

RESEARCH

Open Access



RNA-Seq with a novel glabrous-ZM24*fl* reveals some key lncRNAs and the associated targets in fiber initiation of cotton

Xianyan Zou^{1,2}, Faiza Ali¹, Shuangxia Jin², Fuguang Li^{1,3*} and Zhi Wang^{1,3*}

Abstract

Background: Cotton fiber is an important natural resource for textile industry and an excellent model for cell biology study. Application of glabrous mutant cotton and high-throughput sequencing facilitates the identification of key genes and pathways for fiber development and cell differentiation and elongation. lncRNA is a type of ncRNA with more than 200 nt in length and functions in the ways of chromatin modification, transcriptional and post-transcriptional modification, and so on. However, the detailed lncRNA and associated mechanisms for fiber initiation are still unclear in cotton.

Results: In this study, we used a novel glabrous mutant ZM24*fl*, which is endowed with higher somatic embryogenesis, and functions as an ideal receptor for cotton genetic transformation. Combined with the high-throughput sequencing, fatty acid pathway and some transcription factors such as MYB, ERF and bHLH families were identified the important roles in fiber initiation; furthermore, 3,288 lncRNAs were identified, and some differentially expressed lncRNAs were also analyzed. From the comparisons of ZM24_0 DPA vs ZM24_-2 DPA and *fl*_0 DPA vs ZM24_0 DPA, one common lncRNA MSTRG 2723.1 was found that function upstream of fatty acid metabolism, MBY25-mediated pathway, and pectin metabolism to regulate fiber initiation. In addition, other lncRNAs MSTRG 3390.1, MSTRG 48719.1, and MSTRG 31176.1 were also showed potential important roles in fiber development; and the co-expression analysis between lncRNAs and targets showed the distinct models of different lncRNAs and complicated interaction between lncRNAs in fiber development of cotton.

Conclusions: From the above results, a key lncRNA MSTRG 2723.1 was identified that might mediate some key genes transcription of fatty acid metabolism, MYB25-mediated pathway, and pectin metabolism to regulate fiber initiation of ZM24 cultivar. Co-expression analysis implied that some other important lncRNAs (e.g., MSTRG 3390.1, MSTRG 48719.1, and MSTRG 31176.1) were also showed the different regulatory model and interaction between them, which proposes some valuable clues for the lncRNAs associated mechanisms in fiber development.

Keywords: Glabrous cotton, RNA-Seq, Fiber initiation, Fiber development, lncRNA, Transcription factor

Introduction

Cotton (*Gossypium* spp.) is one of the most important cash crops in the world because its main product fiber is the important natural source for the textile industry. In the four cultivars of *Gossypium* genus (*G. hirsutum*, *G. barbadense*, *G. arboreum*, and *G. raimondii*), *G. hirsutum* (upland cotton) is the most widely planted due to its high yields and adaptability [42]. The period of cotton

*Correspondence: aylifug@caas.cn; wangzhi01@caas.cn

³ Zhengzhou Research Base, State Key Laboratory of Cotton Biology, School of Agricultural Sciences, Zhengzhou University, Zhengzhou 450001, China

Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

fiber development has been classified into four stages: initiation, elongation, secondary cell wall deposition, and maturity of fiber [12]. The first two stages could determine the number and length of fibers, further affecting fiber yields. Consequently, many studies have been documented to explore the underlying genetic mechanisms related to fiber initiation and elongation, contributing to cotton production improvement [16, 17, 19, 29, 65].

Cotton mutants with fibreless, fuzzless, and lintless are good materials for studying the mechanism of fiber initiation development. With the auxin and gibberellin (GA) application in two fibreless mutants of Asian cotton *in vitro* culture, it showed that fiber cells differentiated from ovule epidermis at a temperature lower than 30 degrees, but not above 32 degrees, which indicated the important roles of auxin and GA in fiber development promotion at some specific conditions [2]. SNPs comparison obtained by RNA-Seqs showed that glabrous mutant Xu142*fl* may be the progeny of *G. barbadense*. Based on the F₂ and BC₁ population between TM-1 and Xu142*fl*, the *Li*₃ gene encoding an MYB-MIXTA-like transcription factor was mapped and adjacent to *MYB25-like* in the D12 chromosome [60]. The inheritance evaluation of fuzzless seed in segregation population suggested that the interaction of three loci (*N*₁, *n*₂ and *n*₃) contributed to fuzzless seed [48], among which two loci, *N*₁ and *n*₂, located on a pair of homologous chromosomes A12/D12 [6]. The plants of *N*₁*N*₁ homozygous and *N*₁*n*₁ heterozygous produced fuzzless seeds [48]. The *n*₃ locus that could produce the fibreless seed was identified by genetic analysis of cross progeny between *N*₁*N*₁ and *n*₂*n*₂ [48]. The fourth locus, named *n*₄^t*n*₄^t, was identified from ethyl methanesulfonate (EMS) induced mutation analysis, whose homozygous seed exhibited a partially naked phenotype [3]. All these fiber development defect mutants provide suitable materials for fiber development study.

With the advantage of Next Generation Sequencing (NGS), RNA-Seq as one of the NGS has been widely used to reveal expressions of genes and transcripts, among which some transcripts have been identified as non-coding RNA (ncRNA) because of their limitation of coding proteins. ncRNA includes microRNAs (miRNAs), long non-coding RNAs (lncRNAs), and so on, which have emerged as key regulators of gene expression through their direct and indirect actions on chromatin [23–25]. In *Oryza sativa*, 1,254 differentially expressed lncRNAs (DELs) were identified from BIL progenies [26]. Another RNA-Seq showed that 328 of 444 DELs were associated with meiosis and the low fertility in autotetraploid rice [27]. The lncRNAs were also involved in abiotic stress such as drought and re-watering in *Brassica napus* [46], and osmotic and salt stress in *Medicago truncatula* [53]. The differences in

genes expressions and regulations between fibreless mutants and wild-type have been investigated using omics methods [14, 28, 45, 51]. With fibreless mutant Xu142*fl* and its counterpart Xu142, a previous comparative small RNAome analysis uncovered a possible network of fiber initiation-related miRNAs in cotton ovules, which comprises seven miRNAs expressed in cotton ovules, and each of them bears functional specific targets [51]. Another work showed that 54 miRNAs are differentially expressed in fiber initiation between Xu142*fl* and its wild-type, which are potentially targeted to TFs such as MYB, auxin response factor, and Leucine repeat receptor [45]. Using multi-omics, the differentially expressed genes (1,953), proteins (187), and phosphoproteins (131) were identified by the comparison of Xu142 and Xu142*fl* [28]. Genetic markers including 302 SNPs for fiber development were also developed and validated based on a deep sequencing between Xu142 and Xu142*fl* [28]. In particular, a transcriptomic repertoire revealed that 645 and 651 lncRNAs were preferentially expressed in Xu142*fl* and Xu142, respectively. Further study showed that down-regulating two lncRNAs XLOC_545639 and XLOC_039050 in Xu142 *fl* increased the fiber initials on the ovules, while silencing XLOC_079089 in Xu142 shortened the fiber length [14], indicating the important and diverse roles of lncRNAs in fiber development.

lncRNA is a type of ncRNA with more than 200 nt in length and without protein-coding abilities [4, 70]. Lots of evidence have shown that lncRNAs could regulate genes functions in the ways of chromatin modification, transcriptional and post-transcriptional modification, etc. [32], through which the lncRNAs play vital roles in plant growth, development [67], and response to biotic [68] and abiotic stresses [24, 43, 62]. lncRNAs could be classified into long intergenic non-coding RNAs (lincRNAs), natural antisense non-coding RNAs (lncNAT), sense non-coding lncRNAs, and intronic lncRNAs according to their location in the genome [5, 31, 55]. In cotton fiber development, the detailed regulation mechanism of lncRNA is still ambiguous. In this research, we introduced a natural lintless-fuzzless (ZM24*fl*) mutant from zhongmiansuo24 (ZM24) cultivar, which is easy for transformation [8, 59]. Further, strand-specific transcriptome sequencing was conducted to reveal the differential expression profile of genes in the fiber initiation periods including lint and fuzz initiation stages. The lncRNAs/PC-genes pairs are identified and analyzed to explore the potential key lncRNAs and the corresponding targets contributing to the fiber development differences between the ZM24*fl* and ZM24.

Results

ZM24 (*G. hirsutum*) is a high-quality commercial cotton cultivar bred by the Chinese Cotton Research Institute; this cultivar exhibits excellent somatic embryogenesis potential, which makes it an ideal receptor for cotton genetic transformation [59]. We identified a spontaneous mutation in ZM24 that resulted in the production of fuzzless and lintless seed and designated it as ZM24 fuzzless-lintless (ZM24*fl*), which would be indicated as *fl* in the following study.

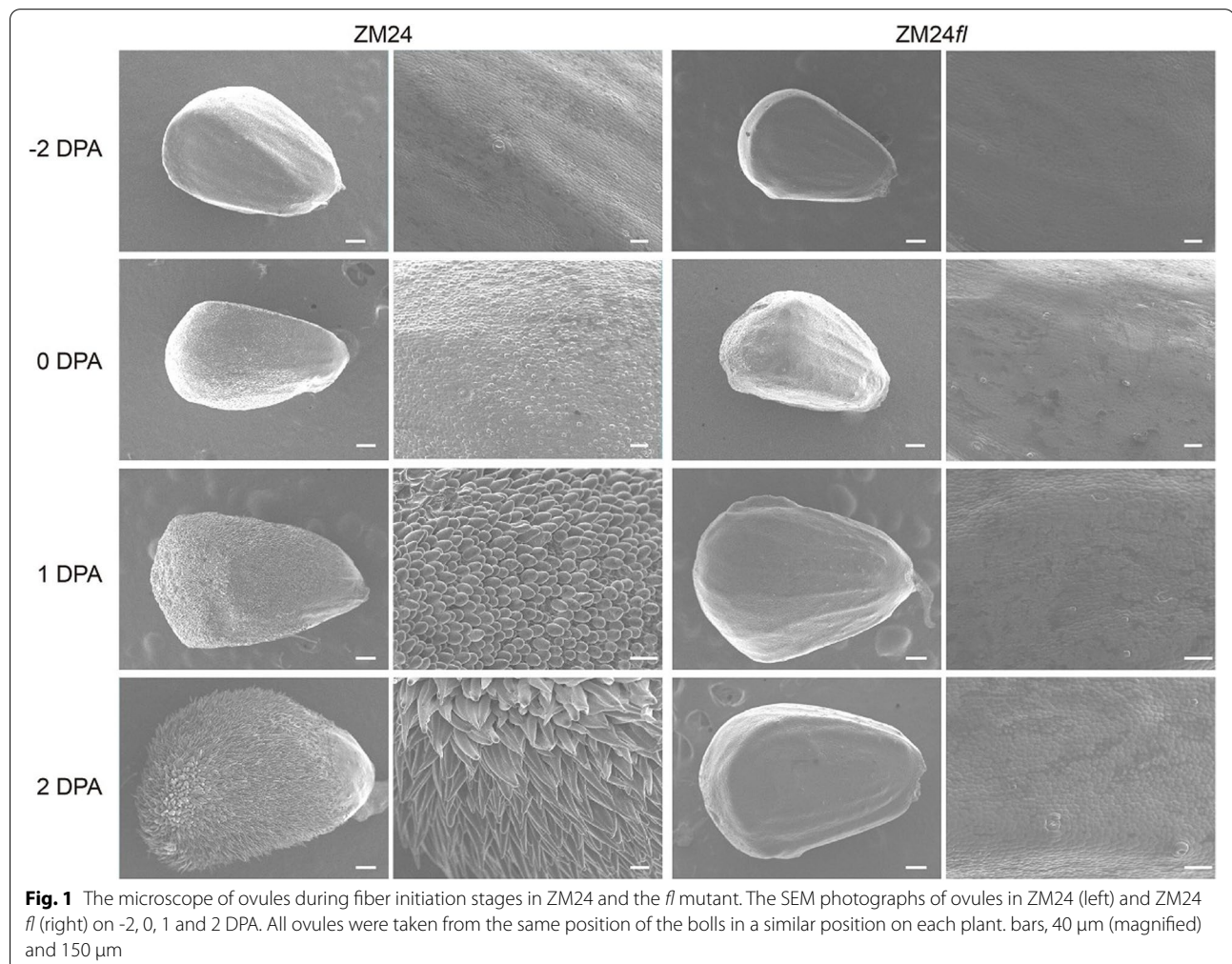
The difference of the fiber cell development between *fl* mutant and ZM24

Firstly, we observed the vegetative development of *fl* and ZM24; the results showed that the development of the organs and tissues such as branches, leaves, bolls, and epidermal hair of leaves, and stems is normal in *fl*, and similar with ZM24, except that the boll size is smaller in *fl* due to fiber development defects (Figure S1).

To reveal the difference of fiber development between *fl* and ZM24 during fiber initiation stages, ovules at -2 day-post-anthesis (DPA), 0 DPA, 1 DPA and 2 DPA from two lines were stripped for observation using scanning electron microscopy. The scanning results showed that there were no significant differences on the epidermis of -2 DPA ovules in both lines (Fig. 1). The epidermis of ovules from 0 to 2 DPA, which indicated the fiber initiation stage, presents the clear differences of fibrous protuberance between ZM24 and *fl*. The fibrous protuberance appeared on the epidermis of 0 DPA ovules in ZM24, and they slowly elongated over time. However, no fiber was observed on the epidermis of ovules of 0 DPA, 1 DPA, and 2 DPA from *fl* (Fig. 1), which indicated that *fl* is an excellent line for fiber cell initiation study.

Identification and characterization of lncRNAs in cotton fiber initiation

lncRNAs have been shown the important roles in fiber development [14]. To investigate the potent lncRNAs



responsible for the fiber initiation, we performed strand-specific RNA sequencing on 6 ovule samples (-2 DPA, 0 DPA, and 5 DPA from ZM24 and *fl*) with twice replicates. A total of 362.86 Gb clean reads were obtained by removing low-quality reads, and all Q30 of RNA-Seq results are greater than 85%, indicating that clean reads were qualified for downstream analysis (Additional file 1). To assess the correlations between biological replicates, the Pearson correlation was calculated using FPKM assays and the r^2 between two replicates of each material was above 0.88, indicating the high correlation and reliable data (Figure S2). An integrated pipeline (Fig. 2a) was used to identify lncRNA in these tissues (see details in materials and methods). Finally, 3,288 lncRNA transcripts were obtained, among which the lincRNAs (2,618), lncNATs (559), sense lncRNAs (78) and intronic lncRNAs (33) accounted for 79.6 %, 17 %, 2.4 %, and 1 %, respectively (Fig. 2b). The pattern of exons in lncRNA transcripts showed that single-exon lncRNAs represented the largest proportion, accounting for 43.8% (Fig. 2c). The physical positions and the class codes of lncRNAs were detailed in Additional file 2.

Differentially expressed lncRNAs and protein-encoding genes between *fl* and ZM24

Furthermore, the clean reads were mapped to ZM24 genomes, and FPKM and counts of transcripts and genes were identified for further analysis with a threshold of $|\log_2(\text{FC})| > 1$ and p -value < 0.05 . In summary, a total of 371 differentially expressed lncRNAs (DELs) and 12,971 DEGs were identified in vertical (*fl* vs ZM24 at the same developmental stages) and horizontal (different development stages in the same line) comparisons, respectively. The FPKM values of the DELs and DEGs were shown in Additional files 3 and 4.

In detail, 87 (32 up-regulated, 55 down-regulated), 27 (5 up-regulated, 22 down-regulated), and 201 (133 up-regulated, 68 down-regulated) lncRNAs were differentially expressed in the -2, 0, and 5 DPA ovules of *fl* compared with that in ZM24, respectively (Fig. 2d), which indicated that many down-regulated lncRNAs in the -2 and 0 DPA ovules of *fl* genes play potential roles in fiber initiation. Correspondingly, there were 452 (347 up- and 105 down-regulated), 389 (25 up- and 364 down-regulated), and 3,986 DEGs (1,378 up- and 2,608 down-regulated) in *fl* compared to ZM24 at the three time

points of fiber initiation stage, respectively (Fig. 2e), these results indicated that the lncRNAs might regulate target genes encoding proteins either negatively or positively.

To further investigate the DEGs between *fl* and ZM24 at different development stages, the DELs and DEGs of “0 DPA vs -2 DPA” and “5 DPA vs 0 DPA” were identified, and the Venn diagram established showed that 21 (10 up- and 11 down-regulated) and 60 (30 up- and 30 down-regulated) DELs were identified in *fl* and ZM24, respectively, in the comparison of 0_DPA vs -2 DPA, of which only 3 common DELs (1 up-regulated, 2 down-regulated) were identified in both comparisons. Similarly, 218 DELs were identified in comparison of 5 DPA vs 0 DPA, in which, 34 and 163 DELs were unique to *fl* and ZM24, respectively. Interestingly, in the other 21 common DELs, 17 and 3 DELs were up- and down-regulated at 5 DPA compared to 0 DPA; only one DELs, MSTRG.3394.1, was down regulated in *fl*_5DPA compared with *fl*_0DPA, but up-regulated in that of ZM24. These results indicated that more lncRNA were involved in the fiber development than that in the ovule development. Consistently, a lot of DEGs were identified from the comparisons (Fig. 2g), suggesting that the dramatic transcription regulation differences between two lines might contribute to the development of the fiber in ZM24. These results proposed that the expression of genes including protein-encoding genes and lncRNA-encoding genes changed dramatically in the two lines, which may contribute to the formation of the glabrous seed of *fl*, however, the underlying mechanism is unclear.

GO enrichment of DEGs

To investigate the functions of DEGs between *fl* and ZM24 in fiber initiation and elongation, the GO enrichment analysis of DEGs was performed to explore the causal pathways or genes responsible for the phenotype differences in two lines. According to the results of GO enrichment, the most significant (corrected p -value < 0.05) classes were identified. On -2 DPA, the biological process of regulation activity and exogenous and endogenous signals such as cold (GO:0009409), fungus (GO:0050832), jasmonic acid (GO:0009753), and chitin (GO:0010200) were mainly enriched in up-regulated genes, whereas the down-regulated DEGs were enriched in the cellular process such as “cytosol” (GO:0005829), “plastid” (GO:0009536), “post-embryonic development” (GO:0009791) in *fl* compared

(See figure on next page.)

Fig. 2 An integrative computational pipeline for the bioinformatics analysis and summary of DELs and DEGs in the RNA-Seq. **a** Bioinformatics methods for the identification of lncRNA. **b** The number of lincRNA, lncNAT, sense lncRNA, and intronic lncRNA. **c** Exon number pattern of four kinds of lncRNA. **d** and **e** Number of differentially expressed lncRNAs and genes (up- and down-regulated) between *fl* and ZM24, respectively, in three time points (-2, 0, and 5 DPA) during fiber initiation. **f** and **g** Venn diagrams showing the common DELs and DEGs between comparisons of 0 DPA vs -2 DPA and 5 DPA vs 0 DPA of *fl* and ZM24, respectively. DELs and DEGs were identified with $|\log_2(\text{fold change})| > 1$, FDR < 0.001

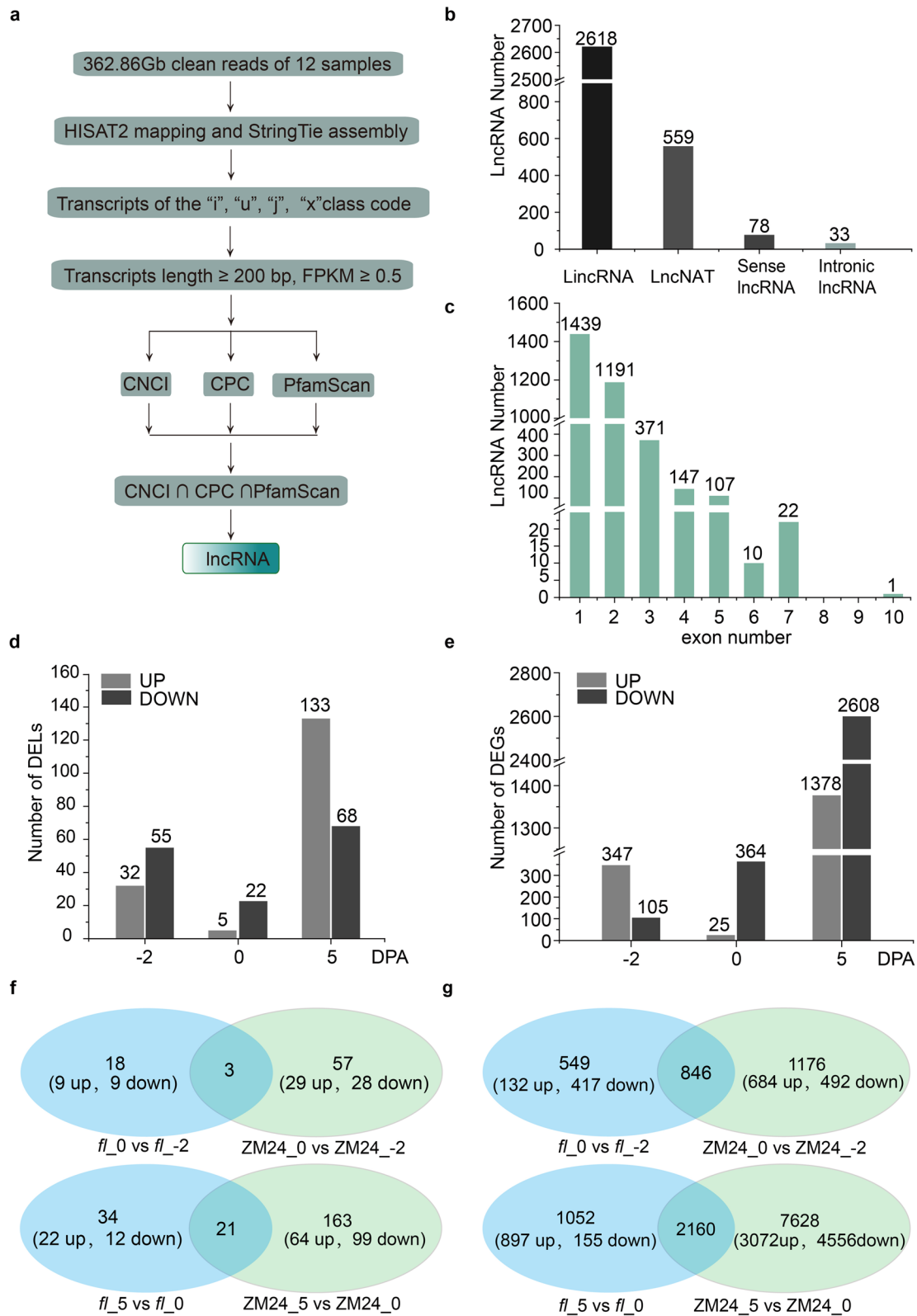
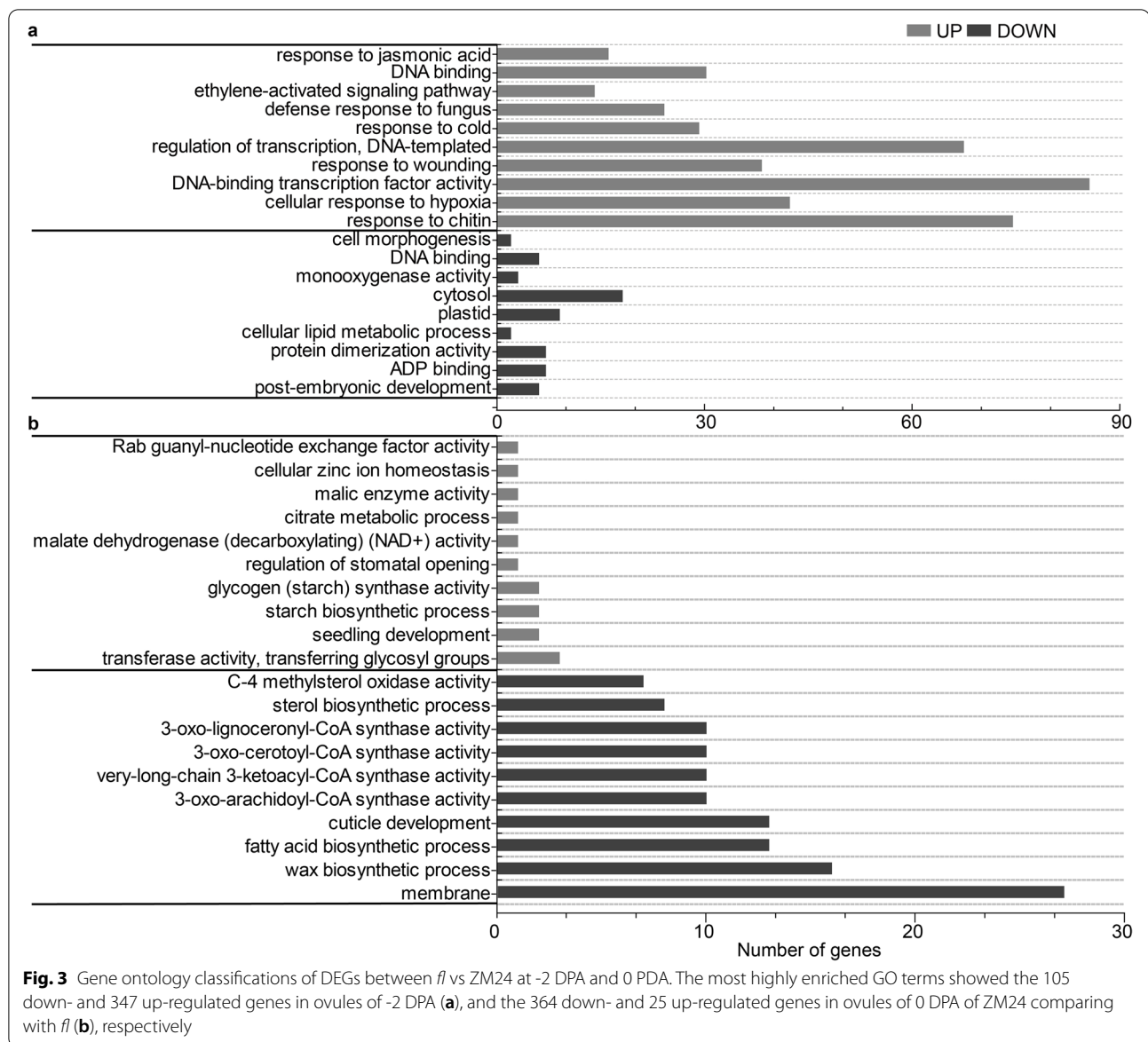


Fig. 2 (See legend on previous page.)



with ZM24 (Fig. 3a). On the day of flowering, the down-regulated genes were enriched in terms of “wax biosynthetic process” (GO:0010025), “cuticle development” (GO:0042335), “3-oxo-lignoceryl-CoA synthase activity” (GO:0102338), “3-oxo-cerotoyl-CoA synthase activity” (GO:0102337), and “fatty acid biosynthetic process” (GO:0006633), which were related to lipid metabolism (Fig. 3b), indicating that cell wall plasticity regulated by lipid metabolism might play important roles in fiber cell initiation. On 5 DPA, the earlier fiber elongation stage, the DEGs of two lines exhibited more significant functions in cellular components (Fig. S3). For example, the down-regulated genes in “*fl_5* vs ZM24_5” comparison was mostly enriched in terms of “plasma membrane” (GO:0005886),

“cytoplasm” (GO:0005737), “integral component of membrane” (GO:0016021), etc., indicating that fiber development requires the expression and regulation of a large number of genes to bring cellular component effect.

Identification of potential genes for fiber initiation

Generally, the phenotype is determined by genotype. Therefore, the reason why the *fl* exhibits a fibreless phenotype is mainly due to the differences in transcriptional regulation that happened during a few days before the flowering. To analyze and obtain genes sets related to fiber initiation, venn diagrams were conducted to show the common target between “ZM24_0 vs *fl_0*” and “ZM24_0 vs ZM24_-2”. Subsequently, one

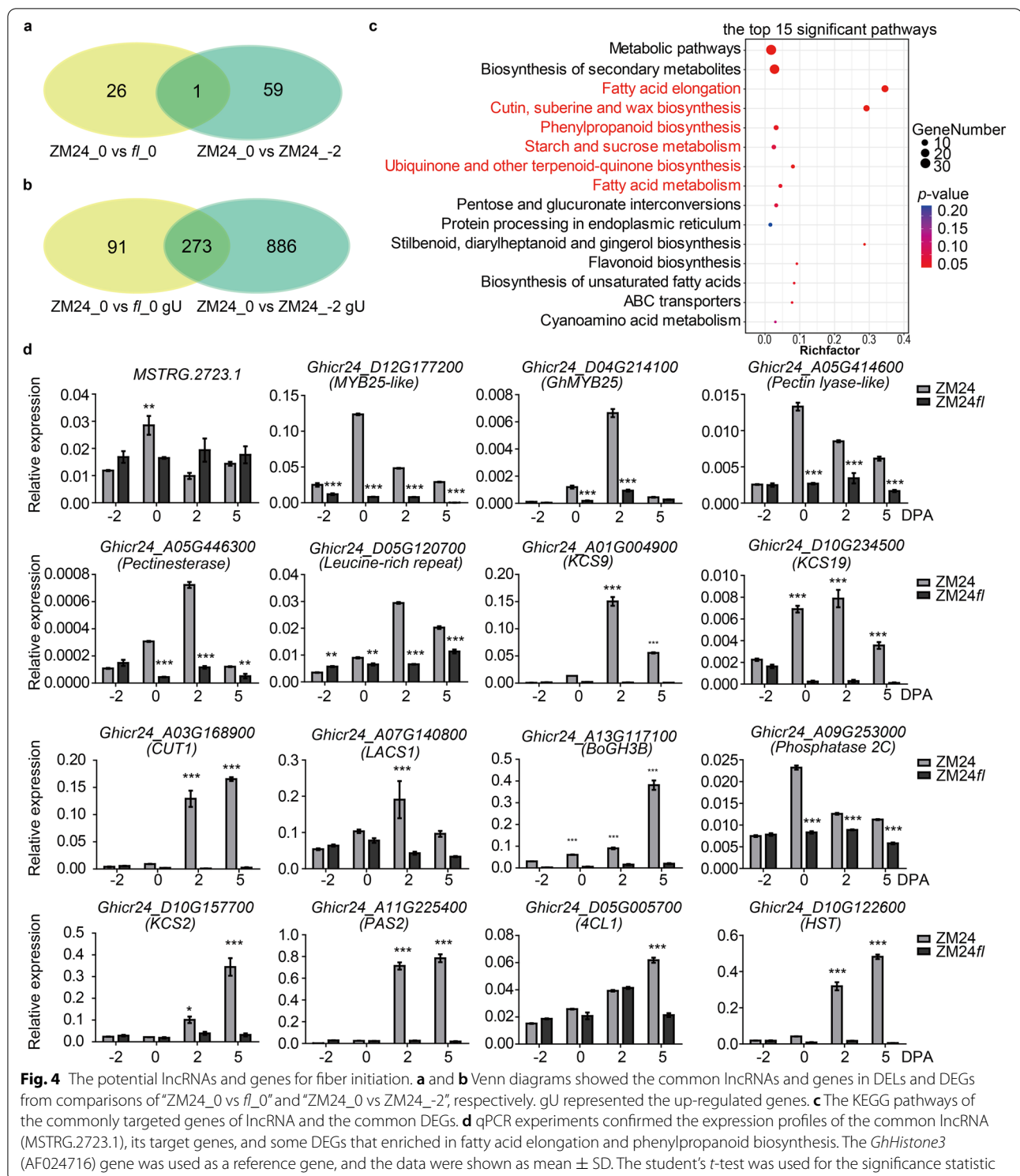


Fig. 4 The potential lncRNAs and genes for fiber initiation. **a** and **b** Venn diagrams showed the common lncRNAs and genes in DELs and DEGs from comparisons of "ZM24_0 vs *fl_0*" and "ZM24_0 vs ZM24_-2", respectively. gU represented the up-regulated genes. **c** The KEGG pathways of the commonly targeted genes of lncRNA and the common DEGs. **d** qPCR experiments confirmed the expression profiles of the common lncRNA (*MSTRG.2723.1*), its target genes, and some DEGs that enriched in fatty acid elongation and phenylpropanoid biosynthesis. The *GhHistone3* (AF024716) gene was used as a reference gene, and the data were shown as mean \pm SD. The student's *t*-test was used for the significance statistic

common DEL, *MSTRG2723.1*, was identified between "ZM24_0 vs *fl_0*" and "ZM24_0 vs ZM24_-2" (Fig. 4a), which showed significantly higher expression in ovules of ZM24_0 DPA compared with *fl_0* DPA and

ZM24_-2 DPA, and targeted to 36 DEGs (Additional files 3 and 5). Moreover, 273 common DEGs between the up-regulated genes in ZM24_0 vs *fl_0* and ZM24_0 vs ZM24_-2 comparisons were identified (Fig. 4b).

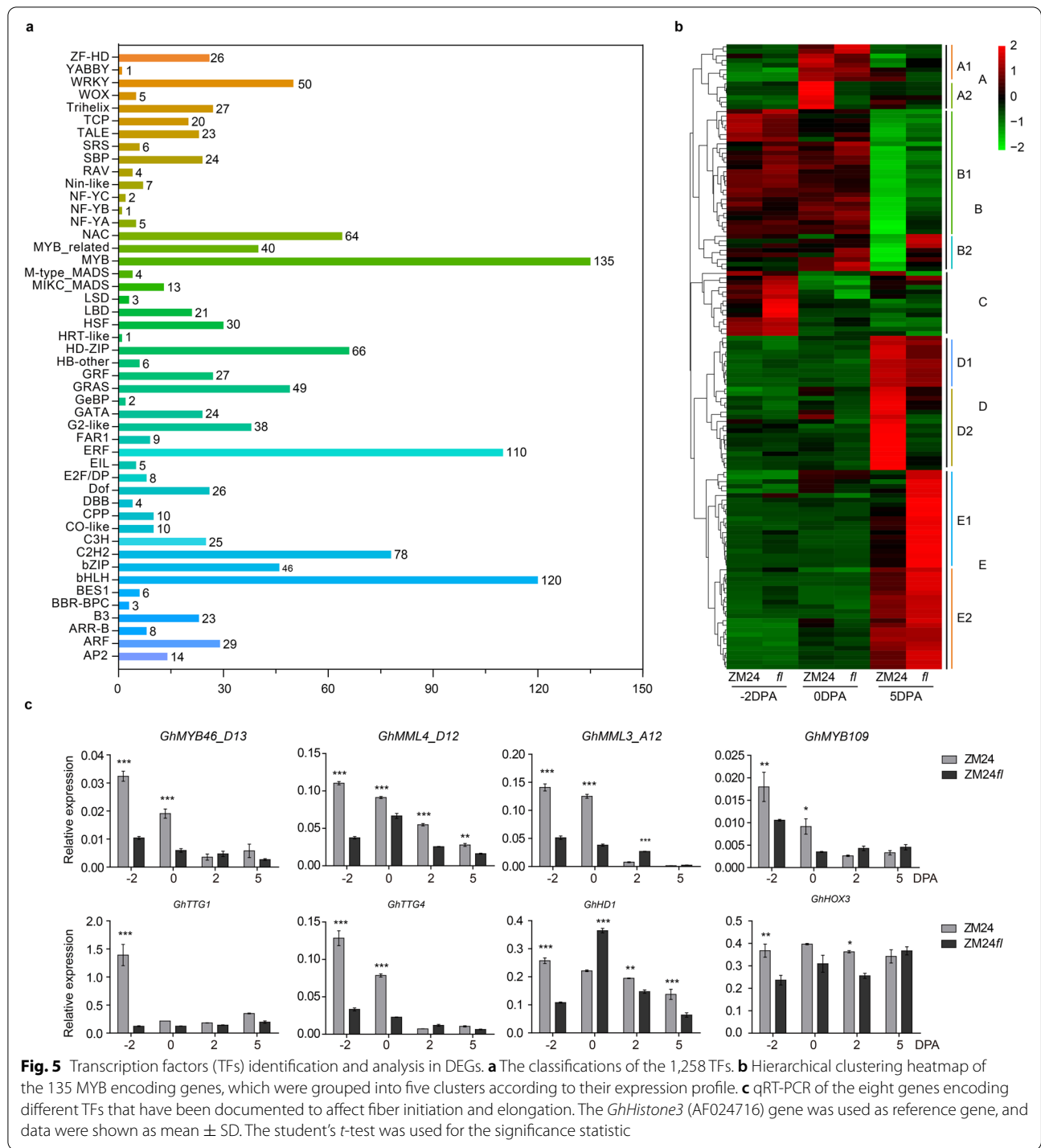
Hence, KEGG enrichment showed that these targeted genes and DEGs were significantly enriched in “fatty acid elongation” (12 genes), “cutin, suberine, and wax biosynthesis” (9 genes), “phenylpropanoid biosynthesis” (5 genes), “starch and sucrose metabolism” (4 genes), “ubiquinone and another terpenoid-quinone biosynthesis” (3 genes) and “fatty acid metabolism” (3 genes) pathways that possibly be related to fiber initiation (Fig. 4c). To validate the RNA-Seq results, the specific primers of the lncRNA MSTRG2723.1, and the randomly selected 11 genes from the targeted genes and DEGs were designed and used for Q-PCR. Relative expression levels of these genes in -2, 0, 2, and 5 DPA ovules were identified and consistent with the RNA-Seq. Of these, the transcription factor coding genes *GhMYB25-like* and *GhMYB25* that were positive to fiber initiation were down-regulated in *fl*. In addition, the genes that were annotated as pectin lyase-like, pectinesterase, and a leucine-rich repeat (LRR) were down-regulated in *fl* compared to ZM24. Furthermore, the encoding 3-ketoacyl-CoA synthase genes *Ghcr24_A01G004900 (KCS9)* and *Ghcr24_D10G234500 (KCS19)*, targeted by MSTRG.2723.1, were significantly down-regulated in *fl* during the initiation stage, and enriched in the fatty acid elongation pathway (Fig. 4d). The down-regulated genes such as *CUT1* and *LONG CHAIN ACYL-COA SYNTHASE 1 (LACS1)* are required for elongation of C24 very-long-chain fatty acid (VLCFA) with a function in wax production [33]. The gene *GhPAS2*, enriched in the fatty acid metabolism pathway, was also characterized as the enzyme that catalyzes VLCFA production [54]. The genes *4-COUMARATE-COA LIGASE 1 (4CL1)* and *O-HYDROXYCINNAMOYLTRANSFERASE (HST)* that were down-regulated in *fl* and enriched in phenylpropanoid biosynthesis pathway, affected the lignin composition [13]. These results indicated that the potential targeted genes involved in VLCFA by lncRNAs may play critical roles in fiber initiation.

Identification of genes encoding transcription factors from the DEGs

Transcription factors (TFs) play vital roles in gene expression regulation and plant development. A considerable number of researches have shown that some TFs play key roles in fiber initiation and elongation [16, 29, 36, 49]. An analysis of TFs from the DEGs set showed that 1,258 genes encode TFs, accounting for 9.7 % (1,258/12,971) DEGs, which were classified into 48 families. The largest group of TFs was the MYB (135) family, followed by the bHLH (120), ERF (110), C2H2 (78), and HD-ZIP (66) families, etc. (Fig. 5a). Of these differentially expressed

TFs, 90 (86 up- and 4 down-regulated), 27 (1 up- and 26 down-regulated), and 279 (179 up- and 99 down-regulated) TFs were identified comparing *fl* with ZM24 at -2, 0, 5 DPA, respectively (Additional file 4 and 6). In horizontal comparison, 210 TFs were differentially expressed in the comparison of “*fl*_-2 vs *fl*_0”, of which the most DE-TFs members were ERF (55) and bHLH (24). Two hundred and two TFs including MYB (27), MYB_related (19), ERFs (18), etc. were identified in 0 DPA compared to -2 DPA in ZM24 ovules. Correspondingly, the number of DE-TFs between 0 vs 5 DPA has reached 397 and 842 in *fl* and ZM24, respectively. In *fl*, the most differentially expressed TFs include MYB (52), NAC (48), HD-ZIP (33), C2H2 (30), bHLH (29), ERF (25), WRKY (23), etc., whereas, the TFs groups of bHLH (83), MYB (82), ERF (52), HD-ZIP (51), C2H2 (50) were the most differentially expressed transcription factor families in ZM24 (Additional file 4).

Given that MYB TFs play key roles in fiber initiation, hierarchical clustering analysis was performed to show the expression profiles of the 135 MYB TFs encoding genes in two lines (Fig. 5b). These MYBs were grouped into 5 main clusters (A to E). The clusters C and A harbored 14 TFs each and genes in them are specifically expressed in -2 and 0 DPA, respectively. Furthermore, 6 genes in A2 were showed higher expression in ZM24 but low expression in *fl* on 0 DPA. Twenty-seven genes clustered in B1 were highly expressed in -2 and 0 DPA, whereas the other 8 MYB TFs in B2 showed high expression in all samples except at 5 DPA of ZM24. In clusters D and E, 11 genes in D1 and 22 genes in E2 were specifically expressed in 5 DPA. In addition, 18 MYBs in group D2 showed highest expression at 5 DPA of ZM24, with the remaining 21 MYBs clustered in E1 showed the highest expression at 5 DPA of *fl*. Several known positive MYB genes that have been reported to regulate fiber initiation [16, 19, 29, 49] were selected for the further qPCR experiment, and most of which showed a down-regulated profile in *fl* (Fig. 5c). In addition, two HD-ZIP TFs, GhHD1, and GhHOX3 that promoted fiber initiation and elongation by mediating accumulation of ethylene and reactive oxygen together [40, 50] were down-regulated in *fl* at -2 DPA, 2 DPA, and 5 DPA. Furthermore, two TRANSPARENT TESTA GLABRA loci (TTG1 and TTG4) that were positive to fiber initiation also showed lower expression in *fl* than that of ZM24 during fiber initiation [18, 56]. These results indicated that different MYBs and other TFs might exert specific roles during the different fiber developmental stages, which help to the identification of distinct TFs and the detailed gene transcription regulation mechanisms underlying the fiber initiation and primary development.



Functions analysis of DELs based on lncRNA-mRNA co-expression network

To further explore the roles of DELs, cis-targeted genes and trans-targeted genes of DELs were identified by calculating PCC values and distances between lncRNAs and mRNAs. Finally, 249 DELs were co-expressed with

5,501 DE_mRNAs. Thereby 70,987 pairs of lncRNA-mRNAs were identified in all the comparisons (Additional file 7). According to the expression profile, the targeted DEGs were summarized as shown in Fig. 6. In ovules of -2, 0, and 5 DPA, 49, 170, and 2,102 targeted mRNAs were down-regulated in *fl* compared with ZM24.

In the comparison of “-2 DPA vs 0 DPA”, 7 (2 up, 5 down) and 50 (42 up, 8 down) DE-mRNAs were identified in *fl* and ZM24, respectively. Similarly, 1,543 (1,517 up, 26 down) and 5,545 (3,863 up, 1,682 down) DE-mRNAs were identified in “*fl_0 vs fl_5*” and “*ZM24_0 vs ZM24_5*”, respectively. Thus, the KEGG enrichment of targeted DE-mRNAs was performed to investigate the functions of DELs. As a result, of the down-regulated targeted mRNAs in *fl* related to ZM24, genes were significantly enriched in fatty acid elongation and lipid metabolism at 0 and 5 DPA. Of the up-regulated targeted mRNAs, DE-mRNAs at 5 DPA related to 0DPA in *fl* were mainly enriched in amino acid metabolism, phenylpropanoid biosynthesis, fatty acid degradation and flavonoid biosynthesis pathways; however, the pathways of fatty acid elongation, lipid metabolism, carbohydrate metabolism, energy metabolism were enriched by DE-mRNAs of that in ZM24 (Additional file 8).

To comprehensively visualize the relationship between lncRNAs and the targeted mRNAs, the co-expression network among DELs, DE-mRNAs, and TFs at 0 DPA were shown using Cytoscape software (Fig. 7). The center lncRNAs MSTRG.2723.1 and MSTRG.3390.1 are more important and independent lncRNAs that regulate many gene transcriptions including some known key TFs coding genes (e.g., *MYB25*, *MYB25-Like*, *bZIP*) in fiber initiation. Other types of lncRNAs such as MSTRG.31176.1

and MSTRG.48719.1 regulate genes transcription reciprocally, in which other different lncRNAs are also involved. Above these imply the diversity and complexity of molecular mechanisms by lncRNAs and provide a possible lncRNA-TFs regulation model in the fiber cell initiation.

Discussion

With the glabrous mutant of cotton and high-throughput technology, generous coding- and non-coding RNAs and associated mechanisms have been identified in the fiber development including TFs, miRNAs and lncRNAs and so on [14, 28, 51, 69]. Many studies on non-coding RNAs in cotton have been limited to small RNAs until now. For instance, a lot of miRNAs specifically expressed during anther development or callus were identified in male sterile cotton as well as cotton somatic embryogenesis [57, 64]. Gong *et al.* revealed the 33 conserved miRNAs families between the A and D genomes [9]. On the genomic level, the expression of 79 miRNAs families was studied and 257 novel miRNAs were identified related to cotton fiber elongation [63]. In addition, two key miR828 and miR858 were proved the roles in the regulation of homoeologous *MYB2* (*GhMYB2A* and *GhMYB2D*) in *G. hirsutum* fiber development [11].

As a kind of long non-coding RNA, lncRNA provides more regulatory mechanisms for gene expression,

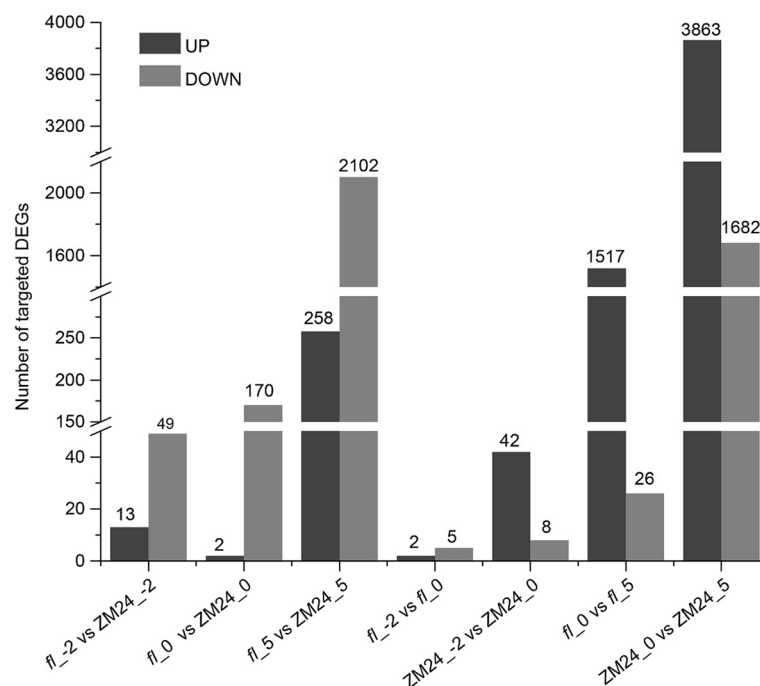
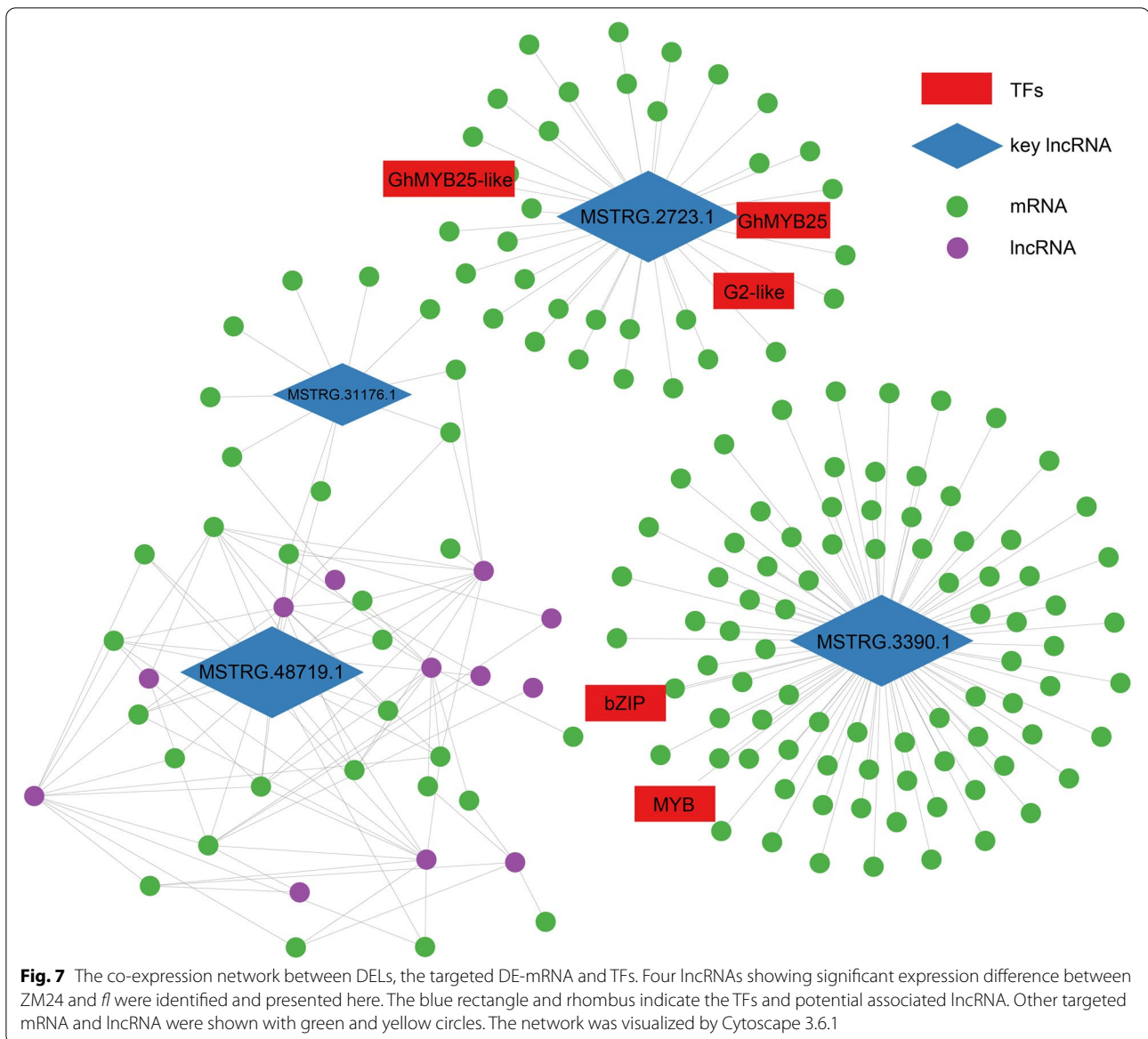


Fig. 6 Number of up- and down-regulated targeted genes of DELs in horizontal and vertical comparisons between *fl* and ZM24. Seven comparisons were established and the differentially expressed lncRNAs were identified and summarized



protein synthesis, chromatin remodeling etc., while it is not clear about the detailed lncRNAs and the underlying mechanism in fiber development. A previous study identified 30,550 lncRNAs loci and 4,718 lncNATs loci, which are rich in repetitive sequences and preferentially expressed in a tissue-specific manner with weak evolutionary conservation. Further, lncRNAs showed overall higher methylation levels, and their expression was less affected by gene body methylation [52]. Using the epidermal cells from the ovules at 0 and 5 DPA from Xu142 and Xu142*fl*, 35,802 lncRNAs and 2,262 circular RNAs (circRNAs) were identified, of which 645 lncRNAs were preferentially expressed in the fibreless mutant Xu142*fl* and 651 lncRNAs were preferentially expressed in the

fiber-attached lines; three lncRNAs XLOC_545639, XLOC_039050, and XLOC_079089 all showed the solid function in fiber development by VIGS assay [14]. Here, a novel glabrous mutant-ZM24*fl*, which showed excellent somatic embryogenesis induction was used to identify the key lncRNA involved in fiber initiation development [59].

Totally, 3,288 lncRNA transcripts were identified from the -2 DPA, 0 DPA and 5 DPA ovules of ZM24 and *fl*, which is significantly different from the number of identified lncRNA in Xu142*fl* [14] and *G. barbadense* L. cv 3-79 [52]. To identify the causal lncRNAs for fiber initiation, some comparisons were built to analyze the differentially expressed genes including lncRNAs and mRNAs

during fiber initiation and earlier elongation. The identified DELs and DEGs in comparisons of 0 DPA vs -2 DPA and 5 DPA vs 0 DPA of ZM24 and *fl* indicated that many lncRNAs and coding genes are involved in the fiber initiation and primary development, while few lncRNAs and coding genes may involve the ovule development. The analysis of the DEGs further showed that fatty acid metabolism, very long strain fatty acid synthesis and sugar metabolism play important roles in the fiber initiation of ZM24, supporting the previous results [15, 39]. Moreover, some MYB family, bHLH type TFs encoding genes were also identified the important roles in fiber initiation, which is in agreement with the function of these TFs in previous research [10, 16, 29, 36, 40, 49, 50]. To uncover the upstream factors such as lncRNAs, we focused on the comparisons of ZM24_0 DPA vs *fl*_0 DPA and ZM24_0 DPA vs ZM24_-2 DPA to find the common lncRNAs which should be a key regulator for fiber initiation. Consequently, one lncRNA MSTRG 2723.1 was obtained, which locates on the A02G (84218766–84219942) encoding a lncNAT and covering the most coding region and partial 3'-terminal untranslated region of *Ghircr24_A02G147600* (Figure S4). The co-expression analysis further identified its potent targets including 3-ketoacyl-CoA synthase, MYB family proteins, phosphatase 2C family proteins, pectin lysase, and some uncharacterized proteins, which may be involved in fiber initiation through fatty acid pathway, cell wall plasticity, MYB-mediated signaling etc. These results provide important clues for the upstream regulatory lncRNAs in fiber initiation and novel information associated with the fiber development regulation network. In addition, MSTRG 3390.1, MSTRG 48719.1, and MSTRG 31176.1 were also identified some positive correlation between fiber development and ovule development. The sequence analysis indicated that these lncRNAs are different from the previous lncRNAs XLOC_545639, XLOC_039050, and XLOC_079089 [14]. The target analysis also implied the possible interaction between different lncRNAs through mediating the common targets, which provide novel clues to explore the regulatory lncRNAs and underlying mechanisms in fiber development. Even with some achievement of lncRNAs, the understanding of the underlying mechanism of lncRNAs regulating targets or chromosome remodeling still needs more work to disclose.

Conclusion

Here, a novel glabrous cotton mutant ZM24*fl* was identified and applied to study the potential lncRNAs for fiber development with high-throughput sequencing. ZM24*fl* is derived from an elite cultivar of ZM24, which possesses high callus induction and somatic embryogenesis ability,

and is endowed with the valuable receptor for cotton genetic transformation [59]. Through the RNA-Seq and analysis in different ovules of ZM24 and *fl*, 3,288 lncRNAs were identified and some differentially expressed lncRNAs responsible for fiber (lint and fuzz) initiation and fiber earlier elongation were shown. Collectively, four lncRNAs MSTRG.2723.1, MSTRG.3390.1, MSTRG.48719.1 and MSTRG.31176.1 were shown potential important roles in fiber development, and the analysis of the target implied that MSTRG 2723.1 may function upstream of fatty acid metabolism, MBY25-mediated pathway, and pectin metabolism to regulate fiber initiation; the co-expression analysis between lncRNAs and targets further indicated the distinct models of different lncRNAs and interaction between lncRNAs, which provide precious information for illumination of the molecular mechanism of lncRNAs in fiber development of cotton.

Materials and methods

Plant Materials

Gossypium hirsutum L. acc. Zhongmiansuo24 (ZM24) and a natural fuzzless-lintless (*fl*) mutant from ZM24 were used and grown under standard field conditions in the Institute of Cotton Research of the Chinese Academy of Agricultural Sciences (Zhengzhou research base, Henan). The ovule tissues were collected from cotton bolls on -2, 0, and 5 DPA using a sterile knife. All materials were frozen in liquid nitrogen immediately and stored at -80 °C for the following experiments.

Microscopic observation of fiber initiation on ovules epidermis

To study the fiber initiation phenotypes of ZM24 and *fl*, the cotton bolls of two lines on -2, 0, 1 and 2 DPA were collected. Then, the ovules were stripped from the bolls in the middle region. Immediate Scanning electron microscopy (Hitachi) was performed to observe the ovule epidermis as described previously [14].

Strand specific libraries construction and sequencing

Total RNAs of each ovule sample was extracted using the RNeasy Pure Plant Kit (Qiagen, Beijing, China) following the manufacturer's instruction. Total RNAs of each sample was quantified and qualified by Agilent 2100 Bio-analyzer (Agilent Technologies, Palo Alto, CA, USA), Nanodrop 2000 (Thermo Fisher Scientific Inc.), and 1% agarose gel. RNA with RIN value above 7 was used for following library construction. The rRNA was removed using the Ribo-Zero™ rRNA removal Kit. The ribosomal depleted RNA was then used for sequencing library preparation according to the manufacturer's protocol (NEBNext® Ultra™ Directional RNA Library

Prep Kit for Illumina®). The cDNA libraries with different indices were multiplexed and loaded on an Illumina HiSeq2500 with 150 base pair (bp) paired-end (PE150) raw reads according to the manufacturer's instruction (Illumina, San Diego, CA, USA). RNA-Seq raw data with accession number SRP285346 was uploaded in the NCBI sequence read archive (<http://www.ncbi.nlm.nih.gov/sra/>) and the accession numbers of the twenty-four runs are SRR12710181-SRR12710192, and SRR12718970-SRR12718981.

Mapping to the reference genome and LncRNAs identification

The raw data in fastq format were filtered with cutadapt (v1.9.1) software [30]. Clean data were obtained by removing reads that contained adapter, poly-N and base with Phred quality < 20 in 3' or 5' end, and the reads of length < 75 bp were removed after filtering. Finally, the GC percentage and Q30 of each sample were calculated using FastQC software (<https://www.babraham.ac.uk/>) and shown in Table S1. Clean data were mapped to the ZM24 genome (<https://github.com/gitmalm/Genome-data-of-Gossypium-hirsutum/>) [66] using HISAT(v2.1.0) [20, 21] software with the parameter "--rna-strandness RF". Transcriptomes of each sample were assembled based on mapped reads and were merged by StringTie software (v2.0) [34, 35]. Transcripts annotation was performed using Cuffcompare [47]. Long non-coding RNA was identified as following steps: 1) transcripts with class codes of "l", "u", "x", "j" representing the intronic transcripts, long intergenic noncoding RNAs (lincRNAs), long noncoding natural antisense transcripts (lncNAT), and the sense transcripts, respectively, were selected. 2) Transcripts with length > 200 bp, coverage > 1, FPKM > 0.5; 3) The CNCI [44], CPC [22] and Pfam-Scan software were used to assess protein-coding ability [7], with the parameter of (CPC score < 0, CNCI score < 0).

Differential expression analysis

The FPKM values and counts of genes and lncRNAs in each sample were calculated using StringTie and Ballgown [35]. Differential expression analyses were conducted by edgeR in R package [37, 38]. The DEGs and DELs were identified with an expression FPKM > 1.0, FDR (false discovery rate < 0.001), and $|\log_2(\text{fold change value})| \geq 1$ between each pairwise comparison.

Co-expression analysis between lncRNA and mRNA

To unveil the potential functions of DELs between the two genotypes, two interaction models of lncRNAs and protein-coding genes (lncRNAs/PC-genes) including cis- and trans-target were analyzed: 1) the Pearson

correlation coefficient (PCC) between differentially expressed lncRNAs and mRNAs were calculated using the OmicShare tools (<https://www.omicshare.com/>) with the expression profiles (FPKM). The lncRNA-mRNA pairs with $|PCC| > 0.95$ and $p\text{-value} < 0.01$ were regarded as trans interaction between lncRNAs and mRNAs. 2) Protein-coding genes with a distance less than 20 kb from the upstream or downstream of lncRNAs were putative cis interaction. The co-expression networks were visualized by Cytoscape 3.6.1 [41].

GO and KEGG

To explore the functions of DEGs and lncRNAs between ZM24 and *fl*, the gene ontology (GO) enrichment was performed using the BLASTP program [1] and GO databases (<http://archive.geneontology.org/latest-lite/>) and (<http://ftp.ncbi.nlm.nih.gov/gene/DATA/>). Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis was performed at KOBAS 3.0 website [58, 61] (<http://kobas.cbi.pku.edu.cn/kobas3>).

Q-PCR analysis

Ovules from bolls at -2, 0, 2, and 5 DPA were collected, and then total RNAs were extracted using the RNAprep Pure Plant Kit (Polysaccharides & Polyphenolics-rich, Tiangen, Beijing, China) following the manufacturer's instruction. Each reverse-transcribed reaction was performed with 1 µg RNA using a transScript® First-Strand cDNA Synthesis SuperMix (AT301-02, TransGen). The real-time PCR was performed on Roche 480 PCR system with a SYBR-Green Real-time PCR SuperMix (AQ101-01, TransGen). The 20 µL reaction volumes in each well contain 1 µL cDNA, 8.2 µL sterile water, 10 µL Mix, and 0.4 µL each of the forward and reverse primers. The Q-PCR procedures were as: pre-incubation of 30 s at 95 °C; followed by denaturation at 95 °C for 10 s, primer annealing at 55 °C for 10 s, and then extension at 72 °C for 30 s; finally, a melting curve at 95 °C for 30 s to check the primer specificity. The *GhHistone3* (AF024716) gene was used as a reference gene. The $2^{-\Delta Ct}$ method was used to calculate the relative expression of each gene, with three technical repetitions and three biological repetitions. Data were shown as mean ± SD. The student's *t*-test was used for the significance statistic. The primer sequences used in the presented study are listed in Additional file 9.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-022-03444-9>.

Additional file 1. RNA-Seq data for 12 samples

Additional file 2. List of lncRNAs identified from ovules of two lines during fiber initiation stage.

Additional file 3. Differentially expressed lncRNAs in different comparisons.

Additional file 4. Differentially expressed genes in different comparisons.

Additional file 5. The targeted genes by lncRNA MSTRG.2723.1.

Additional file 6. Transcription factors identification in the 12,971 DEGs.

Additional file 7. Differentially expressed lncRNA and their targets in all comparisons.

Additional file 8. The KEGG pathways of differentially expressed targeted genes of DELs.

Additional file 9. List of primers used in this research.

Additional file 10: Figure S1. Observation and comparison of *fl* and ZM24 in the different developmental stages and tissues. **a** and **f** The plants architecture of *fl* and ZM24; **b** and **c** The magnifications of white rectangular dotted bolls at 20 DPA from **a** and **f**, respectively. **d** The size comparison of bolls at 15 DPA from *fl* (left) and ZM24 (right). **e** The size and shape of leaves from two lines *fl* (left) and ZM24 (right). **g** and **i** The epidermal hair on the abaxial leaf surface of *fl* and ZM24; **h** The stem epidermal hair of *fl* (up) and ZM24 (down); bars in **a-f**: 2.0 cm; bars in **g-i**: 1000 μ m.

Additional file 11: Figure S2. Heatmap shows the Pearson correlation coefficients among the 12 samples.

Additional file 12: Figure S3. Gene ontology classifications of DEGs in ovules of ZM24 vs *fl* at 5 PDA. The most highly enriched GO terms showed the 1,378 down- and 2,608 up-regulated genes in ovules of ZM24 vs *fl* at 5 DPA.

Additional file 13: Figure S4. The physical location of lncRNA MSTRG.2723.1 on the ZM24 genome. The MSTRG.2723.1 is a natural antisense transcript and overlaps with the gene of *Ghcr24_A02G147600*. Blue and orange rectangles represent exons and introns, respectively. Arrows indicate the direction of transcription.

Acknowledgments

We thank Yong Cheng and Xin Li (Zhengzhou Research Base, Institute of Cotton Research of CAAS) for technical assistance.

Authors' contributions

Zhi Wang designed the study, Xianyan Zou and Faiza Ali wrote the main manuscript text and prepared all figures. Shuangxia Jin and Fuguang Li edited the manuscript. All authors read and approved the final manuscript.

Funding

This study was supported financially by the National Natural Science Foundation of China (32072022 and 31690093), the Creative Research Groups of China (31621005), and Central Public-interest Scientific Institution Basal Research Fund (1610162020010202) for scientific research into non-profit industries.

Availability of data and materials

All the related data and files are presented including the sequences of the primers used in the Q-PCR. RNA-Seq raw data with accession number SRP285346 was uploaded in the NCBI sequence read archive (<http://www.ncbi.nlm.nih.gov/sra/>) and can be accessible with bioproject archive number PRJNA665585 (<http://www.ncbi.nlm.nih.gov/bioproject/>).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹State Key Laboratory of Cotton Biology, Key Laboratory of Biological and Genetic Breeding of Cotton, Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang 455000, China. ²National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China. ³Zhengzhou Research Base, State Key Laboratory of Cotton Biology, School of Agricultural Sciences, Zhengzhou University, Zhengzhou 450001, China.

Received: 15 June 2021 Accepted: 24 January 2022

Published online: 03 February 2022

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Molecular biology*. 1990;215(3):403–10. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
- Beasley CA. Temperature-dependent Response to Indoleacetic Acid Is Altered by NH (4) in Cultured Cotton Ovules. *Plant physiol*. 1977;59(2):203–6. <https://doi.org/10.1104/pp.59.2.203>.
- Bechere E, Turley RB, Auld DL, Zeng LH. A New Fuzzless Seed Locus in an Upland Cotton (*Gossypium hirsutum* L.) Mutant. *Am J Plant Sci*. 2012;3(6):799–804. <https://doi.org/10.4236/ajps.2012.3.6096>.
- Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes & development*. 2011;25(18):1915–27. <https://doi.org/10.1101/gad.17446611>.
- Chekanova JA. Long non-coding RNAs and their functions in plants. *Curr Opin Plant Biol*. 2015;27:207–16. <https://doi.org/10.1016/j.pbi.2015.08.003>.
- Endrizzi JE, Ramsay G. Identification of ten chromosome deficiencies of cotton. *J Heredity*. 1980;1:1. <https://doi.org/10.1093/oxfordjournals.jhered.a109309>.
- Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*. 2016;44(D1):D279–85. <https://doi.org/10.1093/nar/gkv1344>.
- Ge X, Zhang C, Wang Q, Yang Z, Wang Y, Zhang X, et al. iTRAQ protein profile differential analysis between somatic globular and cotyledonary embryos reveals stress, hormone, and respiration involved in increasing plantlet regeneration of *Gossypium hirsutum* L. *J Proteome Res*. 2015;14(1):268–78. <https://doi.org/10.1021/pr500688g>.
- Gong L, Kakrana A, Arikiti S, Meyers BC, Wendel JF. Composition and expression of conserved microRNA genes in diploid cotton (*Gossypium*) species. *Genome Biol Evol*. 2013;5(12):2449–59. <https://doi.org/10.1093/gbe/evt196>.
- Guan XY, Li QJ, Shan CM, Wang S, Mao YB, Wang LJ, et al. The HD-Zip IV gene *GaHOX1* from cotton is a functional homologue of the *Arabidopsis* *GLABRA2*. *Physiologia plantarum*. 2008;134(1):174–82. <https://doi.org/10.1111/j.1399-3054.2008.01115.x>.
- Guan X, Pang M, Nah G, Shi X, Ye W, Stelly DM, et al. miR828 and miR858 regulate homoeologous MYB2 gene functions in *Arabidopsis* trichome and cotton fibre development. *Nat Comm*. 2014;5:3050. <https://doi.org/10.1038/ncomms4050>.
- Haigler CH, Betancur L, Stiff MR, 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>.
- Hoffmann L, Besseau S, Geoffroy P, Ritzenthaler C, Meyer D, Lapierre C, et al. Silencing of hydroxycinnamoyl-coenzyme A shikimate/quinate hydroxycinnamoyltransferase affects phenylpropanoid biosynthesis. *Plant Cell*. 2004;16(6):1446–65. <https://doi.org/10.1105/tpc.020297>.
- Hu H, Wang M, Ding Y, Zhu S, Zhao G, Tu L, et al. Transcriptomic repertoires depict the initiation of lint and fuzz fibres in cotton (*Gossypium hirsutum* L.). *Plant Biotechnol J*. 2018;16(5):1002–12. <https://doi.org/10.1111/pbi.12844>.
- Hu W, Chen L, Qiu X, Wei J, Shen G. AKR2A participates in the regulation of cotton fiber development by modulating biosynthesis of very long chain fatty acids. *Plant Biotechnol J*. 2019;18(2). <https://doi.org/10.1111/pbi.13221>.
- Huang Y, Liu X, Tang K, Zuo K. Functional analysis of the seed coat-specific gene *GbMYB2* from cotton. *Plant Physiol Biochem*. 2013;73:16–22. <https://doi.org/10.1016/j.plaphy.2013.08.004>.

17. Huang G, Huang JQ, Chen XY, Zhu YX. Recent Advances and Future Perspectives in Cotton Research. *Annu Rev Plant Biol.* 2021. <https://doi.org/10.1146/annurev-arplant-080720-113241>.
18. Humphries JA, Walker AR, Timmis JN, Orford SJ. Two WD-repeat genes from cotton are functional homologues of the *Arabidopsis thaliana* TRANSPARENT TESTA GLABRA1 (TTG1) gene. *Plant molecular biology.* 2005;57(1):67–81. <https://doi.org/10.1007/s11103-004-6768-1>.
19. Suo J, Liang X, Pu L, Zhang Y, Xue Y. Identification of GhMYB109 encoding a R2R3 MYB transcription factor that expressed specifically in fiber initials and elongating fibers of cotton (*Gossypium hirsutum* L.). *Biochimica et Biophysica Acta (BBA). Gene Structure and Expression.* 2003;1630(1):25–34. ISSN 0167-4781. <https://doi.org/10.1016/j.bbaexp.2003.08.009>.
20. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods.* 2015;12(4):357–60. <https://doi.org/10.1038/nmeth.3317>.
21. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology.* 2019;37(8):907–15. <https://doi.org/10.1038/s41587-019-0201-4>.
22. Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, et al. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic acids research.* 2007;35(Web Server issue):W345–9. <https://doi.org/10.1093/nar/gkm391>.
23. Kornienko AE, Guenzl PM, Barlow DP, Pauler FM. Gene regulation by the act of long non-coding RNA transcription. *BMC biology.* 2013;11:59. <https://doi.org/10.1186/1741-7007-11-59>.
24. Kowalczyk MS, Higgs DR, Gingeras TR. RNA discrimination. *Nature.* 2012;482(7385):310–1. <https://doi.org/10.1038/482310a>.
25. Kung JT, Colognori D, Lee JT. Long noncoding RNAs: past, present, and future. *Genetics.* 2013;193(3):651–69. <https://doi.org/10.1534/genetics.112.146704>.
26. Li M, Cao A, Wang R, Li Z, Li S, Wang J. Genome-wide identification and integrated analysis of lincRNAs in rice backcross introgression lines (BC (2) F (1,2)). *BMC plant biology.* 2020a;20(1):300. <https://doi.org/10.1186/s12870-020-02508-y>.
27. Li X, Shahid MQ, Wen M, Chen S, Yu H, Jiao Y, et al. Global identification and analysis revealed differentially expressed lincRNAs associated with meiosis and low fertility in autotetraploid rice. *BMC plant biology.* 2020b;20(1):82. <https://doi.org/10.1186/s12870-020-2290-0>.
28. Ma Q, Wu M, Pei W, Wang X, Zhai H, Wang W, et al. RNA-Seq-Mediated Transcriptome Analysis of a Fiberless Mutant Cotton and Its Possible Origin Based on SNP Markers. *PloS one.* 2016;11(3):e0151994. <https://doi.org/10.1371/journal.pone.0151994>.
29. Machado A, Wu Y, Yang Y, Llewellyn DJ, 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>.
30. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.* 2011;17:10–2. <https://doi.org/10.14806/ej.17.1.200>.
31. Mattick JS, Rinn JL. Discovery and annotation of long noncoding RNAs. *Nat Struct Mol Biol.* 2015;22(1):5–7. <https://doi.org/10.1038/nsmb.2942>.
32. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nat Rev Gen.* 2009;10(3):155–9. <https://doi.org/10.1038/nrg2521>.
33. Millar AA, Clemens S, Zachgo S, Giblin EM, Taylor DC, Kunst L. CUT1, an *Arabidopsis* gene required for cuticular wax biosynthesis and pollen fertility, encodes a very-long-chain fatty acid condensing enzyme. *Plant Cell.* 1999;11(5):825–38. <https://doi.org/10.1105/tpc.11.5.825>.
34. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol.* 2015;33(3):290–5. <https://doi.org/10.1038/nbt.3122>.
35. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT StringTie and Ballgown. *Nat Protoc.* 2016;11(9):1650–67. <https://doi.org/10.1038/nprot.2016.095>.
36. Pu L, Li Q, Fan X, Yang W, Xue Y. The R2R3 MYB transcription factor GhMYB109 is required for cotton fiber development. *Genetics.* 2008;180(2):811–20. <https://doi.org/10.1534/genetics.108.093070>.
37. Robinson MD, Smyth GK. Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics.* 2007;23(21):2881–7. <https://doi.org/10.1093/bioinformatics/btm453>.
38. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010;26(1):139–40. <https://doi.org/10.1093/bioinformatics/btp616>.
39. Ruan YL, Llewellyn DJ, Furbank RT, Chourey PS. The delayed initiation and slow elongation of fuzz-like short fibre cells in relation to altered patterns of sucrose synthase expression and plasmodesmata gating in a lintless mutant of cotton. *J Exp Bot.* 2005;56(413):977–84. <https://doi.org/10.1093/jxb/eri091>.
40. Shan C-M, Shangguan X-X, Zhao B, Zhang X-F, Chao L-M, Yang C-Q, et al. Control of cotton fibre elongation by a homeodomain transcription factor GhHOX3. *Nat Comm.* 2014;5:5519. <https://doi.org/10.1038/ncomm56519>.
41. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* 2003;13(11):2498–504. <https://doi.org/10.1101/gr.1239303>.
42. Shao Q, Zhang F, Tang S, Liu Y, Fang X, Liu D, et al. Identifying QTL for fiber quality traits with three upland cotton (*Gossypium hirsutum* L.) populations. *Euphytica.* 2014;198(1):43–58. <https://doi.org/10.1007/s10681-014-1082-8>.
43. Shuai P, Liang D, Tang S, Zhang Z, Ye CY, Su Y, et al. Genome-wide identification and functional prediction of novel and drought-responsive lincRNAs in *Populus trichocarpa*. *J Ex Bot.* 2014;65(17):4975–83. <https://doi.org/10.1093/jxb/eru256>.
44. Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C, et al. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic acids research.* 2013;41(17):e166. <https://doi.org/10.1093/nar/gkt646>.
45. Sun R, Li C, Zhang J, Li F, Ma L, Tan Y, et al. Differential expression of microRNAs during fiber development between fuzzless-lintless mutant and its wild-type allotetraploid cotton. *Scientific reports.* 2017;7(1):3. <https://doi.org/10.1038/s41598-017-00038-6>.
46. Tan X, Li S, Hu L, Zhang C. Genome-wide analysis of long non-coding RNAs (lncRNAs) in two contrasting rapeseed (*Brassica napus* L.) genotypes subjected to drought stress and re-watering. *BMC plant biology.* 2020;20(1):81. <https://doi.org/10.1186/s12870-020-2286-9>.
47. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol.* 2010;28(5):511–5. <https://doi.org/10.1038/nbt.1621>.
48. Turley RB, Kloth RH. Identification of a third fuzzless seed locus in upland cotton (*Gossypium hirsutum* L.). *J Heredity.* 2002;93(5):359–64. <https://doi.org/10.1093/jhered/93.5.359>.
49. Walford SA, Wu Y, Llewellyn DJ, Dennis ES. GhMYB25-like: a key factor in early cotton fibre development. *Plant J.* 2011;65(5):785–97. <https://doi.org/10.1111/j.1365-313X.2010.04464.x>.
50. Walford SA, Wu Y, Llewellyn DJ, Dennis ES. Epidermal cell differentiation in cotton mediated by the homeodomain leucine zipper gene, GhHD-1. *Plant J.* 2012;71(3):464–78. <https://doi.org/10.1111/j.1365-313X.2012.05003.x>.
51. Wang ZM, Xue W, Dong CJ, Jin LG, Bian SM, Wang C, et al. A comparative miRNAome analysis reveals seven fiber initiation-related and 36 novel miRNAs in developing cotton ovules. *Molecular plant.* 2012;5(4):889–900. <https://doi.org/10.1093/mp/ssr094>.
52. Wang M, Yuan D, Tu L, Gao W, He Y, Hu H, et al. Long noncoding RNAs and their proposed functions in fibre development of cotton (*Gossypium* spp.). *New phytologist.* 2015a;207(4):1181–97. <https://doi.org/10.1111/nph.13429>.
53. Wang TZ, Liu M, Zhao MG, Chen R, Zhang WH. Identification and characterization of long non-coding RNAs involved in osmotic and salt stress in *Medicago truncatula* using genome-wide high-throughput sequencing. *BMC plant biology.* 2015b;15:131. <https://doi.org/10.1186/s12870-015-0530-5>.
54. Wang XC, Li Q, Jin X, Xiao GH, Liu GJ, Liu NJ, et al. Quantitative proteomics and transcriptomics reveal key metabolic processes associated with cotton fiber initiation. *J Proteomics.* 2015c;114:16–27. <https://doi.org/10.1016/j.jprot.2014.10.022>.
55. Wang M, Zhao W, Gao L, Zhao L. Genome-wide profiling of long non-coding RNAs from tomato and a comparison with mRNAs associated with the regulation of fruit ripening. *BMC plant biology.* 2018;18(1):75. <https://doi.org/10.1186/s12870-018-1300-y>.

56. Wang Z, Yang Z, Li F. Updates on molecular mechanisms in the development of branched trichome in *Arabidopsis* and nonbranched in cotton. *Plant Biotechnol J*. 2019;17(9). <https://doi.org/10.1111/pbi.13167>.
57. Wei M, Wei H, Wu M, Song M, Zhang J, Yu J, et al. Comparative expression profiling of miRNA during anther development in genetic male sterile and wild type cotton. *BMC plant biology*. 2013;13:66. <https://doi.org/10.1186/1471-2229-13-66>.
58. Wu J, Mao X, Cai T, Luo J, Wei L. KOBAS server: a web-based platform for automated annotation and pathway identification. *Nucleic acids research*. 2006;34(Web Server issue):W720–4. <https://doi.org/10.1093/nar/gkl167>.
59. Wu X, Li F, Zhang C, Liu C, Zhang X. Differential gene expression of cotton cultivar CCRI24 during somatic embryogenesis. *J Plant Physiol*. 2009;166(12):1275–83. <https://doi.org/10.1016/j.jplph.2009.01.012>.
60. Wu H, Tian Y, Wan Q, Fang L, Guan X, Chen J, et al. Genetics and evolution of MIXTA genes regulating cotton lint fiber development. *New Phytologist*. 2018;217(2):883–95. <https://doi.org/10.1111/nph.14844>.
61. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, et al. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res*. 2011;39(Web Server issue):W316–22. <https://doi.org/10.1093/nar/gkr483>.
62. Xin M, Wang Y, Yao Y, Song N, Hu Z, Qin D, et al. Identification and characterization of wheat long non-protein coding RNAs responsive to powdery mildew infection and heat stress by using microarray analysis and SBS sequencing. *BMC plant biol*. 2011;11:61. <https://doi.org/10.1186/1471-2229-11-61>.
63. Xue W, Wang Z, Du M, Liu Y, Liu JY. Genome-wide analysis of small RNAs reveals eight fiber elongation-related and 257 novel microRNAs in elongating cotton fiber cells. *BMC genomics*. 2013;14:629. <https://doi.org/10.1186/1471-2164-14-629>.
64. Yang J, Liu X, Xu B, Zhao N, Yang X, Zhang M. Identification of miRNAs and their targets using high-throughput sequencing and degradome analysis in cytoplasmic male-sterile and its maintainer fertile lines of *Brassica juncea*. *BMC genomics*. 2013;14:9. <https://doi.org/10.1186/1471-2164-14-9>.
65. Yang Z, Zhang C, Yang X, Liu K, Wu Z, Zhang X, et al. PAG1, a cotton brassinosteroid catabolism gene, modulates fiber elongation. *New Phytologist*. 2014;203(2):437–48. <https://doi.org/10.1111/nph.12824>.
66. Yang Z, Ge X, Yang Z, Qin W, Sun G, Wang Z, et al. Extensive intraspecific gene order and gene structural variations in upland cotton cultivars. *Nat Comm*. 2019;10(1):2989. <https://doi.org/10.1038/s41467-019-10820-x>.
67. Zhang YC, Chen YQ. Long noncoding RNAs: new regulators in plant development. *Biochem Biophys Res Commun*. 2013;436(2):111–4. <https://doi.org/10.1016/j.bbrc.2013.05.086>.
68. Zhang L, Wang M, Li N, Wang H, Qiu P, Pei L, et al. Long noncoding RNAs involve in resistance to *Verticillium dahliae*, a fungal disease in cotton. *Plant Biotechnol J*. 2018;16(6):1172–85. <https://doi.org/10.1111/pbi.12861>.
69. Zhao T, Xu X, Wang M, Li C, Li C, Zhao R, et al. Identification and profiling of upland cotton microRNAs at fiber initiation stage under exogenous IAA application. *BMC genomics*. 2019;20(1):421. <https://doi.org/10.1186/s12864-019-5760-8>.
70. Zhou ZY, Li AM, Adeola AC, Liu YH, Irwin DM, Xie HB, et al. Genome-wide identification of long intergenic noncoding RNA genes and their potential association with domestication in pigs. *Genome Biol Evol*. 2014;6(6):1387–92. <https://doi.org/10.1093/gbe/evu113>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

