Co-expression network and comparative transcriptome analysis for fiber initiation and elongation reveal genetic differences in two lines from upland cotton CCRI70 RIL population

Xiao Jiang[1], Liqiang Fan[1,2], Pengtao Li[3], Xianyan Zou[1], Zhen Zhang[1], Senmiao Fan[1], Juwu Gong[1], Youlu Yuan[1,2] and Haihong Shang[1,2]

[1] State Key Laboratory of Cotton Biology, Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang, Henan, China
[2] School of Agricultural Sciences, Zhengzhou University, Zhengzhou, Henan, China
[3] School of Biotechnology and Food Engineering, Anyang Institute of Technology, Anyang, Henan, China

ABSTRACT

Upland cotton is the most widely planted for natural fiber around the world, and either lint percentage (LP) or fiber length (FL) is the crucial component tremendously affecting cotton yield and fiber quality, respectively. In this study, two lines MBZ70-053 and MBZ70-236 derived from *G. hirsutum* CCRI70 recombinant inbred line (RIL) population presenting different phenotypes in LP and FL traits were chosen to conduct RNA sequencing on ovule and fiber samples, aiming at exploring the differences of molecular and genetic mechanisms during cotton fiber initiation and elongation stages. As a result, 249/128, 369/206, 4296/1198 and 3547/2129 up-/down- regulated differentially expressed genes (DGEs) in L2 were obtained at -3, 0, 5 and 10 days post-anthesis (DPA), respectively. Seven gene expression profiles were discriminated using Short Time-series Expression Miner (STEM) analysis; seven modules and hub genes were identified using weighted gene co-expression network analysis. The DEGs were mainly enriched into energetic metabolism and accumulating as well as auxin signaling pathway in initiation and elongation stages, respectively. Meanwhile, 29 hub genes were identified as *14-3-3ω*, *TBL35*, *GhACS*, *PME3*, *GAMMA-TIP*, *PUM-7*, etc., where the DEGs and hub genes revealed the genetic and molecular mechanisms and differences during cotton fiber development.

INTRODUCTION

Cotton is one of the most important cash crops around the world, providing the main natural fiber for the textile industry. Due to its adaptability and yield, upland cotton has been the most widely cultivated *Gossypium* species, which could contribute to almost 95% production of all planted cotton in spite of presenting the relatively ordinary fiber quality (*Yoo & Wendel, 2014*). However, either fiber yield or quality traits of cotton are sensitive

to the environment, belonging to quantitative traits controlled by multi-genes, therefore breeders focus on developing new upland cotton varieties simultaneously performing superior fiber quality and high yield. It would be not only significantly beneficial for upland cotton breeding, but also for the global textile industry (*Kim & Triplett, 2001*).

On the basis of previous researches on cotton, both fiber yield and quality performances are collectively determined by the four developmental stages, which could be classified into: initiation (−3 to 3 DPA), elongation (3 to 23 DPA), secondary wall biosynthesis (20 to 40DPA) and maturity (40 to 50DPA) (*Basra & Malik, 1984*; *Kim & Triplett, 2001*; *Lee et al., 2006*; *Lee, Woodward & Chen, 2007*; *Haigler et al., 2012*). The developmental stages of fiber initiation, elongation and secondary wall biosynthesis prevailingly affect fiber number, length and strength, respectively (*Basra & Malik, 1984*; *Patel et al., 2020*). During fiber initiation, trichome protrusion and enlargement of epidermal cells (*Qin & Zhu, 2011*), and fiber initiation had great impact on the number of lint fibers because the later initials always develop as fuzz fibers (*Lee et al., 2006*). In fiber elongation, fiber cells started primary cell wall biosynthesis alongside pectin biosynthesis genes expressed, which promoted fiber elongation by ethylene signaling pathways (*Pang et al., 2010*). Therefore, to explore the upland cotton agronomy traits of lint percentage (LP) and fiber length (FL), we focused on initiation and elongation stages.

Along with the rapid development of sequencing technology, reference genomes of diploid and allotetraploid *Gossypium* species have been successfully sequenced, constructed and published, which provide a solid foundation for researching on the genetic mechanisms at the genome level (*Paterson et al., 2012*; *Li et al., 2014*; *Li et al., 2015*; *Zhang et al., 2015*; *Hu et al., 2019*; *Wang et al., 2019*; *Yang et al., 2019*; *Huang et al., 2020*). Transcriptome sequencing, known as RNA-seq, provides a suitable procedure to analyze individual gene transcription and the entire transcriptome profile during various stages of fiber development. Concentrating on analysis of differential expressed genes, comparative transcriptome is an efficient tool to scan candidate genes between different samples. In the past few years, numerous studies have used comparative transcriptome analysis on cotton fiber development (*Applequist, Cronn & Wendel, 2001*; *Gilbert et al., 2014*; *Yoo & Wendel, 2014*; *Islam et al., 2016*; *Li et al., 2017a*; *Li et al., 2017b*; *Lu et al., 2017*; *Zou et al., 2018*). However, there were few studies concentrating on fiber initiation and elongation stages or using extreme materials in breeding population.

To explore the parental source of potential alleles, CCRI70 RIL population was developed. In this study, the two lines MBZ70-053 (L1, high-FL) and MBZ70-236 (L2, high-LP) derived from upland cotton RIL population, presenting excellent performances either in cotton yield or in fiber quality trait, were applied to comparative transcriptome analysis using RNA sequencing aimed at revealing the differences on a transcription level between the two lines during fiber initiation and elongation. Through DEG and WGNCA analyses, two *GAMMA-TIP, GhAcs6, Sus4, PME3* and other key candidate genes were identified, which might have great influence on cotton fiber initiation and elongation. All those provided insights and evidences for understanding molecular mechanism of cotton fiber development and differences leading to the negative correlation between quality and yield traits on transcription level that would be beneficial for upland cotton breeding.

MATERIALS AND METHODS

Plant materials

Upland cotton hybrid CCRI70 (F_1), the first national approved higher fiber quality hybridized upland cotton variety for utilizing heterosis between the conventional cotton varieties in China, was developed from two upland cotton cultivars sGK156 and 901-001, which performed superior yield and fiber quality, respectively. While CCRI70 showed excellent performance in fiber quality while moderate performance in lint percentage. To investigate the parental source of potential alleles and to explore the molecular and genetic mechanisms aiming at improving fiber quality and yield, we developed the CCRI70 RIL population (*Zou et al., 2018*; *Deng et al., 2019*). The CCRI70 RIL population was planted in 2015 and 2016 (*Zou et al., 2018*). Phenotypic data in five environments were used in this study, including Anyang of Henan Province in 2015 and 2016 (15AY and 16AY), Linqing of Shandong Province in 2015 and 2016 (15LQ and 16LQ) and Changde of Hunan Province in 2016 (16CD).

In 2017, MBZ70-053 (L1) and MBZ70-236 (l2), designated from $F_{8:9}$ family of CCRI70 RIL population, were used as plant materials for conducting RNA-seq. The two lines were planted under standard field conditions in Anyang Experimental Station (Anyang, Henan, China) (*Zou et al., 2018*). Among the two materials, L1 showed positive extreme-parent performance in FL and negative extreme-parent in LP as well as L2 possessed positive extreme-parent performance in LP and negative extreme-parent in FL. The day of anthesis was marked as 0 DPA. According to the size of the buds and extensive field experience, the flower buds at 3 days before anthesis were recorded as −3 DPA. At −3 and 0 DPA, cotton ovules were collected from ovaries, while fiber samples were collected from bolls at 5 and 10 DPA, respectively. Both the ovule and fiber samples prepared for RNA-seq analysis were collected with three biological repeats and frozen by liquid nitrogen. For convenience, samples of L1 and L2 used in this research were recorded as L1_-3DPA, L1_0DPA, L1_5DPA, L1_10DPA, L2_-3DPA, L2_0DPA, L2_5DPA and L2_10DPA, respectively.

Phenotypic data evaluation in multiple environments

Two RILs and two parents were planted with two replications in five environments across two years and three locations. To evaluate the phenotypic date of FL and LP, thirty mature fully-opened bolls from every plot were harvested to test fiber length using an HVI1000 (Uster Technologies, Switzerland) with HVICC Calibration in the Cotton Quality Supervision, Inspection and Testing Center, Ministry of Agriculture, Anyang, China. Briefly, after the seed cotton samples were weighed and ginned, lint percentage was evaluated.

RNA isolation, cDNA library construction, Illumina deep sequencing and RNA-seq data analysis

Total RNAs of ovule and fiber samples were extracted by RNAprep Pure Plant Kit (Polysaccharides& Polyphenolics-rich, Tiangen, Beijing, China), and RNA degradation and contamination were checked by 1% agarose gel electrophoresis. The RNA concentration was confirmed using NanoDrop 2000 spectrophotometer (Thermo Scientific, Waltham,

MA, USA). RNA purity was detected using the NanoPhotometer spectrophotometer (IMPLEN, CA, USA). The RNA integrity was confirmed using the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, CA, USA). According to the manufacturer's recommendations, an amount of 2 μg RNA per sample was used for transcriptome library construction using Illumina TruSeq^{TM} RNA Sample Preparation Kit (Illumina, San Diego, CA, USA) Totally, 24 libraries were separately sequenced using Illumina Novaseq 6000 sequencing platform with 150 base pair (bp) paired-end (PE) raw reads (BerryGenomics Co., Ltd., Beijing, China).

Subsequently, Trimmomatic software was utilized to process all the generated raw data in Fastq format (*Bolger, Lohse & Usadel, 2014*). Clean data were obtained by removing reads that contained the adapter, poly-N and low-quality reads, of which the reads harbored $\geq 10\%$ unidentified nucleotides (N) and $>20\%$ bases with Phred quality <5. Meanwhile, the GC percentage and Q30 were calculated to finally evaluate the quality of clean data, which were qualified for downstream analysis. HISAT2 v2.1.0 was used to build an index of reference genome (*Pertea et al., 2016*), and the sequence alignment was conducted referring to the *G. hirsutum* genomes (*Wang et al., 2019*) with default parameters, where the reference genome was available at the website http://cotton.hzau.edu.cn/EN/download.php and the CottonGen database (https://www.cottongen.org/). Then the fragments per kilobase of exon per million reads (FPKM) values of genes were quantized by StringTie v1.3.5 (*Pertea et al., 2015*), which were subjected to Pearson correlation coefficient (PCC) for revealing the correlation coefficients between samples. As to the correlation coefficients less than 0.8 among the three biological repeats, the samples would be removed from the dataset.

Furthermore, to identify the genetic differences between the two lines, we employed Samtools v1.4 software to summarize the genotypic data (*Li et al., 2009*; *Li, 2011*) and SNPEff program to annotated the genotypic variants distribution on the reference genome (*Wang et al., 2019*) with default parameter (*Cingolani et al., 2012*).

Differentially expressed gene analysis

Based on the count number of each gene, the DESeq2 R package was employed for identifying differentially expressed genes (DEGs), of which the screening criterion were FDR value <0.05, and $\log_2$ Fold-Change value >1 or <-1 between each pairwise comparison (*Love, Huber & Anders, 2014*). The DEGs were identified through vertical and horizontal comparisons, i.e., at the same developmental stage between the two lines and in the same line between different stages.

To explore the temporal expression profiles of DEGs during the fiber development, Short Time-series Expression Miner (STEM) was conducted to analyze the DEGs expression patterns in two lines (*Ernst, Nau & Bar-Joseph, 2005*). The enrichment analysis of Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) analysis were based on KOBAS 3.0 software, BLASTX, and GO databases (http://archive.geneontology.org/latest-lite/) (*Altschul et al., 1990*; *Wu et al., 2006*; *Xie et al., 2011*), respectively.

Scanning DEGs in quantitative trait locis

To identify the potential candidate alleles in CCRI70 RIL population, we compared the DEGs with previous quantitative trait loci (QTL) result (*Deng et al., 2019*). Mapping the simple sequence repeat (SSR) loci sequences of LP and FL stable QTLs to the new reference genome using bowtie software (*Langmead et al., 2009*). Scanning the DEGs of −3 and 0 DPA in the physical confidence intervals of LP QTLs while comparing the DEGs of 5 and 10 DPA in FL QTLs.

Hub genes identification and co-expression networks construction

WGCNA (weighted gene co-expression network analysis) R package was used for identifying modules and hub (or highly correlated) genes that highly associated with fiber initiation and elongation (*Langfelder & Horvath, 2008*). The topology overlap matrix was built with hierarchical clustering method, and the dynamic tree cut and merged into modules. Among the modules, what containing a coefficient (>0.6) with each sample were identified as key modules and organized for co-expression networks construction. In WGCNA, K_{ME} was a value to describe the eigengene connectivity. In this study, the DEGs with the highest K_{ME} values in each key module were identified as hub genes (*Pei, Chen & Zhang, 2017*). The top 200 pairs of network connections stored in the edges files by weight value were selected to build interaction networks within DEGs, and the hub genes were selected by the basis of module membership (K_{ME}) values, of which the interaction networks were drawn by Cytoscape 3.7.1 (*Shannon et al., 2003*).

Hub genes and DEGs expression pattern validation

To validate the expression pattern, we performed qRT-PCR on hub genes and selected DEGs. The samples of cDNA were synthesized from 1μg of total RNA by using *TransScript*® II All-in-One First-Strand cDNA Synthesis SuperMix for qPCR (TransGen Biotech co., ltd, China). Real-time PCR was performed by using *TransStart*® *Taq* DNA Polymerase (TransGen Biotech co., ltd, China) and LightCycler® 480 II Real-time PCR instrument (Roche, Basel, Switzerland). The specific primers for qRT-PCR were designed referring to qPrimerDB (https://biodb.swu.edu.cn/qprimerdb) (*Lu et al., 2018*). Gene expression levels were calculated according to the $2^{-\Delta\Delta Ct}$ method with three biological (*Livak & Schmittgen, 2001*).

RESULTS

Phenotypic data analysis of the two lines

In the five environments, Lint percentage (Fig. 1A) in L1 and L2 were 34.40% and 42.86% as well as the fiber length were 32.60 mm and 27.51 mm (Fig. 1B), respectively, which indicated that L1 had longer FL while L2 had higher LP (Table S1).

Transcriptome sequencing analysis and correlation of replicate samples

To reveal gene expression during fiber development, we conducted transcriptome sequencing (RNA-seq) on ovule samples at −3 and 0 DPA as well as fiber samples at 5 and 10 DPA (Fig. 1C). A total of 941.90 million clean reads were obtained from 24
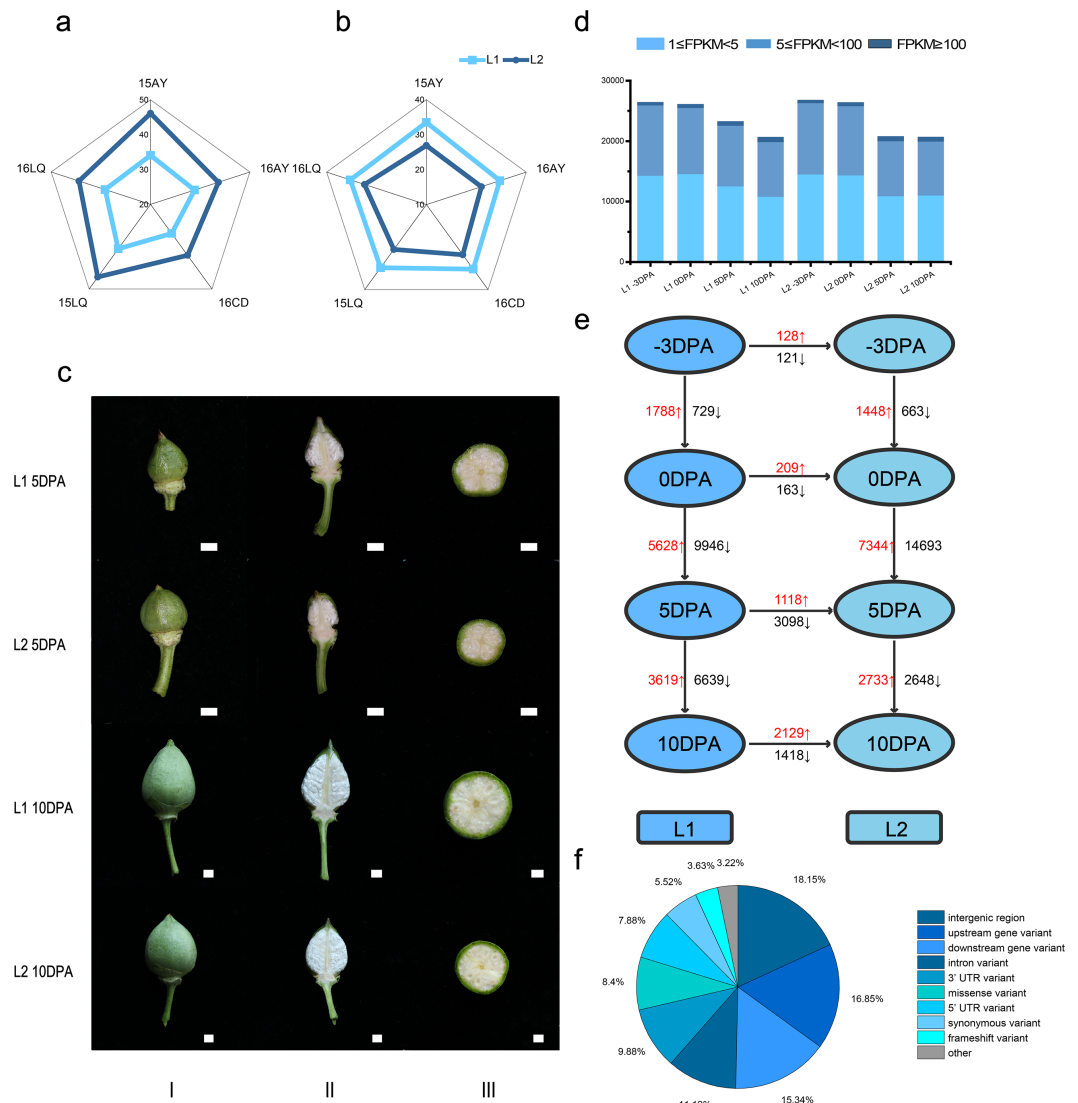
Figure 1 Phenotypic data, statistics for transcript levels at each development stage and vertical of L1 relative to L2 and horizontal comparisons of DEGs. (A) Phenotypic data of lint percentage (LP) trait in five environments of L1 and L2; (B) Phenotypic data of fiber length (FL) trait in five environments of L1 and L2; (C) Photos of bolls of L1 and L2 at 5 and 10 DPA, column I, II and III refer to front photo, longitudinal cut and crosscut of cotton bolls, respectively; (D) Statistics for transcript levels of each sample at each development stage, the numbers of expressed genes were divided by 0.5 <FPKM <5, 5 <FPKM <100 and 100 < FPKM. (E) Vertical and horizontal comparisons showed the DEGs in the same developmental stage between the two lines (L2 relative to L1) and in the same line between different stages. The numbers of upregulated and downregulated genes were marked in red and black, respectively. (F) Variant type distribution.

Full-size ⬚ DOI: 10.7717/peerj.11812/fig-1

libraries with an average of 39.24 million reads per sample. Meanwhile, 91.16% to 95.09% of the Q30 was calculated with an average of 92.97%, while 43.02% to 51.27% of the GC content range was calculated with an average of 45.27%, which indicated the reliability

Jiang et al. (2021), *PeerJ*, DOI 10.7717/peerj.11812

6/28

of RNA-seq data (Table S2 and Fig. S1). The raw data had been submitted to National Genomics Data Center (accession number CRA002982).

Based on FPKM, genes with FPKM value not less than 1 were considered as expressed in this study and the Pearson correlation coefficient was conducted on the 24 samples (Table S3). Consequently, at −3, 0, 5 and 10 DPA, 31456, 31330, 27737 and 25120 expressed genes were identified in L1, respectively. Meanwhile, 31867, 31494, 25215 and 25146 genes were expressed in the L2. The percentages of expressed genes with FPKM values of 1–5, 5–100 and >100 were 54.9%, 43.7% and 1.4%, respectively (Fig. 1D).

Differentially expressed genes analysis

To identify the significantly differentially expressed expressed genes during fiber initiation and elongation, we used DESeq2 R package between each pairwise comparison. Through vertical (Table S4) and horizontal (Table S5) comparisons, after removing the duplicate genes, 30352 unique DEGs (Fig. 1C, Table S6) were identified during cotton fiber development.

At 0 DPA, 369 DEGs were identified, including 206 up-regulated ($\log_2(FC) > 1$) and 163 down-regulated ($\log_2(FC) < -1$), of which the representative DEGs showed high $|\log_2(FC)|$ between L1 and L2 were listed in Table S7. Among up-regulated genes, *Ghir_A05G006080* was enriched in the carbon metabolism pathway and annotated as *NP-GAPDH*, which was involved in catalyzing the oxidation of Ga3P to 3-phosphoglycerate (*Valverde et al., 2005*) and was important in fruit development and energetic metabolism (*Rius et al., 2006*); *Ghir_D08G011800* was annotated as UDP-glycosyltransferase superfamily protein and participates in starch biosynthetic process presenting a direct influence on starch glycan composition (*Ortiz-Marchena et al., 2014*), which might be relevant to accumulating and mobilizing sugars process. Among the down-regulated genes, *Ghir_D13G006000* was annotated as alpha-galactosidase 2, enriched in the galactose metabolism pathway, related with cell wall loosening during cell growth in *Arabidopsis* and barley, it was involved in lengthening the polymers occurring in the wall, upon secretion, or for binding of the XyGs to cellulose (*Peña et al., 2004*) and was specifically localized in the cell wall (*Chrost et al., 2007*); *Ghir_D06G012250* was annotated as disproportionating enzyme 2 and enriched in starch and sucrose metabolism pathway, which could utilize maltose as glucosyl donor and glycogen as acceptor releasing the other hexosyl unit as free glucose that then are further metabolized by the cellular central carbon metabolism (*Andersson & Rådström, 2002*); Le (*Breton et al., 2005*; *Smirnova et al., 2017*).

At 5 DPA, there were 1198 up-regulated and 3098 down-regulated DEGs between L1 and L2. 156 representative ones with high expression (FPKM >5) or high $|\log_2(FC)|$ (>2) are shown in Table S8. Based on KEGG enrichment analysis on the up-regulated DEGs, *Ghir_A13G021680*, *Ghir_A11G013660*, *Ghir_A11G006910*, *Ghir_D11G007650*, *Ghir_A11G025370*, and *Ghir_D05G001330* were enriched into SNARE interactions of vesicular transport pathway, which was participated in endoplasmic reticulum to Golgi vesicle-mediated transport and membrane fusion (*Schiller et al., 2012*; *Bolaños Villegas, Guo & Jauh, 2015*; *Sharma et al., 2017*). In addition, the down-regulated DEGs were identified to be mainly enriched in plant hormone signal transduction, starch and sucrose metabolism

and metabolic pathways, suggesting that in the line with superior fiber quality these pathways play important roles during fiber elongation. *Ghir_D05G014410* was annotated as pectin methylesterase 3 (*PME3*) which was ubiquitously expressed, particularly in vascular tissues and had influence on degree of methylesterification of galacturonic acids. The reaction of demethylesterification decreased the extracellular pH to increase the hydrolytic enzyme activities of enzymes such as poly-galacturonic acid and several pectin enzyme cleavage enzymes (*Wen, Zhu & Hawes, 1999*), when pectin was subject to substantial degradation leading to cell wall structure relaxation and enhancing the growth of cell tips (*Catoire et al., 1998*; *Li et al., 2016*). Meanwhile, *PME3* was also reported affecting the number of adventitious roots (*Guénin et al., 2011*); *Ghir_A13G020210* was annotated as *sucrose synthase 4 (Sus 4)*, where *Sus* was demonstrated to be critically important for cotton fiber initiation and elongation (*Ruan, Llewellyn & Furbank, 2003*). *Ghir_D07G008950*, *Ghir_D10G022670*, *Ghir_A05G005960* and *Ghir_A10G020870* were enriched into phenylalanine metabolism pathway, which was involved in lignin polymer producing and secondary cell wall construction (*Barros et al., 2016*; *Zhou et al., 2017*; *Vanholme et al., 2019*). *Ghir_D05G003750* was annotated as an auxin-responsive factor 7 (*ARF7*), which was required for leaf expansion and/or lateral root induction (*Wilmoth et al., 2005*). There were also some other transcription factors or genes associated with or response to auxin, such as *Ghir_A01G010000* (*ARF5*), *Ghir_D09G022910* (*ATAUX2-11*), *Ghir_D03G003390* (IAA9), *Ghir_D05G022030* (IAA9) and *Ghir_D12G011080* (*SAUR36*) (*Aspuria et al., 2002*; *Fujita et al., 2012*; *Hou, Wu & Gan, 2013*; *Stamm & Kumar, 2013*; *Krogan et al., 2016*) suggesting that auxin signaling pathway is involved in fiber elongation phase in high-FL line.

Genotypic variants analysis

To explore the genetic differences between the two lines, we employed SNPEff program to analyze the genotypic variants distribution on the genome A total of 239493 variants were identified referring to TM-1 (*Wang et al., 2019*), and 40522 genes were involved, where 20181 of them were DEGs.

Among the variants, 20128 (8.40%) arisen as missense variant, 8685 frame shift mutation (3.63%) occurred, 739 variants (0.31%) led to transcription termination and 493 variants (0.21%) occurred losing of stop codon of the transcripts. These variants had of significance impact on the gene functions. In addition, 19760 and 23666 variants were identified at 5′ and 3′ UTR region, respectively, which might have influence on regulation of gene transcription (Table S9, Fig. 1F). Comparing to the DEGs in different developmental stages, 14146 variants were identified in 2662 expressed DEGs (Fig. S4). In the 157 unique significantly up or down regulated DEGs in 5 and 10 DPA (Tables S7 and S8), 652 variants were identified and 78 performed differently in 26 DEGs between two RILs. Due to insertion, *Ghir_A05G005960* occurred frame shift variant. For the SNP variant, transcription termination was occurred in *Ghir_D07G024380* and there were protein sequence mutations in 11 DEGs. There were three variants located (*Ghir_A05G006080*, *Ghir_A04G013900* and *Ghir_A10G010640*) in splice region and one identified on splice acceptor (*Ghir_D03G003390*), which might lead to alternative splicing. Those variants had

significant impact on the protein sequence and functions of DEGs. Besides, 25 variants were located 5′ or 3′ UTR and 15 variants were identified at upstream or downstream that might have effects on the regulation of gene transcription.

Temporal gene expression patterns analysis

To identify the temporal expression profiles, we performed STEM analysis using all the DEGs. 17180 and 17585 DEGs were classified and organized into seven expression profiles with e-values less than 0.001 in L1 (Fig. 2A) and L2 (Fig. 2B), of which the genes in the same profile performed similar expression patterns during fiber development (Table S10). Associated with the fiber development, we focused on the DEGs in profile 14 and 18. In profile 14, the expression had high expression in -3 and 0 DPA, and then was decreasing in fiber samples. While the DEGs in profile 18 had low expression in ovule samples and had a increasing trend in 5 and 10 DPA. The trends of profile 14 and 18 fitted with fiber developmental stages of initiation and elongation, respectively. Profile 14 contained 820 genes in L1 and 401 in L2, while profile 18 identified 1833 genes in L1 and 1788 in L2. Venn diagram was utilized to visualize the gene comparison (Fig. 2C). A total of 681 and 262 DEGs were identified in profile 14 of L1 and L2, with 139 genes expressed in common. Similarly, in profile 18, 1031 and 986 genes were expressed differentially in L1 and L2, while 802 was in common (Fig. 2D).

To investigate the pathways of the common DEGs of the profile 14 and 18, we employed KEGG enrichment analysis (Table S11) and visualized the results with bubble graph. The common DEGs in profile 14 were mainly enriched in the pathways of ribosome, AGE-RAGE signaling pathway, biosynthesis of unsaturated fatty acids and fatty acid metabolism (Fig. 2E). Meanwhile, they were mainly enriched in metabolic pathways such as phagosome, biosynthesis of secondary metabolites, starch and sucrose metabolism and oxidative phosphorylation in profile 18 (Fig. 2F).

Simultaneously, to further explore the specific DEGs in the two lines of profile 14 and 18, we performed GO enrichment analysis and categorized into 35 most frequent GO terms based on biological process, cellular component and molecular function (Table S11). Compared to L1, the specific DEGs of profile 14 in L2 were annotated to the GO terms of GO:0003735 structural constituent of ribosome, GO:0006412 translation, GO:0044212 transcription regulatory region DNA binding, GO:0003676 nucleic acid binding, GO:0009570 chloroplast stroma, GO:0005840 ribosome, GO:0022625 cytosolic large ribosomal subunit, GO:0005773 vacuole and GO:0003729 mRNA binding (Figs. 2G and 2H). As for the GO terms of profile 18 in L1, there were three different enrichment terms compared to those in L2, such as GO:0005622 intracellular, GO:0016787 hydrolase activity and GO:0006355 regulation of transcription, DNA-templated. The GO enrichment analysis results suggested that transcription factors play the different roles between the two RILs during the fiber development (Figs. 2I and 2J).
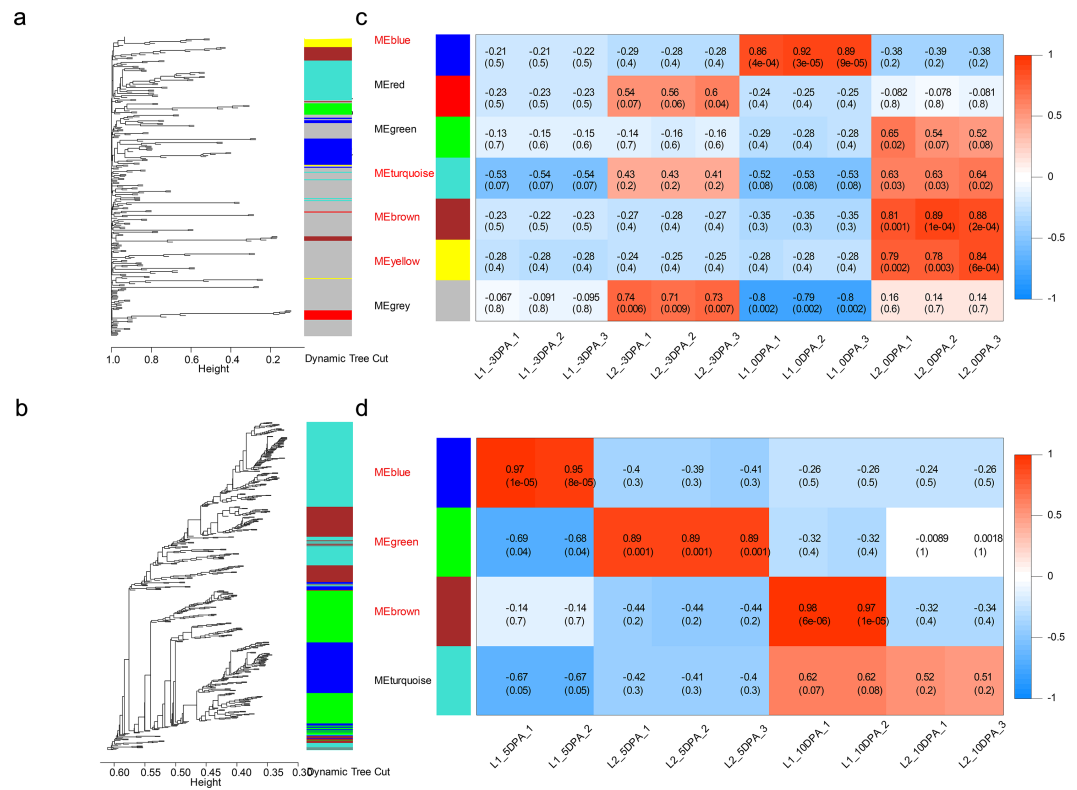
Gene co-expression network analysis and identification of hub genes in correlation networks

To broaden the further insight into the relationship between gene expression and fiber development as well as to identify genes associated with LP and FL, we constructed the
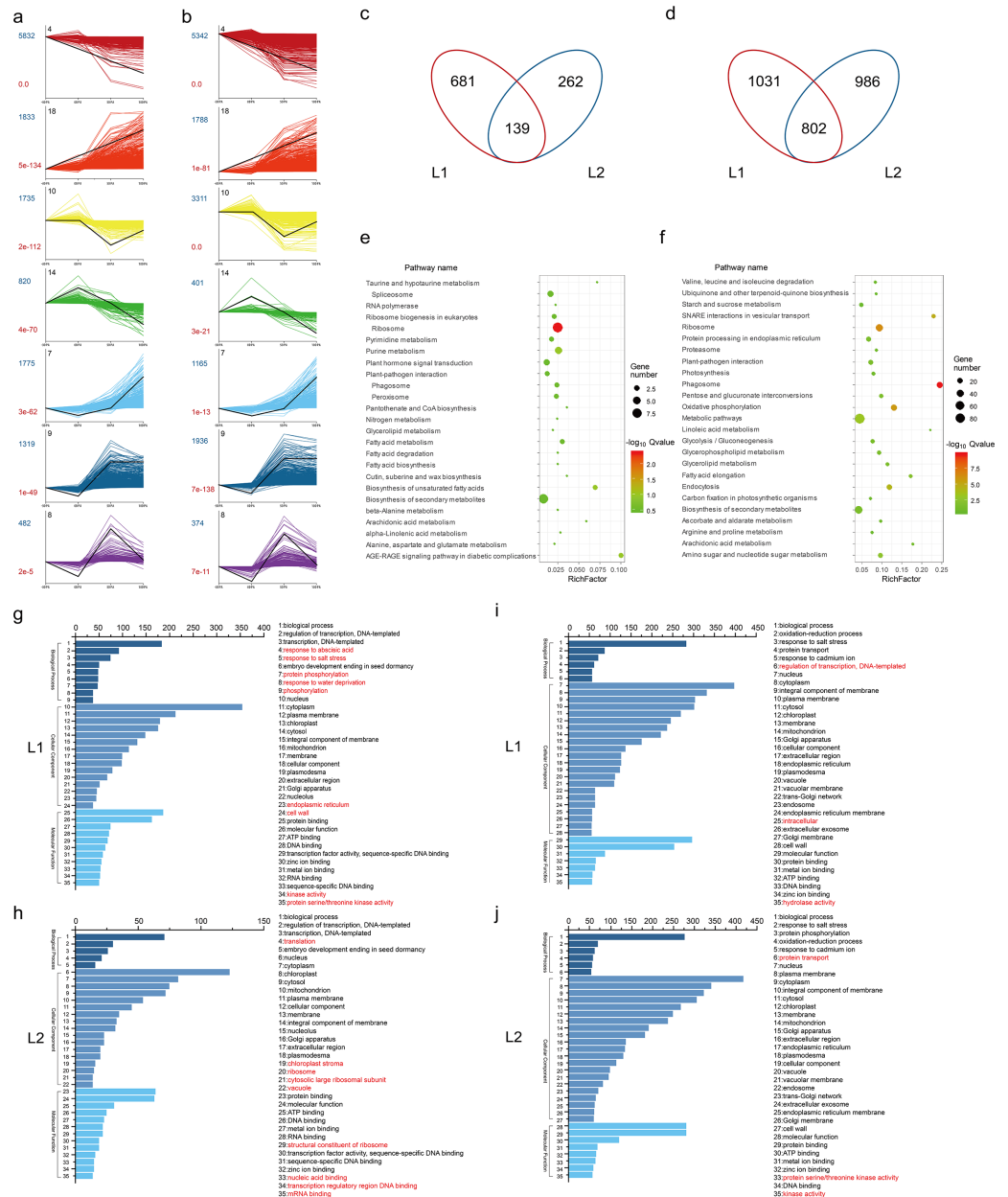
Figure 2 DEGs analysis. (A) Different gene expression profiles in L1 and L2. Each profile presents a gene expression trend. The profile ID, gene number and *P*-value were marked in black, darkblue and darkred, respectively; (B) Venn diagram showed the same and different genes between L1 and L2 in profile 14 and 18, respectively. (C) Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways analysis of the common genes between L1 and L2 in profile 14 and 18. The size and color of bubbles presented the gene number enriched in the pathway and the size of Q-value, respectively. (D) Gene ontology (GO) enrichment analysis of the different genes between L1 and L2 in profile 14 and 18, and the top 35 terms of GO enrichment for 681, 262, 1031 and 986 genes unique to L1 and L2 in profile 14 and 18, respectively. The different terms were marked in red.

Full-size 🖼 DOI: 10.7717/peerj.11812/fig-2

Figure 3 **Weighted gene co-expression network analysis of DEGs at four developmental stages.** Hierarchical dendrogram showing co-expression modules in ovule (A) and fiber (B) samples identified by WGCNA. Each leaf in the tree represents one gene. The major tree was divided into 11 modules in total, where seven modules were classified in ovule (C) and four in fiber (D) samples.

Full-size 🖼 DOI: 10.7717/peerj.11812/fig-3

co-expression networks for the DEGs of ovule (546 genes) and fiber (6976 genes) samples and analyzed through weighted gene co-expression network analysis.

The dynamic tree cut and merged analogous expression patterns into modules (Figs. 3A and 3B). Finally, seven modules and four modules were identified in ovule and fiber samples (Table S12), respectively (Figs. 3C and 3D). Among them, seven key modules were identified. In ovule samples, ovule yellow, brown, turquoise and blue modules were specifically associated with initiation stage of the high-LP line. While fiber brown, blue and green modules were specifically associated with elongation stage of the high-FL line.

In seven key modules and according to K_{ME}, 29 DEGs genes with the highest eigengene connectivity in each module were identified as hub genes (Table 1). All the hub genes performed the K_{ME} values greater than 0.94. The hub genes in ovule yellow module encoding EXORDIUM like 2 protein, SCARECROW-like 21 protein and 1-aminocyclopropane-1-carboxylic acid (acc) synthase 6 (ACS6) protein, where *acs6* was identified involved in cell division and ethylene biosynthesis (*Luo et al., 2014*; *Yin et al., 2019*). *GhACS* and ethylene were playing important roles in cotton fiber development (*Wang et al., 2007*). Hub genes enriched in ovule brown module, encoded ubiquitin family protein, dsRNA-binding protein 2 and natural resistance-associated macrophage protein 1. Additionally, in ovule

Table 1 Candidate hub genes in modules.

Gene Name	KME	ArabidopsisID	Description in *Arabidopsis thaliana*
Fiber brown module			
Ghir_D05G023530	0.99	AT1G77780	Glycosyl hydrolase superfamily protein
Ghir_A08G023930	0.98	AT1G65450	HXXXD-type acyl-transferase family protein
Ghir_D13G004870	0.98		
Ghir_D11G035770	0.98	AT2G36830	gamma tonoplast intrinsic protein
Ghir_A11G034930	0.98	AT2G36830	gamma tonoplast intrinsic protein
Fiber blue module			
Ghir_D08G021430	0.99	AT2G27040	Argonaute family protein
Ghir_D03G003280	0.99	AT5G65700	Leucine-rich receptor-like protein kinase family protein
Ghir_D07G014800	0.99	AT1G01300	Eukaryotic aspartyl protease family protein
Ghir_D05G019650	0.98	AT5G42800	dihydroflavonol 4-reductase
Ghir_A07G015620	0.98	AT2G45290	Transketolase
Fiber green module			
Ghir_A05G020170	0.95	AT5G42655	Disease resistance-responsive (dirigent-like protein) family protein
Ghir_D05G019950	0.95	AT1G45170	outer envelope pore 24B-like protein
Ghir_D11G004790	0.95	AT2G18170	MAP kinase 7
Ghir_A03G022510	0.94	AT1G29500	SAUR-like auxin-responsive protein family
Ovule yellow module			
Ghir_D13G021030	0.98		
Ghir_D05G022870	0.97	AT5G64260	EXORDIUM like 2
Ghir_A04G002270	0.95	AT2G04890	SCARECROW-like 21
Ghir_A12G024510	0.95	AT4G11280	1-aminocyclopropane-1-carboxylic acid (acc) synthase 6
Ovule brown module			
Ghir_A03G022700	0.98	AT2G17200	ubiquitin family protein
Ghir_D08G021550	0.98	AT2G28380	dsRNA-binding protein 2
Ghir_D05G012040	0.96	AT1G80830	natural resistance-associated macrophage protein 1
Ovule turquoise module			
Ghir_A01G001190	0.97	AT1G78300	general regulatory factor 2
Ghir_A07G020020	0.96	AT5G27260	Myb/SANT-like DNA-binding domain protein
Ghir_A09G017620	0.96	AT5G01620	TRICHOME BIREFRINGENCE-LIKE 35
Ghir_A05G042810	0.94	AT5G54130	Calcium-binding endonuclease/exonuclease/phosphatase family
Ovule blue module			
Ghir_A08G021080	0.98	AT2G30090	Acyl-CoA N-acyltransferases (NAT) superfamily protein
Ghir_A06G019930	0.96	AT3G15820	phosphatidic acid phosphatase-related / PAP2-related
Ghir_A11G031900	0.96	AT2G44260	Plant protein of unknown function (DUF946)
Ghir_D13G001750	0.94	AT3G11210	SGNH hydrolase-type esterase superfamily protein

turquoise module (Fig. 4A), the hub genes were annotated as general regulatory factor 2 (GRF2), myb/SANT-like DNA-binding domain protein, TRICHOME BIREFRINGENCE-LIKE 35 (TBL35) protein and Calcium-binding endonuclease/exonuclease/phosphatase family protein, of which *Ghir_A09G017620* annotated as *TBL35*, may participate in xylan
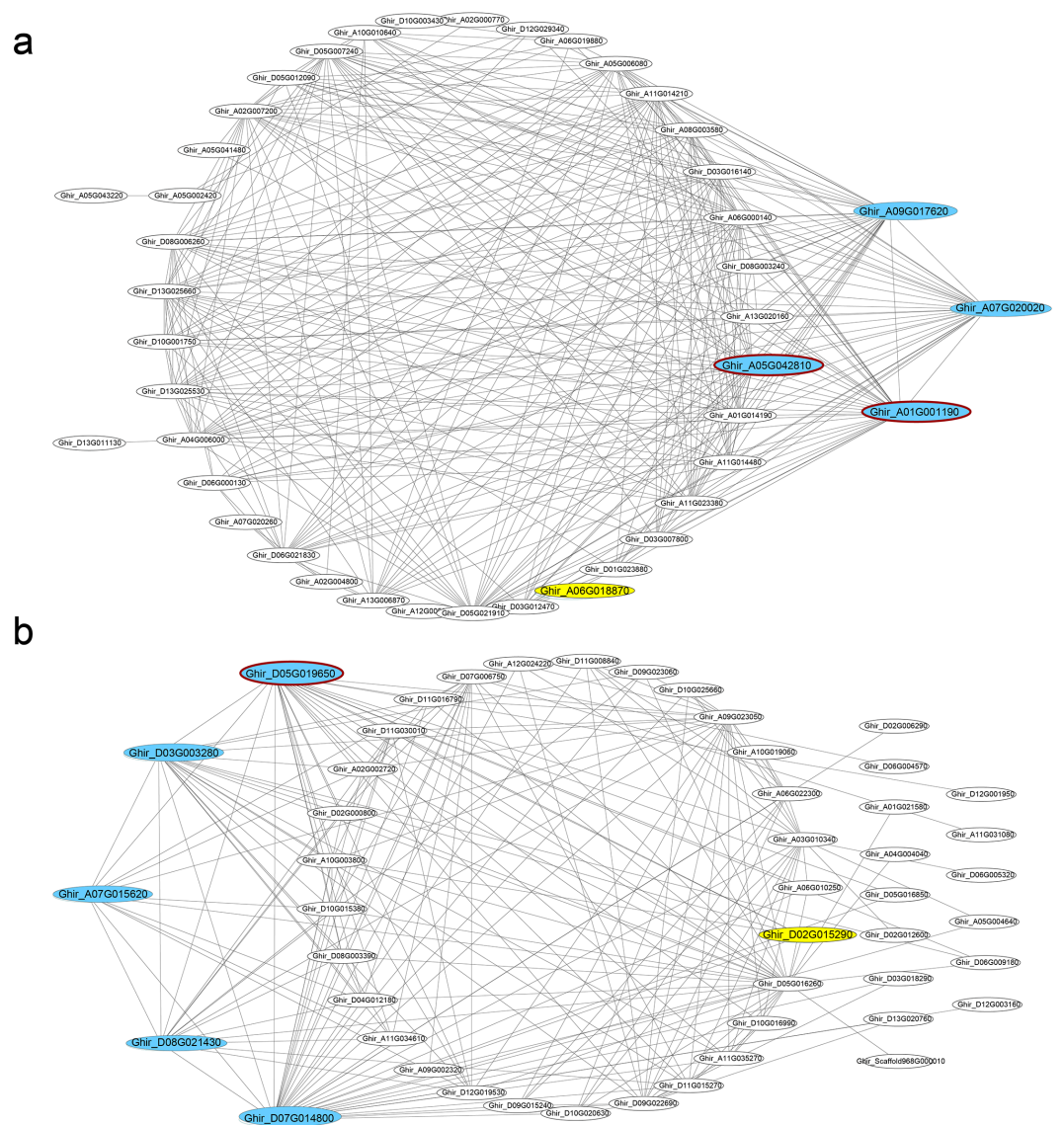
Figure 4 **Co-expression networks.** Hub genes and transcription factors of (A) ovule turquoise module and (B) fiber blue module, which are highlighted in cyan and yellow, respectively. The hub genes circled dark red shared genotypic differences.

Full-size 🖼 DOI: 10.7717/peerj.11812/fig-4

acetylation (*Yuan et al., 2016*). *Ghir_A01G001190* was annotated as G-box binding factor GF14 omega encoding a $14-3-3$ protein, which was reported to be located at the regions of the plant that comprise dividing cells and involved in plant cell cycle (*Sorrell et al., 2003*).

In the modules of fiber tissues at 5 and 10 DPA, hub genes were critically associated with the fiber elongation stage. The hub genes in blue module at 5 DPA encoded an argonaute family protein, a leucine-rich receptor-like protein kinase family protein, a eukaryotic aspartyl protease family protein, dihydroflavonol 4-reductase and transketolase, which was involved in vascular development (*Qian et al., 2018*). Similarly, the green module contained the hub genes encoding a disease resistance-responsive (dirigent-like protein)

family protein, an outer envelope pore 24B-like protein, a SAUR-like auxin-responsive protein and a MAP kinase 7. At 10 DPA, the hub genes in brown module were annotated with a glycosyl hydrolase superfamily protein, a HXXXD-type acyl-transferase family protein and two gamma tonoplast intrinsic proteins. Among the hub genes in fiber blue (Fig. 4B) and brown modules, *Ghir_D11G035770* and *Ghir_A11G034930* were both identified as hub genes in fiber brown module and were annotated as gamma tonoplast intrinsic protein (GAMMA-TIP1), which is mainly expressed in vessel-flanking cells of vascular bundles (*Beebo et al., 2009*) and confirmed to mediate unbalanced water content in leaves (*Zhu et al., 2014*). *Gh γ TIP1* during fiber cell elongation played important roles in supporting the rapid influx of water into vacuoles during cotton fiber cell expansion (*Liu et al., 2008*). Besides, *Ghir_A01G001290*, sharing high eigengene connectivity, was annotated as APUM-7 translation factor, and *Ghir_A01G005740* encoded a domain of unknown function (DUF1218) family protein, whose homologous gene was knocked-out showing a reduction in total xylem (*Ubeda-Tomas et al., 2007*). The protein containing DUF1218 domain played important role in xylogenesis and/or secondary cell wall formation (*Mewalal et al., 2016*).

Scanning the genotypic variants in the hub genes, 161 variants were identified in the transcripts of hub genes and 51 variants performed differently between the two RILs. Due to the variants, transcription termination of *Ghir_A05G042810* and *Ghir_D13G001750* occurred in L2 and frame shift mutation of *Ghir_A01G001190* occurred in L1 with one nucleotide deletion, which might have huge impact on the protein sequences. In addition, one SNP variant located on the intron of *Ghir_A05G042810* that might lead to alternative splicing and changing the protein sequence. Causing by SNP variant, missense variants occurred in another five hub genes that changed the protein primary structure (*Ghir_A06G019930*, *Ghir_D05G019650*, *Ghir_D05G019950*, *Ghir_D05G022870* and *Ghir_A05G042810*). 14 variants were identified at 5′ or 3′ UTR and 16 were located at downstream that might have effects on the regulation of gene transcription.

qRT-PCR expression pattern validation

To validate the expression pattern, we performed qRT-PCR on important DEGs and hub genes using the primers according to qPrimerDB (Table S13). The expression pattern validation result of the 29 hub genes from ovule yellow (Figs. 5A–5D), ovule brown (Figs. 5E–5G), ovule turquoise (Figs. 5H–5K), ovule blue (5L– 5O), fiber brown (Figs. 5P–5T), fiber blue (Figs. 5U–5Y) and fiber green (Figs. 5Z–5CC) modules was similar to RNA-seq result. While the 11 key DEGs performed the similar result (Figs. 5 and 5D–5NN) (Table S14). Those results indicated the RNA-seq result is reliable.

DISCUSSION

Transcriptome sequencing of two extrame-parent RILs provided new insight for exploring the expression profile in fiber initiation and elongation stages

To explore the genetic and molecular mechanisms during fiber initiation and elongation, we selected two extreme RILs form CCRI70 RIL population and conducted RNA-seq. In recent
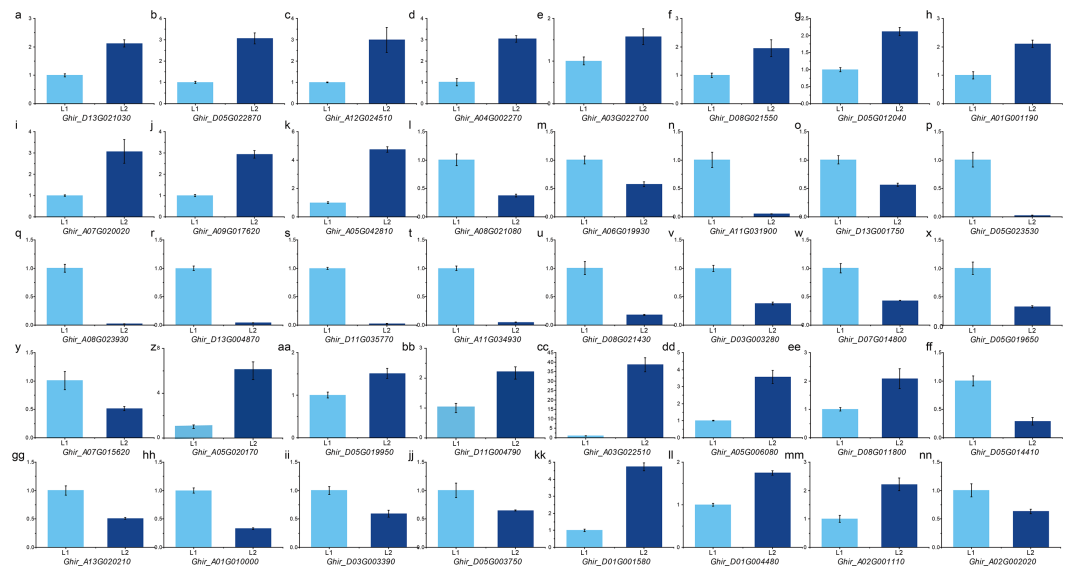
Figure 5 qPCR analysis of hub genes and key DEGs. qPCR results of the hub genes in ovule yellow (A–D), brown (E–G), turquoise (H–K) and blue (L–O) modules as well as the hub genes in fiber brown (P–T), blue (U–Y) and green (Z–CC) modules. And the key DEGs during fiber development (BB–JJ) and the DEGs in QTLs (KK–NN).

years, RNA-seq has been largely applied into cotton fiber development researches (*Gilbert et al., 2014*; *Yoo & Wendel, 2014*; *Islam et al., 2016*; *Li et al., 2017b*; *Li et al., 2017a*; *Lu et al., 2017*; *Zou et al., 2019*) as well as under biotic (*Xu et al., 2011*; *Zhang et al., 2017*) and abiotic stresses (*Zhang et al., 2016a*). Referring to the nearly studies of fiber development using RNA-seq, elongation and secondary wall biosynthesis were more often concentrated in, while the initiation stage was seldom focused on. In our work, ovule and fiber samples collected during cotton fiber initiation and elongation were conducted RNA-seq, which provided plenty of valuable data for investigating and revealing the differences of genetic and molecular mechanisms during initiation and elongation on mRNA level. CCRI70 is a breeding hybrid with excellent performance in fiber length and moderate performance in lint percentage. The two extreme RILs provide an ideal model to investigate the candidate genes coming from and the differences in initiation and elongation stages. What's more, what good alleles related to lint percentage and fiber length were found would also benefit upland cotton breeding to improve yield and fiber quality simultaneously. In our study, 941.90 million clean reads were obtained from 24 libraries, with an average of 39.24 million per sample. Among the 24 libraries, the GC% and Q30 were 45.27% and 92.97% on average, respectively, which indicates that the quality of the RNA-seq data is reliable. However, three samples in the third biological replicate showed low correlation (<0.8), which might be affected by environmental and other factors. That provided ideal and fine basis for exploring the transcriptional differences during fiber initiation and elongation stages, which prevailingly affect the lint percentage and fiber length traits, respectively.

DEGs revealing the transcriptional differences in initiation and elongation phases

Due to the initiation and elongation affected the number and length of fibers, which had influences on LP and FL traits, respectively. DEGs of each period were obtained and used for identifying the candidate genes that would reveal the genetic basis of fiber development and provide an insight into the molecular mechanism for the negative correlation between quality and yield traits. At fiber initiation stage, up-regulated genes in high-LP line L2 were mainly enriched in pentose and glucuronate interconversions, carbon metabolism, biosynthesis of secondary metabolites and metabolic pathways. *Ghir_A05G006080* (Fig. 5DD) might play the role like *NP-GAPDH*, a cytosolic non-phosphorylating NADP-dependent GAPDH that catalyzes the oxidation of Ga3P to 3-phosphoglycerate (*Valverde et al., 2005*; *Rius et al., 2006*). *Ghir_D08G011800* (Fig. 5EE) might be involved in starch biosynthetic process that had a direct influence on starch glycan composition (*Ortiz-Marchena et al., 2014*). At 0 DPA, the up-regulated genes were mainly enriched in the energetic metabolism and accumulating as well as mobilizing sugars process, implying that energetic metabolism and sugar transport may participate in fiber initiation and have effect on the number of cotton fiber. Among the up-regulated genes in high-FL line L1 at 5 DPA, *Ghir_D05G014410* (Fig. 5FF), annotated as *PME3*, had influence on degree of methylesterification of galacturonic acids (*Wen, Zhu & Hawes, 1999*; *Guénin et al., 2011*). Pectin was subject to substantial degradation leading to cell wall structure relaxation and enhancing the growth of cell tips (*Catoire et al., 1998*; *Li et al., 2016*). During cotton fiber development, *PME* played significant physiological role by influencing the chemical properties of pectin (*Liu, Talbot & Llewellyn, 2013*). A *sucrose synthase 4* (*Sus 4*), *Ghir_A13G020210* (Fig. 5GG), was specifically expressed in L1 with high FPKM, where it played a major role in metabolic regulation and sugar signaling, and silencing *Sus* expression led to a fiberless seed phenotype. *Sus* was demonstrated to be significantly important for cotton fiber development, and suppression of sucrose synthase gene expression repressed cotton fiber cell initiation, elongation, and seed development (*Ruan, Llewellyn & Furbank, 2003*). In addition, *Ghir_A01G010000* (Fig. 5HH), *Ghir_D03G003390* (Fig. 5II), *Ghir_D05G003750* (Fig. 5JJ) and some other genes were annotated as transcription factors or genes related to or responding to auxin demonstrated that auxin regulate fiber development and auxin signaling was shown to be important for fiber initiation and elongation (Samuel (*Yang et al., 2006*; *Gou et al., 2007*; *Liu et al., 2012*; *Wang et al., 2013*; *Zhang et al., 2016b*). Overexpressing *iaaM*, critically important for auxin biosynthesis, led to enhanced initiation and increased fiber length (*Zhang et al., 2011*). The DEGs result suggested that *Sus4, PME3* and auxin signaling pathway play important roles in fiber elongation stage. In the two materials, the DEGs were identified and enriched into energetic metabolism and sugar transport pathway in initiation stage, while the DEGs were enriched into auxin signaling pathway in rapid elongation stage.

In addition, comparing the DEGs with CCRI70 previous QTL result, 14 DEGs were located in LP or FL QTLs (*Deng et al., 2019*). Among them, *Ghir_D01G001580* (Fig. 5KK) and *Ghir_D01G004480* (Fig. 5LL) were up-regulated in L2 at 0 DPA and detected in LP QTL. *Ghir_D01G001580* was annotated as ATXR-2 that was involved in cellular dedifferentiation

(*Lee, Park & Seo, 2017*). ATXR2-ARF-LBD axis was key for the epigenetic regulation of callus formation in *Arabidopsis*. *Ghir_D01G004480* was annotated as *Ku70* that was involved in repair of DNA double-stranded breaks and telomere regulation (*Tamura et al., 2002*) demonstrated to be required for the maintenance of chromosome stability and normal developmental growth in rice (*Hong et al., 2010*). Besides, *Ghir_A02G001110* (Fig. 5MM) and *Ghir_A02G002020* (Fig. 5NN) were identified in FL stable QTL. *Ghir_A02G002020* was annotated as xyloglucan endotransglucosylase/hydrolases 16 (*XTH16)* and had a higher expression in L1 at 10 DPA. XTHs worked on xyloglucan-cellulose network, modified the cell wall via enzymatic mechanisms (*Nishitani & Tominaga, 1992*; *Rose et al., 2002*) and were required during plant growth in cell wall modification (*Campbell & Braam, 1999*; *Rose et al., 2002*). It was confirmed that XTHs were necessary in cell wall restructuring during cellular expansion, which fueled rapid petiole elongation (*Sasidharan et al., 2010*). XTH16 was also involved in radish taproot thickening (*Yu et al., 2015*). *Ghir_A02G001110* was annotated as *IQD13* and was interacting with both microtubules and the plasma membrane. It specifically promoted cortical microtubule rescue, which consequently increased cortical microtubule density (*Sugiyama et al., 2017*). All above DEGs provided insight into the differences of molecular mechanism during fiber development on transcription level, which would be also beneficial for cotton breeding.

Hub genes identified by WGCNA may have significantly impact on lint percentage and fiber length

In this study, WGCNA was performed to identify hub genes and modules, which were highly associated with cotton fiber initiation and elongation. To investigate the influences of the hub genes on fiber yield and quality traits during fiber development, multiple comparisons with the previous studies were performed. At 5 DPA, *Ghir_A07G015620*, sharing high protein sequence identification with *Gh_A07G1360*, showed highly correlation with boll weight and seed index traits in *Zhang et al.*'s (*2020*) report. In *Song et al.*'s (*2019*) study, *Ghir_D05G012040* (*Gh_D05G1139*) identified in the module associated with high-LP line L2 at 0 DPA was locating in the QTL related to lint percentage trait. *Ghir_A08G023930* (*Gh_A08G2014*) and *Ghir_A09G017620* (*Gh_A09G2422*), identified in Fiber brown module and Ovule turquoise module, were detected in FL QTL in *Naoumkina et al.*'s (*2019*) report. In *Sun et al.*'s (*2017*) and *Liu et al.*'s (*2018*) studies, *Ghir_D03G003280* (*Gh_D03G0303*) was reported to have influence on fiber length by GWAS analysis. At 0 DPA, *Ghir_A12G024510* was annotated as 1-aminocyclopropane-1-carboxylate synthase 6 in *G. hirsutum* (*GhACS6*), which was identified as the key enzyme involved in ethylene biosynthesis and was considered critically important for cotton fiber elongation (*Wang et al., 2007*). Meanwhile, *Ghir_A03G022700* showed highly protein sequence identity with a ubiquitin family protein that could interact with and responds to the degradation of *GbPDF1*. *GbPDF1* was confirmed playing a critical role in cotton fiber development and required in fiber initiation, where *PDF1*-silenced cotton showed retarded fiber initiation and had shorter fibers or lower lint percentage (*Deng et al., 2012*). It was hub genes result suggested that ethylene is significantly important for cotton fiber development. At 10 DPA, *Ghir_D05G023530* shared high identity with endo-1,3-beta-glucanase, which was reported

to be involved in secondary wall synthesis accompanying the deposition of cellulose in growing cotton fiber cells (*Shimizu et al., 1997*), implying that it might have influence on fiber strength. During fiber development, hub genes played important roles and had impact on fiber yield and quality traits. Therefore, the functions and genetic mechanisms of the hub genes were worthy for further exploring.

CONCLUSIONS

In summary, the two extreme *G. hirsutum* RILs selected from CCRI70 RIL population were conducted transcriptome research on fiber initiation and elongation, aiming to understand the parental source of potential alleles and the differential molecular mechanisms associated with LP and FL. As a result, 249/128, 369/206, 4296/1198 and 3547/2129 up-/down-regulated DEGs were obtained during fiber development at -3, 0, 5 and 10 DPA, respectively. According to TM-1, 239493 genotypic variants were identified and 40522 genes were involved. Based on KEGG enrichment analysis on the DEGs at 0 and 5 DPA, galactose metabolism, auxin signaling pathway and etc. were significant enriched into. By STEM, profile 14 and 18 were considered highly associated with cotton fiber initiation and elongation, of which genes expressed differently were analyzed by gene ontology analysis while genes expressed in common were enriched by KEGG. KEGG analysis revealed that the DEGs were involved in the pathways of ribosome, AGE-RAGE signaling pathway, biosynthesis of unsaturated fatty acids and fatty acid metabolism of profile 14. Genes in profile 18 were enriched into the pathways of metabolic pathways, phagosome, biosynthesis of secondary metabolites, starch and sucrose metabolism and oxidative phosphorylation. Co-expression network analysis by using WGCNA identified 29 hub genes in four fiber developmental time points. *Ghir_A03G022700* was annotated as a ubiquitin family protein that could interact with and responds to the degradation of *GbPDF1*, which was considered being critically important and required in fiber initiation. *Ghir_A12G024510* annotated as *GhACS* was identified as the key enzyme involved in ethylene biosynthesis and was considered critically important for cotton fiber elongation. These findings would provide insights into the molecular mechanism of the fiber development, which would be the genetic basis to improve the yield and fiber quality simultaneously of upland cotton breeding.

ADDITIONAL INFORMATION AND DECLARATIONS

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Xiao Jiang conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.
- Liqiang Fan and Senmiao Fan performed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.
- Pengtao Li performed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Xianyan Zou analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Zhen Zhang conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Juwu Gong performed the experiments, prepared figures and/or tables, and approved the final draft.
- Youlu Yuan and Haihong Shang conceived and designed the experiments, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The raw data is available at the National Genomics Data Center (NGDC): CRA002982.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.11812#supplemental-information.

REFERENCES

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215(3)**:403–410 DOI 10.1016/S0022-2836(05)80360-2.

Andersson U, Rådström P. 2002. β-Glucose 1-phosphate-interconverting enzymes in maltose-and trehalose-fermenting lactic acid bacteria. *Environmental Microbiology* **4(2)**:81–88 DOI 10.1046/j.1462-2920.2002.00268.x.

Applequist WL, Cronn R, Wendel JF. 2001. Comparative development of fiber in wild and cultivated cotton. *Evolution & Development* **3(1)**:3–17 DOI 10.1046/j.1525-142x.2001.00079.x.

Aspuria E, Ooura C, Chen G, Uchimiya H, Oono Y. 2002. GFP accumulation controlled by an auxin-responsive promoter as a non-destructive assay to monitor early auxin response. *Plant Cell Reports* **21(1)**:52–57 DOI 10.1007/s00299-002-0484-6.

Barros J, Serrani-Yarce JC, Chen F, Baxter D, Venables BJ, Dixon RA. 2016. Role of bifunctional ammonia-lyase in grass cell wall biosynthesis. *Nature Plants* **2(6)**:16050 DOI 10.1038/nplants.2016.50.

Basra AS, Malik C. 1984. Development of the cotton fiber. In: Bourne GH, Danielli JF, Jeon KW, eds. *International review of cytology*. vol 89. Academic Press, 65–113.

Beebo A, Thomas D, Der C, Sanchez L, Leborgne-Castel N, Marty F, Schoefs B, Bouhidel K. 2009. Life with and without AtTIP1; 1, an Arabidopsis aquaporin preferentially localized in the apposing tonoplasts of adjacent vacuoles. *Plant Molecular Biology* **70(1-2)**:193–209 DOI 10.1007/s11103-009-9465-2.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30(15)**:2114–2120 DOI 10.1093/bioinformatics/btu170.

Breton YLe, Pichereau V, Sauvageot N, Auffray Y, Rince A. 2005. Maltose utilization in Enterococcus faecalis. *Journal of Applied Microbiology* **98(4)**:806–813 DOI 10.1111/j.1365-2672.2004.02468.x.

Campbell P, Braam J. 1999. Xyloglucan endotransglycosylases: diversity of genes, enzymes and potential wall-modifying functions. *Trends in Plant Science* **4(9)**:361–366 DOI 10.1016/s1360-1385(99)01468-5.

Catoire L, Pierron M, Morvan C, du Penhoat CH, Goldberg RJJoBC. 1998. Investigation of the action patterns of pectinmethylesterase isoforms through kinetic analyses and NMR spectroscopy. Implications In cell wall expansion. *The Journal of Biological Chemistry* **273(50)**:33150–33156 DOI 10.1074/jbc.273.50.33150.

Chrost B, Kolukisaoglu U, Schulz B, Krupinska K. 2007. An α-galactosidase with an essential function during leaf development. *Planta* **225(2)**:311–320 DOI 10.1007/s00425-006-0350-9.

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* **6(2)**:80–92 DOI 10.4161/fly.19695.

Deng F, Tu L, Tan J, Li Y, Nie Y, Zhang X. 2012. GbPDF1 is involved in cotton fiber initiation via the core cis-element HDZIP2ATATHB2. *Plant Physiology* **158(2)**:890–904 DOI 10.1104/pp.111.186742.

Deng X, Gong J, Liu A, Shi Y, Gong W, Ge Q, Li J, Shang H, Wu Y, Yuan Y. 2019. QTL mapping for fiber quality and yield-related traits across multiple generations in segregating population of CCRI 70. *Journal of Cotton Research* **2(1)**:13 DOI 10.1186/s42397-019-0029-y.

Ernst J, Nau GJ, Bar-Joseph Z. 2005. Clustering short time series gene expression data. *Bioinformatics* **21(suppl_1)**:i159–i168 DOI 10.1093/bioinformatics/bti1022.

Fujita K, Horiuchi H, Takato H, Kohno M, Suzuki S. 2012. Auxin-responsive grape Aux/IAA9 regulates transgenic Arabidopsis plant growth. *Molecular Biology Reports* **39(7)**:7823–7829 DOI 10.1007/s11033-012-1625-9.

Gilbert MK, Kim HJ, Tang Y, Naoumkina M, Fang DD. 2014. Comparative transcriptome analysis of short fiber mutants Ligon-lintless 1 and 2 reveals common mechanisms pertinent to fiber elongation in cotton (Gossypium hirsutum L.). *PLOS ONE* **9(4)**:e95554 DOI 10.1371/journal.pone.0095554.

Gou JY, Wang LJ, Chen SP, Hu WL, Chen XY. 2007. Gene expression and metabolite profiles of cotton fiber during cell elongation and secondary cell wall synthesis. *Cell Research* **17(5)**:422–434 DOI 10.1038/sj.cr.7310150.

Guénin S, Mareck A, Rayon C, Lamour R, Ndong YAssoumou, Domon JM, Sénéchal F, Fournet F, Jamet E, Canut H. 2011. Identification of pectin methylesterase 3 as a basic pectin methylesterase isoform involved in adventitious rooting in Arabidopsis thaliana. *New Phytologist* **192(1)**:114–126 DOI 10.1111/j.1469-8137.2011.03797.x.

Haigler CH, Betancur L, Stiff MR, Tuttle JR. 2012. Cotton fiber: a powerful single-cell model for cell wall and cellulose research. *Frontiers in Plant Science* **3**:104 DOI 10.3389/fpls.2012.00104.

Hong JP, Byun MY, An K, Yang SJ, An G, Kim WT. 2010. OsKu70 is associated with developmental growth and genome stability in rice. *Plant Physiology* **152(1)**:374–387 DOI 10.1104/pp.109.150391.

Hou K, Wu W, Gan S-S. 2013. SAUR36, a small auxin up RNA gene, is involved in the promotion of leaf senescence in Arabidopsis. *Plant Physiology* **161(2)**:1002–1009 DOI 10.1104/pp.112.212787.

Hu Y, Chen J, Fang L, Zhang Z, Ma W, Niu Y, Ju L, Deng J, Zhao T, Lian J, Baruch K, Fang D, Liu X, Ruan YL, Rahman MU, Han J, Wang K, Wang Q, Wu H, Mei G, Zang Y, Han Z, Xu C, Shen W, Yang D, Si Z, Dai F, Zou L, Huang F, Bai Y, Zhang Y, Brodt A, Ben-Hamo H, Zhu X, Zhou B, Guan X, Zhu S, Chen X, Zhang T. 2019. Gossypium barbadense and Gossypium hirsutum genomes provide insights into the origin and evolution of allotetraploid cotton. *Nature Genetics* **51(4)**:739–748 DOI 10.1038/s41588-019-0371-5.

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

Islam MS, Fang DD, Thyssen GN, Delhom CD, Liu Y, Kim HJ. 2016. Comparative fiber property and transcriptome analyses reveal key genes potentially related to high fiber strength in cotton (Gossypium hirsutum L.) line MD52ne. *BMC Plant Biology* **16(1)**:36 DOI 10.1186/s12870-016-0727-2.

Kim HJ, Triplett BA. 2001. Cotton fiber growth in planta and in vitro. Models for plant cell elongation and cell wall biogenesis. *Plant Physiology* **127(4)**:1361–1366 DOI 10.1104/pp.010724.

Krogan NT, Marcos D, Weiner AI, Berleth T. 2016. The auxin response factor MONOPTEROS controls meristem function and organogenesis in both the shoot and root through the direct regulation of PIN genes. *New Phytologist* **212(1)**:42–50 DOI 10.1111/nph.14107.

Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9(1)**:559 DOI 10.1186/1471-2105-9-559.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10(3)**:R25 DOI 10.1186/gb-2009-10-3-r25.

Lee JJ, Hassan OS, Gao W, Wei NE, Kohel RJ, Chen X-Y, Payton P, Sze S-H, Stelly DM, Chen ZJ. 2006. Developmental and gene expression analyses of a cotton naked seed mutant. *Planta* **223(3)**:418–432 DOI 10.1007/s00425-005-0098-7.

Lee JJ, Woodward AW, Chen ZJ. 2007. Gene expression changes and early events in cotton fibre development. *Annals of Botany* **100(7)**:1391–1401 DOI 10.1093/aob/mcm232.

Lee K, Park O-S, Seo PJ. 2017. < em > Arabidopsis < em > ATXR2 deposits H3K36me3 at the promoters of < em > LBD < em > genes to facilitate cellular dedifferentiation. *Science Signaling* **10(507)**:eaan0316 DOI 10.1126/scisignal.aan0316.

Li F, Fan G, Lu C, Xiao G, Zou C, Kohel RJ, Ma Z, Shang H, Ma X, Wu J. 2015. Genome sequence of cultivated Upland cotton (Gossypium hirsutum TM-1) provides insights into genome evolution. *Nature Biotechnology* **33(5)**:524 DOI 10.1038/nbt.3208.

Li F, Fan G, Wang K, Sun F, Yuan Y, Song G, Li Q, Ma Z, Lu C, Zou C. 2014. Genome sequence of the cultivated cotton Gossypium arboreum. *Nature Genetics* **46(6)**:567 DOI 10.1038/ng.2987.

Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27(21)**:2987–2993 DOI 10.1093/bioinformatics/btr509.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. 1000 genome project data processing subgroup, 2009. The sequence alignment/map format and samtools. *Bioinformatics* **25(16)**:2078–2079 DOI 10.1093/bioinformatics/btp352.

Li W, Shang H, Ge Q, Zou C, Cai J, Wang D, Fan S, Zhang Z, Deng X, Tan Y, Song W, Li P, Koffi PK, Jamshed M, Lu Q, Gong W, Li J, Shi Y, Chen T, Gong J, Liu A, Yuan Y. 2016. Genome-wide identification, phylogeny, and expression analysis of pectin methylesterases reveal their major role in cotton fiber development. *BMC Genomics* **17(1)**:1000–1000 DOI 10.1186/s12864-016-3365.

Li P-T, Wang M, Lu Q-W, Ge Q, Liu A-Y, Gong J-W, Shang H-H, Gong W-K, Li J-W. 2017a. Comparative transcriptome analysis of cotton fiber development of Upland cotton (Gossypium hirsutum) and Chromosome Segment Substitution Lines from G. hirsutum × G. barbadense. *BMC Genomics* **18(1)**:705 DOI 10.1186/s12864-017-4077-8.

Li X, Wu M, Liu G, Pei W, Zhai H, Yu J, Zhang J. 2017b. Identification of candidate genes for fiber length quantitative trait loci through RNA-Seq and linkage and physical mapping in cotton. *BMC Genomics* **18(1)**:427 DOI 10.1186/s12864-017-3812-5.

Liu D, Tu L, Wang L, Li Y, Zhu L, Zhang X. 2008. Characterization and expression of plasma and tonoplast membrane aquaporins in elongating cotton fibers. *Plant Cell Reports* **27(8)**:1385–1394 DOI 10.1007/s00299-008-0545-6.

Liu K, Sun J, Yao L, Yuan Y. 2012. Transcriptome analysis reveals critical genes and key pathways for early cotton fiber elongation in Ligon lintless-1 mutant. *Genomics* **100(1)**:42–50 DOI 10.1016/j.ygeno.2012.04.007.

Liu Q, Talbot M, Llewellyn DJ. 2013. Pectin methylesterase and pectin remodelling differ in the fibre walls of two gossypium species with very different fibre properties. *PLOS ONE* **8(6)**:e65131-e65131 DOI 10.1371/journal.pone.0065131.

Liu R, Gong J, Xiao X, Zhang Z, Li J, Liu A, Lu Q, Shang H, Shi Y, Ge Q. 2018. GWAS analysis and QTL identification of fiber quality traits and yield components in upland cotton using enriched high-density SNP markers. *Frontiers in Plant Science* **9**:1067 DOI 10.3389/fpls.2018.01067.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\delta\delta CT}$ Method. *Methods* **25(4)**:402–408 DOI 10.1006/meth.2001.1262.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15(12)**:550 DOI 10.1186/s13059-014-0550-8.

Lu K, Li T, He J, Chang W, Zhang R, Liu M, Yu M, Fan Y, Ma J, Sun W, Qu C, Liu L, Li N, Liang Y, Wang R, Qian W, Tang Z, Xu X, Lei B, Zhang K, Li J. 2018. qPrimerDB: a thermodynamics-based gene-specific qPCR primer database for 147 organisms. *Nucleic Acids Research* **46(D1)**:D1229–D1236 DOI 10.1093/nar/gkx725.

Lu Q, Shi Y, Xiao X, Li P, Gong J, Gong W, Liu A, Shang H, Li J, Ge Q. 2017. Transcriptome analysis suggests that chromosome introgression fragments from sea island cotton (Gossypium barbadense) increase fiber strength in upland cotton (Gossypium hirsutum). *G3: Genes, Genomes, Genetics* **7(10)**:3469–3479 DOI 10.1534/g3.117.300108.

Luo X, Chen Z, Gao J, Gong Z. 2014. Abscisic acid inhibits root growth in Arabidopsis through ethylene biosynthesis. *The Plant Journal* **79(1)**:44–55 DOI 10.1111/tpj.12534.

Mewalal R, Mizrachi E, Coetzee B, Mansfield SD, Myburg AA. 2016. The Arabidopsis domain of unknown function 1218 (DUF1218) containing proteins, MODIFYING WALL LIGNIN-1 and 2 (At1g31720/MWL-1 and At4g19370/MWL-2) function redundantly to alter secondary cell wall lignin content. *PLOS ONE* **11(3)**:e0150254 DOI 10.1371/journal.pone.0150254.

Naoumkina M, Thyssen GN, Fang DD, Jenkins JN, McCarty JC, Florane CB. 2019. Genetic and transcriptomic dissection of the fiber length trait from a cotton (Gossypium hirsutum L.) MAGIC population. *BMC Genomics* **20(1)**:112 DOI 10.1186/s12864-019-5427-5.

Nishitani K, Tominaga R. 1992. Endo-xyloglucan transferase, a novel class of glycosyl-transferase that catalyzes transfer of a segment of xyloglucan molecule to another xyloglucan molecule. *The Journal of Biological Chemistry* **267(29)**:21058–21064 DOI 10.1016/S0021-9258(19)36797-3.

Ortiz-Marchena MI, Albi T, Lucas-Reina E, Said FE, Romero-Campero FJ, Cano B, Ruiz MT, Romero JM, Valverde F. 2014. Photoperiodic control of carbon distribution during the floral transition in Arabidopsis. *The Plant Cell* **26**(**2**):565–584 DOI 10.1105/tpc.114.122721.

Pang CY, Wang H, Pang Y, Xu C, Jiao Y, Qin YM, Western TL, Yu SX, Zhu YX. 2010. Comparative proteomics indicates that biosynthesis of pectic precursors is important for cotton fiber and Arabidopsis root hair elongation. *Molecular & Cellular Proteomics: MCP* **9**(**9**):2019–2033 DOI 10.1074/mcp.M110.000349.

Patel JD, Huang X, Lin L, Das S, Chandnani R, Khanal S, Adhikari J, Shehzad T, Guo H, Roy-Zokan EM, Rong J, Paterson AH. 2020. The Ligon lintless-2 short fiber mutation is located within a terminal deletion of chromosome 18 in cotton. *Plant Physiology* **183**(**1**):277–288 DOI 10.1104/pp.19.01531.

Paterson AH, Wendel JF, Gundlach H, Guo H, Jenkins J, Jin D, Llewellyn D, Showmaker KC, Shu S, Udall J. 2012. Repeated polyploidization of Gossypium genomes and the evolution of spinnable cotton fibres. *Nature* **492**(**7429**):423–427 DOI 10.1038/nature11798.

Peña MJ, Ryden P, Madson M, Smith AC, Carpita NC. 2004. The galactose residues of xyloglucan are essential to maintain mechanical strength of the primary cell walls in Arabidopsis during growth. *Plant Physiology* **134**(**1**):443–451 DOI 10.1104/pp.103.027508.

Pei G, Chen L, Zhang W. 2017. WGCNA application to proteomic and metabolomic data analysis. *Methods in Enzymology* **585**:135–158 DOI 10.1016/bs.mie.2016.09.016.

Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols* **11**(**9**):1650 DOI 10.1038/nprot.2016.095.

Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology* **33**(**3**):290 DOI 10.1038/nbt.3122.

Qian P, Song W, Yokoo T, Minobe A, Wang G, Ishida T, Sawa S, Chai J, Kakimoto T. 2018. The CLE9/10 secretory peptide regulates stomatal and vascular development through distinct receptors. *Nature Plants* **4**(**12**):1071–1081 DOI 10.1038/s41477-018-0317-4.

Qin YM, Zhu YX. 2011. How cotton fibers elongate: a tale of linear cell-growth mode. *Current Opinion in Plant Biology* **14**(**1**):106–111 DOI 10.1016/j.pbi.2010.09.010.

Rius SP, Casati P, Iglesias AA, Gomez-Casati DF. 2006. Characterization of an Arabidopsis thaliana mutant lacking a cytosolic non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase. *Plant Molecular Biology* **61**(**6**):945–957 DOI 10.1007/s11103-006-0060-5.

Rose JK, Braam J, Fry SC, Nishitani K. 2002. The XTH family of enzymes involved in xyloglucan endotransglucosylation and endohydrolysis: current perspectives and a new unifying nomenclature. *Plant & Cell Physiology* **43**(**12**):1421–1435 DOI 10.1093/pcp/pcf171.

Ruan YL, Llewellyn DJ, Furbank RT. 2003. Suppression of sucrose synthase gene expression represses cotton fiber cell initiation, elongation, and seed development. *The Plant Cell* **15(4)**:952–964 DOI 10.1105/tpc.010108.

Samuel Yang S, Cheung F, Lee JJ, Ha M, Wei NE, Sze SH, Stelly DM, Thaxton P, Triplett B, Town CD, Jeffrey Chen Z. 2006. Accumulation of genome-specific transcripts, transcription factors and phytohormonal regulators during early stages of fiber cell development in allotetraploid cotton. *The Plant Journal: for Cell and Molecular Biology* **47(5)**:761–775 DOI 10.1111/j.1365-313X.2006.02829.x.

Sasidharan R, Chinnappa CC, Staal M, Elzenga JTM, Yokoyama R, Nishitani K, Voesenek LACJ, Pierik R. 2010. Light quality-mediated petiole elongation in Arabidopsis during shade avoidance involves cell wall modification by xyloglucan endotransglucosylase/hydrolases. *Plant Physiology* **154(2)**:978–990 DOI 10.1104/pp.110.162057.

Schiller M, Massalski C, Kurth T, Steinebrunner I. 2012. The Arabidopsis apyrase AtAPY1 is localized in the Golgi instead of the extracellular space. *BMC Plant Biology* **12(1)**:123 DOI 10.1186/1471-2229-12-123.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* **13(11)**:2498–2504 DOI 10.1101/gr.1239303.

Sharma SS, Yamamoto K, Hamaji K, Ohnishi M, Anegawa A, Sharma S, Thakur S, Kumar V, Uemura T, Nakano A. 2017. Cadmium-induced changes in vacuolar aspects of Arabidopsis thaliana. *Plant Physiology and Biochemistry* **114**:29–37 DOI 10.1016/j.plaphy.2017.02.017.

Shimizu Y, Aotsuka S, Hasegawa O, Kawada T, Sakuno T, Sakai F, Hayashi T. 1997. Changes in levels of mRNAs for cell wall-related enzymes in growing cotton fiber cells. *Plant and Cell Physiology* **38(3)**:375–378 DOI 10.1093/oxfordjournals.pcp.a029178.

Smirnova J, Fernie AR, Spahn CM, Steup M. 2017. Photometric assay of maltose and maltose-forming enzyme activity by using 4-alpha-glucanotransferase (DPE2) from higher plants. *Analytical Biochemistry* **532**:72–82 DOI 10.1016/j.ab.2017.05.026.

Song C, Li W, Pei X, Liu Y, Ren Z, He K, Zhang F, Sun K, Zhou X, Ma X, Yang D. 2019. Dissection of the genetic variation and candidate genes of lint percentage by a genome-wide association study in upland cotton. *Theoretical and Applied Genetics* **132(7)**:1991–2002 DOI 10.1007/s00122-019-03333-0.

Sorrell DA, Marchbank AM, Chrimes DA, Dickinson JR, Rogers HJ, Francis D, Grierson CS, Halford NG. 2003. The Arabidopsis $14-3-3$ protein, $GF14\,\omega$, binds to the Schizosaccharomyces pombe Cdc25 phosphatase and rescues checkpoint defects in the $rad24-$ mutant. *Planta* **218(1)**:50–57 DOI 10.1007/s00425-003-1083-7.

Stamm P, Kumar PP. 2013. Auxin and gibberellin responsive Arabidopsis SMALL AUXIN UP RNA36 regulates hypocotyl elongation in the light. *Plant Cell Reports* **32(6)**:759–769 DOI 10.1007/s00299-013-1406-5.

Sugiyama Y, Wakazaki M, Toyooka K, Fukuda H, Oda Y. 2017. A novel plasma membrane-anchored protein regulates xylem cell-wall deposition through microtubule-dependent lateral inhibition of rho GTPase domains. *Current Biology: CB* **27**(**16**):2522–2528 DOI 10.1016/j.cub.2017.06.059.

Sun Z, Wang X, Liu Z, Gu Q, Zhang Y, Li Z, Ke H, Yang J, Wu J, Wu L. 2017. Genome-wide association study discovered genetic variation and candidate genes of fibre quality traits in Gossypium hirsutum L. *Plant Biotechnology Journal* **15**(**8**):982–996 DOI 10.1111/pbi.12693.

Tamura K, Adachi Y, Chiba K, Oguchi K, Takahashi H. 2002. Identification of Ku70 and Ku80 homologues in Arabidopsis thaliana: evidence for a role in the repair of DNA double-strand breaks. *The Plant Journal: For Cell and Molecular Biology* **29**(**6**):771–781 DOI 10.1046/j.1365-313x.2002.01258.

Ubeda-Tomas S, Edvardsson E, Eland C, Singh SK, Zadik D, Aspeborg H, Gorzsàs A, Teeri TT, Sundberg B, Persson P, Bennett M, Marchant A. 2007. Genomic-assisted identification of genes involved in secondary growth in Arabidopsis utilising transcript profiling of poplar wood-forming tissues. *Physiologia Plantarum* **129**:415–428 DOI 10.1111/j.1399-3054.2006.00817.x.

Valverde F, Ortega JM, Losada M, Serrano A. 2005. Sugar-mediated transcriptional regulation of the Gap gene system and concerted photosystem II functional modulation in the microalga Scenedesmus vacuolatus. *Planta* **221**(**6**):937–952 DOI 10.1007/s00425-005-1501-0.

Vanholme R, De Meester B, Ralph J, Boerjan W. 2019. Lignin biosynthesis and its integration into metabolism. *Current Opinion in Biotechnology* **56**:230–239 DOI 10.1016/j.copbio.2019.02.018.

Bolaños Villegas P, Guo C-L, Jauh G-Y. 2015. Arabidopsis Qc-SNARE genes BET11 and BET12 are required for fertility and pollen tube elongation. *Botanical Studies* **56**(**1**):21 DOI 10.1186/s40529-015-0102-x.

Wang M, Tu L, Yuan D, Zhu D, Shen C, Li J, Liu F, Pei L, Wang P, Zhao G, Ye Z, Huang H, Yan F, Ma Y, Zhang L, Liu M, You J, Yang Y, Liu Z, Huang F, Li B, Qiu P, Zhang Q, Zhu L, Jin S, Yang X, Min L, Li G, Chen LL, Zheng H, Lindsey K, Lin Z, Udall JA, Zhang X. 2019. Reference genome sequences of two cultivated allotetraploid cottons, Gossypium hirsutum and Gossypium barbadense. *Nature Genetics* **51**(**2**):224–229 DOI 10.1038/s41588-018-0282-x.

Wang MY, Zhao PM, Cheng HQ, Han LB, Wu XM, Gao P, Wang HY, Yang CL, Zhong NQ, Zuo JR, Xia GX. 2013. The cotton transcription factor TCP14 functions in auxin-mediated epidermal cell differentiation and elongation. *Plant Physiology* **162**(**3**):1669–1680 DOI 10.1104/pp.113.215673.

Wang X, Zhang Y, Zhang J, Cheng C, Guo X. 2007. Molecular characterization of a transient expression gene encoding for 1-aminocyclopropane-1-carboxylate synthase in cotton (Gossypium hirsutum L). *BMB Reports* **40**(**5**):791–800 DOI 10.5483/BMBRep.2007.40.5.791.

Wen F, Zhu Y, Hawes MC. 1999. Effect of pectin methylesterase gene expression on pea root development. *Plant Cell* **11**(**6**):1129–1140 DOI 10.1105/tpc.11.6.1129.

Wilmoth JC, Wang S, Tiwari SB, Joshi AD, Hagen G, Guilfoyle TJ, Alonso JM, Ecker JR, Reed JW. 2005. NPH4/ARF7 and ARF19 promote leaf expansion and auxin-induced lateral root formation. *The Plant Journal* **43(1)**:118–130 DOI 10.1111/j.1365-313X.2005.02432.x.

Wu J, Mao X, Cai T, Luo J, Wei L. 2006. KOBAS server: a web-based platform for automated annotation and pathway identification. *Nucleic Acids Research* **34(suppl_2)**:W720–W724 DOI 10.1093/nar/gkl167.

Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, Kong L, Gao G, Li C-Y, Wei L. 2011. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Research* **39(suppl_2)**:W316–W322 DOI 10.1093/nar/gkr483.

Xu L, Zhu L, Tu L, Liu L, Yuan D, Jin L, Long L, Zhang X. 2011. Lignin metabolism has a central role in the resistance of cotton to the wilt fungus Verticillium dahliae as revealed by RNA-Seq-dependent transcriptional analysis and histochemistry. *Journal of Experimental Botany* **62(15)**:5607–5621 DOI 10.1093/jxb/err245.

Yang Z, Ge X, Yang Z, Qin W, Sun G, Wang Z, Li Z, Liu J, Wu J, Wang Y. 2019. Extensive intraspecific gene order and gene structural variations in upland cotton cultivars. *Nature Communications* **10(1)**:1–13 DOI 10.1038/s41467-018-07882-8.

Yin J, Zhang X, Zhang G, Wen Y, Liang G, Chen X. 2019. Aminocyclopropane-1-carboxylic acid is a key regulator of guard mother cell terminal division in Arabidopsis thaliana. *Journal of Experimental Botany* **70(3)**:897–908 DOI 10.1093/jxb/ery413.

Yoo M-J, Wendel JF. 2014. Comparative evolutionary and developmental dynamics of the cotton (Gossypium hirsutum) fiber transcriptome. *PLOS Genetics* **10(1)**:e1004073 DOI 10.1371/journal.pgen.1004073.

Yu R, Wang Y, Xu L, Zhu X, Zhang W, Wang R, Gong Y, Limera C, Liu L. 2015. Transcriptome profiling of root microRNAs reveals novel insights into taproot thickening in radish (Raphanus sativus L). *BMC Plant Biology* **15**:30 DOI 10.1186/s12870-015-0427-3.

Yuan Y, Teng Q, Zhong R, Ye Z-H. 2016. Roles of Arabidopsis TBL34 and TBL35 in xylan acetylation and plant growth. *Plant Science* **243**:120–130 DOI 10.1016/j.plantsci.2015.12.007.

Zhang F, Zhu G, Du L, Shang X, Cheng C, Yang B, Hu Y, Cai C. 2016a. Genetic regulation of salt stress tolerance revealed by RNA-Seq in cotton diploid wild species, Gossypium davidsonii. *Scientific Reports* **6**:20582 DOI 10.1038/srep20582.

Zhang M, Zeng J-Y, Long H, Xiao Y-H, Yan X-Y. 2016b. Auxin regulates cotton fiber initiation via GhPIN-mediated auxin transport. *Plant and Cell Physiology* **58(2)**:385–397 DOI 10.1093/pcp/pcw203.

Zhang M, Zheng X, Song S, Zeng Q, Hou L, Li D, Zhao J, Wei Y, Li X, Luo M, Xiao Y, Luo X, Zhang J, Xiang C, Pei Y. 2011. Spatiotemporal manipulation of auxin biosynthesis in cotton ovule epidermal cells enhances fiber yield and quality. *Nat Biotechnol* **29(5)**:453–458 DOI 10.1038/nbt.1843.

Zhang T, Hu Y, Jiang W, Fang L, Guan X, Chen J, Zhang J, Saski CA, Scheffler BE, Stelly DM. 2015. Sequencing of allotetraploid cotton (Gossypium hirsutum L. acc.

TM-1) provides a resource for fiber improvement. *Nature Biotechnology* **33(5)**:531 DOI 10.1038/nbt.3207.

Zhang W, Zhang H, Liu K, Jian G, Qi F, Si N. 2017. Large-scale identification of Gossypium hirsutum genes associated with Verticillium dahliae by comparative transcriptomic and reverse genetics analysis. *PLOS ONE* **12(8)**:e0181609 DOI 10.1371/journal.pone.0181609.

Zhang Z, Li J, Jamshed M, Shi Y, Liu A, Gong J, Wang S, Zhang J, Sun F, Jia F, Ge Q, Fan L, Zhang Z, Pan J, Fan S, Wang Y, Lu Q, Liu R, Deng X, Zou X, Jiang X, Liu P, Li P, Iqbal MS, Zhang C, Zou J, Chen H, Tian Q, Jia X, Wang B, Ai N, Feng G, Wang Y, Hong M, Li S, Lian W, Wu B, Hua J, Zhang C, Huang J, Xu A, Shang H, Gong W, Yuan Y. 2020. Genome-wide quantitative trait loci reveal the genetic basis of cotton fiber quality and yield-related traits in aG. hirsutumrecombinant inbred linepopulation. *Plant Biotechnology Journal* **18(1)**:239–253 DOI 10.1111/pbi.13191.

Zhou M, Zhang K, Sun Z, Yan M, Chen C, Zhang X, Tang Y, Wu Y. 2017. LNK1 and LNK2 corepressors interact with the MYB3 transcription factor in phenylpropanoid biosynthesis. *Plant Physiology* **174(3)**:1348–1358 DOI 10.1104/pp.17.00160.

Zhu D, Wu Z, Cao G, Li J, Wei J, Tsuge T, Gu H, Aoyama T, Qu L-J. 2014. Translucent green, an ERF family transcription factor, controls water balance in Arabidopsis by activating the expression of aquaporin genes. *Molecular Plant* **7(4)**:601–615 DOI 10.1093/mp/sst152.

Zou X, Gong J, Duan L, Jiang X, Zhen Z, Fan S, Ge Q, Liu A, Gong W, Li J, Shi Y, Wang Y, Fan L, Liu R, Lei K, Zhang Q, Shang H, Yuan Y. 2018. High-density genetic map construction and QTL mapping for fiber strength on Chr24 across multiple environments in a CCRI70 recombinant inbred lines population. *Euphytica* **214(6)**:102 DOI 10.1007/s10681-018-2177-4.

Zou X, Liu A, Zhang Z, Ge Q, Fan S, Gong W, Li J, Gong J, Shi Y, Tian B, Wang Y, Liu R, Lei K, Zhang Q, Jiang X, Feng Y, Zhang S, Jia T, Zhang L, Yuan Y, Shang H. 2019. Co-expression network analysis and hub gene selection for high-quality fiber in Upland cotton (Gossypium hirsutum) using RNA sequencing analysis. *Gene* **10(2)**:119 DOI 10.3390/genes10020119.