single cells originating from the ovule epidermis, undergo a complex developmental program, entailing five sequential yet overlapping stages: initiation, elongation, transition, secondary cell wall (SCW) thickening, and maturation (Haigler et al. 2012). These complex, coordinated stages are crucial for fiber production, as initiation determines the number of epidermal cells that develop into fibers, while elongation and SCW thickening determine the final length and strength of each fiber (Yang et al. 2014). Given the importance of these stages,

the past two decades have witnessed considerable progress towards elucidating the principal pathways and genes orchestrating fiber initiation and elongation, mostly focusing on the regulatory role of transcription factors (TF) (Huang et al. 2021).

The initiation of cotton fibers shares a similar mechanism with the development of *Arabidopsis* leaf trichome, regulated by the intricate MYB-bHLH-WDR (MBW) transcriptional complex (Wang et al. 2019; Zhang et al. 2019; Wen et al. 2023). A specific MIXTA-like MYB TF, *GhMYB25-like*, serves as a pivotal switch in this context, with suppression leading to abnormal fiber cell initiation and fiberless seeds (Walford et al. 2011). The elongation phase of cotton fiber development is distinctive among plant cells, involving special factors and mechanisms that confer extraordinary fiber length and growth rate. Many TFs, including HD-ZIP, TCP, WRKY, and ARF, are integral to modulating this phase (Wen et al. 2022). Notably, the cotton HD-ZIP family TF *GhHOX3* promotes fiber elongation by upregulating transcription of the cell wall loosening protein genes *GhRDL1* and *GhEXPA1* (Shan et al. 2014), while a fiber-preferential WRKY TF GhWRKY16 directly activates the transcription of *GhHOX3*, a MYB family TF (*GhMYB109*), and a cellulose synthase gene (*GhCesA6D-D11*) (Wang et al. 2021b). TEOSINTE BRANCHED, CYCLOIDEA AND PCF 14 (GhTCP14) mediates cotton fiber elongation by directly activating the expression of auxin-responsive gene *GhIAA3* and auxin transporter genes *GhPIN2* and *GhAUX1* (Wang et al. 2013). Beyond their involvement in elongation, *GhHOX3, GhTCP14*, and *GhWRKY16* have also been confirmed to positively regulate fiber initiation (Qin et al. 2022; Wen et al. 2023). Transitioning from elongation to SCW thickening, several TFs, such as *GhFSN1* (Zhang et al. 2018) and five MYB family TF (*GhMYB1* (Yadav et al. 2017), *GhMYBL1* (Sun et al. 2015), *GhMYB7* (Huang et al. 2016), *GhMYB46_D9*, and *GhMYB46_D13* (Huang et al. 2019) have been reported to positively regulate SCW thickening. *GhTCP4* and a Class II KNOX TF (*GhKNL1*) function both in fiber elongation and SCW thickening; however, interestingly, GhKNL1 represses genes promoting elongation and SCW cellulose deposition, whereas GhTCP4 coordinates the suppression of fiber elongation through its interaction with *GhHOX3* to activate SCW synthesis (Cao et al. 2020; Wang et al. 2022). These results underscore the nuanced regulatory roles of TFs and intricate dynamics across different stages. Despite extensive research into TF-mediated regulatory networks of cotton fiber development, these studies have been primarily conducted in a gene-by-gene fashion, leaving relatively

unexplored a more comprehensive understanding of the dynamic interactions among networks of genes governing fiber development.

Organ development in plants is intricate, depending on the precise timing and spatial regulation of gene expression, a process captured by complex gene regulatory networks (GRNs) (Haque et al. 2019; Jones and Vandepoele 2020; Vandepoele and Kaufmann 2023). These networks represent the full suite of interactions between TFs and their target genes, where TFs bind to specific DNA sequences known as TF binding sites (TFBSs) and regulate the transcription of downstream targets. Central to GRNs are hub TFs, which, due to their large number of target and/or regulating genes, are crucial for the integration and dissemination of regulatory signals across the network (Barabási and Oltvai 2004; Levine and Davidson 2005). Identifying hub genes in plant GRNs offers a clear roadmap for pinpointing master regulators and unraveling interconnections essential for biological processes and developmental programs (Gaudinier and Brady 2016; Haque et al. 2019; Jones and Vandepoele 2020). These network components, when modulated, can enhance plant productivity or resilience, often yielding more significant influence over complex phenotypes than manipulating individual genes alone (Springer et al. 2019). Therefore, the construction and mining of GRNs is key for increasing the predictive power of genome engineering approaches aimed at agronomic traits for crop improvement.

GRN construction methods can be broadly categorized into two main approaches differentiated by the source of information utilized: data-driven methods and prior knowledge-based methods. Data-driven methods leverage high-throughput experimental techniques to unveil physical interactions between TFs and their target genes. These techniques includes: (1) Chromatin immunoprecipitation sequencing (ChIP-seq) (Furey 2012) which identifies genomic sites bound by a given TF *in vivo*; (2) DNA-affinity purification sequencing (DAP-seq) (O'Malley et al. 2016) captures DNA bound by the *in vitro* expressed TF; and (3) yeast one-hybrid assay (Taylor-Teeples et al. 2015), which identifies physical interactions between TFs and their potential DNA binding sites. Additionally, chromatin accessibility assays (Song and Crawford 2010; Buenrostro et al. 2015; Zhao et al. 2020), including DNase-I hypersensitive site sequencing (DNase-seq), assay for transposase-accessible chromatin with sequencing (ATAC-seq), and MNase hypersensitive sequencing (MH-seq), have also been applied to characterize *cis*-regulatory elements as potential transcription factor binding sites (TFBSs) at a genome-wide

scale, thereby revealing regulatory relationships between TFs and target genes. Despite the substantial increase in information about regulatory sequences and interactions offered by these assays, inherent technical challenges and cost still pose limitations for studying large numbers of TFs (Kulkarni and Vandepoele 2020). Consequently, only a few plant species, such as *Arabidopsis* and maize, have constructed GRNs based on large-scale experimental data of regulatory interactions (Taylor-Teeples et al. 2015; Gaudinier et al. 2018; Tu et al. 2020; Tang et al. 2021).

In contrast to data-driven methods, prior knowledge-based methods for GRN construction integrate existing biological knowledge, drawing from scientific literature, known biological pathways, functional gene ontology categories, and various knowledge databases of gene-to-gene relationships (Linde et al. 2015). For instance, resources like PlantTFDB (https://planttfdb.gao-lab.org/) and PlantRegMap (https://plantregmap.gao-lab.org/) serve as integrated platforms for plant regulatory data and analysis, which systematically screens for functional TFBSs and regulatory interactions in plants. PlantRegMap, in particular, curates additional functional and evolutionary annotations, such as expression profiles and multiple-species comparisons, along with corresponding literature references, resulting in generation of regulatory maps for the main lineages of angiosperms, effectively representing their preliminary GRNs (Tian et al. 2020a). However, these general GRNs do not account for differences in gene regulation relationships across different tissues, developmental stages, or conditions.

Integration methods often combine both data-driven and prior knowledge-based approaches, leveraging expression data underlying specific states to refine GRNs built based on existing knowledge, or vice versa. Computational algorithms that infer GRNs from gene expression data include correlation and information theory-based methods, probabilistic graphical models, and machine learning (Haque et al. 2019). Correlations and mutual information methods assume that co-expression is an indicator of coregulation and deterministically controlled by upstream regulators. Probabilistic graphical models consider gene expression as random variables with a certain probability distribution over different tissues and conditions. Machine learning algorithms, such as ensemble decision trees and support vector machines, are trained on expression data to predict regulatory relationships between genes. In recent years, these inference methods have been employed to construct GRNs and identify important genes and regulatory

relationship involved in plant growth and developmental processes, such as photomorphogenesis in *Arabidopsis* (Balcerowicz et al. 2021), abiotic and disease responses in wheat (Ramírez-González et al. 2018), nitrogen-deficiency responses in rice (Ueda et al. 2020), as well as Kranz anatomy development in maize and rice (Chang et al. 2019). More recently, as demonstrated for spike phenotypic traits in wheat and flowering time regulation in maize (Chen et al. 2023; Han et al. 2023), GRN inference has been improved by integrating heterogeneous -omics and functional validation data for a more comprehensive understanding of the biomolecular networks. Despite the critical importance of cotton fiber development to its success as a major crop species, a comprehensive GRN that unravels the intricate molecular mechanisms underlying fiber traits is still lacking.

In this study, we employed three distinct inference methods to construct GRNs utilizing transcriptome data from 401 samples. Notably, we validated the robustness and efficacy of resulting GRNs through rigorous integration with prior knowledge-based regulatory maps, DAP-seq data, and additional transcriptomic datasets from gene perturbation experiments. Through this integrative analysis, we identified novel transcription factors crucial for orchestrating fiber development. We further validated the functional significance of a homoeologous pair of top-ranked G2-like TF genes (*GhMYS1_A10* and *GhMYS1_D10*) in the GRN, revealing their potential regulatory mechanisms in fiber development.

An additional important dimension of our study is that it addresses the fate of duplicated GRNs in an allopolyploid plant, that is, one that contains two co-resident genomes. *Gossypium hirsutum* contains the descendant genomes of both its A-genome and D-genome ancestors (each n=13), and thus has an AD-genome with an additive (n=26) chromosome number. This evolutionary history raises the possibilities of revealing the fate of duplicated GRN dynamics following allopolyploid evolution, a prominent process in plant evolution (Hu et al. 2021; Viot and Wendel 2023). Here, we elucidate subgenomic control over fiber expression at both the genic co-expression and GRN levels, providing insights into the regulatory landscape of fiber development in an allopolyploid contect. Finally, we provide an integrative network resource and demonstrate its utility in enhancing our understanding of cotton fiber development, thereby facilitating targeted interventions to modulate fiber traits.

Results

A fiber gene expression atlas of the Upland cotton *G. hirsutum*

We compiled a dataset of 473 Upland cotton (*G. hirsutum*) fiber transcriptomes from 12 RNA-seq studies (Tuttle et al. 2015; Zhang et al. 2015, 2021a; Hinchliffe et al. 2016; Lu et al. 2017; Bao et al. 2019; Hu et al. 2019; Sun et al. 2019a; Huang et al. 2020; Li et al. 2020; He et al. 2021) (Supplementary Table S1). These samples spanned key fiber developmental stages from 0 to 30 days post-anthesis (dpa), including fiber initiation, elongation, transition, and secondary cell wall (SCW) synthesis (Figure 1A, Supplementary Fig. S1). To ensure specificity to fiber cells, 12 samples from 0 to 3 dpa obtained from whole ovules were excluded. After quality screening based on a unique mapping rate higher than 70% and outlier removal through principal component analysis (PCA), a final set of 413 high-quality samples was obtained with Q20 above 93.09% (Supplementary Table S2). Further refinement using principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE) led to the removal of another 12 outlier samples, resulting in a final dataset of 401 samples (Supplementary Fig. S2, Supplementary Table S2).

Based on the standardized gene expression by TPM (transcripts per million), both PCA t-SNE identified two distinct clusters of fiber samples: one comprising 329 samples from 5 to 15 dpa and another with 57 samples from 19 to 30 dpa, while 15 samples from 18 dpa exhibited an intermediate distribution (Figure 1B, C). This observation indicates the pronounced transcriptional distinction of the fiber cell from about 19 dpa as it becomes intensely committed to SCW synthesis. Categorization of the three earlier stages was less clear, likely due to genetic variation and variation in growing conditions or collection techniques across studies. Notably, the inclusion of natural green-fiber cotton varieties highlights developmental differences that can distinguish accessions. That is, among the 15 samples representing 18 dpa fiber, the 12 samples derived from white-fiber producing accessions clustered with the 5 to 15 dpa samples (circa 50 on PC1), whereas the 3 green-fiber (variety Xincai 7) samples clustered with the 19 to 30 dpa samples, suggesting that the green-fiber accessions transition to SCW synthesis sooner than the white-fiber accessions represented here and underscoring the potential for temporal differences in development among cotton varieties.

To ensure the reliability of this dataset, we examined expression patterns of 192 fiber-related genes known for their roles in cotton and/or *Arabidopsis* trichome development (Supplementary Table S3). These genes were classified into three groups based on their average TPM values per dpa (Supplementary Fig. S3). Group I, comprising 93 genes, displayed high expression levels early during fiber elongation, featuring well-known elongation-associated genes like *GhMYB25* (Machado et al. 2009), *GhMYB25-like* (Walford et al. 2011), *GhTCP4* (Cao et al. 2020), *GhPIN3a* (Zeng et al. 2019), *GhHOX3* (Shan et al. 2014), *GhHD1* (Walford et al. 2012), *GhCaM7* (Tang et al. 2014), *GhWRKY16* (Wang et al. 2021b), and *GhBZR1* (Zhou et al. 2015). Group II, containing 29 genes exhibiting higher expression during SCW synthesis at later time points, included established SCW genes such as *GhBZR3* (Shi et al. 2022), *GhKNL1* (Gong et al. 2014), *GhSWEET12* (Sun et al. 2019b), *GhFSN1* (Zhang et al. 2018), and *GhMYB46_D13* (Huang et al. 2019). Group III comprised 70 genes with expression profiles peaking at various time points between 5 and 30 DPA. The expression patterns observed here closely align with previous reports (Supplementary Table S3); that is, 77% (105 out of 136) of the genes surveyed exhibited the expected expression profiles, providing robust validation of our gene expression atlas. The few inconsistencies observed were primarily attributed to missing data (i.e., lack of later time point data or large gaps between time points) in earlier studies. For example, several genes (including *GhGA20ox1* (Xiao et al. 2010), *GhTUA9* (Li et al. 2007), *GhMYB212* (Sun et al. 2019b), *GhACO1* (Wei et al. 2022), *GhMAH1* (Ma et al. 2022), *GhMYB5_A12* (Wang et al. 2021a), and *GhCPC* (Liu et al. 2015)), which were previously compared only between 5 and 15 dpa, exhibited continuous expression changes in 5 to 30 dpa based on our comprehensive expression profiles. Additionally, our dataset revealed that several well-known fiber initiation genes, including *GhMYB25* (Machado et al. 2009), *GhPIN6* (Zhang et al. 2017b), *GhPIN3a* (Zeng et al. 2019), *GhBZR3* (Shi et al. 2022), and *GhSWEET12* (Sun et al. 2019b), exhibited high expression levels in later stages of development that were not previously examined. This suggests that these genes may have regulatory roles beyond fiber initiation, highlighting insights enabled by our comprehensive data analysis.

Co-expression gene network analysis reveals fiber developmental dynamics

To explore the transcriptional dynamics of cotton fiber development, we employed weighted co-expression gene network analysis (WGCNA) on fiber-expressed genes. Opting for a filtering

criteria of TPM>0 in 30% of the samples (see Methods and Supplementary Fig. S4), we identified and included 57,151 genes for further analysis, representing 76.3% of the total genome expressed in fibers, consistent with previous reports (Hovav et al. 2008a; Yoo and Wendel 2014; Gallagher et al. 2020). The subsequent WGCNA analysis categorized 34,075 genes into 20 co-expression modules, varying in size from 109 to 7,360 module gene members (Supplementary Fig. S5). The seven largest modules, ME1 (turquoise), ME2 (blue), ME3 (brown), ME4 (tan), ME5 (green), ME6 (black), and ME7 (red), collectively accounted for 87.7% (29,884 genes) of all co-expressed genes. The remaining 23,076 genes, which could not be assigned to any modules, were grouped into a grey module, indicating no discernible co-expression relationships.

To examine phenotypic associations, we correlated module eigengenes (MEs) with fiber development for 14 time points between 5 and 30 dpa (inclusive; Figure 2A). Pearson correlation analysis showed significant associations with fiber development for the majority of modules, treating the time points as a binary categorical variable (Figure 2A) or a numeric variable (DPA; Supplementary Fig. S5B). ANOVA of MEs revealed significant developmental changes for seventeen modules, excluding ME6, ME17, and ME20 (MEs ~ DPA, ANOVA $P < 0.05$) (Figure 2B).

Notably, the green module displayed the highest correlations with DPA ($r = 0.80$, $P = 2e\text{-}91$), exhibiting a gradually increasing expression profile along fiber development. This module consists of 4015 genes and 223 TFs, and it was enriched with GO terms related to cell wall development, such as plant-type secondary cell wall biogenesis, cell wall polysaccharide biosynthetic process, hemicellulose metabolic process, tube morphogenesis, and cell wall macromolecule biosynthetic process (Figure 2D). Conversely, the brown, blue, and tan modules showed strong negative correlations ($r = -0.61$~-0.85, $P < 2e\text{-}41$), corresponding to decreasing expression along fiber development (Figure 2B-C). The tan module in particular showed significant enrichment of GO terms associated with cotton fiber development, encompassing processes like very long-chain fatty acid metabolism, microtubule organization, cell tip growth, pectin biosynthesis, and polymeric cytoskeletal fiber processes (Figure 2D; Supplementary Fig. S6; Supplementary Table S4).

The turquoise module, the largest with 7360 genes (including 984 TFs), peaked in expression at 25 dpa. Significant enrichment with GO terms including core promoter sequence-specific DNA binding, DNA-binding transcription activator activity, and RNA polymerase II-specific were observed. The red module, consisting of 2232 genes (including 50 TFs) and without significant correlation with DPA, were enriched with diverse GO functions (Supplementary Table S4).

Exploring the roles of hormone signaling pathways in regulating cotton fiber development (Huang et al. 2021), we observed significant functional enrichment in key modules. The brown module with expression peaking at 5 dpa revealed a strong association with auxin (IAA)-activated signaling pathways (Supplementary Fig. S7A), consistent with the known function of IAA-activated signaling pathways in promoting fiber initiation and elongation. Surprisingly, brassinosteroid (BR)-related signaling pathways, known to regulate fiber initiation and elongation, were enriched in both the tan module (early peaking at 5 dpa) and the turquoise module (late peaking at 25 dpa). This introduces a novel perspective on the impact of BR post-fiber elongation, which has not been reported previously (Supplementary Fig. S7B-C). Additionally, the red and turquoise modules, which exhibited more complex and dynamic gene expression patterns across development, were enriched for BR, jasmonic acid, gibberellin, and ethylene-related signaling pathways (Supplementary Fig. S7C), warranting further investigation into their functional implications. Besides these extensively studied phytohormones, the turquoise module also showed significant enrichment of abscisic acid, cytokinins, and salicylic acid-related signaling pathways (Supplementary Fig. S7D), While these pathways are recognized for their roles in plant growth and development (Santner and Estelle 2009), their specific impact on cotton development remains underexplored.

Construction and evaluation of the cotton fiber gene regulatory networks

To infer regulatory interactions beyond co-expression relationships between genes, we systematically constructed gene regulatory networks (GRNs) using three distinct inference methods: Corto, GENIE3, and dynGENIE3. Leveraging the 57,151 fiber-expressed genes derived from 401 RNA-seq samples, we evaluated the regulatory relationships between 3,638 transcription factors (TFs) and their putative target genes. Both GENIE3 and dynGENIE3 inferences were confined to the top one million TF-target interactions (edges) for comparison,

297 retaining over twice the number of genes (nodes) from the GENIE3 network than from the
298 dynGENIE3 network (54,237 and 25,441, respectively). Although Corto inferred only 232,943
299 TF-target interactions (edges), it retained a comparable number of nodes to GENIE3 (56,052),
300 resulting in the densest and most clustered network topology among the three methods, followed
301 by GENIE3 and then dynGENIE3 (Table 1). Because differences in GRN construction can lead
302 to different inferences, we evaluated GRN quality for each method based on existing and newly
303 generated data, as listed below.

304 We first assessed the ability of each GRN to capture documented TF-target interactions based on
305 systematic literature mining in plants, as assembled into the PlantRegMap (Jin et al. 2015).
306 These known regulatory relationships were projected onto cotton orthologs to generate the
307 cottonRegMap. Among the three GRN methods, GENIE3 outperformed Corto and dynGENIE3,
308 recovering the highest percentage of known interactions (14.98% vs. 14.58% and 13.85%,
309 respectively), although the range among these percentages is relatively small. We also note that
310 these seemingly low percentages of interactions reflect the non-specific nature of cottonRegMap,
311 which involves prior knowledge assembled from various plants and is not specific to cotton
312 fibers. Without a true gold-standard dataset for validation, we employed a permutation test to
313 determine the expected number of interactions captured by chance. Both GENIE3 and Corto
314 captured more interactions than the expected 14.37% of interactions (bootstrapping $P < 0.05$),
315 demonstrating their utility in capturing biological information for cotton fiber (Figure 3A). For
316 subsequent analyses, we integrated prior biological knowledge by retaining only the regulatory
317 interactions predicted by GENIE3, dynGENIE3, and Corto that were also supported by
318 cottonRegNet. This approach allowed us to retain the relative topological patterns between
319 methods (Table 1). These integrated networks were designated as cGENIE3, cdynGENIE3, and
320 cCorto, respectively.

321 Using a second approach, we assessed the ability of each GRN to recover known fiber-related
322 functional genes and TFs (Supplementary Table S3) previously reported in the literature. We
323 curated 192 fiber-related functional genes, of which the cGENIE3 network contained 155 (80%),
324 the cdynGENIE3 network contained 114 (59%), and the cCorto network contained 91 (47%) of
325 the genes on the list. In terms of the percentage of known genes among total network nodes,
326 cdynGENIE3 exhibited the highest percentage (0.45%, 117 of 25,441), followed by cGENIE3

(0.37%, 155 of 41,757) and cCorto (0.36%, 91 of 25,245). We extended this to assess whether known TFs were enriched among the highly ranked TF regulators in each GRN. Gene set enrichment analysis (GSEA) showed that the curated TFs were significantly enriched at the top of the cGENIE3 network; specifically, a leading-edge subset comprising 77 TFs was identified as the most significant contributors to this enrichment (Figure 3B, Supplementary Table S5). In contrast, known TFs were not enriched at the top of cdynGENIE3 and were randomly distributed in rank in the cCorto network (Supplementary Fig. S8). These results suggested that cGENIE3 has stronger prediction power for key TFs compared to cdynGENIE3 and cCorto.

We further validated the GRN-inferred TF-target relationships for two top ranked (homoeologous) TFs using physical evidence from DNA-affinity purification sequencing (DAP-seq), an *in vitro* genome-wide assay of TF-DNA binding (O'Malley et al. 2016). The homoeologous G2-like TFs *GhMYS1_A10* (*Gohir.A10G036400*) and *GhMYS1_D10* (*Gohir.D10G037100*) were among the most confident (highest-ranked) regulators in all three networks; therefore, these genes, were independently assayed for genome-wide binding sites using DAP-seq (Figure 3C, Supplementary Table S6). These assays yielded 227,117 and 141,945 peaks for *GhMYS1_A10* and *GhMYS1_D10*, respectively, with approximately 8.27% and 6.73% of the peaks located within 2kb of the transcription start site for 10,132 and 9,363 genes (Supplementary Fig. S9A-D). Among these genes, 7,784 and 6,773 were expressed in fibers and identified as targets for *GhMYS1_A10* and *GhMYS1_D10*, respectively (Supplementary Fig. S9F-E). Examination of the overlap in target genes between DAP-seq and each GRN revealed a significant association for both TFs in cGENIE3 (hypergeometric test *p*-values of 2.24e-08 and 3.56e-06) and for *GhMYS1_D10* only in cdynGENIE3 ($p = 0.0471$); no significant association was found for the cCorto GRN (Figure 3C, Supplementary Fig. S9F-G). These results were reiterated when we compared DAP-seq for an additional gene (*GhBES1.4*) with each GRN. *GhBES1.4* is a known core TF in the BR signaling pathway that positively regulates fiber elongation (Liu et al. 2023), yet it was ranked eighteenth by different GRN methods. Using a published DAP-seq dataset for *GhBES1.4* (Liu et al. 2023), we found significant overlaps between the 1214 fiber-expressed target genes of *GhBES1.4* inferred by DAP-seq (Supplementary Fig. S9G) and the cGENIE3 and cdynGENIE3 GRNs, but not the cCorto GRN (Figure 3C).

Our fourth approach utilized published RNA-seq data from TF mutants or transgenic lines to evaluate the accuracy of GRN inferences, by assessing how well the predicted regulatory interactions in the GRNs corresponded to the differentially expressed genes (DEGs) observed in these TF mutants or transgenic lines. Specifically, we identified 3,508 DEGs in *GhWRKY16* RNAi lines, 1,422 in *GhBES1.4* RNAi lines, and 1,790 in *GhBES1.4* overexpression lines compared to wild-type plants, most of which (96.2-99.0%) were expressed in the fiber dataset evaluated here (Supplementary Fig. S10A-C). These DEGs likely represent downstream targets of the TFs perturbed in each respective experiment, and are thus useful to validate our GRN predictions. For *GhWRKY16*, which is a WRKY TF known for promoting fiber initiation and elongation (Wang et al. 2021b), we found significant overlap between the 3,472 DEGs identified from the RNAi line comparison and the *GhWRKY16*-target relationships found in the cGENIE3 GRN (hypergeometric test p = 3.90e-06; Figure 4C); in contrast, cCorto and cdynGENIE3 inferred only two and zero DEG targets, respectively. Conversely, the 1,477 DEGs detected in the *GhBES1.4* exhibited significant overlap with the *GhBES1.4*-targets only recovered for the cdynGENIE3 GRN (p = 0.005751457). Notably, no significant overlap was found between the DEGs from the *GhBES1.4* RNAi lines and any of the networks (Figure 3C). Combined with the DAP-seq evaluation, these results suggest that both GENIE3 and dynGENIE3 outperform Corto in predicting regulatory targets for specific TFs, notwithstanding the inherent variance depending on the TF and experimental context.

Our final assessment correlated the trait fiber length with key fiber TFs inferred by the GRNs. Using the top 77 TFs ranked by each GRN method, Pearson correlation analysis between their expression levels in 15 DPA fiber and mature fiber length revealed the highest phenotypic correlations were found with TFs implicated in the cGENIE3 network, followed by cdynGENIE3 and then cCorto (Figure 3D, Supplementary Table S5, 7-8). All networks showed significantly higher correlations with phenotype than did all 3638 TFs expressed in fibers (Figure 3D).

Performance evaluation of GRN inferences in the case of cotton cellulose synthesis

In this case study, we evaluated three GRN methods by focusing on their ability to predict regulatory relationships involved in cellulose synthesis in cotton fiber, aiming to further validate

386 their predictive power and highlight novel findings. Cotton fiber is composed primarily of
387 cellulose, accounting for over 90% of its composition at maturity (Haigler et al. 2012).
388 Performing a genome-wide analysis of the cellulose synthase (CesA) gene family, we identified
389 27 CesA genes in the *G. hirsutum* genome and divided them into six classes, consistent with the
390 previous reports (Supplementary Fig. S11) (Zhang et al. 2021c; Wen et al. 2022). Thirteen
391 *GhCesAs* were highly expressed during fiber elongation *via* primary cell wall (PCW) synthesis,
392 and seven were linked to SCW formation after 15 dpa. The remaining seven *GhCesAs* genes
393 exhibit relatively low expression levels throughout fiber development and were considered of
394 unknown function (Figure 4A).

395 Inspecting the GRN-inferred TF-target relationships involving the fiber development related
396 *GhCesAs*, we next compared how well each GRN method represents these genes and known
397 regulator relationships (Figure 4B). The cGENIE3 network effectively identified all 20 *GhCesAs*
398 as targets and predicted 71 regulatory TFs (Supplementary Fig. S12), resulting in the largest
399 cellulose synthesis subnetwork (Figure 4C). In contrast, cdynGENIE3 identified only 13
400 *GhCesAs* regulated by 43 TFs, notably missing all of the SCW *GhCesAs* (Figure 4B). Likewise,
401 cCorto identified even fewer (11) *GhCesAs*, again missing all SCW *GhCesAs*, and finding only 8
402 TFs as regulators (Figure 4B; Supplementary Fig. S13-14). In addition to predicting the greatest
403 number of relationships, cGENIE3 also recovered regulatory relationships verified by prior
404 studies, whereas cdynGENIE3 and cCorto did not. For example, the NAC TFs family genes
405 *GhFSN1_A12* and *GhFSN1_D12* were predicted by cGENIE3 to regulate *GhCesA4* and
406 *GhCesA7*, consistent with their differential expression patterns in *GhFSN1* overexpression lines
407 compared to the wild-type cotton plants that suggest the same regulatory relationship (Zhang et
408 al. 2018). Likewise, *GhWRKY16_D06* was a predicted regulator of *GhCesA7_D7*, aligning with
409 its known role in regulating *GhCesAs* during fiber initiation and elongation (Wang et al. 2021b)
410 (Figure 4C). GO enrichment results showed that the 71 regulatory TFs predicted by cGENIE3
411 were significantly enriched in plant-type cell wall modification, regulation of secondary cell wall
412 biogenesis, and xylem development (Figure 4F). The results collectively suggested that the
413 cGENIE3 network presents a higher predictive power for cellulose synthesis compared to
414 cdynGENIE3 and cCorto.

Closer examination of the cGENIE3 network revealed two distinct yet interconnected network components (Figure 4C). The smaller component I consists of 11 PCW-related *GhCesAs* (3 *GhCesA1*, 4 *GhCesA3*, and 4 *GhCesA6*) and 27 regulatory TFs. Most of these TFs primarily exhibited peak expression early during PCW synthesis, with 3 exceptions that peaked later. The larger component II includes 7 SCW-related *GhCesAs* (2 *GhCesA8*, 3 *GhCesA7*, and 2 *GhCesA4*) and 38 TFs. Fewer than half of the TFs in this component exhibited concordant expression with their target *GhCesAs.* Among those disconcordant TFs peaking early during PCW formation, the homoeologous pair of top-ranked G2-like TFs described above, *GhMYS1_A10* and *GhMYS1_D10*, were identified (Figure 4C and E). Combining trait association results and expression patterns, *GhMYS1_A10* and *GhMYS1_D10* emerge as potential novel TFs that may positively regulate fiber elongation by promoting PCW formation while inhibiting SCW formation (further explored later; Figure 4C). In addition to more diverse TF expression patterns, component II is enriched for SCW-related genes and is denser and more interconnected than component I, which is enriched for PCW-related genes (Figure 4C-E); this distinction reflects the intricate gene regulatory control underlying the transition from fiber elongation to cell wall thickening. The two components were interconnected through 6 TFs that regulate both PCW-related and SCW-related *GhCesAs*. These findings underscore the utility of GRN interrogation in characterizing key regulators and functions in cotton fiber development.

Regarding At and Dt homoeologous relationships, we identified five TF homoeolog pairs and two *GhCesA* homoeolog pairs in component I, and six TF homoeolog pairs and two *GhCesA* homoeolog pairs in component II. These homoeolog pairs present in the same component accounted for 44.4% (8 of 18) *GhCesAs* and 33.8% (22 of 65) TFs, representing functional conservation or redundant regulatory relationships between homoeologs (Supplementary Table S9). This duplicated nature of allopolyploid gene networks, along with the identification of new master regulators, is discussed next.

The allopolyploid nature of cotton fiber GRN

Understanding the allopolyploid nature of *G. hirsutum* (2n= 4x = 52; AADD genome) is essential for unraveling the regulatory basis of cotton fiber development. The ascertainment of orthologous-homoeologous relationships among the polyploid A-subgenome (At) and D-

subgenome (Dt) genes and their parental A- and D-genome diploids provides a foundation for understanding the evolutionary dimension of duplicated gene regulation during cotton fiber development. We used 22,889 homoeologous pairs that were previously characterized into single-copy orthologous-homoeolog groups (scOGs; each containing a single representative for At and Dt) (Hu et al. 2023) to evaluate the evolutionary outcomes for genes inherited from parental diploids and maintained in duplicate post allopolyploidization. The remaining genes (13,229 At; 15,895 Dt) were categorized into variable-copy orthologous-homoeolog groups (vcOGs), possibly reflecting genetic variation between parental diploids and/or accrued post allopolyploidy. Against this backdrop, we leveraged the network perspectives of gene expression, co-expression, and regulatory interactions, to assess the contributions of the A- versus D- subgenomes for both scOG and vcOG categorization of homoeologous gene pairs during the dynamic process of fiber development.

Proportion of fiber-expressed genes. Of the 57,151 fiber-expressed genes (76.3% of the total genome), the A-subgenome contains fewer fiber-expressed genes compared to the D-subgenome (28,004 At vs 29,147 Dt), although this is a higher percentage of the total number of At genes versus Dt (77.53% vs 75.15%; chi-square test $P = 0.008523$). These fiber-expressed genes were further categorized into (1) 19,213 paired scOGs where both At and Dt were expressed; (2) 1,597 unpaired scOGs where only one homoeolog was expressed in fibers (749 At and 848 Dt); and (3) 17,128 vcOG genes (8,042 At and 9,086 Dt) (Table 2: I&II). Gene expressed in fibers represented a significantly higher proportion of the scOG category versus the vcOG category (87.4% vs. 58.5%, respectively; chi-square test $P = 2.2e\text{-}16$). Between subgenomes, a higher percentage of vcOG At genes was expressed in fiber versus vcOG Dt genes (60.6% vs 57.2%), while the percentages were comparable for scOGs (87.2% vs 87.7% At and Dt genes, respectively). Thus, the higher percentage of expressed gene content in the A-subgenome was mainly attributable to the higher proportion of vcOGs At genes expressed in fiber.

Overall expression levels. Comparing expression levels between A- and D-subgenomes revealed a subtle pattern with slightly higher expression of Dt genes (Figure 5A), consistent with previous reports in cotton fibers (Hovav et al. 2008b; You et al. 2023). This expression imbalance was consistently observed for the scOG gene set, whereas the vsOG gene set exhibited the opposite pattern (i.e., higher expression of At genes; Figure 5A). Notably, paired scOGs exhibited the

highest expression levels for both At and Dt genes ("scOG pair": mean TPM of At 18.53 and Dt 19.96), followed by vcOG genes ("vcOG", At 15.98 and Dt 14.07), and then singleton scOGs with only one homoeolog expressed exhibiting the lowest expression levels ("scOG unpair": At 1.42 and Dt 1.00).

Homoeolog expression bias (HEB). Analysis of HEB, where homoeolog expression statistically varies between duplicates, revealed 8,981 A-biased and 9,153 D-biased pairs among the 19,213 scOG homoeolog pairs, numbers that are not statistically different ($P = 0.3451$), and aligning with previous results (Zhang et al. 2015). Intriguingly, A-biased pairs displayed higher expression levels and larger variation across samples compared to D-biased pairs (Figure 5B). However, D-biased pairs exhibited significantly more expression differences than the A-biased pairs (i.e., Dt-At > At-Dt; Supplementary Fig. S15). This resulted in an overall higher gene expression of scOGs in the D subgenome than the A subgenome, despite the presence of more A-biased versus D-biased pairs.

Co-expression modular HEB. The co-expression gene network analysis clustered 25,751 fiber-expressed genes (12,816 At and 12,935 Dt genes) into 20 co-expression modules. Approximately 48.8% of module member genes were paired in modules as homoeologous pairs (6,280 pairs; Table 2: III), indicating substantial functional conservation. The remaining 51.2% of module genes were present in different modules for At and Dt, suggesting functional divergence in terms of co-expression patterns (Supplementary Table S10). Proportions of homoeologous TF pairs in the same module were significantly higher than other homoeologous gene pairs (53.0% vs 48.8%; chi-square test P = 2.2e-16), indicating a higher level of functional conservation between TF homoeologs (Table 2: III; Table S10). Investigating modular HEB for the homoeologous pairs within the same module revealed an absence of significant imbalance of HEB toward either subgenome (Supplementary Table S11). This observation is consistent with the overall pattern of 19,213 scOG pairs. The expression level differences between At and Dt genes across modules (Figure 5C) can be mostly attributed to the expression differences between A-biased and D-biased pairs (Figure 5D; Supplementary Fig. S16). Notably, within the tan module corresponding to fiber elongation, a significantly higher |At-Dt| difference was observed in D-biased then A-biased pairs, implying that the D subgenome might exert a greater effect on fiber elongation than the A subgenome.

Subgenomic asymmetry in fiber GRN. Taking the cGENIE3 network as an example, we evaluated the subgenomic contributions to regulatory nodes and edges within the inferred regulatory network. For nodes (Table 2: IV), a higher percentage of At genes was recovered in the network compared to the Dt genes (57.0% vs 54.6%; chi-square test $P = 0.0005271$), and this biased pattern was mainly caused by the target genes (TGs: 56.7% vs. 54.3%; chi-square test $P = 0.0004837$), particularly the scOG ones (39.6% vs. 35.6%; chi-square test $P = 1.757e-06$). The proportion of TFs with both At and Dt homoeologs present in GRN was significantly higher than that of target genes ("TFs in scOG pair" 88.1% vs "TGs in scOG pair" 77.9%; chi-square test P = 2.2e-16). Depending on whether the TF-TG regulatory links were inferred within or between subgenomes, network edges were classified into four categories: two intra-subgenome classes within either subgenome (38,704 At-At and 35,997 Dt-Dt) and two inter-subgenome classes (40,116 At-Dt and 35,032Dt-At) (Table 2: V). The observed ratio of these four edge classes (1.10:1.03:1.15:1.00) significantly deviated from expected proportions assuming full network connectivity from TF to TG nodes (1.01:1.03:1.04:1.00; chi-square test, $P < 2.2e-16$). The intra-subgenome At-At and inter-subgenome At-Dt edges were observed more frequently than expected, indicating a biased regulatory role of At TFs compared to Dt TFs in the fiber gene regulatory network (GRN) (Table 2: V). Finally, we assessed the extent of functional conservation between homoeologs in the GRN, differentiating their roles as TFs or TGs. We observed a significantly higher proportion (15.8%) of edges targeting paired TG homoeologs (i.e., regulatory role as TFs targeting conserved *cis* binding sites) compared to the proportion (6.4%) of edges regulated by paired TF homoeologs (i.e., TGs being regulated by conserved *trans* TF proteins) (Table 2: V; Supplementary Table S12). This suggests that functional divergence between homoeologs in the GRN is more likely to occur in *trans* rather than in *cis*. Consistent patterns were observed in cottonRegNet and other GRNs (Supplementary Table S12).

Exploring novel regulators of cotton fiber development by GRN inference.

We next utilized the best performing cGENIE3 network to identify inter-connections among previously characterized fiber-related genes. Of the curated list of 192 known fiber-related genes (Supplementary Table S3), 154 were present in the network, with 657 directed network edges pointing to them from various TFs. This yielded a seeded network termed kGRN, comprising 432 nodes and 657 edges (Figure 6A, Supplementary Table S13).

Within the kGRN, ten known fiber-related genes function as TFs, regulating other known genes. Eight of these TFs, *GhTCP14*, *GhMYB46_D13*, *GhARF2b*, *GhFSN1_A12*, *GhWRKY16*, *GhGT2*, and a homoeologous pair of *GhMYB30*, have been functionally validated in fiber development (Supplementary Table S14) (Wang et al. 2013, 2021b; Zhang et al. 2018, 2021b; Huang et al. 2019; Tian et al. 2022; Wu et al. 2023). *GhMYB7_A12* and *GhJMJ12_D12*, identified in previous GWAS studies, were significantly associated with fiber strength and/or length (Wang et al. 2017; Liu et al. 2020). Among the remaining 187 TFs with unknown roles in cotton fibers, 97 have *Arabidopsis* orthologs with proven roles in cell wall development or involvement in signaling pathways regulating cotton fiber elongation (Supplementary Table S15), suggesting them as potential candidates for future molecular validation.

By integrating network clustering results with co-expression module annotation, the kGRN was partitioned into three subnetworks (Figure 6A). Subnetwork I, forming a loosely connected periphery on the left, prominently featured co-expressed TFs and target genes from the turquoise module. The target genes of these turquoise module TFs were identified across multiple co-expression modules, suggesting a broad spectrum of regulatory effects amplified by the fluctuating gene expression patterns spanning from 5 to 30 dpa, potentially involving various signaling pathways acting at the inter-modular level (Figure 2A).

Subnetwork II, situated on the right periphery, consists of most of the well-characterized TFs from the brown module, orchestrating key aspects of fiber initiation, including *GhHOX3*, *GhHD1*, *GhMYB25-like*, and *GhWD40* (Figure 6B). Of particular interest is *GhHOX3*, simultaneously regulated by both the brown and tan module TFs (Figure 6B), aligning with its multifaceted role in fiber initiation and elongation functions, respectively (Shan et al. 2014; Qin et al. 2022).

Subnetwork III, centrally located and densely interconnected, weaves together regulatory relationships between TFs and known function target genes in the green, yellow, and blue modules. Among the hub TFs regulating numerous target genes, two, *GhFSN1_A12* (Zhang et al. 2018) and *GhMYB30* (Wu et al. 2023), have been previously characterized. *GhFSN1_A12*, encoding a NAC TF, acts as a positive regulator of fiber SCW thickening by activating a series of known SCW-related genes (Zhang et al. 2018). *GhMYB30,* among the latest characterized

members of cotton MYB TFs, was found to regulate cotton fiber development by inhibiting the expression of *GhMYB46*, which was also verified in kGRN (Supplementary Fig. S17) (Wu et al. 2023). Other uncharacterized hub TFs also include *GhMYB73* and a homoeologous pair of *GhMYB60*, offering promising candidates given the well-documented roles of MYB TFs in fiber initiation, elongation, and SCW synthesis.

Focusing on the top hub TFs in kGRN subnetwork III, *GhMYS1_A10* and *GhMYS1_D10* (Figure 6C), represent a homoeologous pair of G2-like TF *MYS1* (*MYB-SHAQKYF 1*). These TFs have known functions in *Arabidopsis* wax biosynthesis and drought tolerance (Liu et al. 2022). Among the 26 targets of *GhMYS1_A10* and 21 targets of *GhMYS1_D10*, 18 target genes were commonly regulated by both, indicating substantial redundancy between homoeologous genes. Among their common targets are several known functional genes including *GhPIN3a*, *GbTCP*, *GhFSN1_A12*, *GhCesA8_D10,* and *GhCesA8_A10* (Figure 6C). In conclusion, known functional genes and their upstream TFs reflect a complex GRN of fiber elongation and SCW synthesis and also helped us identify nine highly connected TFs as candidate regulators of fiber elongation.

Functional validation of *GhMYS1* reveals its positive role in fiber elongation

Based on the top rankings of *GhMYS1A10* and *GhMYS1D10* in fiber GRNs and their significant trait associations, we selected this homoeologous pair of TFs for functional analysis (Figure 6 and Table S5-6). Comparative expression analysis showed significant upregulation of these genes at 15 dpa in cultivated versus wild cotton and in elite long-fiber versus short-fiber cotton accessions (Figure 7A and B), indicating a potential link with domestication and breeding improvements. Given the 97.29% similarity in the coding regions of *GhMYS1_A10* and *GhMYS1_D10*, VIGS primers were designed to simultaneously silence both genes. VIGS-mediated silencing successfully reduced the expression of both *GhMYS1* genes (Figure 7C), resulting in significantly shorter fibers in pCLCrVA: *GhMYS1* (23.8 mm) compared to pCLCrVA:00 control plants (28.5 mm; P =0.001352) (Figure 7D-E).

The joint analysis of DAP-seq results and cGENIE3 predictions identified five genes regulated by GhMYS1_A10 and four genes regulated by GhMYS1_D10. Among these, *GhPIN3a* and *GhWLIMa* were common targets regulated by both (Supplementary Table S16). *GhTBL7*,

GhVIN1, and *GhCesA7_D05* were exclusively targeted by GhMYS1_A10, while *GaMYB2*, and *GbAAR3* was only regulated by GhMYS1_D10 (Supplementary Table S16). Interestingly, *GaMYB2* and three genes including *GhTBL7*, *GhVIN1*, and *GhCesA7_D05* were target genes of GhMYS1_A10 and GhMYS1_D10, respectively, in at least one method of the DAP-seq and cGENIE3. To verify these regulatory relationships, we conducted Dual-luciferase reporter assay (LUC) using the promoter sequences of *GhPIN3*, *GhTBL7*, *GaMYB2*, and *GhCesA7_D05*. The LUC results showed that GhMYS1_A10 could activate *GhPIN3a* while repressing *GhMYB2*, *GhTBL7*, and *GhCesA7_D05*, even though *GhMYB2* was only predicted by GRN (Figure 7F). Consistent with the joint prediction, GhMYS1_D10 activated the expression of *GhPIN3a* and *GhMYB2* (Figure 7G). However, despite GRN and DAP-seq predictions identifing *GhTBL7* and *GhCesA7_D05* as target genes of GhMYS1_D10, respectively, the LUC experiment did not confirm these regulatory relationships, indicating potential false positives (Figure 7G, Supplementary Table S16). The common and discordant regulation of target genes by *GhMYS1_A10* and *GhMYS1_D10* highlights both functional redundancy and differentiation of these homoeologs during fiber development. These findings suggest that *GhMYS1* is a novel transcription factor regulating fiber development, potentially by modulating auxin and positively regulating fiber elongation by suppressing the expression of secondary wall formation-related genes.

Discussion

Leveraging GRN inferences for cotton fiber development study

Over the last two decades, conventional molecular genetic analyses have elucidated nearly 200 genes that are important for cotton fiber development, providing valuable insights into the genetic regulation of this process (Huang et al. 2021; Wen et al. 2023). However, previous efforts have often focused on individual genes and limited gene-to-gene inter-connections, resulting in simplified, linear, or limited local networks that fail to capture the complexity of the comprehensive genome-wide GRN governing fiber development. This limitation hinders the full exploration and discovery of the intricate biological networks governing cotton fiber development.

To address this challenge, we leveraged large-scale transcriptome datasets and a wealth of functional gene resources to construct comprehensive genome-wide GRNs for cotton fiber development. We rigorously compared and validated three distinct GRN inference methods against prior knowledge-based regulatory maps, known fiber-related functional genes, DAP-seq data, RNA-seq data from gene perturbation experiments, and phenotypic correlation analyses. This integrative approach demonstrated a carefully crafted, step-by-step process of network evaluation and optimization.

An important consideration in our study was the selection of gene expression thresholds for network construction. Conventional approaches often use generic cutoffs (e.g., greater than 1 RPKM/TPM) (Zhou et al. 2020; Chen et al. 2023), which are prone to exclude transcriptional factors (TFs) and other functional genes with low transcript abundance (Ghaemmaghami et al. 2003; Vaquerizas et al. 2009). In our study, we screened various thresholding options and set the final cutoff at TPM greater than 0 in 30% of samples (see Method or Supplementary Fig. S4), which retained 76.3% of total genes and 72.2% of TFs, ensuring the inclusion of 98.3% of known functional genes crucial for network construction. This approach enabled us to capture a diverse range of genes dynamically expressed across different stages of fiber development, such as *GhHD1* (Walford et al. 2012), *GhMYB25-like* (Walford et al. 2011), *GhWD40* (Tian et al. 2020b), facilitating the identification of the modules related to fiber initiation, even in the absence of fiber samples from 0-4 DPA (Figure 2C, Figure 6B).

GENIE3, an ensemble machine-learning algorithm based on random forests, has demonstrated superior performance in the DREAM4 and DREAM5 GRN reconstruction challenges (Huynh-Thu et al. 2010; Marbach et al. 2012). It has been extensively employed to understand the transcriptional regulation mechanism of plant traits in *Arabidopsis*, rice, wheat, and maize (Walley et al. 2016; Ezer et al. 2017; Huang et al. 2018; Ramírez-González et al. 2018; Shibata et al. 2018; Harrington et al. 2020; Ueda et al. 2020). Given that the fiber transcriptome data in this study consists of 14 time points, dynGENIE3, an adaptation of the original GENIE3 for time series data (Huynh-Thu and Geurts 2018; Balcerowicz et al. 2021), was also used for GRN construction. Additionally, we included another method, Corto (Mercatelli et al. 2020), due to its resemblance to the well-established ARACNe algorithm, which was among the early

demonstrated GRN applications known for its ability to infer direct regulatory interactions by eliminating indirect effects (Margolin et al. 2006).

To obtain an approximate "gold standard" for evaluating the performance of GRN predictions, we leveraged the prior knowledge of *Arabidopsis* regulatory interactions (AtRegMap) to assemble a cottonRegMap through orthologous relationships (Wu et al. 2021). Although the limited availability of fiber cistrome data (i.e., TF ChIP-seq or DAP-seq) hinders constructing GRNs directly from empirical evidences of TF-target relationships, we integrated DAP-seq results and transcriptomic data from perturbation experiments for key TFs such as *GhBES1.4* (Liu et al. 2023), *GhWRKY16* (Wang et al. 2021b), *GhMYS1_A10*, and *GhMYS1_D10* to reinforced the reliability of our GRN predictions. In addition to validating the capture of known regulatory relationships, we also considered the inclusion and network centrality of known fiber functional genes, concluding that GENIE3 exhibited the strongest predictive power for known regulatory relationships and key TFs. Notably, these evaluation approaches were applied with appropriate statistical tests (e.g., permutation tests), considering the different network sizes resulting from the three GRN inference methods (with node numbers of 54,237, 25,441, and 56,052 inferred by GENIE3, dynGENIE3, and Corto, respectively).

To exemplify insights gained from this integrative approach, we focussed on the developmentally important process of cellulose synthesis. Although the number of regulator and *GhCesAs* genes captured in networks differed across methods (Table 1 and Figure 5A), cGENIE3 captured the most functionally relevant regulatory relationships in CesA biosynthesis networks. For example, *GhFSN1_A12*'s negative role in suppressing fiber elongation by promoting secondary cell wall (SCW) biosynthesis (Zhang et al. 2018) was evident in its regulation of several *GhCesA* genes involved in SCW formation, including *GhCesA8_D10*, *GhCesA07_D5*, and *GhCesA4_D07*(Figure 4C). Similarly, *GhHOX3_D12*'s involvement in fiber initiation and elongation (Shan et al. 2014; Qin et al. 2022) was supported by its regulation of *GhCesA* genes involved in primary cell wall (PCW) synthesis, such as *GhCesA6_A06* and *GhCesA1_D05-1*(Figure 4C). Notably, homologous genes of SND2 and SND4, key NAC TFs of SCW synthesis in *Arabidopsis* (Taylor-Teeples et al. 2015; Zhong et al. 2021), were also identified as key regulators targeting multiple SCW-related cellulose (Figure 4C). Furthermore, novel regulatory relationships uncovered by cGENIE3, such as GhMYS1_A10's regulation on

GhCesA7_D5, were experimentally validated through LUC experiments (Figure 4C, Figure 6C and Figure 7F). These findings underscore the potential of GRN to elucidate molecular mechanisms underlying key TF-gene interactions in fiber development.

Novel TFs regulating fiber development

GRNs are invaluable tools for predicting the function of TFs. By utilizing the seeded network kGRN of known fiber genes constructed by cGENIE3, we not only confirmed eight previously known TFs (*GhTCP14*, *GhMYB46_D13*, *GhARF2b*, *GhFSN1_A12*, *GhWRKY16*, *GhGT2*, a homoeologous pair of *GhMYB30*) (Wang et al. 2013, 2021b; Zhang et al. 2018, 2021b; Huang et al. 2019; Tian et al. 2022; Wu et al. 2023), but also identified 185 novel TFs regulating known fiber genes (Figure 5A, Supplementary Table S15). Included were a pair of GhMYS1 TFs, GhMYS1_A10, and GhMYS1_D10, that are predicted to regulate 26 and 21 known genes, respectively, and which are ranked highly by prediction scores across all three GRN inference methods, suggesting their important role in fiber development (Supplementary Table S6). This hypothesis is supported by the significant association between fiber traits and gene expression (Supplementary Table S5), where the expression level of this gene pair at 15 dpa fiber is markedly higher in domesticated and elite varieties compared to wild and short-fibered varieties (Figure 7A-B). Experimental validation revealed that silencing *GhMYS1_A10* and *GhMYS1_D10* simultaneously led to a significant reduction in cotton fiber length, underscoring their role in fiber elongation during domestication and breeding processes.

Previous research indicated that MYS1 affects cuticular wax content by down-regulating genes related to wax biosynthesis when overexpressed in *Arabidopsis*, leading to increased contents of primary alcohols, alkanes, and total wax (Liu et al. 2022). With very long-chain fatty acids (VLCFAs) serving as the precursors for wax biosynthesis (Kunst and Samuels 2009) and acting upstream of the ethylene signaling pathway (Huang et al. 2021; Wen et al. 2023), we speculate that MYS1's role in fiber development involves mediating VLCFA content. Although no significant changes in the content of VLCFAs were detected in the MYS1 overexpression transgenic *Arabidopsis*, MYS1 was co-expressed with several 3-ketoacyl-CoA synthases (KCSs) involved in VLCFA biosynthesis (Liu et al. 2022). Our cGENIE3 results showed that *GhMYS1_A10* and *GhMYS1_D10* simultaneously regulate VLCFA biosynthesis-related genes

GhKCS13 (Shi et al. 2022) and *GhKCS10* (Yang et al. 2023) (Figure 6C), suggesting that *GhMYS1* may affect fiber elongation by regulating VLCFA biosynthesis.

Auxin plays a well-documented positive role in fiber initiation and elongation (Huang et al. 2021; Wen et al. 2023). *GhPIN3a*, an auxin efflux carrier, mediates fiber initiation by establishing hormone gradients in ovule epidermal cells and fibroblast cells (Zhang et al. 2017a; Zeng et al. 2019). Both GRN and DAP-seq results indicated that *GhMYS1_A10* and *GhMYS1_D10* regulate *GhPIN3*, a regulatory relationship further validated by LUC assays (Figure 5C and Figure 7F-G). Additionally, GRN identified four GhCesAs related to secondary wall formation, including *GhCesA7_A05*, *GhCesA7_D05*, *GhCesA8_A10*, and *GhCesA8_D10*, which were regulated by *GhMYS1_A10* and/or *GhMYS1_D10* (Figure 5C). DAP-seq and LUC experiments confirmed the negative regulatory relationship between *GhMYS1A* and *GhCesA7_D05* (Figure 5C and Figure 7E).

Overall, our study validates that *GhMYS1_A10* and *GhMYS1_D10* positively regulate fiber elongation by controlling auxin transport and VLCFA synthesis while inhibiting SCW formation. GRN and DAP-seq results indicate that these TFs regulate numerous genes involved in fiber development (Figure 5C and Supplementary Table S5), suggesting a more complex regulation than previously anticipated. Beyond GhMYS1_A10 and GhMYS1_D10, further exploration of other top-ranking regulators identified by GRN could provide valuable insights for improving fiber quality.

Asymmetric subgenome contribution to fiber gene expression and network properties

Previous cotton research suggested that the D subgenome exhibits dominant expression (i.e. imbalance of more D-biases than A-biases) and therefore may play a more important role overall than the A subgenome during fiber development and in response to domestication selection (Wang et al. 2017; Ma et al. 2018; Li et al. 2020; You et al. 2023). We note that differences in accessions used, fiber stage, sample numbers and calculation methods among studies have led to varying reports of homoeolog expression bias (HEB), including imbalance favoring D-biased homoeolog pairs (Hovav et al. 2008b; Pei et al. 2022; You et al. 2023) and imbalance (Yoo and Wendel 2014; Zhang et al. 2015; Mei et al. 2021). Against this backdrop of variation, our study

found no significant imbalance between A- and D- biased homoeolog expression based on 401 high-quality transcriptome datasets (Supplementary Table S11 and Figure 5A-B).

Beyond the perspective offered by biased homoeolog expression, our analysis explored the nuances of asymmetric duplicated gene expression. Notable findings include a higher number of At than Dt fiber-expressed genes a slightly higher overall transcript abundance of Dt than At genes; and more highly expressed A-biased homoeologous pairs but with lower expression differences (i.e., |At-Dt|) compared to the D-biased homoeologous pairs (Supplementary Table S11 and Figure 5). These subtle and nuanced features and their connections prompted us to speculate that it is the larger expression differences in D-biased homoeologous pairs that contribute to the higher overall transcript abundance of Dt genes, thus leading to the D subgenome exhibiting a disproportionate expression level, which has not been shown in previous studies. These nuanced features enrich our understanding of subgenome contributions to gene expression.

A particularly important methodological consideration is that the analysis of duplicated gene expression, in cotton and other allopolyploid systems (Grover et al. 2012; Bird et al. 2021; Birchler and Yang 2022), typically encompasses single-copy homoeologous gene pairs (or sets, scOGs) derived from the inference of homoeologous relationships. In cotton, even with high-quality genomes and using the latest approaches to orthology inference, such as pSONIC (Conover et al. 2021) and GENESPACE (Lovell et al. 2022), the inferred proportion of single-copy homoeolog groups range from 52% to 73% of the total genomic genic content, meaning that a substantial proportion of genes are missing from analyses of duplicated gene expression patterns. Here we specifically included these variable-copy gene groups (vcOGs) to examine subgenomic contributions and detect previously overlooked patterns. For example, we found that the average expression level of vcOGs in the A subgenome is significantly higher than that in the D subgenome, contrary to the results of scOG (Figure 5A). This finding highlights the importance of considering vcOGs in addition to scOGs when studying gene expression in polyploid systems. It is likely that epigenetic modifications, including DNA methylation and histone modifications, which affect gene expression in polyploid plants (Song and Chen 2015), might also be explored for vcOGs to further our understanding of subgenomic contributions to

allopolyploid gene expression. This comprehensive approach will provide a more detailed picture of how gene expression is regulated in polyploid systems.

Perhaps more important than the genic perspective, with respect to the genomic duplication that accompanies allopolyploidy, is that provided by gene co-expression network and regulatory network analyses. These analyses permit the exploration of the joint as well as separate contributions of the A- and D- subgenomes to fiber development, from the standpoint of a more biologically realistic network perspective. Co-expression relationships are often inferred to reflect genes with similar or biologically associated functions (Rhee and Mutwil 2014). Our study shows that scOGs present in the same module account for 48.8% of network genes (Table 2: III), which is higher than the proportions reported for other studies of cotton (Gallagher et al. 2020; Jareczek et al. 2023) and wheat (37.4%) (Ramírez-González et al. 2018). For example, in two previous studies of fiber co-expression gene network construction based on 24 wild and domesticated fiber samples, the proportion of scOGs present in the same module was 20.2-36.1% in *G. hirsutum* and 23.5% in *Gossypium. barbadense* (*G. barbadense*), suggesting that the majority of homoeologous gene pairs are in separate modules in the polyploid network (Gallagher et al. 2020; Jareczek et al. 2023). This discrepancy can likely be attributed to the different RNA-seq samples used. Compared to these earlier studies, our inclusion of more RNA-seq samples, primarily from *G. hirsutum* cultivars, could have resulted in a more connected and denser fiber network due to the effect of domestication, as previously suggested in cotton (Bao et al. 2019; Gallagher et al. 2020) and in other plants (Alonge et al. 2020; Groen et al. 2020). Consequently, we inferred more homoeolog pairs into the same modules, estimating a higher level of functional conservation or closer functional association of homoeologs. Additionally, our larger sample size reduces noise in module assignment, as variable data are more prone to placing homoeologs into different modules. Beyond the overall network structure, our results revealed modular-level features specific to associated functions. For example, the tan and green modules, which were highly expressed during the fiber elongation and SCW thickening stages, showed obvious D and A subgenome biases, respectively. These results further enriched our understanding of the contributions of different subgenomes to fiber development, providing insights that could not be discerned from a single-gene perspective.

Compared to co-expression relationships, TF-TG regulatory relationships inferred by GRNs allow for an examination of subgenomic contributions, including intra-subgenomic interactions (At-At and Dt-Dt) and inter-subgenomic interactions (At-Dt and Dt-At) as previously proposed (Hu and Wendel 2019). This aspect has been explored using three-dimensional genomic interaction (Hi-C) and expression quantitative trait locus (eQTL) methods (Li et al. 2020; Wang et al. 2018). Wang et al. (2018) characterized 3D genome architectures, revealing that inter-subgenomic interactions (At-Dt) accounted for approximately half of all interactions in tetraploid cottons (45.5% in *G. hirsutum* and 47.1% in *G. barbadense*), indicating an equivalent amount of inter- and intra-subgenomic interactions, consistent with our findings. Further, Li et al. (2020) used eQTL analysis on 15 dpa fiber transcriptomes from 251 *G. hirsutum* accessions, identifying 15,330 eQTLs associated with 9,282 genes. They found that the proportion of inter-subgenomic eQTLs was higher in the A subgenome (52.6%) than in the D subgenome (46.5%), suggesting a more prominent regulatory role of At regulators on Dt genes, consistent with our findings. However, they also observed that 44.3% of eGenes in the A-subgenome are regulated by eQTLs in the D-subgenome, whereas only 23.4% of eGenes in the D-subgenome have eQTL regulation in the A-subgenome. This highlights unequal transcriptional regulation patterns between the two subgenomes. An expanded study by You et al. (2023) using fiber transcriptomes from 376 *G. hirsutum* accessions across five time points identified 53,854 cis-eQTLs and 23,811 trans-eQTLs, revealing genetic variants associated with gene expression during fiber development. This larger dataset offers a promising avenue to further delineate inter- and intra-subgenomic regulatory effects and compare them with GRN results. As neither Hi-C nor eQTL analyses directly refined the interaction relationships between TFs and TGs, further analysis integrating eQTL and Hi-C data is needed to obtain TF-TG regulatory relationships and compare them with GRN-based regulatory relationships.

One question of broad interest regarding the functional genomics of allopolyploids is the extent to which duplicated TFs and TGs are functionally conserved in a GRN. A key result emerging from the present work is that the proportion of TG homoeologs simultaneously regulated by any given TF is significantly higher than the proportion of TF homoeologs co-regulating any given downstream genes (e.g., 15.8% vs. 6.4% in cGENIE3; Table 2: V and Supplemental Table S12). This indicates a higher level of conservation in TG promoter *cis*-regulatory sites than in TF *trans*

functions. In other words, the *trans*-regulatory roles of TFs diversify faster between homoeologs than does the *cis* landscape of their TG binding sites. This finding is consistent with the experimentally validated notion that *trans*-regulatory mutations have a larger target size compared to *cis*-regulatory mutations in yeast (Siddiq and Wittkopp 2022), hence evolving faster. Further experimental studies in cotton are needed to explore the functional and phenotypic implications of these regulatory variants. For example, in the homoeologous pair of *GhMYS1* genes, DAP-seq results demonstrated both functional conservation and divergence regarding a few target genes with known fiber-related functions. One caveat is that our VIGS experiments can only simultaneously silence both copies due to high sequence identity. Future directions include perturbation experiments targeting individual homoeologs to examine the phenotypic outcomes of disrupting network interactions.

In summary, we constructed comprehensive GRNs using a diverse collection of public RNA-seq datasets for cotton fibers. These rigorously evaluated fiber GRNs enabled us to infer numerous potential regulatory factors controlling fiber development. These include well-studied TFs such as *GhTCP14, GhFSN1_A12, GhWRKY16_D06*, and *GhMYB30*, as well as many TFs with uncharacterized functions. Experimental verification further revealed a key regulatory role of an uncharacterized pair of *GhMYS1* genes in fiber development. Our study reveals subgenomic asymmetries that either accompanied or evolved subsequent to allopolyploidization, including a global expression difference of D-biased homoeolog pairs that underlies the dominant expression of the D subgenome, and further demonstrated multidimensional characteristic of subgenomic asymmetry from the perspective of co-expression and regulatory networks. These findings elucidate the complex gene regulatory network of cotton fiber development, providing insights into the phenomenon of allopolyploidy and offering a resource for exploring genes related to fiber elongation and enhancing cotton fiber quality through breeding.

Methods

RNA-Seq data collection and processing

Twelve public cotton fiber RNA-seq datasets comprising 473 samples representing 16 time points of *Gossypium hirsutum* were downloaded from the National Center for Biotechnology

Information (NCBI) SRA depository (Supplementary Table S1). Raw reads were preprocessed using fastp (v0.20.1) (Chen et al. 2018) to remove adapters and low-quality reads. Clean reads were aligned to the reference genome *G. hirsutum* var. TM-1 UTX_v2.1(Chen et al. 2020) using Hisat2 (v2.2.1) with default settings (Kim et al. 2015), and transcript abundances were quantified as transcripts per million (TPM) using StringTie (v2.2.1) (Pertea et al. 2015). Dimensionality reduction and visualization of gene expression profiles were conducted through principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE) in R v4.0.5 (R core Team 2020). The following sample filter criteria were applied to ensure a high-quality dataset: 1) samples were exclusively from fiber tissue, specifically excluding ovular and fibreless mutant samples; 2) samples with a unique mapping rate below 70% were discarded; 3) only uniquely mapped reads were used for TPM calculation; and 4) outlier samples were identified and removed based on PCA and t-SNE.

Weighted gene co-expression gene network analysis (WGCNA)

A gene co-expression network was constructed using the WGCNA package in R (Langfelder and Horvath 2008) with data from the surviving 401 RNA-seq samples and 57,151 genes. Briefly, the TPM data was used to generate an adjacency matrix based on signed Pearson correlations between all gene pairs powered to an optimized soft thresholding of 28. The adjacency matrix considering gene-to-gene connection strength in isolation was then used to calculate a topological overlap matrix (TOM), which considered each pair of genes in relation to all other genes. Genes with highly similar expression patterns were clustered into co-expression modules, using parameters minModuleSize of 100 and mergeCutHeight of 0.25. Genes belonging to the same co-expression module were assigned the same module color, while genes that cannot be clustered into any of the co-expression modules were labeled grey.

Construction of gene regulatory networks (GRNs)

Three distinct inference strategies were used to construct fiber gene regulatory networks, including GENIE3 (Huynh-Thu et al. 2010), dynGENIE3 (Huynh-Thu and Geurts 2018), and Corto (Mercatelli et al. 2020). Each method requires both a user-provided list of transcription factors (TFs) and gene expression data to enable inference of directed network connections

(edges) from TFs to target genes. A total of 5,048 TFs were identified from the *Gossypium hirsutum* var. TM-1 reference genome (Chen et al. 2020) with PlantTFDB (Jin et al. 2017). Among these, 3,638 fiber-expressed TFs were used as the input TFs to predict targets from all 57,151 fiber-expressed genes. The resulting TF-target predictions were filtered to retain the top one million connections as output GRNs for subsequent analyses, consistent with the thresholding applied in previous studies (Ramírez-González et al. 2018; Harrington et al. 2020). For Corto, which inferred fewer than one million connections, no filtering was applied.

Corto is a correlation-based GRN inference method, implemented as a fast and lightweight R package that resembles the well-established pipeline of ARACNe algorithm (Margolin et al. 2006). Given the normalized TPM data as a gene expression matrix and a list of TFs as centroids, Corto infers direct TF-target relationships through optimized pairwise Pearson correlation. Data Processing Inequality (DPI) on correlation triplets and bootstrapping were applied to evaluate the significance of edges, using the parameters *nbootstraps*=10 and p=0.05.

GENIE3 is a machine learning-based approach for GRN inference implemented in R (Huynh-Thu et al. 2010). This method was recognized as the best-performing algorithm in the DREAM4 In Silico Multifactorial challenge (Greenfield et al. 2010) and the DREAM5 Network Inference challenge (Marbach et al. 2012). GENIE3 utilizes the Random Forests tree ensemble algorithm to solve a regression problem for each gene in the given expression dataset, determining how the expression patterns of input TFs predict the expression of the target gene. The importance measure of a TF in predicting the target gene expression serves as the weight for the TF-target regulatory link. GENIE3 was executed using the same gene expression matrix and TF list as input, with default parameters.

Dynamical GENIE3 (dynGENIE3) is an adaptation of the original GENIE3 method that was designed for GRN inference from time series data alone or in conjunction with steady-state data. This semi-parametric model accounts for the dependence between time points by modeling the temporal changes in gene expression with ordinary differential equations (ODEs). In each ODE, the transcription function is learned using a nonparametric Random Forests Model. The fiber gene expression matrix of 57,151 genes and 401 samples was reformatted into two distinct datasets, steady-state and time series, used together as input. The steady-state dataset

encompassed 251 samples of 15 days post anthesis (dpa) fibers from Li et al. (2020), focusing on a cultivar population. The time series dataset was constructed using RNA-seq data sourced from other studies (Supplementary Table S1) with at least 3 time points involved: TPM values at each time point were averaged across these studies to obtain the expression profiles spanning 14 time points; genes with a TPM value of 0 in more than two time points were removed, leading to the final inclusion of 1011 TFs and 24,331 other genes. Using both the steady-state and time series data jointly as input, dynGenie3 was executed with default parameters.

Evaluation of GRN inference

For the performance evaluation of the GRN inference methods, five independent strategies were employed:

I. Homology-based cotton Transcriptional Regulatory Map (cottonRegMap): Serving as a benchmark dataset for validating predicted regulatory links by the above GRN inference methods, this map was constructed by adapting the regulatory prediction approach of PlantRegMap (https://plantregmap.gao-lab.org/) to represent an ensemble list of known regulatory interactions in plants. Briefly, FIMO from the MEME software suite (Bailey et al. 2009) was used to scan TF binding sites in the cotton gene promoters (i.e., 2000 bp upstream of the transcriptional start sites) using a significant threshold of *p*-value <1e-5 with Fisher's exact test. Regulatory interactions between *Arabidopsis* TFs and cotton gene promoters were assigned if one or more binding sites of a TF were found in the promoter of a gene. Based on the orthologous relationships between 619 *Arabidopsis* TFs and 2,267 *G. hirsutum* TFs (1129 from the At subgenome and 1138 from the Dt subgenome), the TF-target relationships were fully projected onto the *G. hirsutum* genome to form the cottonRegMap, consisting of 53,878,120 TF-target interactions.

II. Cotton TFs with confirmed roles in fiber development: A curated set of 54 TFs with known functions in fiber development was compiled (Supplementary Table 3). Gene set enrichment analysis (GSEA) was used to test if these curated TFs were enriched among the highly ranked TF regulators in each GRN.

III. Physical regulatory relationships based on DAP-seq data: To ground truth the predicted interactions by GRN inference, DNA-affinity purification sequencing (DAP-seq) was performed on a pair of homoeologous G2-like TFs, *GhMYS1_A10* and *GhMYS1_D10*. These TFs were selected based on their consistently high regulator ranking across different GRN inference methods (see Results section for details). Additionally, published DAP-seq data for an *EMS-SUPPRESSOR1v(BES1)/BRASSINAZOLE-RESISTANT1* (*BZR1*) family TF *GhBES1.4* (Liu et al. 2023) was incorporated for validation analysis, which also exhibited high rankings in our GRN inferences. The physical regulatory relationships mapped by DAP-seq were used to validate the GRN prediction by intersecting and significance testing.

IV. RNA-seq analysis of mutants or overexpression lines: We utilized RNA-seq data from TF mutant and overexpression lines to assess the function prediction of candidate TFs in fiber development GRNs. Specifically, RNA-seq datasets for *GhWRKY16* (Wang et al. 2021b) *GhBES1.4* (Liu et al. 2023) reported from previous studies were downloaded. Differential expression analysis was conducted to compare transgenic lines with wild-type controls. The resulting differentially expressed genes (DEGs) were considered potential targets regulated by respective TFs under perturbation conditions, thereby validating the gene targets predicted by GRN. DEGs were identified using DESeq2 (Love et al. 2014) with criteria set at an absolute fold change >1 and the *P*-values <0.05 corrected by the Benjamini-Hochberg method (Benjamini and Hochberg 1995).

V. Fiber traits phenotypic association with gene expression. Corresponding to the 251 RNA-seq samples from 15 dpa fibers (Li et al. 2020), fiber traits from the same natural population were provided by Professor Maojun Wang of Huazhong Agricultural University. The best linear unbiased predictions (BLUPs) of five fiber traits (fiber length, strength, elongation, uniformity, and micronaire value) across the four environments were estimated using the lme4 package in R (Bates et al. 2014). Pearson correlation coefficients were estimated between gene expression levels in 15 dpa fibers and phenotypic variation across the population of 251 cultivars.

DAP-seq experiments

For the homoeologous TF pair of *GhMYS1_A10* and *GhMYS1_D10*, DAP-seq experiments were performed following the protocol developed by Bartlett et al (Bartlett et al. 2017). Genomic DNA (gDNA) was extracted from 10 dpa fiber of the *G. hirsutum* cultivar TM-1 using the CTAB method. The extracted gDNA was fragmented using a Covaris M220 focused-ultrasonicator (Woburn, Massachusetts, USA) to achieve an average fragment size of 200 bp. These gDNA fragments were used to construct an affinity purification library using the NGS0602-MICH TLX DNA-Seq Kit (Bluescape Hebei Biotech Co., Ltd, Baoding, China). The TF coding sequences were cloned into pFN19K HaloTag T7 SP6 Flexi vector. The TNT SP6 coupled wheat germ extract system (Promega, Wisconsin, USA) was used to express the HALO-tagged TFs in 50 µL reactions, which were incubated for 2 hours at 37 °C. The expressed proteins were directly captured using Magne HaloTag Beads (Promega) and subsequently incubated with the affinity purification library to isolate the TF-DNA binding complexes. The enriched TF-bound gDNA fragments were then eluted from the HaloTag beads, amplified by PCR, and sequenced on the NovaSeq 6000 platform. Two independent biological replicates were conducted for each TF, along with one negative control using a mock DAP-seq library without adding the expressed protein during the HaloTag beads incubation. The DAP-seq raw data have been deposited in the Genome Sequence Archive in National Genomics Data Center, China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (GSA: CRA029084 and CRA029060) that are publicly accessible at https://ngdc.cncb.ac.cn/gsa.

DAP-seq data analysis

Raw DAP-seq reads were pre-processed by removing reads containing adapters and low-quality reads using fastp (v0.20.1) (Chen et al. 2018). Clean reads were aligned to the *G. hirsutum* reference genome using Bowtie 2 (v2.4.5) (Langmead and Salzberg 2012). To identify DAP-seq peaks, MACS2 (v2.2.7.1) peak calling was performed with default parameters (Zhang et al. 2008). Identified peaks from two biological/technical duplicated samples were merged using IDR (v2.0.4.2) to assess the reliability of peak identification (Li et al. 2011). The ChIPseeker R package (v1.40.0) was used for peak annotation in relation to adjacent genes (Yu et al. 2015). Genes with significant peaks (q-value <0.05) within 2000 bp upstream of the transcription start site (TSS) were considered as target genes of the *in vitro* expressed TFs.

Functional enrichment analysis

Gene functions were annotated based on the eggNOG databases (Huerta-Cepas et al. 2019). Gene Ontology (GO) enrichment analyses were performed using the ClusterProfiler R package (v3.18.1) (Yu et al. 2012). Only GO terms with *P*-values below 0.05 were considered as significantly enriched. GO enrichment results were visualized using aPEAR (v1.0.0) (Kerseviciute and Gordevicius 2023) in R.

Dual-luciferase (LUC) reporter assay

The 2000 bp promoters of *GhMYB2*, *GhTBL4*, *GhTBL7*, *GhCesA7_D05*, and *GhPIN3a* were cloned using primers listed in Supplementary Table S17 and inserted into the pGreenII 0800-LUC vector. The full-length coding sequences of *GhMYS1_A10* and *GhMYS1_D10* were cloned into the pGreenII 62-SK vector. Resulting plasmids were transduced into *Agrobacterium tumefaciens* strain GV3101, and the LUC reporter assay was performed as previously described (Xie et al. 2017). The pGreen II 0800-LUC and pGreenII 62-SK were used as internal controls. After injecting a mixture of the fusion constructs of pGreenII 62-SK and pGreenII 0800-LUC in a 1:1 ratio into tobacco leaves for 3 days, quantitative analysis of luciferase activity was performed using a Dual-Luciferase Reporter Assay System (E1910, Promega, USA), following the manufacturer's instructions. All experiments were performed in three independent replicates.

Virus-induced gene silencing (VIGS) of *GhMYS1*

The cotton leaf crumple virus (CLCrV)-based vectors were used to perform VIGS assays (Gu et al. 2014). To simultaneously silence both *GhMYS1_A10* and *GhMYS1_D10*, a 300 bp coding sequence conserved between homoeologs was designed and inserted into the pCLCrV-A vector to generate the *pCLCrV: GhMYS1* construct. The positive recombinant plasmid of *pCLCrV*: *GhMYS1* and *pCLCrV:00* was subsequently transferred into *Agrobacterium tumefaciens* strain LBA4404 by electroporation. Primers used in vector constructions were listed in Supplementary Table S16. The auxiliary vector pCLCrVB was used to facilitate the intercellular movement of CLCrV DNA. After cultivating *Agrobacterium* colonies containing pCLCrVB, pCLCrV: *GhMYS1*, and pCLCrV: 00 vectors on a shaker at 28 °C for 24 h, *Agrobacterium* cells were collected by centrifugation and resuspended in solution (10 mM MgCl2, 10 mM MES, and 200

mM acetosyringone) to achieve OD600 = 1.2. The *A. tumefaciens* strains containing pCLCrV: *GhMYS1* and pCLCrV: 00 were mixed with pCLCrVB in equal proportions. The resulting mixture was then injected into the cotyledons of 10-day-old seedlings of *G. hirsutum* variety TM-1 using a 1 ml headless syringe. After 24 h of incubation in darkness at 24°C, all plants were transferred to a constant temperature lightroom for cultivation (25°C, 16 hours/day, 8 hours/night). Five plants were injected for each vector, consisting of three biological replicates. The expression of *GhMYS1* was examined in 15 dap fiber of pCLCrV: *GhMYS1* and pCLCrV: 00 cotton plants through RT-qPCR to determine the silencing efficacy.

Genomic single-copy orthologous-homoeolog groups (scOGs) gene identification

scOGs analysis was carried out by pSONIC software which uses MCScanX and OrthoFinder to infer species pairwise collinearity blocks and identify a high-confidence set of singleton orthologs, respectively (Conover et al. 2021). A total of 22,889 pairs of homologous genes were characterized into scOGs. The remaining 13,229 and 15,895 genes without unique correspondence in At and Dt were named variable copy ortholog groups (vcOGs).

Subgenomic expression and homoeolog expression bias (HEB) analysis

Because not all of the 45,778 genes placed in scOGs were among the 57,151 fiber-expressed genes, some scOGs were represented in expression data by only the At or Dt homoeolog. Consequently, we further categorized the 22,889 scOGs as either “paired” or “unpaired” based on whether both homoeologs were expressed (scOG paired) or if only one homoeolog was expressed (scOG unpaired). To analyze the expression levels of genes contained within the expressed OGs (vcOGs, scOG paired, and scOG unpaired genes) between the two subgenomes, we compared the average TPM values of 57,151 expressed genes across 401 samples using a two-sided Wilcoxon signed-rank test. For HEB analysis, if the TPM between scOGs in one sample exhibited a more than 2-fold change, the gene pair was identified as a biased homoeologous gene pair in that sample. We utilized a chi-square test and corrected the P value using the Benjamini-Hochberg method to compare the expression bias in 401 samples between the At and Dt subgenomes. When the number of samples with an A or D subgenome bias

exceeded the number of samples with a D or A subgenome bias, and FDR ≤ 0.05, we considered that there was an A or D subgenome bias.

Acknowledgments

We thank members from the Hu Lab and the Ma Lab for helpful discussions. We thank Dr. Vân Anh Huynh-Thu for discussion in GRN construction, the Research IT unit at Iowa State University (https://researchit.las.iastate.edu/) for computational support. This project was supported by the Innovation Program of Chinese Academy of Agricultural Sciences (CAAS-CSIAF-202402) and the National Natural Science Foundation of China (32072111) to GH, the Guangdong Basic and Applied Basic Research (Grant No. 2022A1515110758), and the U.S. Department of Agriculture ARS 58-6066-0-066 NACA "Genomics of Malvaceae" to CEG.

Author Contributions

G.H. and X.M. conceived the project. X.X. conducted data analysis and wrote the paper, with inputs from JC for ortholog and homoeolog detection. D.Z. conducted data analysis and performed the VIGS and LUC assays. C.E.G., J.F.W., X.M., and G.H., revised the manuscript. All authors read and approved the final manuscript.

Figure Legends

Figure 1. Cotton fiber transcriptomic datasets for this study. (**A**) Timeline displaying the stages represented by the 12 studies used to generate a dataset of 401 fiber RNA-seq samples for an in-depth exploration of cotton fiber development. Fiber elongation, transition, and SCW synthesis stages are indicated by red, blue, and green bars, respectively, and each line represents one existing dataset. This color scheme is applied consistently across all figures here. (**B**) Principal component analysis (PCA) of 57,151 gene expression profiles. PC1 and PC2 captured 16.8% and 11.5% of variance, respectively. (**C**) T-distributed stochastic neighbor embedding (t-SNE) was also employed for dimension reduction and visualization of the fiber expression landscape.

Figure 2. Phenotypic and functional associations of co-expression gene modules during fiber development. (A) For the 20 co-expression gene modules identified by weighted gene co-expression network analysis (WGCNA), heatmap represents Pearson correlation coefficients and

P-values (cell color and text, respectively) between the module eigengenes (MEs, by row) and fiber developmental stages treated as the binary categorical variable (by column). (**B**) ANOVA of MEs (by column) by fiber developmental stages treated as a numeric variable (MEs, by row). Heatmap cell color and text represent Pearson correlation coefficients and *P*-values, respectively. **(C)** Heatmap of z-score normalized MEs for the seven largest modules across fourteen fiber developmental time points. (**D**) Gene Ontology (GO) enrichment analysis of the seven largest modules, displaying the top two most significant interconnected GO clusters terms each. Different colors represent corresponding modules.

Figure 3. Evaluation of fiber GRN inferences. (**A**) Histogram presents the bootstrap distribution (n=1000) of cottonRegMap TF-target relationships as captured by chance. Red, blue, and green lines represent the cottonRegMap TF-target relationships inferred by GENIE3, dynGENIE3, and Corto, respectively. Both GENIE3 and Corto inferred significantly more interactions outside the bootstrap distribution. (**B**) GSEA of known functional TFs among TFs rankings inferred by cGENEI3. The enrichment score reflects the degree of over-representation of a set of 54 known functional TFs at the top of the ranked TFs identified by cGENIE3. The red dashed line indicates that these known functional TFs were significantly enriched at the top 77 ranking TFs. (**C**) Heatmap of overlapping target genes between empirical evidence (columns) and GRN inferences (rows). WRKY16, with GRN inferences for cGENIE3, cdynGENIE3, and cCorto. Each cell represents the number of overlaps and the significance of the corresponding hypergeometric test. DAP-seq results of *GhMYS1_A10*, *GhMYS1_D10*, and *GhDES1.4* as well as RNA-seq results of *GhDES1.4* and *GhWRKY16* were shown. (D) The correlation between expression variation of 77 hub TFs and fiber length was significantly higher than that of 3,638 TFs expressed in fibers. Five different percentages ranks were divided according to the correlation between TF and fiber length, where 0% to 100% represent increasing correlation.

Figure 4. GRN performance in cotton cellulose synthesis. (**A**) Categorization of *GhCesAs* based on gene expression patterns during cotton fiber development. Heatmap presents TPM expression levels in the long-fiber variety J02 and the short-fiber cotton variety ZRI015. Three hierarchical clusters correspond to PCW-related, SCW-related, and unknown *GhCesAs.* (**B**) The number of CesA genes, regulator transcription factors (TFs), and regulatory relationships identified by cGENIE3, cdynGENIE3, and cCorto. (**C**) Cellulose synthesis-related subnetwork

inferred by cGENIE3. Square and round nodes represent *GhCesAs* and TFs, respectively, which are connected by directed edges indicating the TF-target relationships inferred. Red and blue node colors represent the categorization of PCW-related and SCW-related genes based on expression patterns during fiber development, respectively. Two network components were detected corresponding to PCW (left) and SCW (right), which were co-regulated by six TFs in the middle. (**D**) Ranking *GhCesAs* by in-degree (i.e., number of incoming linking) from all TFs inferred by cGENIE3. (**E**) Ranking cellulose synthesis related TFs by out-degree (i.e., number of outward links) to target *GhCesAs*. (**F**) Enriched GO terms associated with the 71 TFs inferred by cGENIE3.

Figure 5. Expression level analysis of homoeologous gene pairs. (**A**) Gene expression levels compared between At and Dt homoeologs for all 57,151 fiber-expressed genes ("all genes"), 22,889 homoeologous pairs characterized into single-copy ortho-homoeolog groups ("scOGs"), the remaining 13,229 At and 15,895 Dt genes uncategorized ("vcOGs"), 19,213 scOGs with both At and Dt expressed in fiber ("scOGs pair"), and 17,028 scOGs with only one homoeolog expressed in fibers ("scOGs unpair"). (**B**) Gene expression levels compared for scOGs pairs exhibiting homoeolog expression bias (HEB). (**C**) Absolute expression differences compared between A-biased and D-biased scOGs. (**C**) Expression comparisons for scOGs present within the same co-expression modules identified by WGCNA. (**D**) Absolute expression differences compared between A-biased and D-biased scOGs in co-expression modules. Statistical significance was determined using a two-sided Wilcoxon rank-sum test. ***$P< 0.001$.

Figure 6. GRN built based on known function genes and their directly regulated TF in fiber. (**A**) GRN of known functional genes and their regulated TFs. Known functional genes and TFs are shown as circles and rhombus, respectively. Different colors indicate the modules where genes and TFs are located in the co-expression network. (**B**) Novel TFs in brown module regulate *GhHOX3*, *GhHD1*, *GhMYB25-like*, and *GhWD40* involved in fiber initiation. (**C**) Network of known functional genes regulated by *GhMYS1_A10* and *GhMYS1_D10*.

Figure 7 *GhMYS1* positively regulates fiber elongation. (A) Expression pattern analysis of *GhMYS1_A10* and *GhMYS1_D10* in wild and domestication cotton accession from 5 to 25 days post-anthesis (dpa). (**B**) Expression pattern analysis of *GhMYS1_A10* and *GhMYS1_D10* in long-

fiber (J02) and short-fiber (ZRI105) varieties from 5 to 25 dpa. (**C**) Relative expression levels measured by qRT-PCR showed reduced *GhMYS1* expression in 10 dpa fibers from pCLCrVA: *GhMYS1* cotton plants relative to pCLCrVA: 00 plants. (**D**) Significantly shorter mature fiber length in pCLCrVA: *GhMYS1* versus pCLCrVA: 00 plants. (**E**) Phenotype of mature fibers in pCLCrVA: 00 and pCLCrVA: *GhMYS1* plants. bar = 1 cm. (**F-G**) Transient dual-luciferase (LUC) reporter assay testing interactions between *GhMYS1_A10* (**F**) and *GhMYS1_D10* (**G**), and the promoters of *GhPIN3a*, *GhCesA7_D05*, *GhTBL7*, and *GhMYB2*. Expression of Renilla luciferase (REN) was used as an internal control. Values given are mean ± SD (n = 4). Relative LUC activity obtained with the empty plasmid (none) was set to 1. Statistically significant differences between groups as determined by Student's t-test. **P*< 0.05 and ***P*< 0.01.

Tables

Table 1. Fiber gene regulatory networks constructed.

Table 2. Subgenomic contribution to fiber-expressed genes.

Supplementary data

Supplementary Figure S1. Number of RNA-seq samples representing each time point before (left) and after (right) quality control.

Supplementary Figure S2. Dimensionality reduction and visualization of gene expression profiles for the 413 public RNA-seq samples passing quality control before removing 12 outlier samples.

Supplementary Figure S3. Expression analysis of 192 fiber-related functional genes clustered into three groups

Supplementary Figure S4. Criterion testing for filtering fiber-expressed genes.

Supplementary Figure S5. Weighted gene co-expression network analysis of 57,151 fiber-expressed genes.

Supplementary Figure S6. Enriched GO terms of the seven largest modules as illustrated by an UpSet plot.

Supplementary Figure S7. Plant hormone-related GO pathways enriched in the brown (A), tan (B), turquoise (C), and red (D) modules.

Supplementary Figure S8. GSEA shows enrichment of known functional TFs in TFs identified by cdynGENIE3(A), and cCorto(B).

Supplementary Figure S9. Evaluation of GRN inferences by DAP-seq.

Supplementary Figure S10. Evaluation of GRN inferences by RNA-seq.

Supplementary Figure S11. Genome-wide characterization of CesA coding genes in *G. hirsutum*.

Supplementary Figure S12. Expression pattern analysis of TFs regulating cellulose synthase identified by cGENIE3 in long fiber and short fiber cotton varieties.

Supplementary Figure S13. cdynGENIE3 predicted GRN for cotton cellulose synthesis.

Supplementary Figure S14. cCorto predicted GRN for cotton cellulose synthesis.

Supplementary Figure S15. Absolute expression differences compared between A-biased and D-biased scOGs.

Supplementary Figure S16. Gene expression levels compared for scOGs pairs exhibiting homoeolog expression bias (HEB) in co-expression modules.

Supplementary Figure S17. GRN of known functional genes regulated by *GhMYB30_A07* and *GhMYB30_D07*.

Supplementary Table S1. RNA-seq datasets from 12 studies were used in this study.

Supplementary Table S2. Summary statistics of 413 RNA-seq samples passing quality filters.

Supplementary Table S3. A curated list of 192 fiber-related genes with known functions.

Supplementary Table S4. Significantly enriched GO terms of seven largest co-expression modules identified by WGCNA.

Supplementary Table S5. The association analysis between the five fiber traits and the expression in 15 dpa fiber of 77 hub genes identified by cGENIE3.

Supplementary Table S6. A comprehensive ranking of TFs based on target gene numbers among cGENIE3, cCorto, and cdynGENIE3.

Supplementary Table S7. The association analysis between the five fiber traits and the expression in 15 dpa fiber of 77 hub genes identified by cdynGENIE3.

Supplementary Table S8. The association analysis between the five fiber traits and the expression in 15 dpa fiber of 77 hub genes identified by cCorto.

Supplementary Table S9. homologous transcription factor and gene pairs in cellulose synthesis-related subnetwork inferred by cGENIE3.

Supplementary Table S10. The number of expressed paired and unpaired scOGs in different modules.

Supplementary Table S11. Homoeolog expression bias by module.

Supplementary Table S12. Estimates of regulatory functional conservation between homoeologs in GRNs.

Supplementary Table S13. The information about 432 nodes and 657 edges in kGRN.

Supplementary Table S14. The detail of eight known fiber-related TFs in kGRN that directly regulate other known genes.

Supplementary Table S15. Functional information of homologous genes in Arabidopsis thaliana of 195 upstream transcription factors in kGRN.

Supplementary Table S16. know-function target genes of GhMYS1_A10 and GhMYS1_D10 identified by cGENIE3 and DAP-seq.

Supplementary Table S17. The primers used in this study.

References

Alonge M, Wang X, Benoit M, Soyk S, Pereira L, Zhang L, Suresh H, Ramakrishnan S, Maumus F, Ciren D, et al. Major Impacts of Widespread Structural Variation on Gene Expression and Crop Improvement in Tomato. Cell. 2020:**182**(1):145–161.e23. https://doi.org/10.1016/j.cell.2020.05.021

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, and Noble WS. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 2009:**37**(Web Server issue):W202–8. https://doi.org/10.1093/nar/gkp335

Balcerowicz M, Mahjoub M, Nguyen D, Lan H, Stoeckle D, Conde S, Jaeger KE, Wigge PA, and Ezer D. An early-morning gene network controlled by phytochromes and cryptochromes regulates photomorphogenesis pathways in Arabidopsis. Mol Plant. 2021:**14**(6):983–996. https://doi.org/10.1016/j.molp.2021.03.019

Bao Y, Hu G, Grover CE, Conover J, Yuan D, and Wendel JF. Unraveling cis and trans regulatory evolution during cotton domestication. Nat Commun. 2019:**10**(1):5399. https://doi.org/10.1038/s41467-019-13386-w

Barabási A-L and Oltvai ZN. Network biology: understanding the cell's functional organization. Nat Rev Genet. 2004:**5**(2):101–113. https://doi.org/10.1038/nrg1272

Bartlett A, O'Malley RC, Huang S-SC, Galli M, Nery JR, Gallavotti A, and Ecker JR. Mapping genome-wide transcription-factor binding sites using DAP-seq. Nat Protoc. 2017:**12**(8):1659–1672. https://doi.org/10.1038/nprot.2017.055

Bates D, Mächler M, Bolker B, and Walker S. Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:14065823. 2014.

Benjamini Y and Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological). 1995:**57**(1):289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Birchler JA and Yang H. The multiple fates of gene duplications: Deletion, hypofunctionalization, subfunctionalization, neofunctionalization, dosage balance constraints, and neutral variation. Plant Cell. 2022:**34**(7):2466–2474. https://doi.org/10.1093/plcell/koac076

Bird KA, Niederhuth CE, Ou S, Gehan M, Pires JC, Xiong Z, VanBuren R, and Edger PP. Replaying the evolutionary tape to investigate subgenome dominance in allopolyploid Brassica napus. New Phytol. 2021:**230**(1):354–371. https://doi.org/10.1111/nph.17137

Buenrostro JD, Wu B, Chang HY, and Greenleaf WJ. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr Protoc Mol Biol. 2015:**109**:21.29.1–21.29.9. https://doi.org/10.1002/0471142727.mb2129s109

Cao J-F, Zhao B, Huang C-C, Chen Z-W, Zhao T, Liu H-R, Hu G-J, Shangguan X-X, Shan C-M, Wang L-J, et al. The miR319-Targeted GhTCP4 Promotes the Transition from Cell Elongation to Wall Thickening in Cotton Fiber. Mol Plant. 2020:**13**(7):1063–1077. https://doi.org/10.1016/j.molp.2020.05.006

Chang Y-M, Lin H-H, Liu W-Y, Yu C-P, Chen H-J, Wartini PP, Kao Y-Y, Wu Y-H, Lin J-J, Lu M-YJ, et al. Comparative transcriptomics method to infer gene coexpression networks and its applications to maize and rice leaf transcriptomes. Proc Natl Acad Sci U S A. 2019:**116**(8):3091–3099. https://doi.org/10.1073/pnas.1817621116

Chen S, Zhou Y, Chen Y, and Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018:**34**(17):i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Chen Y, Guo Y, Guan P, Wang Y, Wang X, Wang Z, Qin Z, Ma S, Xin M, Hu Z, et al. A wheat integrative regulatory network from large-scale complementary functional datasets enables trait-associated gene discovery for crop improvement. Mol Plant. 2023:**16**(2):393–414. https://doi.org/10.1016/j.molp.2022.12.019

Chen ZJ, Sreedasyam A, Ando A, Song Q, De Santiago LM, Hulse-Kemp AM, Ding M, Ye W, Kirkbride RC, Jenkins J, et al. Genomic diversifications of five *Gossypium* allopolyploid species and their impact on cotton improvement. Nat Genet. 2020:**52**(5):525–533. https://doi.org/10.1038/s41588-020-0614-5

Conover JL, Sharbrough J, and Wendel JF. pSONIC: Ploidy-aware Syntenic Orthologous Networks Identified via Collinearity. G3 . 2021. https://doi.org/10.1093/g3journal/jkab170

Ezer D, Shepherd SJK, Brestovitsky A, Dickinson P, Cortijo S, Charoensawan V, Box MS, Biswas S, Jaeger KE, and Wigge PA. The G-Box Transcriptional Regulatory Code in Arabidopsis. Plant Physiol. 2017:**175**(2):628–640. https://doi.org/10.1104/pp.17.01086

Furey TS. ChIP–seq and beyond: new and improved methodologies to detect and characterize

protein–DNA interactions. Nat Rev Genet. 2012:**13**(12):840–852. https://doi.org/10.1038/nrg3306

Gallagher JP, Grover CE, Hu G, Jareczek JJ, and Wendel JF. Conservation and divergence in duplicated fiber coexpression networks accompanying domestication of the polyploid *Gossypium hirsutum* L. G3. 2020:**10**(8):2879–2892. https://doi.org/10.1534/g3.120.401362

Gaudinier A and Brady SM. Mapping Transcriptional Networks in Plants: Data-Driven Discovery of Novel Biological Mechanisms. Annu Rev Plant Biol. 2016:**67**:575–594. https://doi.org/10.1146/annurev-arplant-043015-112205

Gaudinier A, Rodriguez-Medina J, Zhang L, Olson A, Liseron-Monfils C, Bågman A-M, Foret J, Abbitt S, Tang M, Li B, et al. Transcriptional regulation of nitrogen-associated metabolism and growth. Nature. 2018:**563**(7730):259–264. https://doi.org/10.1038/s41586-018-0656-3

Ghaemmaghami S, Huh W-K, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK, and Weissman JS. Global analysis of protein expression in yeast. Nature. 2003:**425**(6959):737–741. https://doi.org/10.1038/nature02046

Gong S-Y, Huang G-Q, Sun X, Qin L-X, Li Y, Zhou L, and Li X-B. Cotton KNL1, encoding a class II KNOX transcription factor, is involved in regulation of fibre development. J Exp Bot. 2014:**65**(15):4133–4147. https://doi.org/10.1093/jxb/eru182

Greenfield A, Madar A, Ostrer H, and Bonneau R. DREAM4: Combining genetic and dynamic information to identify biological networks and dynamical models. PLoS One. 2010:**5**(10):e13397. https://doi.org/10.1371/journal.pone.0013397

Groen SC, Ćalić I, Joly-Lopez Z, Platts AE, Choi JY, Natividad M, Dorph K, Mauck WM 3rd, Bracken B, Cabral CLU, et al. The strength and pattern of natural selection on gene expression in rice. Nature. 2020:**578**(7796):572–576. https://doi.org/10.1038/s41586-020-1997-2

Grover CE, Gallagher JP, Szadkowski EP, Yoo MJ, Flagel LE, and Wendel JF. Homoeolog expression bias and expression level dominance in allopolyploids. New Phytol. 2012:**196**(4):966–971. https://doi.org/10.1111/j.1469-8137.2012.04365.x

Gu Z, Huang C, Li F, and Zhou X. A versatile system for functional analysis of genes and microRNAs in cotton. Plant Biotechnol J. 2014:**12**(5):638–649. https://doi.org/10.1111/pbi.12169

Haigler CH, Betancur L, Stiff MR, and Tuttle JR. Cotton fiber: a powerful single-cell model for cell wall and cellulose research. Front Plant Sci. 2012:**3**:104. https://doi.org/10.3389/fpls.2012.00104

Han L, Zhong W, Qian J, Jin M, Tian P, Zhu W, Zhang H, Sun Y, Feng J-W, Liu X, et al. A multi-omics integrative network map of maize. Nat Genet. 2023:**55**(1):144–153. https://doi.org/10.1038/s41588-022-01262-1

Hao J, Tu L, Hu H, Tan J, Deng F, Tang W, Nie Y, and Zhang X. GbTCP, a cotton TCP transcription factor, confers fibre elongation and root hair development by a complex regulating system. J Exp Bot. 2012:**63**(17):6267–6281. https://doi.org/10.1093/jxb/ers278

Haque S, Ahmad JS, Clark NM, Williams CM, and Sozzani R. Computational prediction of gene regulatory networks in plant growth and development. Curr Opin Plant Biol. 2019:**47**:96–105. https://doi.org/10.1016/j.pbi.2018.10.005

Harrington SA, Backhaus AE, Singh A, Hassani-Pak K, and Uauy C. The Wheat GENIE3 Network Provides Biologically-Relevant Information in Polyploid Wheat. G3 . 2020:**10**(10):3675–3686. https://doi.org/10.1534/g3.120.401436

He S, Sun G, Geng X, Gong W, Dai P, Jia Y, Shi W, Pan Z, Wang J, Wang L, et al. The genomic basis of geographic differentiation and fiber improvement in cultivated cotton. Nat Genet. 2021:**53**(6):916–924. https://doi.org/10.1038/s41588-021-00844-9

Hinchliffe DJ, Condon BD, Thyssen G, Naoumkina M, Madison CA, Reynolds M, Delhom CD, Fang DD, Li P, and McCarty J. The GhTT2_A07 gene is linked to the brown colour and natural flame retardancy phenotypes of Lc1 cotton (Gossypium hirsutum L.) fibres. J Exp Bot. 2016:**67**(18):5461–5471. https://doi.org/10.1093/jxb/erw312

Hovav R, Udall JA, Chaudhary B, Hovav E, Flagel L, Hu G, and Wendel JF. The evolution of spinnable cotton fiber entailed prolonged development and a novel metabolism. PLoS Genet. 2008a:**4**(2):e25. https://doi.org/10.1371/journal.pgen.0040025

Hovav R, Udall JA, Chaudhary B, Rapp R, Flagel L, and Wendel JF. Partitioned expression of duplicated genes during development and evolution of a single cell in a polyploid plant. Proc Natl Acad Sci U S A. 2008b:**105**(16):6191–6195. https://doi.org/10.1073/pnas.0711569105

Huang G, Huang J-Q, Chen X-Y, and Zhu Y-X. Recent Advances and Future Perspectives in Cotton Research. Annu Rev Plant Biol. 2021:**72**:437–462. https://doi.org/10.1146/annurev-arplant-080720-113241

Huang G, Wu Z, Percy RG, Bai M, Li Y, Frelichowski JE, Hu J, Wang K, Yu JZ, and Zhu Y. Genome sequence of Gossypium herbaceum and genome updates of Gossypium arboreum and Gossypium hirsutum provide insights into cotton A-genome evolution. Nat Genet. 2020:**52**(5):516–524. https://doi.org/10.1038/s41588-020-0607-4

Huang J, Chen F, Wu S, Li J, and Xu W. Cotton GhMYB7 is predominantly expressed in developing fibers and regulates secondary cell wall biosynthesis in transgenic Arabidopsis. Sci China Life Sci. 2016:**59**(2):194–205. https://doi.org/10.1007/s11427-015-4991-4

Huang J, Guo Y, Sun Q, Zeng W, Li J, Li X, and Xu W. Genome-Wide Identification of R2R3-MYB Transcription Factors Regulating Secondary Cell Wall Thickening in Cotton Fiber Development. Plant Cell Physiol. 2019:**60**(3):687–701. https://doi.org/10.1093/pcp/pcy238

Huang J, Zheng J, Yuan H, and McGinnis K. Distinct tissue-specific transcriptional regulation revealed by gene regulatory networks in maize. BMC Plant Biol. 2018:**18**(1):111. https://doi.org/10.1186/s12870-018-1329-y

Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms

1349 and 2502 viruses. Nucleic Acids Res. 2019:**47**(D1):D309–D314.
1350 https://doi.org/10.1093/nar/gky1085
1351 **Hu G, Grover CE, Jareczek J, Yuan D, Dong Y, Miller E, Conover JL, and Wendel JF**.
1352 Evolution and Diversity of the Cotton Genome. . In. Cotton Precision Breeding, M-U-
1353 Rahman, Y Zafar, and T Zhang, eds. (Springer International Publishing: Cham), pp. 25–78.
1354 https://doi.org/10.1007/978-3-030-64504-5_2
1355 **Hu G, Grover C, Vera D, Lung P-Y, Girimurugan S, Miller E, Conover J, Ou S, Xiong X,**
1356 **Zhu D, et al.** Evolutionary dynamics of chromatin structure and duplicate gene expression
1357 in diploid and allopolyploid cotton. 2023. https://doi.org/10.21203/rs.3.rs-3373364/v1
1358 **Hu G and Wendel JF**. *Cis–trans* controls and regulatory novelty accompanying
1359 allopolyploidization. New Phytol. 2019:**221**(4):1691–1700.
1360 https://doi.org/10.1111/nph.15515
1361 **Hu Y, Chen J, Fang L, Zhang Z, Ma W, Niu Y, Ju L, Deng J, Zhao T, Lian J, et al.**
1362 Gossypium barbadense and Gossypium hirsutum genomes provide insights into the origin
1363 and evolution of allotetraploid cotton. Nat Genet. 2019:**51**(4):739–748.
1364 https://doi.org/10.1038/s41588-019-0371-5
1365 **Huynh-Thu VA and Geurts P**. dynGENIE3: dynamical GENIE3 for the inference of gene
1366 networks from time series expression data. Sci Rep. 2018:**8**(1):3384.
1367 https://doi.org/10.1038/s41598-018-21715-0
1368 **Huynh-Thu VA, Irrthum A, Wehenkel L, and Geurts P**. Inferring regulatory networks from
1369 expression data using tree-based methods. PLoS One. 2010:**5**(9).
1370 https://doi.org/10.1371/journal.pone.0012776
1371 **Jareczek JJ, Grover CE, Hu G, Xiong X, Arick MA Ii, Peterson DG, and Wendel JF**.
1372 Domestication over Speciation in Allopolyploid Cotton Species: A Stronger Transcriptomic
1373 Pull. Genes. 2023:**14**(6). https://doi.org/10.3390/genes14061301
1374 **Jin J, He K, Tang X, Li Z, Lv L, Zhao Y, Luo JC, Gao G.** An Arabidopsis transcriptional
1375 regulatory map reveals distinct functional and evolutionary features of novel transcription
1376 factors. Mol Biol Evol. 2015: **32**(7):1767–1773. https://doi.org/10.1093/molbev/msx245
1377 **Jin J, Tian F, Yang D-C, Meng Y-Q, Kong L, Luo J, and Gao G**. PlantTFDB 4.0: toward a
1378 central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Res.
1379 2017:**45**(D1):D1040–D1045. https://doi.org/10.1093/nar/gkw982
1380 **Jones DM and Vandepoele K**. Identification and evolution of gene regulatory networks:
1381 insights from comparative studies in plants. Curr Opin Plant Biol. 2020:**54**:42–48.
1382 https://doi.org/10.1016/j.pbi.2019.12.008
1383 **Kerseviciute I and Gordevicius J**. aPEAR: an R package for autonomous visualisation of
1384 pathway enrichment networks. bioRxiv. 2023:2023.03.28.534514.
1385 https://doi.org/10.1101/2023.03.28.534514
1386 **Kim D, Langmead B, and Salzberg SL**. HISAT: a fast spliced aligner with low memory
1387 requirements. Nat Methods. 2015:**12**(4):357–360. https://doi.org/10.1038/nmeth.3317
1388 **Kulkarni SR and Vandepoele K**. Inference of plant gene regulatory networks using data-driven

methods: A practical overview. Biochim Biophys Acta Gene Regul Mech. 2020:**1863**(6):194447. https://doi.org/10.1016/j.bbagrm.2019.194447

Kunst L and Samuels L. Plant cuticles shine: advances in wax biosynthesis and export. Curr Opin Plant Biol. 2009:**12**(6):721–727. https://doi.org/10.1016/j.pbi.2009.09.009

Langfelder P and Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008:**9**:559. https://doi.org/10.1186/1471-2105-9-559

Langmead B and Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012:**9**(4):357–359. https://doi.org/10.1038/nmeth.1923

Levine M and Davidson EH. Gene regulatory networks for development. Proc Natl Acad Sci U S A. 2005:**102**(14):4936–4942. https://doi.org/10.1073/pnas.0408031102

Li L, Wang X-L, Huang G-Q, and Li X-B. Molecular characterization of cotton GhTUA9 gene specifically expressed in fibre and involved in cell elongation. J Exp Bot. 2007:**58**(12):3227–3238. https://doi.org/10.1093/jxb/erm167

Linde J, Schulze S, Henkel SG, and Guthke R. Data- and knowledge-based modeling of gene regulatory networks: an update. EXCLI J. 2015:**14**:346–378. https://doi.org/10.17179/excli2015-168

Li Q, Brown JB, Huang H, and Bickel PJ. Measuring reproducibility of high-throughput experiments. The annals of applied. 2011.

Liu B, Zhu Y, and Zhang T. The R3-MYB gene GhCPC negatively regulates cotton fiber elongation. PLoS One. 2015:**10**(2):e0116272. https://doi.org/10.1371/journal.pone.0116272

Liu L, Chen G, Li S, Gu Y, Lu L, Qanmber G, Mendu V, Liu Z, Li F, and Yang Z. A brassinosteroid transcriptional regulatory network participates in regulating fiber elongation in cotton. Plant Physiol. 2023:**191**(3):1985–2000. https://doi.org/10.1093/plphys/kiac590

Liu Q, Huang H, Chen Y, Yue Z, Wang Z, Qu T, Xu D, Lü S, and Hu H. Two Arabidopsis MYB-SHAQKYF transcription repressors regulate leaf wax biosynthesis via transcriptional suppression on DEWAX. New Phytol. 2022:**236**(6):2115–2130. https://doi.org/10.1111/nph.18498

Liu W, Song C, Ren Z, Zhang Z, Pei X, Liu Y, He K, Zhang F, Zhao J, Zhang J, et al. Genome-wide association study reveals the genetic basis of fiber quality traits in upland cotton (Gossypium hirsutum L.). BMC Plant Biol. 2020:**20**(1):395. https://doi.org/10.1186/s12870-020-02611-0

Li Z, Wang P, You C, Yu J, Zhang X, Yan F, Ye Z, Shen C, Li B, Guo K, et al. Combined GWAS and eQTL analysis uncovers a genetic regulatory network orchestrating the initiation of secondary cell wall development in cotton. New Phytol. 2020:**226**(6):1738–1752. https://doi.org/10.1111/nph.16468

Lovell JT, Sreedasyam A, Schranz ME, Wilson M, Carlson JW, Harkess A, Emms D, Goodstein DM, and Schmutz J. GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. Elife. 2022:**11**. https://doi.org/10.7554/eLife.78526

Love MI, Huber W, and Anders S. Moderated estimation of fold change and dispersion for

RNA-seq data with DESeq2. Genome Biol. 2014:**15**(12):550. https://doi.org/10.1186/s13059-014-0550-8

Lu Q, Shi Y, Xiao X, Li P, Gong J, Gong W, Liu A, Shang H, Li J, Ge Q, et al. Transcriptome Analysis Suggests That Chromosome Introgression Fragments from Sea Island Cotton (Gossypium barbadense) Increase Fiber Strength in Upland Cotton (Gossypium hirsutum). G3. 2017:**7**(10):3469–3479. https://doi.org/10.1534/g3.117.300108

Machado A, Wu Y, Yang Y, Llewellyn DJ, and Dennis ES. The MYB transcription factor GhMYB25 regulates early fibre and trichome development. Plant J. 2009:**59**(1):52–62. https://doi.org/10.1111/j.1365-313X.2009.03847.x

Ma J, Jiang Y, Pei W, Wu M, Ma Q, Liu J, Song J, Jia B, Liu S, Wu J, et al. Expressed genes and their new alleles identification during fibre elongation reveal the genetic factors underlying improvements of fibre length in cotton. Plant Biotechnol J. 2022:**20**(10):1940–1955. https://doi.org/10.1111/pbi.13874

Marbach D, Costello JC, Küffner R, Vega NM, Prill RJ, Camacho DM, Allison KR, DREAM5 Consortium, Kellis M, Collins JJ, et al. Wisdom of crowds for robust gene network inference. Nat Methods. 2012:**9**(8):796–804. https://doi.org/10.1038/nmeth.2016

Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Dalla Favera R, and Califano A. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics. 2006:**7 Suppl 1**(Suppl 1):S7. https://doi.org/10.1186/1471-2105-7-S1-S7

Ma Z, He S, Wang X, Sun J, Zhang Y, Zhang G, Wu L, Li Z, Liu Z, Sun G, et al. Resequencing a core collection of upland cotton identifies genomic variation and loci influencing fiber quality and yield. Nat Genet. 2018:**50**(6):803–813. https://doi.org/10.1038/s41588-018-0119-7

Mei H, Qi B, Han Z, Zhao T, Guo M, Han J, Zhang J, Guan X, Hu Y, Zhang T, et al. Subgenome Bias and Temporal Postponement of Gene Expression Contributes to the Distinctions of Fiber Quality in Gossypium Species. Front Plant Sci. 2021:**12**:819679. https://doi.org/10.3389/fpls.2021.819679

Mercatelli D, Lopez-Garcia G, and Giorgi FM. corto: a lightweight R package for gene network inference and master regulator analysis. Bioinformatics. 2020:**36**(12):3916–3917. https://doi.org/10.1093/bioinformatics/btaa223

O'Malley RC, Huang S-SC, Song L, Lewsey MG, Bartlett A, Nery JR, Galli M, Gallavotti A, and Ecker JR. Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. Cell. 2016:**165**(5):1280–1292. https://doi.org/10.1016/j.cell.2016.04.038

Pei L, Huang X, Liu Z, Tian X, You J, Li J, Fang DD, Lindsey K, Zhu L, Zhang X, et al. Dynamic 3D genome architecture of cotton fiber reveals subgenome-coordinated chromatin topology for 4-staged single-cell differentiation. Genome Biol. 2022:**23**(1):45. https://doi.org/10.1186/s13059-022-02616-y

Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, and Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol.

2015:**33**(3):290–295. https://doi.org/10.1038/nbt.3122

Qin Y, Sun M, Li W, Xu M, Shao L, Liu Y, Zhao G, Liu Z, Xu Z, You J, et al. Single-cell RNA-seq reveals fate determination control of an individual fibre cell initiation in cotton (Gossypium hirsutum). Plant Biotechnol J. 2022:**20**(12):2372–2388. https://doi.org/10.1111/pbi.13918

Ramírez-González RH, Borrill P, Lang D, Harrington SA, Brinton J, Venturini L, Davey M, Jacobs J, van Ex F, Pasha A, et al. The transcriptional landscape of polyploid wheat. Science. 2018:**361**(6403). https://doi.org/10.1126/science.aar6089

Rhee SY and Mutwil M. Towards revealing the functions of all genes in plants. Trends Plant Sci. 2014:**19**(4):212–221. https://doi.org/10.1016/j.tplants.2013.10.006

Santner A and Estelle M. Recent advances and emerging trends in plant hormone signalling. Nature. 2009:**459**(7250):1071–1078. https://doi.org/10.1038/nature08122

Shan C-M, Shangguan X-X, Zhao B, Zhang X-F, Chao L-M, Yang C-Q, Wang L-J, Zhu H-Y, Zeng Y-D, Guo W-Z, et al. Control of cotton fibre elongation by a homeodomain transcription factor GhHOX3. Nat Commun. 2014:**5**:5519. https://doi.org/10.1038/ncomms6519

Shibata M, Breuer C, Kawamura A, Clark NM, Rymen B, Braidwood L, Morohashi K, Busch W, Benfey PN, Sozzani R, et al. GTL1 and DF1 regulate root hair growth through transcriptional repression of ROOT HAIR DEFECTIVE 6-LIKE 4 in Arabidopsis. Development. 2018:**145**(3). https://doi.org/10.1242/dev.159707

Shi Z, Chen X, Xue H, Jia T, Meng F, Liu Y, Luo X, Xiao G, and Zhu S. GhBZR3 suppresses cotton fiber elongation by inhibiting very-long-chain fatty acid biosynthesis. Plant J. 2022:**111**(3):785–799. https://doi.org/10.1111/tpj.15852

Siddiq MA and Wittkopp PJ. Mechanisms of regulatory evolution in yeast. Curr Opin Genet Dev. 2022:**77**:101998. https://doi.org/10.1016/j.gde.2022.101998

Song L and Crawford GE. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. Cold Spring Harb Protoc. 2010:**2010**(2):db.prot5384. https://doi.org/10.1101/pdb.prot5384

Song Q and Chen ZJ. Epigenetic and developmental regulation in plant polyploids. Curr Opin Plant Biol. 2015:**24**:101–109. https://doi.org/10.1016/j.pbi.2015.02.007

Springer N, de León N, and Grotewold E. Challenges of Translating Gene Regulatory Information into Agronomic Improvements. Trends Plant Sci. 2019:**24**(12):1075–1082. https://doi.org/10.1016/j.tplants.2019.07.004

Sun S, Xiong X-P, Zhu Q, Li Y-J, and Sun J. Transcriptome Sequencing and Metabolome Analysis Reveal Genes Involved in Pigmentation of Green-Colored Cotton Fibers. Int J Mol Sci. 2019a:**20**(19). https://doi.org/10.3390/ijms20194838

Sun W, Gao Z, Wang J, Huang Y, Chen Y, Li J, Lv M, Wang J, Luo M, and Zuo K. Cotton fiber elongation requires the transcription factor GhMYB212 to regulate sucrose transportation into expanding fibers. New Phytol. 2019b:**222**(2):864–881. https://doi.org/10.1111/nph.15620

Sun X, Gong S-Y, Nie X-Y, Li Y, Li W, Huang G-Q, and Li X-B. A R2R3-MYB transcription factor that is specifically expressed in cotton (Gossypium hirsutum) fibers affects secondary cell wall biosynthesis and deposition in transgenic Arabidopsis. Physiol Plant. 2015:**154**(3):420–432. https://doi.org/10.1111/ppl.12317

Tang M, Li B, Zhou X, Bolt T, Li JJ, and Cruz N. A genome-scale TF–DNA interaction network of transcriptional regulation of Arabidopsis primary and specialized metabolism. Mol Syst Biol. 2021.

Tang W, Tu L, Yang X, Tan J, Deng F, Hao J, Guo K, Lindsey K, and Zhang X. The calcium sensor GhCaM7 promotes cotton fiber elongation by modulating reactive oxygen species (ROS) production. New Phytol. 2014:**202**(2):509–520. https://doi.org/10.1111/nph.12676

Taylor-Teeples M, Lin L, de Lucas M, Turco G, Toal TW, Gaudinier A, Young NF, Trabucco GM, Veling MT, Lamothe R, et al. An Arabidopsis gene regulatory network for secondary cell wall synthesis. Nature. 2015:**517**(7536):571–575. https://doi.org/10.1038/nature14099

Tian F, Yang D-C, Meng Y-Q, Jin J, and Gao G. PlantRegMap: charting functional regulatory maps in plants. Nucleic Acids Res. 2020a:**48**(D1):D1104–D1113. https://doi.org/10.1093/nar/gkz1020

Tian Y, Du J, Wu H, Guan X, Chen W, Hu Y, Fang L, Ding L, Li M, Yang D, et al. The transcription factor MML4_D12 regulates fiber development through interplay with the WD40-repeat protein WDR in cotton. J Exp Bot. 2020b:**71**(12):3499–3511. https://doi.org/10.1093/jxb/eraa104

Tian Z, Zhang Y, Zhu L, Jiang B, Wang, H, Gao R, Friml J, Xiao G. Strigolactones act downstream of gibberellins to regulate fiber cell elongation and cell wall thickness in cotton (Gossypium hirsutum). Plant Cell, 34(12): 4816–4839. https://doi.org/10.1093/plcell/koac270

Tuttle JR, Nah G, Duke MV, Alexander DC, Guan X, Song Q, Chen ZJ, Scheffler BE, and Haigler CH. Metabolomic and transcriptomic insights into how cotton fiber transitions to secondary wall synthesis, represses lignification, and prolongs elongation. BMC Genomics. 2015:**16**:477. https://doi.org/10.1186/s12864-015-1708-9

Tu X, Mejía-Guerra MK, Valdes Franco JA, Tzeng D, Chu P-Y, Shen W, Wei Y, Dai X, Li P, Buckler ES, et al. Reconstructing the maize leaf regulatory network using ChIP-seq data of 104 transcription factors. Nat Commun. 2020:**11**(1):5089. https://doi.org/10.1038/s41467-020-18832-8

Ueda Y, Ohtsuki N, Kadota K, Tezuka A, Nagano AJ, Kadowaki T, Kim Y, Miyao M, and Yanagisawa S. Gene regulatory network and its constituent transcription factors that control nitrogen-deficiency responses in rice. New Phytol. 2020:**227**(5):1434–1452. https://doi.org/10.1111/nph.16627

Vandepoele K and Kaufmann K. Characterization of Gene Regulatory Networks in Plants Using New Methods and Data Types. Methods Mol Biol. 2023:**2698**:1–11.

https://doi.org/10.1007/978-1-0716-3354-0_1

Vaquerizas JM, Kummerfeld SK, Teichmann SA, and Luscombe NM. A census of human transcription factors: function, expression and evolution. Nat Rev Genet. 2009:**10**(4):252–263. https://doi.org/10.1038/nrg2538

Viot CR and Wendel JF. Evolution of the Cotton Genus, Gossypium, and Its Domestication in the Americas. CRC Crit Rev Plant Sci. 2023:**42**(1):1–33. https://doi.org/10.1080/07352689.2022.2156061

Walford S-A, Wu Y, Llewellyn DJ, and Dennis ES. GhMYB25-like: a key factor in early cotton fibre development. Plant J. 2011:**65**(5):785–797. https://doi.org/10.1111/j.1365-313X.2010.04464.x

Walford S-A, Wu Y, Llewellyn DJ, and Dennis ES. Epidermal cell differentiation in cotton mediated by the homeodomain leucine zipper gene, GhHD-1. Plant J. 2012:**71**(3):464–478. https://doi.org/10.1111/j.1365-313X.2012.05003.x

Walley JW, Sartor RC, Shen Z, Schmitz RJ, Wu KJ, Urich MA, Nery JR, Smith LG, Schnable JC, Ecker JR, et al. Integration of omic networks in a developmental atlas of maize. Science. 2016:**353**(6301):814–818. https://doi.org/10.1126/science.aag1125

Wang M, Tu L, Lin M, Lin Z, Wang P, Yang Q, Ye Z, Shen C, Li J, Zhang L, et al. Asymmetric subgenome selection and cis-regulatory divergence during cotton domestication. Nat Genet. 2017:**49**(4):579–587. https://doi.org/10.1038/ng.3807

Wang M, Wang P, Lin M, Ye Z, Li G, Tu L, Shen C, Li J, Yang Q, and Zhang X. Evolutionary dynamics of 3D genome architecture following polyploidization in cotton. Nat Plants. 2018:**4**(2):90–97. https://doi.org/10.1038/s41477-017-0096-3

Wang M-Y, Zhao P-M, Cheng H-Q, Han L-B, Wu X-M, Gao P, Wang H-Y, Yang C-L, Zhong N-Q, Zuo J-R, et al. The cotton transcription factor TCP14 functions in auxin-mediated epidermal cell differentiation and elongation. Plant Physiol. 2013:**162**(3):1669–1680. https://doi.org/10.1104/pp.113.215673

Wang N, Ma Q, Wu M, Pei W, Song J, Jia B, Liu G, Sun H, Zang X, Yu S, et al. Genetic variation in MYB5_A12 is associated with fibre initiation and elongation in tetraploid cotton. Plant Biotechnol J. 2021a:**19**(10):1892–1894. https://doi.org/10.1111/pbi.13662

Wang N-N, Li Y, Chen Y-H, Lu R, Zhou L, Wang Y, Zheng Y, and Li X-B. Phosphorylation of WRKY16 by MPK3-1 is essential for its transcriptional activity during fiber initiation and elongation in cotton (Gossypium hirsutum). Plant Cell. 2021b:**33**(8):2736–2752. https://doi.org/10.1093/plcell/koab153

Wang Y, Li Y, Gong S-Y, Qin L-X, Nie X-Y, Liu D, Zheng Y, and Li X-B. GhKNL1 controls fiber elongation and secondary cell wall synthesis by repressing its downstream genes in cotton (Gossypium hirsutum). J Integr Plant Biol. 2022:**64**(1):39–55. https://doi.org/10.1111/jipb.13192

Wang Z, Yang Z, and Li F. Updates on molecular mechanisms in the development of branched trichome in Arabidopsis and nonbranched in cotton. Plant Biotechnol J. 2019:**17**(9):1706–1722. https://doi.org/10.1111/pbi.13167

Wei X, Li J, Wang S, Zhao Y, Duan H, and Ge X. Fiber-specific overexpression of GhACO1 driven by E6 promoter improves cotton fiber quality and yield. Ind Crops Prod. 2022:**185**:115134. https://doi.org/10.1016/j.indcrop.2022.115134

Wen X, Chen Z, Yang Z, Wang M, Jin S, Wang G, Zhang L, Wang L, Li J, Saeed S, et al. A comprehensive overview of cotton genomics, biotechnology and molecular biological studies. Sci China Life Sci. 2023. https://doi.org/10.1007/s11427-022-2278-0

Wen X, Zhai Y, Zhang L, Chen Y, Zhu Z, Chen G, Wang K, and Zhu Y. Molecular studies of cellulose synthase supercomplex from cotton fiber reveal its unique biochemical properties. Sci China Life Sci. 2022:**65**(9):1776–1793. https://doi.org/10.1007/s11427-022-2083-9

Wu A, Lian B, Hao P, Fu X, Zhang M, Lu J, Ma L, Yu S, Wei H, and Wang H. GhMYB30-GhMUR3 affects fiber elongation and secondary wall thickening in cotton. Plant J. 2023. https://doi.org/10.1111/tpj.16523

Wu T, Goh H, Azodi C, Krishnamoorthi S, Liu M, Urano D. Evolutionarily conserved hierarchical gene regulatory networks for plant salt stress response. Nat Plants. 2021.7(6):787-799. https://doi.org/10.1038/s41477-021-00929-7

Xiao Y-H, Li D-M, Yin M-H, Li X-B, Zhang M, Wang Y-J, Dong J, Zhao J, Luo M, Luo X-Y, et al. Gibberellin 20-oxidase promotes initiation and elongation of cotton fibers by regulating gibberellin synthesis. J Plant Physiol. 2010:**167**(10):829–837. https://doi.org/10.1016/j.jplph.2010.01.003

Xie Y, Liu Y, Wang H, Ma X, Wang B, Wu G, and Wang H. Phytochrome-interacting factors directly suppress MIR156 expression to enhance shade-avoidance syndrome in Arabidopsis. Nat Commun. 2017:**8**(1):348. https://doi.org/10.1038/s41467-017-00404-y

Yadav VK, Yadav VK, Pant P, Singh SP, Maurya R, Sable A, and Sawant SV. GhMYB1 regulates SCW stage-specific expression of the*GhGDSL*promoter in the fibres of*Gossypium hirsutum*L. Plant Biotechnol J. 2017:**15**(9):1163–1174. https://doi.org/10.1111/pbi.12706

Yang Z, Liu Z, Ge X, Lu L, Qin W, Qanmber G, Liu L, Wang Z, and Li F. Brassinosteroids regulate cotton fiber elongation by modulating very-long-chain fatty acid biosynthesis. Plant Cell. 2023:**35**(6):2114–2131. https://doi.org/10.1093/plcell/koad060

Yang Z, Zhang C, Yang X, Liu K, Wu Z, Zhang X, Zheng W, Xun Q, Liu C, Lu L, et al. PAG1, a cotton brassinosteroid catabolism gene, modulates fiber elongation. New Phytol. 2014:**203**(2):437–448. https://doi.org/10.1111/nph.12824

Yoo MJ and Wendel JF. Comparative evolutionary and developmental dynamics of the cotton (Gossypium hirsutum) fiber transcriptome. PLoS Genet. 2014:**10**(1):e1004073. https://doi.org/10.1371/journal.pgen.1004073

You J, Liu Z, Qi Z, Ma Y, Sun M, Su L, Niu H, Peng Y, Luo X, Zhu M, et al. Regulatory controls of duplicated gene expression during fiber development in allotetraploid cotton. Nat Genet. 2023:**55**(11):1987–1997. https://doi.org/10.1038/s41588-023-01530-8

Yu G, Wang L-G, Han Y, and He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012:**16**(5):284–287.

https://doi.org/10.1089/omi.2011.0118

Yu G, Wang L-G, and He Q-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics. 2015:**31**(14):2382–2383. https://doi.org/10.1093/bioinformatics/btv145

Zeng J, Zhang M, Hou L, Bai W, Yan X, Hou N, Wang H, Huang J, Zhao J, and Pei Y. Cytokinin inhibits cotton fiber initiation by disrupting PIN3a-mediated asymmetric accumulation of auxin in the ovule epidermis. J Exp Bot. 2019:**70**(12):3139–3151. https://doi.org/10.1093/jxb/erz162

Zhang B, Chopra D, Schrader A, and Hülskamp M. Evolutionary comparison of competitive protein-complex formation of MYB, bHLH, and WDR proteins in plants. J Exp Bot. 2019:**70**(12):3197–3209. https://doi.org/10.1093/jxb/erz155

Zhang D, Chen C, Wang H, Niu E, Zhao P, Fang S, Zhu G, Shang X, and Guo W. Cotton Fiber Development Requires the Pentatricopeptide Repeat Protein GhIm for Splicing of Mitochondrial nad7 mRNA. Genetics. 2021a:**217**(1):1–17. https://doi.org/10.1093/genetics/iyaa017

Zhang J, Huang G-Q, Zou D, Yan J-Q, Li Y, Hu S, and Li X-B. The cotton (Gossypium hirsutum) NAC transcription factor (FSN1) as a positive regulator participates in controlling secondary cell wall biosynthesis and modification of fibers. New Phytol. 2018:**217**(2):625–640. https://doi.org/10.1111/nph.14864

Zhang M, Zeng J-Y, Long H, Xiao Y-H, Yan X-Y, and Pei Y. Auxin Regulates Cotton Fiber Initiation via GhPIN-Mediated Auxin Transport. Plant Cell Physiol. 2017a:**58**(2):385–397. https://doi.org/10.1093/pcp/pcw203

Zhang T, Hu Y, Jiang W, Fang L, Guan X, Chen J, Zhang J, Saski CA, Scheffler BE, Stelly DM, et al. Sequencing of allotetraploid cotton (Gossypium hirsutum L. acc. TM-1) provides a resource for fiber improvement. Nat Biotechnol. 2015:**33**(5):531–537. https://doi.org/10.1038/nbt.3207

Zhang X, Cao J, Huang C, Zheng Z, Liu X, Shangguan X, Wang L, Zhang Y, and Chen Z. Characterization of cotton ARF factors and the role of GhARF2b in fiber development. BMC Genomics. 2021b:**22**(1):202. https://doi.org/10.1186/s12864-021-07504-6

Zhang X, Xue Y, Guan Z, Zhou C, Nie Y, Men S, Wang Q, Shen C, Zhang D, Jin S, et al. Structural insights into homotrimeric assembly of cellulose synthase CesA7 from Gossypium hirsutum. Plant Biotechnol J. 2021c:**19**(8):1579–1587. https://doi.org/10.1111/pbi.13571

Zhang Y, He P, Yang Z, Huang G, Wang L, Pang C, Xiao H, Zhao P, Yu J, and Xiao G. A Genome-Scale Analysis of the PIN Gene Family Reveals Its Functions in Cotton Fiber Development. Front Plant Sci. 2017b:**8**:461. https://doi.org/10.3389/fpls.2017.00461

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008:**9**(9):R137. https://doi.org/10.1186/gb-2008-9-9-r137

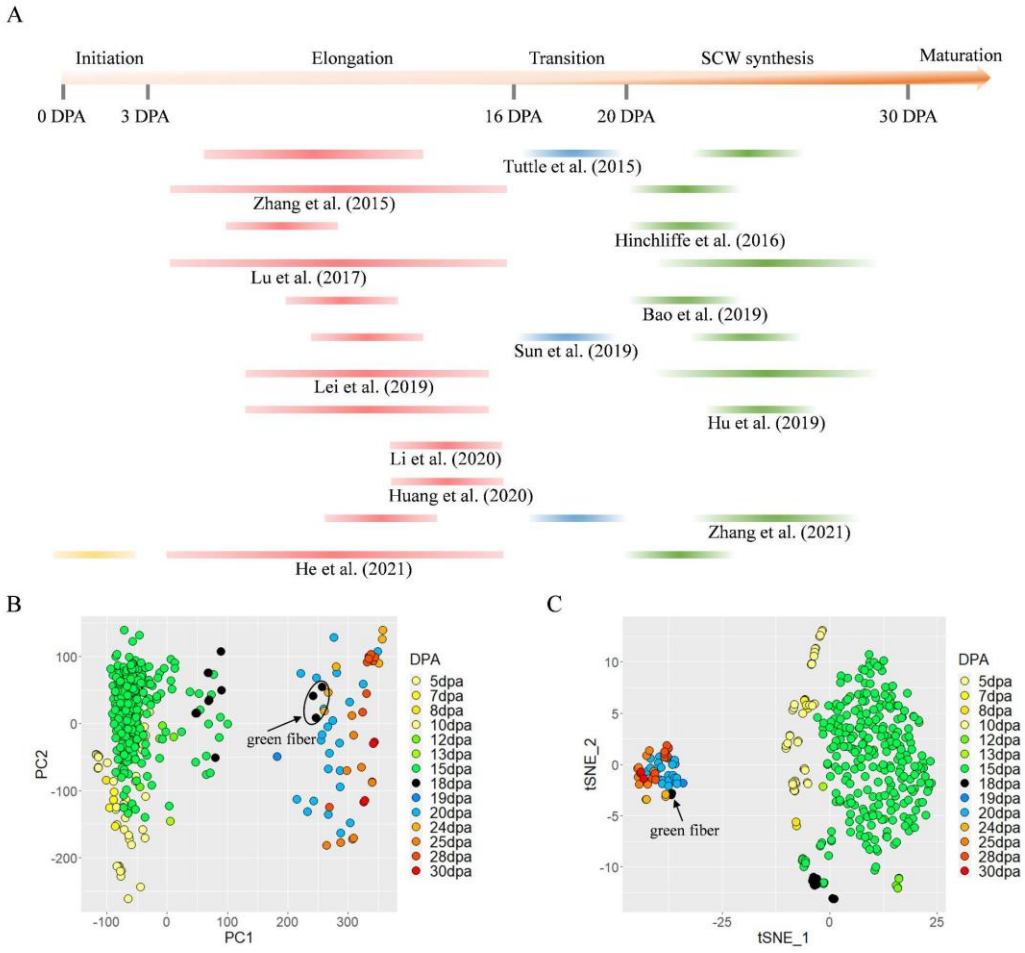
Zhao H, Zhang W, Zhang T, Lin Y, Hu Y, Fang C, and Jiang J. Genome-wide MNase

hypersensitivity assay unveils distinct classes of open chromatin associated with H3K27me3 and DNA methylation in Arabidopsis thaliana. Genome Biology. 2020:**21**(1). https://doi.org/10.1186/s13059-020-1927-5

Zhong R, Lee C, Haghighat M, and Ye Z-H. Xylem vessel-specific SND5 and its homologs regulate secondary wall biosynthesis through activating secondary wall NAC binding elements. New Phytol. 2021:**231**(4):1496–1509. https://doi.org/10.1111/nph.17425

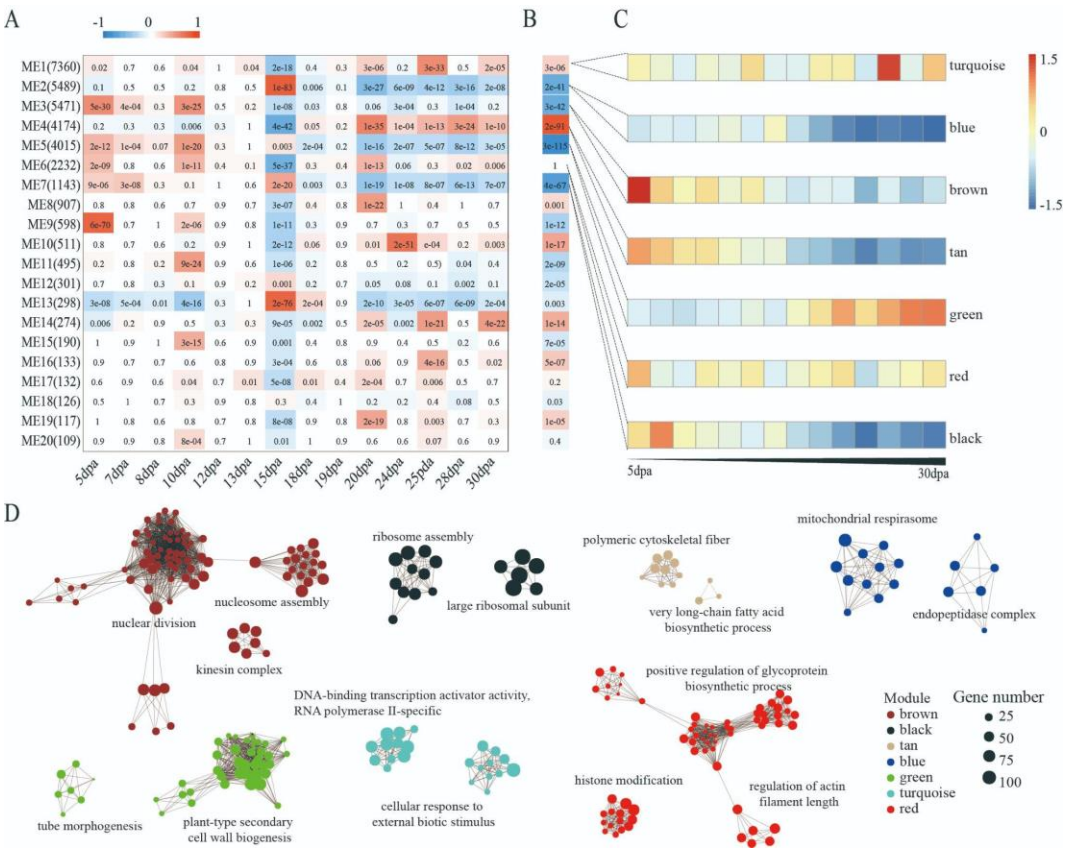
Zhou P, Li Z, Magnusson E, Gomez Cano F, Crisp PA, Noshay JM, Grotewold E, Hirsch CN, Briggs SP, and Springer NM. Meta Gene Regulatory Networks in Maize Highlight Functionally Relevant Regulatory Interactions. Plant Cell. 2020:**32**(5):1377–1396. https://doi.org/10.1105/tpc.20.00080

Zhou Y, Zhang Z-T, Li M, Wei X-Z, Li X-J, Li B-Y, and Li X-B. Cotton (Gossypium hirsutum) 14-3-3 proteins participate in regulation of fibre initiation and elongation by modulating brassinosteroid signalling. Plant Biotechnol J. 2015:**13**(2):269–280. https://doi.org/10.1111/pbi.12275



1701

1702 **Figure 1. Cotton fiber transcriptomic datasets for this study.** (**A**) Timeline displaying the
1703 stages represented by the 12 studies used to generate a dataset of 401 fiber RNA-seq samples for
1704 an in-depth exploration of cotton fiber development. Fiber elongation, transition, and SCW
1705 synthesis stages are indicated by red, blue, and green bars, respectively, and each line represents
1706 one existing dataset. This color scheme is applied consistently across all figures here. (**B**)
1707 Principal component analysis (PCA) of 57,151 gene expression profiles. PC1 and PC2 captured
1708 16.8% and 11.5% of variance, respectively. (**C**) T-distributed stochastic neighbor embedding (t-
1709 SNE) was also employed for dimension reduction and visualization of the fiber expression
1710 landscape.



1711

1712 **Figure 2. Phenotypic and functional associations of co-expression gene modules during**
1713 **fiber development. (A)** For the 20 co-expression gene modules identified by weighted gene co-
1714 expression network analysis (WGCNA), heatmap represents Pearson correlation coefficients and
1715 *P*-values (cell color and text, respectively) between the module eigengenes (MEs, by row) and
1716 fiber developmental stages treated as the binary categorical variable (by column). (**B**) ANOVA
1717 of MEs (by column) by fiber developmental stages treated as a numeric variable (MEs, by row).
1718 Heatmap cell color and text represent Pearson correlation coefficients and *P*-values, respectively.
1719 **(C)** Heatmap of z-score normalized MEs for the seven largest modules across fourteen fiber
1720 developmental time points. (**D**) Gene Ontology (GO) enrichment analysis of the seven largest
1721 modules, displaying the top two most significant interconnected GO clusters terms each.
1722 Different colors represent corresponding modules.

1723

1724

1725

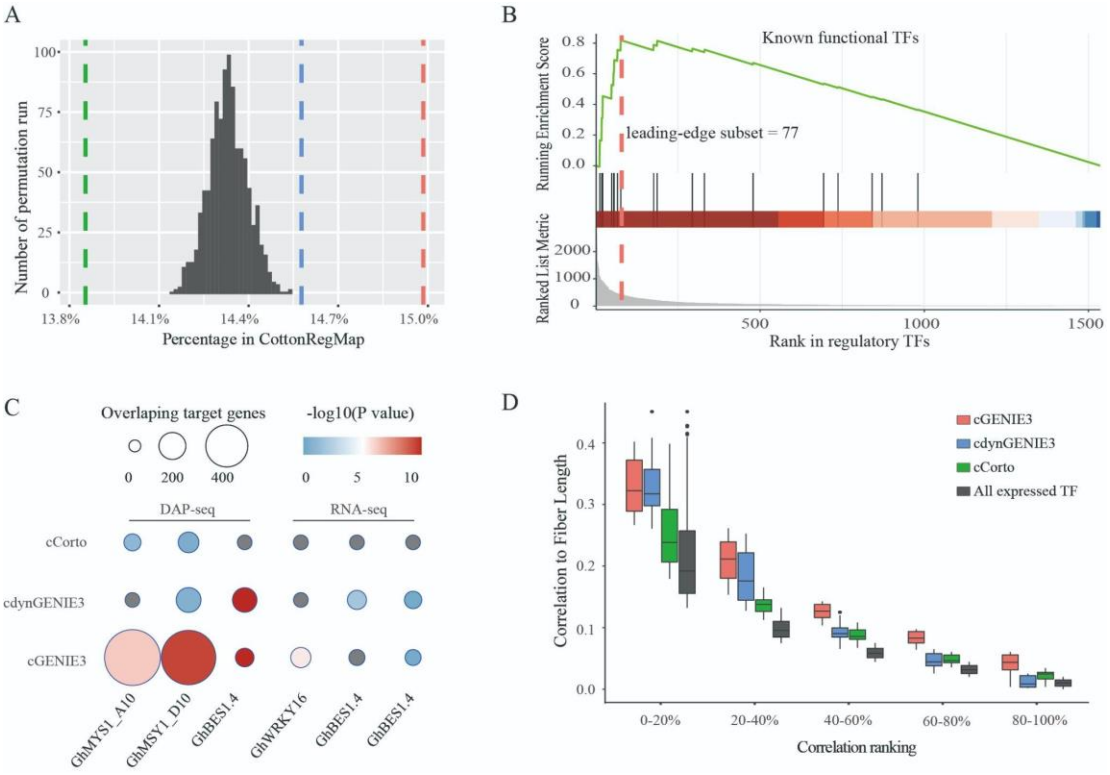


Figure 3. Evaluation of fiber GRN inferences. (**A**) Histogram presents the bootstrap distribution (n=1000) of cottonRegMap TF-target relationships as captured by chance. Red, blue, and green lines represent the cottonRegMap TF-target relationships inferred by GENIE3, dynGENIE3, and Corto, respectively. Both GENIE3 and Corto inferred significantly more interactions outside the bootstrap distribution. (**B**) GSEA of known functional TFs among TFs rankings inferred by cGENEI3. The enrichment score reflects the degree of over-representation of a set of 54 known functional TFs at the top of the ranked TFs identified by cGENIE3. The red dashed line indicates that these known functional TFs were significantly enriched at the top 77 ranking TFs. (**C**) Heatmap of overlapping target genes between empirical evidence (columns) and GRN inferences (rows). WRKY16, with GRN inferences for cGENIE3, cdynGENIE3, and cCorto. Each cell represents the number of overlaps and the significance of the corresponding hypergeometric test. DAP-seq results of *GhMYS1_A10*, *GhMYS1_D10*, and *GhDES1.4* and RNA-seq results of *GhDES1.4* and *GhWRKY16* were shown. (D) The correlation between expression variation of 77 hub TFs and fiber length was significantly higher than that of 3,638 TFs expressed in fibers. Five different percentages ranks were divided according to the correlation between TF and fiber length, where 0% to 100% represent increasing correlation.

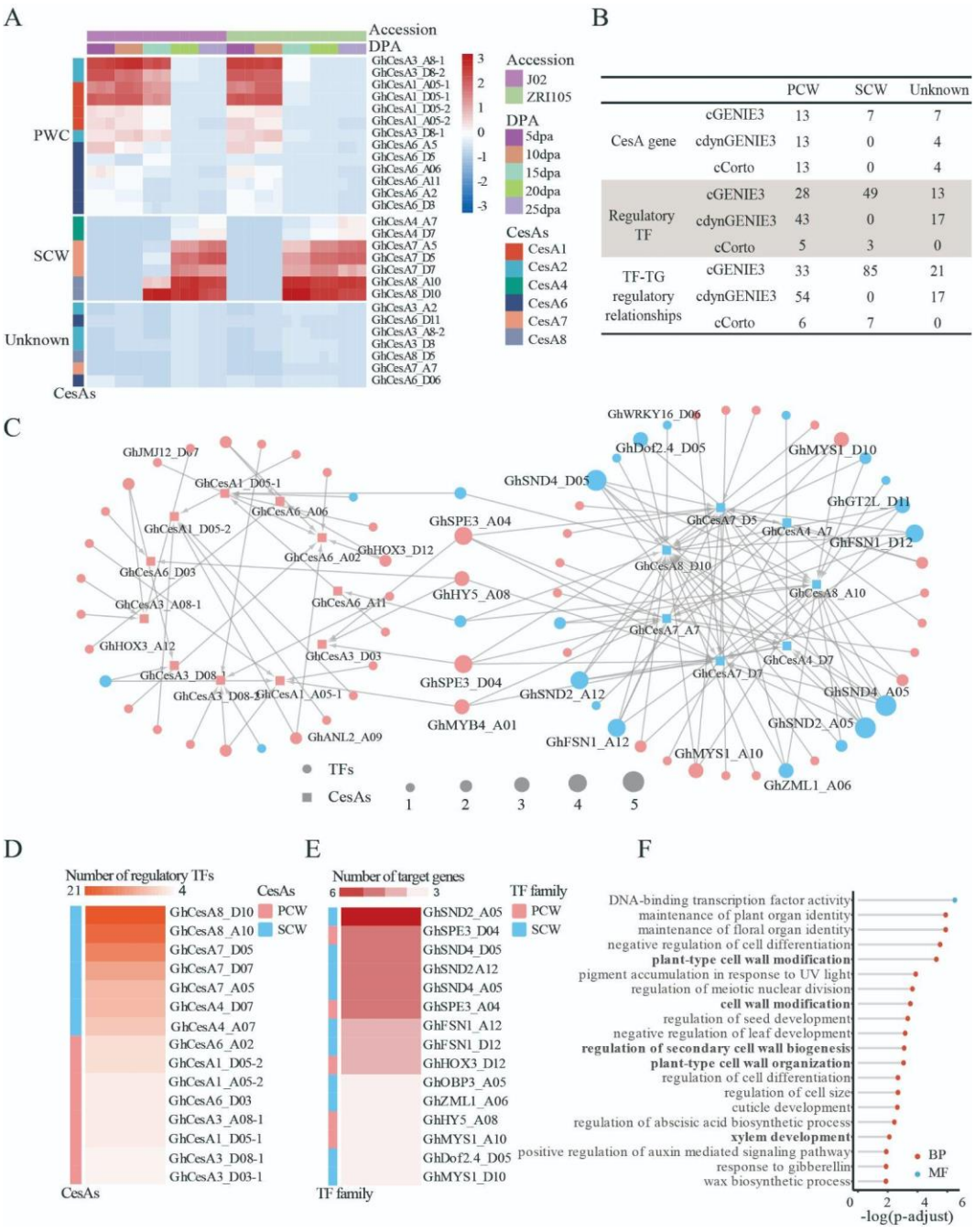


Figure 4. GRN performance in cotton cellulose synthesis. (**A**) Categorization of *GhCesAs* based on gene expression patterns during cotton fiber development. Heatmap presents TPM expression levels in the long-fiber variety J02 and the short-fiber cotton variety ZRI015. Three hierarchical clusters correspond to PCW-related, SCW-related, and unknown *GhCesAs.* (**B**) The number of CesA genes, regulator transcription factors (TFs), and regulatory relationships identified by cGENIE3, cdynGENIE3, and cCorto. (**C**) Cellulose synthesis-related subnetwork

inferred by cGENIE3. Square and round nodes represent *GhCesAs* and TFs, respectively, which are connected by directed edges indicating the TF-target relationships inferred. Red and blue node colors represent the categorization of PCW-related and SCW-related genes based on expression patterns during fiber development, respectively. Two network components were detected corresponding to PCW (left) and SCW (right), which were co-regulated by six TFs in the middle. (**D**) Ranking *GhCesAs* by in-degree (i.e., number of incoming linking) from all TFs inferred by cGENIE3. (**E**) Ranking cellulose synthesis related TFs by out-degree (i.e., number of outward links) to target *GhCesAs*. (**F**) Enriched GO terms associated with the 71 TFs inferred by cGENIE3.

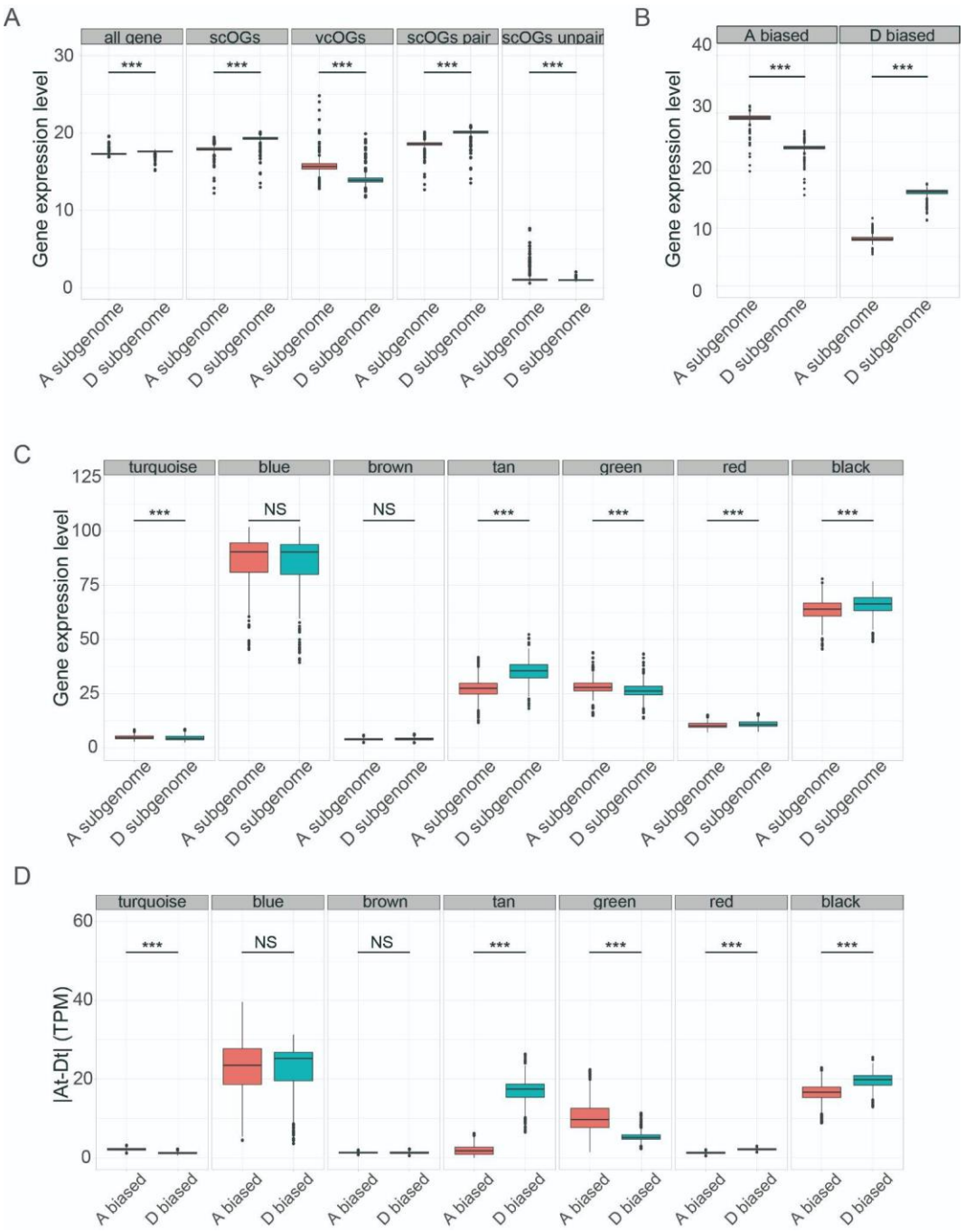


Figure 5. Expression level analysis of homoeologous gene pairs. (**A**) Gene expression levels compared between At and Dt homoeologs for all 57,151 fiber-expressed genes ("all genes"), 22,889 homoeologous pairs characterized into single-copy ortho-homoeolog groups ("scOGs"), the remaining 13,229 At and 15,895 Dt genes uncategorized ("vcOGs"), 19,213 scOGs with both At and Dt expressed in fiber ("scOGs pair"), and 17,028 scOGs with only one homoeolog

expressed in fibers ("scOGs unpair"). (**B**) Gene expression levels compared for scOGs pairs exhibiting homoeolog expression bias (HEB). (**C**) Absolute expression differences compared between A-biased and D-biased scOGs. (**C**) Expression comparisons for scOGs present within the same co-expression modules identified by WGCNA. (**D**) Absolute expression differences compared between A-biased and D-biased scOGs in co-expression modules. Statistical significance was determined using a two-sided Wilcoxon rank-sum test. ***$P< 0.001$.

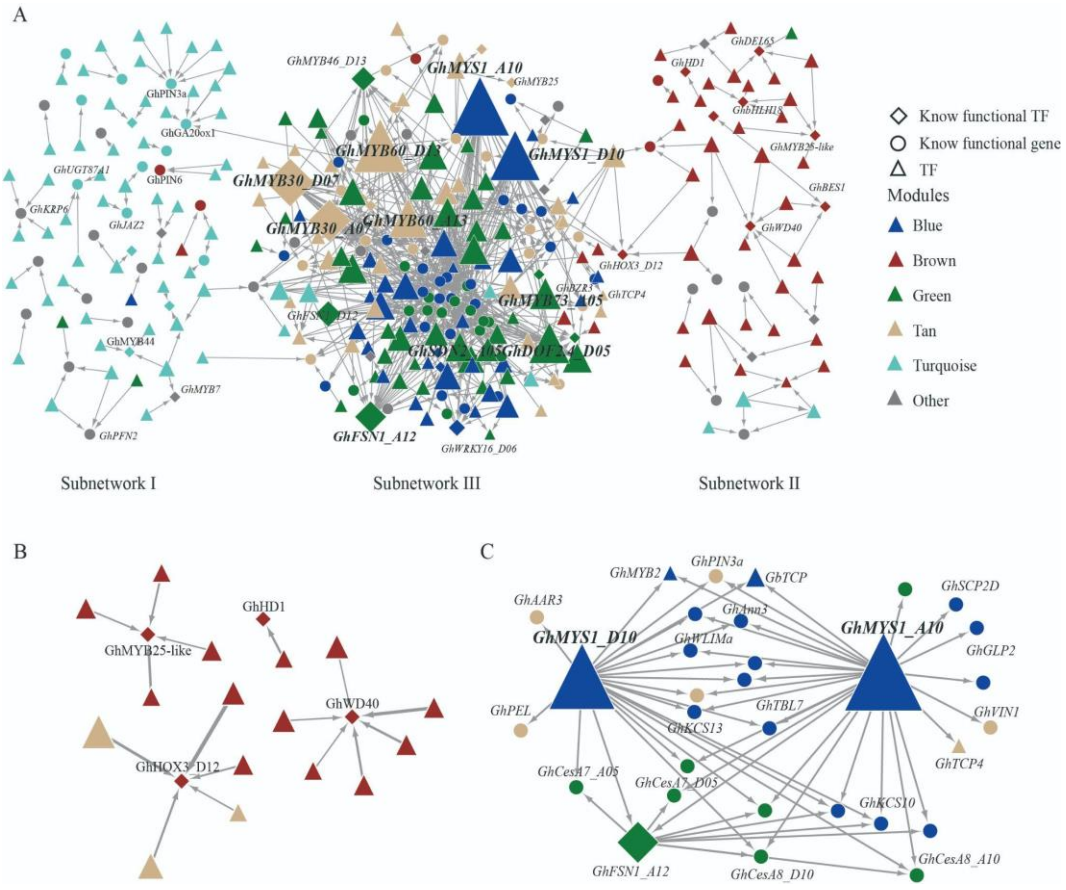
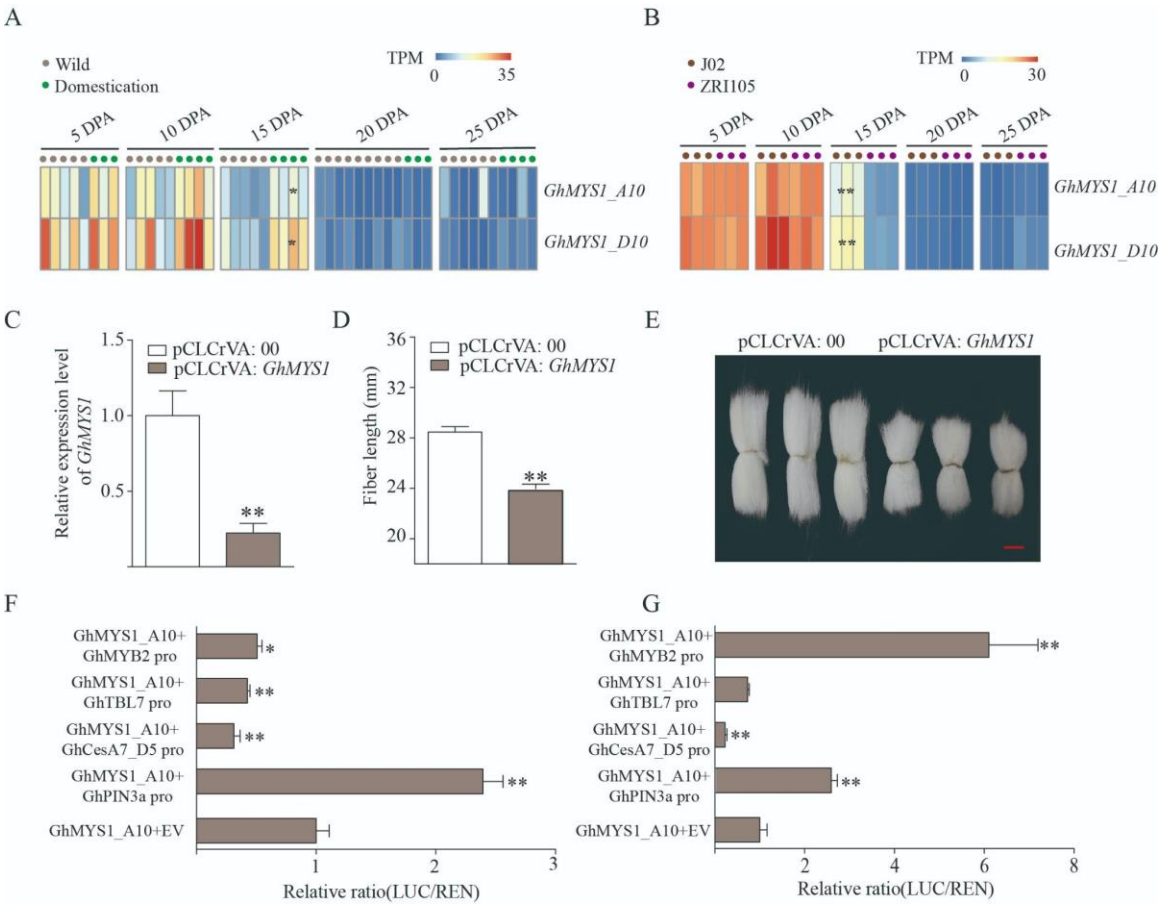


Figure 6. GRN built based on known function genes and their directly regulated TF in fiber. (**A**) GRN of known functional genes and their regulated TFs. Known functional genes and TFs are shown as circles and rhombus, respectively. Different colors indicate the modules where genes and TFs are located in the co-expression network. (**B**) Novel TFs in brown module regulate *GhHOX3*, *GhHD1*, *GhMYB25-like*, and *GhWD40* involved in fiber initiation. (**C**) Network of known functional genes regulated by *GhMYS1_A10* and *GhMYS1_D10*.



1780

1781 **Figure 7 *GhMYS1* positively regulates fiber elongation.** (**A**) Expression pattern analysis of
1782 *GhMYS1_A10* and *GhMYS1_D10* in wild and domestication cotton accession from 5 to 25 days
1783 post-anthesis (dpa). (**B**) Expression pattern analysis of *GhMYS1_A10* and *GhMYS1_D10* in long-
1784 fiber (J02) and short-fiber (ZRI105) varieties from 5 to 25 dpa. (**C**) Relative expression levels
1785 measured by qRT-PCR showed reduced *GhMYS1* expression in 10 dpa fibers from pCLCrVA:
1786 *GhMYS1* cotton plants relative to pCLCrVA: 00 plants. (**D**) Significantly shorter mature fiber
1787 length in pCLCrVA: *GhMYS1* versus pCLCrVA: 00 plants. (**E**) Phenotype of mature fibers in
1788 pCLCrVA: 00 and pCLCrVA: *GhMYS1* plants. bar = 1 cm. (**F-G**) Transient dual-luciferase
1789 (LUC) reporter assay testing interactions between *GhMYS1_A10* (**F**) and *GhMYS1_D10* (**G**), and
1790 the promoters of *GhPIN3a*, *GhCesA7_D05*, *GhTBL7*, and *GhMYB2*. Expression of Renilla
1791 luciferase (REN) was used as an internal control. Values given are mean ± SD (n = 4). Relative
1792 LUC activity obtained with the empty plasmid (none) was set to 1. Statistically significant
1793 differences between groups as determined by Student's t-test. **P*< 0.05 and ***P*< 0.01.

1794 **Table 1. Fiber gene regulatory networks constructed.**

GRN		CottonRegNet	GENIE3	Corto	dynGENIE3	cGENIE3	cCorto	cdynGENIE3
nodes		74,902	54,237	56,052	25,441	41,757	25,245	25,076
edges		53,878,120	1,000,000	232,943	1,000,000	149,849	33,956	138,639
TF in network	At	2,485	1,796	1,723	546	1,577	729	542
	Dt	2,550	1,829	1,725	527	1,544	723	526
TF as regulatory nodes	At	1,088	1,780	1,723	546	773	729	248
	Dt	1,087	1,805	1,725	527	763	723	242
Target genes	At	36,118	26,633	25,809	12,633	20.462	11,646	12,440
	Dt	38,784	27,486	26,795	12,808	21,025	12,147	12,603
known functional genes		191	181	185	117	155	91	114
network density		0.0096035	0.000340	0.000074	0.001545	0.000086	0.000053	0.000221
clustering coefficient		0.0449141	0.031020	0.000000	0.012217	0.007420	0.000000	0.002181

1795
1796
1797
1798
1799
1800
1801
1802

1803 **Table 2. Subgenomic contribution to fiber-expressed genes.**

	Total	**At**	**Dt**	**[1]Subgenome contribution**
I. All genes in the reference genome	**74,902**	**36,118**	**38,784**	**At < Dt**
scOG genes	45,778	22,889	22,889	-
scOG genes	29,124	13,229	15,895	At < Dt
II. Fiber expressed genes	**57,151**	**28,004**	**29,147**	
(% of all genes)	**(76.3%)**	**(77.5%)**	**(75.2%)**	**(At > Dt)**
scOG genes: paired	38,426 (83.9%)	19,213 (83.9%)	19,213 (83.9%)	-
scOG genes: unpaired	1,597 (3.5%)	749 (3.3%)	848 (3.7%)	At < Dt
scOG genes	17,128 (58.5%)	8,042 (60.8%)	9,086 (57.2%)	At < Dt (At > Dt)
III. Genes assigned to co-expression modules	**25,751**	**12,816**	**12,935**	**(At > Dt)**
(% of all genes)	**(34.4%)**	**(35.5%)**	**(33.4%)**	
homoeologs in the same module	12,560; 48.8%	6,280	6,280	-
homoeologs NOT in the same module	13,191; 51.2%	6,536	6,655	-
homoeologs TF in the same module	1042; 53.0%	521	521	-
homoeologs TF NOT in the same module	924; 47.0%	448	476	-
know functional gene in the same module	52	26	26	
know functional gene NOT in the same module	66	33	33	
IV. Nodes in cGENIE3 network	**41,757**	**20,578**	**21,179**	(At > Dt)
(% of all genes)	**(55.7%)**	**(57.0%)**	**(54.6%)**	
regulators (TFs)	1,536	773	763	-
TFs in scOG pair; % scOG pairs	1,146; 88.1%	573	573	-

TFs NOT in scOG pair; % scOG pairs	155; 11.9%	81	74	-
scOG TF	235	119	116	
target genes (TGs):	41,514	20,462 (56.7%)	21,052 (54.3%)	(At > Dt)
TGs in scOG pair; % scOG pairs	23,858; 77.9%	11,929	11,929	-
TGs NOT in scOG pair; % scOG pairs	6,759; 22.1%	3,291	3,468	-
scOG genes; % of all scOG gene	10,897	5,242; 39.6%	5,655; 35.4%	(At > Dt)
V. Edges in cGENIE3 network	**149,849**			
intra-subgenome (average TG number per TF)	74,701	At to At: 38,704 (50.0)	Dt to Dt: 35,997 (47.2)	At > Dt
inter-subgenome (average TG number per TF)	75,148	At to Dt: 40,116 (51.9)	Dt to At: 35,032 (45.9)	At to Dt > Dt to At
[2] TF: regulatory conservation	15.8%	15.7%	15.8%	-
[2] TG: regulatory conservation	6.4%	6.3%	6.5%	-

[1] Significantly different contribution between subgenomes was shown when Chi-square test $P < 0.05$.

[2] For given TFs (e.g. At TFs), regulatory conservation measures the percentage of their edges targeting paired At and Dt TGs among all edges. For given TGs, regulatory conservation measures the percentage of their edges regulated by paired At and Dt TFs. Full conservation is represented by 1, while no conservation is represented by 0.