


RESEARCH ARTICLE

Open Access



# Long non-coding RNAs and their potential functions in Ligon-lintless-1 mutant cotton during fiber development

Haron Salih<sup>1,2</sup>, Wenfang Gong<sup>1</sup>, Shoupu He<sup>1</sup>, Wang Xia<sup>1</sup>, Magwanga Richard Odongo<sup>1</sup> and Xiongming Du<sup>1\*</sup> 

## Abstract

**Background:** Long non-coding RNAs (LncRNAs) are part of genes, which are not translated into proteins and play a vital role in plant growth and development. Nevertheless, the presence of LncRNAs and how they function in Ligon-lintless-1 mutant during the early cessation of cotton fiber development are still not well understood. In order to investigate the function of LncRNAs in cotton fiber development, it is necessary and important to identify LncRNAs and their potential roles in fiber cell development.

**Results:** In this work, we identified 18,333 LncRNAs, with the proportion of long intergenic noncoding RNAs (LincRNAs) (91.5%) and anti-sense LncRNAs (8.5%), all transcribed from Ligon-lintless-1 (Li1) and wild-type (WT). Expression differences were detected between Ligon-lintless-1 and wild-type at 0 and 8 DPA (day post anthesis). Pathway analysis and Gene Ontology based on differentially expressed LncRNAs on target genes, indicated fatty acid biosynthesis and fatty acid elongation being integral to lack of fiber in mutant cotton. The result of RNA-seq and RT-qPCR clearly singles out two potential LncRNAs, LNC\_001237 and LNC\_017085, to be highly down-regulated in the mutant cotton. The two LncRNAs were found to be destabilized or repressed by ghr-miR2950. Both RNA-seq analysis and RT-qPCR results in Ligon-lintless-1 mutant and wild-type may provide strong evidence of LNC\_001237, LNC\_017085 and ghr-miR2950 being integral molecular elements participating in various pathways of cotton fiber development.

**Conclusion:** The results of this study provide fundamental evidence for the better understanding of LncRNAs regulatory role in the molecular pathways governing cotton fiber development. Further research on designing and transforming LncRNAs will help not only in the understanding of their functions but will also in the improvement of fiber quality.

**Keywords:** Identification, Comparative analysis, LncRNAs, Ligon-lintless-1 mutant, Wild-type, Cell fiber development

## Background

Cotton fiber is one of the most important renewable resources of the textile industry worldwide. Cotton fiber develops from a single cell, and the nature of cell elongation provides an excellent model for testing gene expressions that relate to fiber development [1]. Fiber initiation occurs at the anthesis period, it emerges from the ovule epidermal cells [2, 3]. During the initiation period of cotton fiber development, only about 25% of cells on the ovule epidermis develop into a spinnable

cotton fiber [1]. The cotton fiber developmental process is divided into 4 overlapping developmental stages: fiber initiation, fiber cell elongation, secondary cell wall formation and fiber cell maturation [1]. Initiation, elongation and secondary cell wall of cotton fiber have a great impact on the quantity, length and fineness of fiber, which are the main factors determining the lint yield quality [3]. The elongation stage of cotton fiber begins immediately after initiation stage and continues for around 3 weeks before the fiber cell switches to intensive deposition of secondary cell wall [4, 5]. Cotton fiber elongation depends on the genotype and prevailing environmental conditions [5]. Fiber developmental stage is controlled by a multi-complex interaction of genes

\* Correspondence: [dujefrey8848@hotmail.com](mailto:dujefrey8848@hotmail.com)

<sup>1</sup>Institute of Cotton Research, Chinese Academy of Agricultural Science (ICR, CAAS)/State Key Laboratory of Cotton Biology, Anyang 455000, China  
Full list of author information is available at the end of the article



rather than one gene effect, but the lack of evidence at molecular level regarding the network of transcriptional regulatory elements and genes that relate to cotton fiber development is one of the key setbacks in understanding the molecular components to improve fiber length [6]. Cotton fiber mutants are valuable materials for studying the molecular genetics and physiological processes of cell fiber development [7–11]. Recently, several fiber mutants have been revealed and used as model plants to study cotton fiber development, one of which is the monogenic and dominant mutant also known as the Ligon-lintless-1 [12]. The Ligon-lintless-1 mutant has an abnormal morphological appearance including, distorted leaves, stem and extreme reduction in the length of lint fiber about 4–6 mm at maturity stage [12]. Previous studies showed that the Li1 gene is located in the Dt sub-genome of *Gossypium hirsutum* on chromosome 22 (Dt 04) [13–15]. Gene expression analysis in the ovule of Ligon-lintless-1 mutant as compared to the wild-type, only few genes showed differential expression during the initiation stage of cotton fiber development [16, 17]. High expression levels of some secondary cell wall synthesis-related genes, such as tubulin, sucrose synthase and expansin were significantly expressed in Ligon-lintless-1 at the early stages of fiber development. However, they were highly expressed in wild-type cotton during the fiber elongation stage, probably elucidating the mechanism underlying the genotype of the Ligon-lintless-1 mutant [7, 15]. Previous studies using microarray technique and RNA-seq methods have attempted to identify key genes, which involved in regulating fiber elongation in the Ligon-lintless-1 mutant [7, 17–20]. To understand the phenomenon of early cessation of fiber elongation in Ligon-lintless-1 mutant as compared to the wild-type, it is paramount to look into the molecular mechanisms or elements that are required to regulate cotton fiber development.

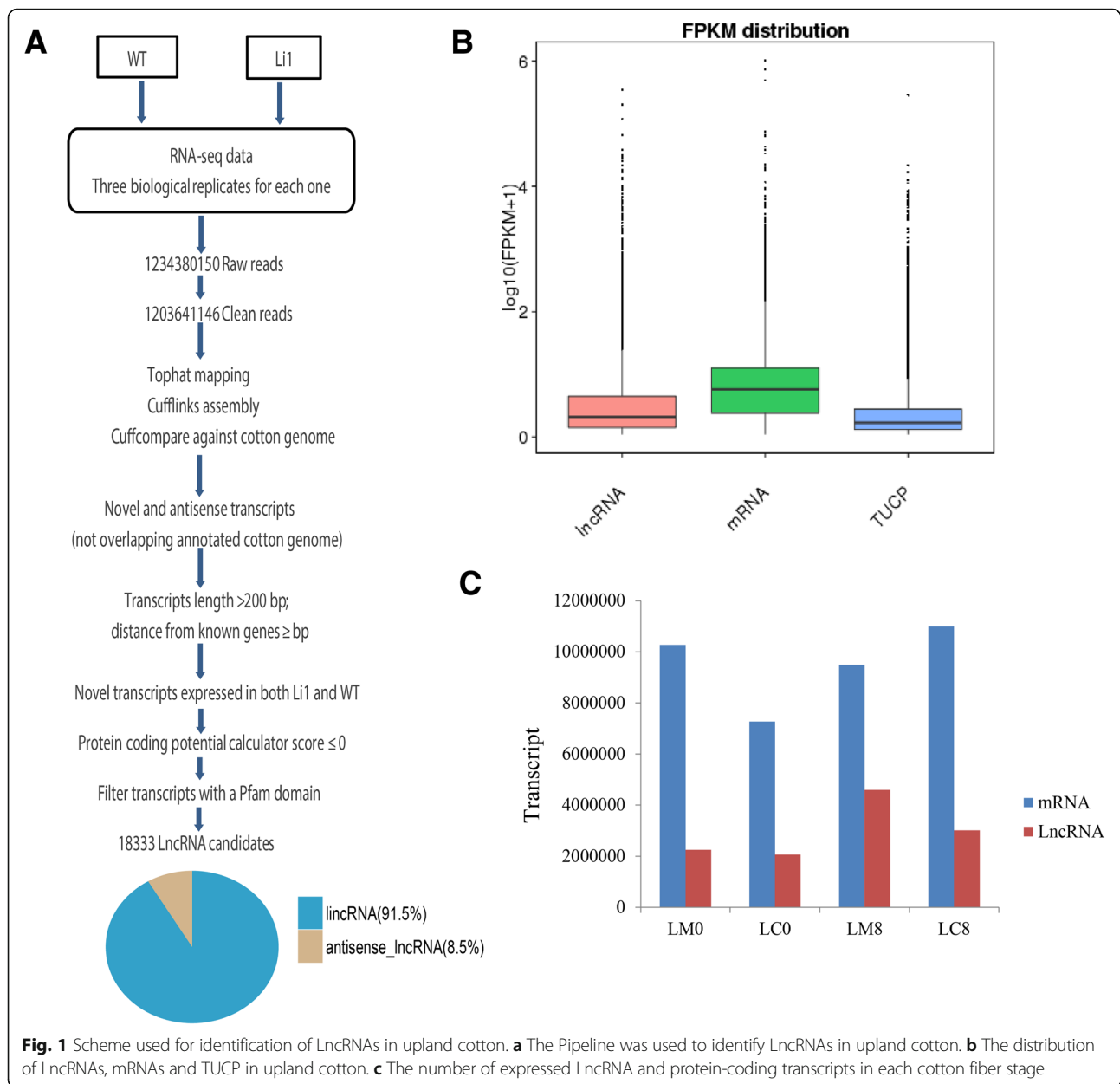
Long non-coding RNAs (LncRNAs) are defined as part of a non-coding RNAs longer than 200 bp (base pair) that do not have the capacity for coding proteins, which are involved in regulation of many biological regulatory processes [21]. According to their genomic location and context, LncRNAs can be divided into long noncoding natural antisense transcripts (LncNATs), long intergenic non-coding RNAs (LincRNAs), overlapping LncRNAs and long intronic non-coding RNAs [22]. Recent studies on plant LncRNAs have linked them to the vital biological processes including; controlling flowering period [23], gene silencing mechanism, abiotic stress tolerance, important developmental pathways [24–27] and cotton fiber development [28]. Many functions of plant LncRNAs are largely unknown except some LncRNAs such as COOLAIR (cold induced long antisense intragenic RNA), COLDAIR (cold assisted intronic noncoding RNA) and

PHOSPHATE1. COOLAIR and COLDAIR play the important role in regulating vernalization in Arabidopsis [29] while PHOSPHATE1 regulates phosphate homeostasis and control photoperiodic male sterility in rice [30]. In previous studies, several LncRNAs have been found to be involved in the regulation of epigenetic like RNA-directed DNA methylation and chromatin modification [28]. Recent investigation on noncoding RNAs in Ligon-lintless-1 and its wild-type is only limited to short noncoding RNAs (microRNAs) [31]. Naoumkina et al. 2016, identified an important small RNAs (microRNAs) expressed during fiber development in both Ligon-lintless-1 and Ligon-lintless-2 mutants as compared to their wild-type. The molecular mechanisms or elements underlying the LncRNAs related to cotton fiber development during initiation and early elongation stages are not well-known, thus using Ligon-lintless-1 mutant and its wild-type to identify and analyze differentially expressed long noncoding RNAs, which will provide a new dimension of understanding the cotton fiber development process. In this study, the sequencing of long non-coding RNA libraries constructed from cotton fiber development stages of Ligon-lintless-1 and wild-type was done. The identification of LncRNAs was analyzed in reference to *Gossypium hirsutum* TM-1 genome. We identified a total of 18,333 LncRNAs through five steps of filtration, of which 91.5% were LincRNAs and 8.5% were anti-sense LncRNAs. The functional prediction of long noncoding RNAs (LncRNAs) and their expressions as involved in cotton fiber development were examined. We investigated putative functional LncRNA candidates by differential expression analysis and co-expression network construction during cotton fiber development between Ligon-lintless-1 mutant and its wild-type.

## Results

### Identification of long noncoding RNAs in Ligon-lintless-1 and wild-type

Cotton fiber initiation and rapid elongation stages are crucial stages with greater impact on fiber quantity and fiber length. It is essential to understand the various molecular pathways involved in the development of cotton fiber. Several proteins or genes have been identified with functional roles in fiber development [16, 20, 28, 32, 33]. For the identification of LncRNAs in Ligon-lintless-1 and wild-type, two critical stages of fiber development, initiation and elongation stages which are represented as 0 and 8 DPA were the points of investigation. Biological triplicates were done through the whole Transcriptome by Illumina sequencing, generating about 1.2 billion clean reads (Fig. 1a and Additional file 1: Table S1). Each clean read (RNA-seq clean data) was mapped to the entire *G. hirsutum* genome (TM-1) [34] independently using both Tophat2 (v2.0.9) software [35] and Bowtie 2



[36]. The transcripts generated from each RNA-seq data were assembled using both Cufflinks (v 2.1.1) [37] and Scripture (beta2) software [38]. In order to reduce transcriptional isoforms noise, only those transcripts assembled and were found in two or more samples by two tools reserved for subsequent analyses. All transcripts were pooled and merged to make the final transcriptome using Cuffmerge [37]. After generation of the final transcriptome, Cufflinks was employed to reconstruct the cotton transcriptome followed by transcript abundance assembly and analysis of differential isoform. Five stepped filtering out of the frequently expressed transcripts with FPKM scores < 0.5 (2 for single-exon transcripts) in all samples, transcripts without strand sense

information and all transcripts that overlapped with annotated genes were removed; we recovered 78.39% of mRNAs in the dataset. The efficiency recovery of known protein-coding genes revealed that the preformed dataset was suitable for the retrieval of novel transcribed regions of the upland cotton genome (Additional file 1: Table S1). We only retained novel transcripts with the following specifications; not overlapping with known genes in a sense, longer than 200 nucleotides and expressed (for multiple-exon transcripts FPKM  $\geq 0.5$  and for single-exon transcripts FPKM  $\geq 2$ ). We further estimated the coding potential for the remaining transcripts using Coding Potential Calculator (CPC) [39] and Coding-Non-Coding Index (CNCI) [40] to predict the

transcript coding potential. Only, the transcripts with CPC scores  $> 0$  were removed (Fig. 1a). Furthermore, we used HMMER to confirm removal of protein-coding transcripts and checking of each transcript with CPC and CNCI score  $< 0$  in all three reading frames to eliminate transcripts which encoded any protein domains family belonging to Pfam database [41]. Finally, we identified 18,333 constantly expressed LncRNAs, with proportions of 16,777 of LincRNAs and 1,556 of anti-sense LncRNAs, counting for 91.5 and 8.5% respectively (Fig. 1a and Additional file 1: Table S1).

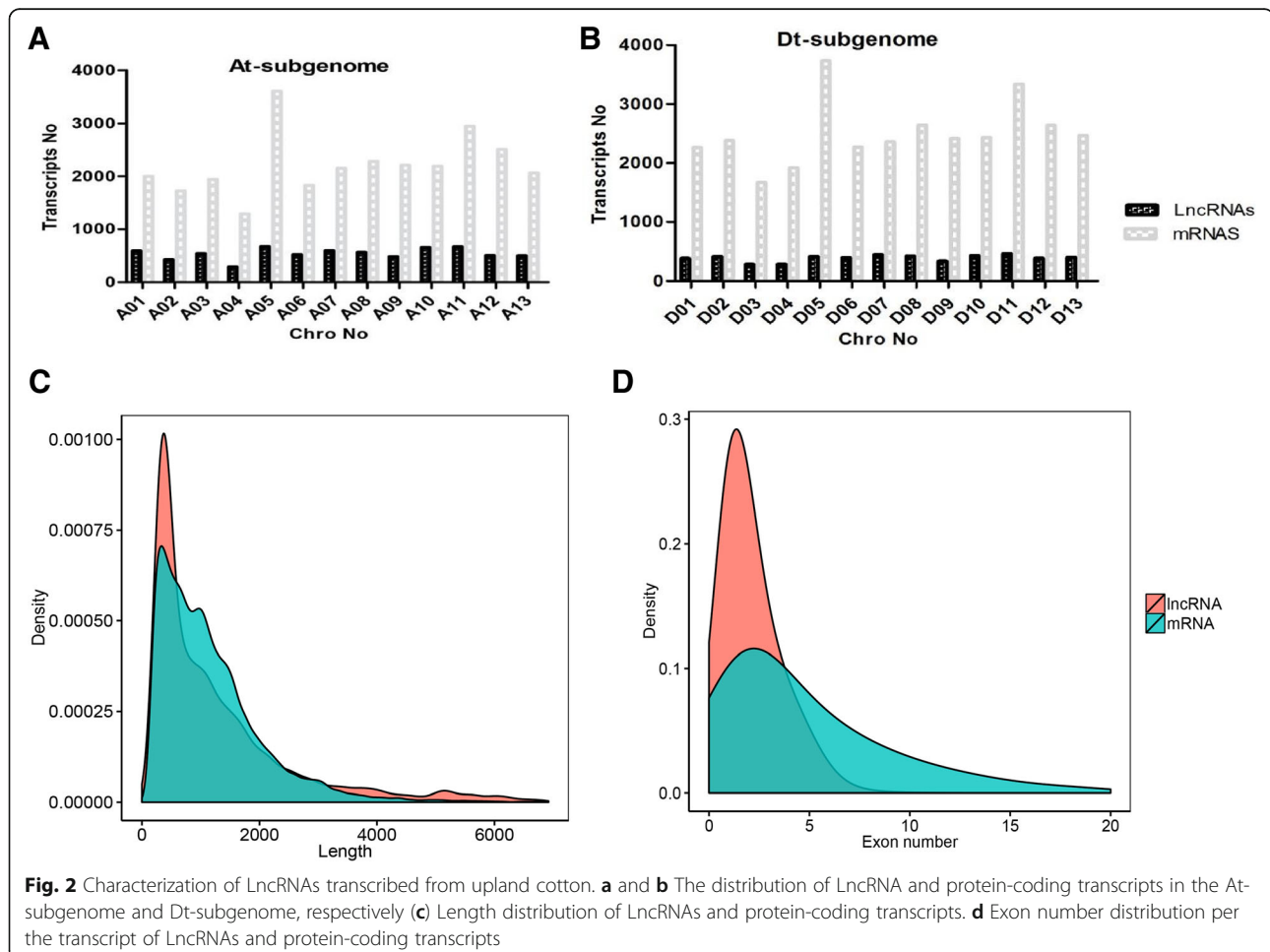
### The distribution of LncRNAs and mRNAs in the upland cotton genome

To further investigate the distribution of LncRNAs and mRNAs in upland cotton, the sequence of the upland cotton genome provided a valuable tool [34]. We observed the distribution of LncRNAs on the upland cotton genome and identified 7,064, 5,089 and 6,280 LncRNAs, which were transcribed from the At-sub genome, Dt-sub genome and scaffolds (undefined chromosomes), respectively (Fig. 2a & b). In addition, we

found that the lengths of LncRNAs were largely varied from 200 to 3,589 bp and transcribed from all upland cotton chromosomes and scaffolds. However, a total of 28,749, 35,514 and 9,219 mRNAs were transcribed from the At-genome, Dt-genome and scaffolds, respectively (Fig. 2a & b). Generally, mRNAs lengths varied from 150 to 428,268 bp and also distributed from all upland cotton chromosomes (Fig. 2c). Additionally, the highest densities of LncRNAs and mRNAs were mapped to the At-sub genome A05 and A11, whereas the least densities were detected in At-sub genome A02 and A04. In the Dt-sub genome, the highest densities of LncRNAs were detected in chromosome D11, and mRNAs were found in chromosomes D05 and D11. Upland cotton LncRNAs had fewer exons than that in mRNAs (protein-coding genes) (Fig. 2d) and LncRNAs and mRNAs had unequal distribution across the upland cotton chromosomes.

### Differentially expressed LncRNAs at different stages of cotton fiber development

We examined the expression levels of LncRNAs related to cotton fiber development at 0 and 8 DPA using



FPKM. LncRNAs were expressed at different levels in fiber development. To address whether the expression pattern of the LncRNAs is linked to a specific stage of cotton fiber development, we analyzed the differential expression levels of LncRNAs of each transcript between Ligon-lintless-1 mutant and near isogonic wild-type. At 8 DPA, higher numbers of LncRNAs were differentially expressed in Ligon-lintless-1 mutant than at 0 DPA (Additional file 2: Table S2 and Additional file 3: Table S3). We also identified differences in the expression profiles of LncRNAs and other transcript groups in the Ligon-lintless-1 and wild-type samples (Fig. 1b). The LncRNAs expression profiles in cotton fiber development were carried out by Cuffdiff which provided statistical procedures of determining the differential expression in a digital transcript of LncRNAs expression data using a model based on the negative binomial distribution [37]. A total of 266 and 407 LncRNAs were differentially expressed at 0 and 8 DPA respectively (Fig. 3a). Among them, 222 and 241 LncRNAs were up-regulated and 44 and 166 LncRNAs were down-regulated at 0 and 8 DPA respectively, which gave an indication of LncRNAs being either positively or negatively involved in the initiation and elongation stages of fiber development. In addition, there were more LncRNAs up-regulated than down-regulated in Ligon-lintless-1 mutant, which further giving proof of their negative role in fiber development. Only 22 common LncRNAs were found at 0 and 8 DPA while the rest, 244 and 385 LncRNAs were stage-specific to 0 and 8 DPA, respectively (Fig. 3a). More recently, actin-1 gene, GhACT\_L11 (Gh\_D04G0865), has been found to be regulating fiber development [42], in references to this gene we found one LncRNA, LNC\_012557 with differential expression only at 0 DPA. LNC\_012557 had a higher transcription level in Ligon-lintless-1 but not in wild-type.

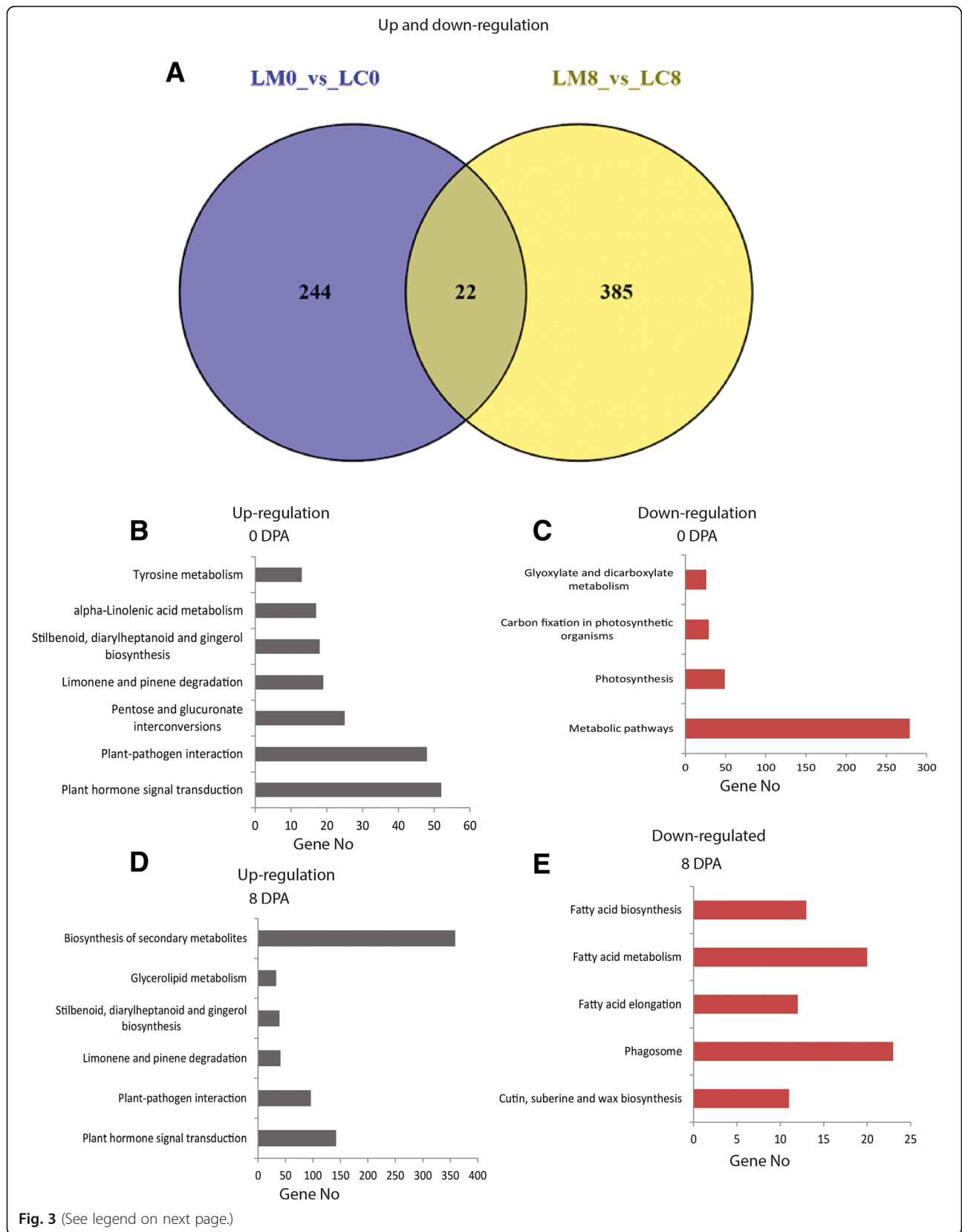
#### Common differentially expressed LncRNAs in cotton fiber development

Cotton fiber initiation and elongation are key stages to determine the number and length of the fiber cell development. Lint of cotton fiber is supposed to appear on the day of anthesis (0 DPA) and continuously elongate until three weeks after the day post anthesis [9]. To predict putative functional LncRNAs related to cotton fiber development during initiation and elongation stages, the expression of 22 common LncRNAs were found at 0 and 8 DPA (Fig. 3a). The result showed that 19 LncRNAs were up-regulated in Ligon-lintless-1 mutant as compared to the wild-type during initiation stage of cotton fiber (0 DPA) while only 3 LncRNAs (LNC\_017608, LNC\_012210 and LNC\_007516) were down-regulated in cotton fiber development at 0 DPA (Table 1). Specifically, these LncRNAs might be needed to

maintain normal fiber development in wild-type. However, 18 LncRNAs were down-regulated at 8 DPA and only 4 LncRNAs (LNC\_007516, LNC\_008336, LNC\_017132 and LNC\_017608) were up-regulated during the elongation stage of fiber development at 8 DPA. The expression of these LncRNAs might in part trigger the early cessation of cotton fiber development in the Ligon-lintless-1 mutant. These LncRNAs have significantly higher transcription levels in normal cotton fiber than in mutant cotton fiber at 8 DPA ( $q$ -value  $< 0.05$ ), but lower in transcription levels in Ligon-lintless-1 mutant than wild-type at 0 DPA ( $q$ -value  $< 0.05$ ) (Table 1).

#### Functional analysis of differentially expressed LncRNAs

To investigate the specific functions of LncRNAs, we systematically predicted the possible targets of LncRNAs based on co-location and co-expression. According to the genome position of the LncRNAs and mRNAs (co-located), the nearest mRNAs around each LncRNA at downstream and upstream locations within 10 kb were investigated. The co-expressed genes were selected based on their expression correlation coefficients for the gene ontology (GO) enrichment analyses. The targets of differentially expressed LncRNAs were classified into different groups by gene ontology annotations as biological process, molecular function and cellular component. The GO analysis of LncRNAs targets based on co-location and co-expression at 0 and 8 DPA showed that some significantly enriched genes were mainly involved in the biological processes, such as protein phosphorylation, phosphorylation, cellular protein modification, protein modification and macromolecule modification, which were up-regulated in Ligon-lintless-1 mutant as compared to wild-type. As can be seen, the biological process and metabolic process were up-regulated at 8 DPA and down-regulated at 0 DPA (Fig. 4a & b). Phosphorylation is a process mediated by protein kinases to activate critical cellular pathways such as metabolism, cell division and cell differentiation during initiation stage in cotton fiber development [43]. Interestingly, Fatty acid biosynthetic process, methionine metabolic process, development involved in symbiotic interaction, post-embryonic morphogenesis, nodule morphogenesis and post-embryonic development were all down-regulated at 8 DPA in the Ligon-lintless-1 mutant (Fig. 4b). This result indicated clearly that these genes involved in biological process and metabolic process during the elongation stage might be responsible for the early cessation of fiber development. In the molecular function categories, kinase activity, protein kinase activity, phosphotransferase activity and transferase activity were significantly up-regulated in cell fiber development at 0 and 8 DPA. Protein kinase activity plays an important role in signal transduction through phosphorylation



(See figure on previous page.)

**Fig. 3** Differentially expressed LncRNAs expressed in Ligon-lintless-1 and wild-type. **a** A Venn diagram showing LncRNAs that was up and down-regulated in Ligon-lintless-1 and wild-type. **b** and **(c)** KEGG enrichment analysis of differentially expressed LncRNA-targets genes was up-regulated at 0 and 8 DPA, respectively. **d** and **(e)** KEGG enrichment analysis of differentially expressed LncRNA-target genes were analyses of differentially expressed LncRNA-target genes were down-regulated at 0 and 8 DPA, respectively

process during cotton fiber development [44]. Catalytic activity, heterocyclic compound binding, organic cyclic compound binding, carbohydrate binding, polysaccharide binding and pattern binding were highly up-regulated in Ligon-lintless-1 at 0 DPA (Fig. 4a). For the molecular component, membrane, membrane part, integral to membrane, thylakoid intrinsic to membrane and thylakoid part were down-regulated at the initiation stage (0 DPA), and only three gene ontology of apoplast, extracellular matrix and proteinaceous extracellular were up-regulated at 8 DPA (Fig. 4a & b).

We further applied all differentially expressed LncRNAs for the possible target genes using KEGG. The pathway analysis of co-located and co-expressed genes was enriched in plant hormone signal transduction, plant-pathogen interaction, pentose and glucuronate

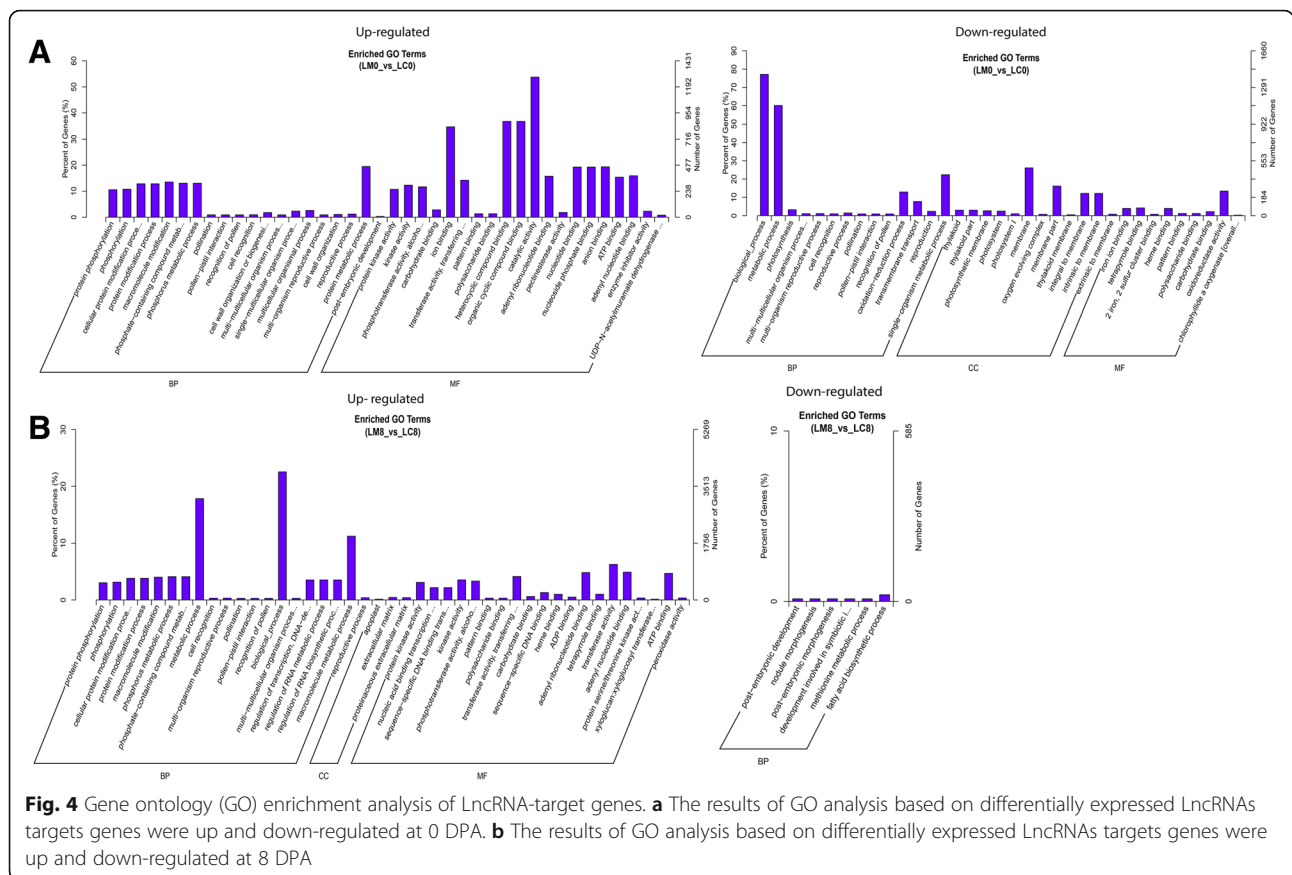
**Table 1** Common differentially expressed LncRNAs in Ligon-lintless-1 mutant as compared to wild-type during cotton fiber development

LncRNAs ID	log2 (fold change 0 DPA Li1/WT)	Q value	log2 (fold change 8 DPA Li1/WT)	Q value
LNC_002075	2.41	5.00E-05	-1.87	0.00045
LNC_015258	2.49	0.0001	-1.76	0.01705
LNC_007516	-1.59	0.00015	5.54	0.0026
LNC_015379	2.75	0.00015	-2.33	0.00035
LNC_015166	2.46	5.00E-05	-1.57	5.00E-05
LNC_012210	-1.88	0.00355	-1.89	0.0186
LNC_015540	3.20	0.0059	-1.96	5.00E-05
LNC_014831	1.52	5.00E-05	-2.12	5.00E-05
LNC_010975	2.93	5.00E-05	-2.17	0.00085
LNC_001352	3.49	5.00E-05	-3.29	5.00E-05
LNC_008874	2.17	0.00435	-1.62	0.00395
LNC_001734	3.81	5.00E-05	-3.08	0.00035
LNC_001543	1.56	0.0086	-1.90	0.0042
LNC_008391	2.97	5.00E-05	-2.52	5.00E-05
LNC_008336	2.51	0.00105	3.27	0.00485
LNC_008289	1.75	0.00025	-1.88	0.0005
LNC_008291	1.76	5.00E-05	-1.85	5.00E-05
LNC_006271	1.98	0.00025	-1.61	0.00045
LNC_017132	1.60	0.0154	inf	0.00305
LNC_017608	-1.84	0.0001	2.56	5.00E-05
LNC_005974	1.53	0.00385	-2.52	0.0001
LNC_003053	1.84	5.00E-05	-1.84	5.00E-05

interconversions, limonene and pinene degradation, stilbenoid, diarylheptanoid and gingerol biosynthesis, alpha-linolenic acid metabolism and tyrosine metabolism pathways, which were up-regulated in the Ligon-lintless-1 mutant at 0 DPA. On the other hand, metabolic pathways, photosynthesis, carbon fixation in photosynthetic organisms, glyoxylate and dicarboxylate metabolism were down-regulated in Ligon-lintless-1 mutant at 0 DPA (Fig. 3b & c). At 8 DPA, the co-located and co-expressed genes were mainly involved in biosynthesis of secondary metabolites, plant hormone signals transduction, plant-pathogen interaction, limonene and pinene degradation, stilbenoid, diarylheptanoid and glycerolipid metabolism were up-regulated in Ligon-lintless-1. In contrast, the phagosome (23 genes), fatty acid metabolism (12 genes), fatty acid elongation (21 genes), fatty acid biosynthesis (13 genes) and cutin, suberine and wax biosynthesis (11 genes) were down-regulated at 8 DPA in the mutant cotton (Fig. 3d & e). These results suggested that the regulation of hormones and some enzymes catalyze metabolic processes, such as fatty acids, biosynthesis of secondary metabolites, transport process and glycerolipid metabolism. They may possibly play an important role during initiation and elongation stages of cotton fiber development. Based on the GO and KEGG analysis, we identified some potentially key target genes which were enriched in fatty acid biosynthesis, fatty acid metabolism, fatty acid elongation and transport process were down-regulated in Ligon-lintless-1 as compared to the wild-type during the elongation stage of cotton fiber development. These possible target genes provide new insight into the role of LncRNAs in cotton fiber developmental stages. In addition, plant hormone metabolism could also have effect on cotton fiber development at elongation stage, for example, a large number of LncRNAs were found to be involved in the metabolism of auxin, ethylene, abscisic acid (ABA) and gibberellin [45]. The majority of LncRNAs involved in the above hormones metabolism were found to be up-regulated in Ligon-lintless-1 mutant.

#### Interaction determination between miRNAs and LncRNAs

The interaction between miRNAs and LncRNAs is important for the determination of functional patterns of LncRNAs because of the destabilization and repressive nature of miRNAs [46]. LncRNA functions and stabilities can be disrupted by miRNAs [47]. To examine LncRNAs and miRNAs interactions, the differentially



expressed LncRNAs were submitted to the psRNATarget website in order to obtain the miRNAs specifically targeting the various differentially expressed LncRNAs [48]. Out of 673 LncRNAs, 233 were found to be targeted by 58 miRNAs (Additional file 4: Table S4). A number of miRNAs detected such as miR156, miR160, miR162, miR164, miR166, miR167, miR169, miR172, miR390, miR393, miR394 and miR156 were found to have temporal differential expression patterns with the peak at -2 DPA [49]. In cotton, it was previously observed that miR166, miR167, miR172 and miR2949 were highly expressed in the ovule [50]. Twenty (20) LncRNAs were found to be either destabilized or repressed by miR156; this could possibly explain their role in fiber development in mutant cotton. This result is consistent with the expression analysis of miR156 and others in fiber initiation at -2 DPA in Xu-142 (mutant) and Xu-142 (wild type), in which miR156, miR157, miR159 and miR160 among others were found to be significantly up-regulated in mutant cotton as compared to the wild-type [51]. Similarly, miR156 was found to have a temporal differential expressions pattern with the peak at -2 DPA. In this study, we identified 14 miRNAs which were differentially expressed during cotton fiber development in Ligon-lintless-1 as compared to wild-

type. Among these miRNAs, ghr-miR2950 (targeting to fatty acid hydrolase), ghr-miR7506, ghr-miR7504a (targeting to cleave carbohydrate-Active enzymes family), ghr-miR7484a (targeted to cleave Ring/U-box domain containing protein) and ghr-miR7502, ghr-miR7499, ghr-miR7492a, ghr-miR399e (targeted to cleave AP2/B3-like transcriptional factor), ghr-miR156c (targeted to cleave squamosal promoter-binding protein), ghr-miR169b and ghr-miR7496a were down-regulated in Ligon-lintless-1 whereas ghr-miR7510a (targeted to cleave peroxidase super family protein) and ghr-miR7506, ghr-miR7498 and ghr-miR7497 were up-regulated in Ligon-lintless-1 during cotton fiber development, indicating that cotton mutation strongly disturbed their expressions. In contrast, ghr-miR7510b (targeted to cleave peroxidase super family protein) and ghr-miR7492a were up-regulated during the initiation stage (0 DPA), and down-regulated during elongation stage (8 DPA) in the Ligon-lintless-1 mutant (Additional file 5: Table S5). In addition, ghr-miR2950 was predicted to target 8 down-regulated LncRNAs. In this context, the two, LNC\_017085 and LNC\_001237 were significantly down-regulated in Ligon-lintless-1 mutant by more than 5-fold and 4-fold, respectively (Additional file 5: Table S5). The interaction between the two aforementioned LncRNAs and ghr-





014975 were up-regulated at 8 DPA (Fig. 6a and b). The results of the down-regulated LncRNAs validation, showed a deviation from the RNA-seq, in which three, LNC\_001237, LNC\_017085, and LNC\_012941 were highly expressed in the wild-type at 5, 8 and 10 DPA but LNC\_017026 and LNC\_014788 had low expression in wild-type compared to the mutant type at 8, 10 and 15 DPA (Fig. 6a). The down-regulation of LNC\_001237 and LNC\_017085 could be attributed to the destabilizing or repressive nature of ghr-miR2950. The up-regulated LncRNAs such as LNC\_013268, LNC\_015152, LNC\_015147, LNC\_014473 and LNC\_014975 were highly expressed in Ligon-lintless-1 mutant but showed low expression levels in the wild-type at 5, 8 and 15 DPA. This result is consistent with the whole RNA-seq analysis in which all the 5 LncRNAs (LNC\_001237, LNC\_017026, LNC\_017085, LNC\_012941 and LNC\_014788) were significantly expressed in Ligon-lintless-1 compared to the wild-type at 8 DPA. Accordingly, the up-regulated LncRNAs in Ligon-lintless-1 mutant might be responsible for the early cessation of the fiber elongation stage (Fig. 6b). RT-qPCR analysis showed that all of the up-regulated LncRNAs were highly expressed in Ligon-lintless-1 mutant but not in wild-type, this validates the negative role of LncRNAs in cotton fiber development during elongation stage (Fig. 6b).

## Discussion

Cotton fiber initiation and rapid elongation stages are crucial stages with greater impact on fiber quantity and fiber length. It is essential to understand the various molecular pathways involved in the development of cotton fiber. Several proteins or genes have been identified with functional roles in fiber development [16, 20, 28, 46, 47]. However, the molecular elements which are responsible for early cessation of the fiber elongation process in Ligon-lintless-1 mutant are still largely not understood. LncRNAs were first discovered in 1991, highlighting certain regulatory pathways involved in different functions of the organism [48]. In recent years, there have been an upsurge in the number of studies on LncRNAs in relation to their roles in various biological processes across plants and animals [21, 24, 25, 27, 28, 49–52, 54, 60–63]. This study aimed at carrying out systematic identification and analysis of differentially expressed LncRNAs in Ligon-lintless-1 mutant and wild-type of *G. hirsutum*. In general, the expression levels of LncRNAs showed significantly lower than that in the protein-coding transcripts (Fig. 1b & c), which conformed with a previous report [28]. Transcriptome analysis in *Gossypium arboresum* and *Gossypium barbadense* showed that the two types of LncRNAs, the lincRNAs and antisense (LncRNAs) had less exons [28, 64]. The total number of long non-coding RNAs identified were fewer compared

to earlier findings in *G. barbadense* [28], the difference in the number of LncRNAs, could possibly be due to the genome differences. To further understand the roles of LncRNAs in Ligon-lintless-1 mutant and wild-type, expression profiling was done at 0 and 8 DPA. Additionally, the results of this study revealed that the majority of differentially expressed LncRNAs were precisely expressed at the specific stage of development, indicating the functional divergence of LncRNAs in cotton fiber development. The expressions of LncRNAs are temporal-specific, highly conserved LncRNAs and are actively regulated and do function predominantly in embryonic development reported on the study of the evolution of LncRNAs repertoires and expression patterns in tetrapods [65]. Interestingly, we found that the LncRNAs were time specific in terms of their expression patterns. In rice, the LncRNAs have unique characteristics compared to those of Arabidopsis in terms of their expression patterns. These groups of LncRNAs have tissue and or stage specificity in their expression profiling patterns [56]. Expression levels of a number of LncRNAs in Ligon-lintless-1 mutant were significantly varied from those in wild-type at 0 and 8 DPA (Additional file 3: Table S3).

Gene Ontology and KEGG pathway analyses based on differentially expressed LncRNAs on target genes identified fatty acid biosynthesis as an inhibitory process related to cotton fiber development (Figs. 3e and 4b). Fatty acid biosynthesis and fatty acid elongation were down-regulated in the Ligon-lintless-1 mutant at 8 DPA contrary to the previous finding in which fatty acid biosynthesis and fatty acid elongation were found to be up-regulated during cotton fiber elongation [66, 57]. Fatty acid biosynthesis and fatty acid elongation do promote trichome cell elongation through stimulation of ethylene biosynthesis [67]. The enzymes that catalyze fatty acid metabolism in Ligon-lintless-1 mutant might indirectly or directly be involved in cotton fiber elongation. This finding is in agreement with a previous report which stated that the majority of genes involved in the synthesis of very long chain fatty acid biosynthesis were down-regulated in the Ligon-lintless-1 mutant [59]. Gene Ontology and KEGG pathway analyses based on differentially expressed LncRNAs on target genes identified fatty acid biosynthesis as an inhibitory process related to cotton fiber development (Figs. 3e and 4b).

In addition, a greater percentage of the differentially expressed LncRNAs in Ligon-lintless-1 were specifically targeted by miRNA, which is an indication of their possible roles in the development process of cotton fiber. Previous several studies reported the influence of miRNAs in cotton fiber development [31, 50]. The high expression level of ghr-miR2950 in the wild-type may play a key role in the elongation stage of fiber. The expression analysis of RNA seq showed that ghr-miR2950



was significantly expressed in wild-type as compared to Ligon-lintless-1 at 8 DPA [31]. It was found that ghr-miR2950 exhibited a higher expression level during the fiber elongation stage than at ovulation stage [49]. In plant cell division, ghr-miR2950 was found to be highly expressed in tall-culm mutant cotton, possibly explaining the plant height [60]. The finding of this research that fatty acid hydroxylase was the putative target of miR2950. Fatty acid hydroxylase was significantly up-regulated at 10 DPA compared to the wild type, further explaining its significant contribution to fiber length in domesticated cotton, [61]. Ring/U-box domain containing protein had an interaction with ghr-miR7484a, which was down-regulated in Ligon-lintless-1. Previously, it was indicated that, ghr-miR7484 had a role in regulating cotton fiber development through targeting MYB in cotton [50]. This implied that ghr-miR7484a might be involved in regulation of cotton fiber development during elongation stage. RING/U-box super family proteins were involved in the transitional developmental stages of cotton fiber [62]. The expression pattern of ghr-miR7510b was found to target peroxidase super family protein which was up-regulated at 0 DPA but down-regulated at 8 DPA in Ligon-lintless-1 mutant. Peroxidase was found to directly correlates to the accumulation of reactive oxygen species which related to cotton fiber elongation [63]. The high expression levels of ghr-miR7510b and ghr-miR7492a in Ligon-lintless-1 at 0 DPA were to trigger the initiation of cotton fiber while their lower expression levels during elongation stages at 8 DPA initiated the cessation of cotton fiber elongation.

To further investigate the differential expression of LNC\_017085 and LNC\_001237 with high potential targets of many mRNAs, we only focused on the significantly expressed mRNA with a high potential role in fiber development. This result provides a strong indication that genes related to cell wall metabolism could be playing key roles in cotton fiber development. Pectin methylesterase is the enzyme which is involved in cell wall structure of plants through deposition of the pectin [68]. In cotton, pectin methylesterase could play an important role in cell fiber wall structure of cotton fiber through pectin deposition [69]. Lipids are involved in the regulation of cotton fiber elongation stage through facilitating the transportation of phosphatidylinositol [70]. During the elongation stage of cotton fiber development, the continuous synthesis of lipids and protein transport are the key steps to promote the dynamic change in enlargement of plasma membrane and vacuoles [71]. Both RNA-seq analysis and RT-qPCR result in Ligon-lintless-1 mutant and wild-type provide strong evidence of LNC\_001237, LNC\_017085 and ghr-miR2950 being integral molecular elements that are likely involved in the various pathways of cotton fiber development.

## Conclusion

The findings of this study provide novel insights into the better understanding of LncRNAs regulatory role in the molecular pathways regulating cotton fiber development. Further research on designing and transforming the LncRNAs will help not only in understanding their functions but will also help in improving fiber quality.

## Methods

### Plant material and RNA extraction

*G. hirsutum*, Ligon-lintless-1 mutant (Li2Li2) and its wild-type (li2li2) were derived from self-pollinated Ligon-lintless-1 plants of at least more 6 generations in our lab [72]. The Ligon-lintless-1 seeds were provided by Kohel (USA) and used in this research study for LncRNA analysis. The two cotton genotypes were grown in the experimental field at the Cotton Research Institute, Chinese Academy of Agricultural Sciences (ICR, CAAS) Anyang city, Henan province, China. One day before anthesis, cotton flowers were tied and tagged for self-pollination. A total of 198 Ligon-lintless-1 mutant Li2Li2 plants and 128 wild-type li2li2 plants were used for samples collection. Cotton bolls were collected at 0 (ovules) and 8 DPA (Fiber) during cotton fiber development. Biological triplicates for each sample were collected during cotton fiber development at 0 and 8 DPA from Ligon-lintless-1 and wild-type. Briefly, cotton fibers separated from the developing ovules using glass bead shearing method/liquid nitrogen. The harvested samples were rapidly frozen in liquid nitrogen and kept at  $-80^{\circ}\text{C}$  until analysis. Cotton fibers were initial ground in liquid nitrogen using a mortar and pestle, and the gridded samples were rapidly transferred into tubes (2 ml micro) with 700  $\mu\text{L}$  lysis/binding buffer. RNA samples were extracted using TRIzol (Invitrogen). RNA contamination and degradation were evaluated on 1% agarose gel. RNA purity and integrity were tested using the NanoDrop-Photometer<sup>®</sup> spectrophotometer RNA and Nano 6000 Assay Kit of the Bioanalyzer 2100 system, separately. Finally, the concentration of RNA was checked using Qubit<sup>®</sup> RNA Assay Kit in Qubit<sup>®</sup> 2.0 Fluorometer.

### Library preparation for lncRNA sequencing

A total of 3  $\mu\text{g}$  RNA per sample was used as input material for the RNA sample preparations. At the initial stage, ribosomal RNA was removed by Epicenter Ribozero<sup>™</sup> rRNA Removal Kit (Epicenter, USA), and rRNA free residue was cleaned up by ethanol precipitation. The sequencing libraries were prepared using the rRNA-depleted RNA by NEBNext<sup>®</sup> Ultra<sup>™</sup> Directional RNA Library Prep Kit for Illumina<sup>®</sup> (NEB, USA) following the manufacturer's specifications. Fragmentation was performed using divalent cations under raised temperature in NEBNext First-Strand Synthesis Reaction Buffer. The

first-strand cDNA was produced using random hexamer primer and M-MuLV Reverse Transcriptase (RNaseH-) and the second-strand cDNA production was subsequently conducted using RNase H and DNA polymerase I. In the reaction buffer, dTTP and dNTPs were changed by dUTP. The rest of the overhangs were successfully transformed into blunt ends through exonuclease/polymerase enzyme activities. After adenylation of 3' ends of DNA fragments, NEBNext Adaptor with the hairpin loop organizations were ligated to make for hybridization. In order to identify cDNA fragments of specially 150~200 base pairs, the library fragments were cleaned using AMPure XP system (Beckman Coulter, Beverly, USA). Then 3  $\mu$ l USER Enzyme (NEB, USA) was used with size selected adaptor-ligated cDNA at 37 °C for 15 min followed by 5 min at 95 °C before PCR. PCR was then performed with Phusion High-Fidelity DNA polymerase using universal PCR primers and index (X) Primer. The final, harvests were purified (AMPure XP system) and library quality was assessed on the Agilent Bioanalyzer 2100 system.

#### Bioinformatic identification of upland cotton LncRNAs

Raw data of fastq format were processed through in-house perl scripts to obtain clean data. The clean data (clean reads) were obtained by removing the adapter and low-quality reads (quality score > Q20). The clean datasets with high quality were mapped independently with reference to the upland cotton genome, using both Tophat2 (v2.0.9) software [35] and Bowtie 2 (version 2.2.9) [36]. Reference genome and gene model annotation files were downloaded from genome website (<http://mas-cotton.njau.edu.cn/html/Data>) [34]. Cufflinks (v2.1.1) [37] and Scripture (beta2) [38] in a reference-based approach were applied to assemble the transcriptomes which were generated by Tophat2 (v2.0.9) and Bowtie 2. All transcriptomes from four cotton fiber development samples were pooled and merged to perform a comprehensive transcriptome using Cuffmerge. The CUFFCOMPARE procedure was used to compare all the transcript assemblies to the upland cotton genome. Then we adapted five steps of filtrations to identify LncRNAs from transcriptome assemblies: (1) transcripts assembled by two methods were detected in less than two samples were discarded; (2) transcripts without strand sense information and short transcripts without mapping coverage of more than half of the transcript length were removed; (3) transcripts overlapped with mRNA on the same strand, transcripts with FPKM (fragments per kilo-base of exon per million fragments mapped)  $\geq 0.5$  (2 for single-exon transcripts) and short transcripts with length less than 200 bp were eliminated; (4) transcripts with potential encoding hit were excluded by CPC (Coding Potential Calculator) and CNCI (Coding-Non-Coding Index) [40] and (5) the remaining transcripts

were queried against the Pfam scan (Pfam databases) to remove all transcripts with known protein-coding domains (E-value 0.001) [41]; transcripts without capability of coding proteins were considered as LncRNAs with length  $\geq 200$  bp.

#### Gene expression analysis

Cuffdiff (v2.1.1) was used to calculate FPKMs of both LncRNAs and coding genes in each sample [37]. Gene FPKMs were calculated by summing the FPKMs of transcripts in each gene group. FPKM value is calculated based on the length of the fragments and reads count mapped to this fragment.

#### Target gene prediction

Determination of target gene prediction of LncRNA was acting on neighboring genes (co-location). We searched the neighboring coding genes around every LncRNA at upstream and downstream locations within 10 kb to investigate their possible functions. Co-expression of the target gene prediction in LncRNA was identified by their expression level. The prediction method of co-location and co-expression of the target genes were done according to the previous studies [73, 74]. Pearson correlation coefficient ( $r_p$ ) was applied to elucidate the expression relationship between LncRNAs and mRNAs pairs using R tool. LncRNAs co-expressed target mRNAs (genes) were identified with ( $R^2$ )  $\geq 0.95$ . We clustered the genes from various samples with WGCNA [75] for the search of common expression modules and the analysis of their function depending on the functional enrichment analysis.

#### Differential expression analysis

Cuffdiff provides statistical routines for determining differential expression in a digital transcript or gene expression data using a model based on the negative binomial distribution [37]. Only transcripts that had FPKM more than 1 were considered to be expressed. The differentially expressed transcripts were identified with q-value < 0.05 and a fold-change  $\geq 1.5$  or  $\leq -1.5$  between the Ligon-lintless-1 mutant and wild-type.

#### GO and KEGG enrichment analysis

GO (Gene Ontology) analyses of the differential expression (DE) of genes or the LncRNA target genes were performed by the GO seq R package, where the length bias of each gene was corrected. GO term analysis was considered significant at a corrected *P*-value < 0.05 were. KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway is a database resource to understand the functions of the biological organization. KOBAS software was used to test the statistical enrichment of the DE genes or the LncRNA target genes in KEGG pathways with corrected *P*-value < 0.05.

### Small RNA sequencing and processing

Small RNA libraries preparation and sequencing were conducted by LC Science (Houston, TX). Biological triplicates of RNA samples extracted from cotton fiber development at 0 and 8 DPA were combined together for preparation of Ligon-lintless-1 and wild-type microRNA libraries, separately. The micro RNA libraries were constructed using 1 µg of total RNA based on the Tru-Seq® microRNA sample preparation guide (Illumina). The general process comprised the following: firstly, the total RNA was ligated to RNA 3' and RNA 5' adapters. Secondly, the cDNA constructs were created by reverse transcription followed by PCR; the construction was based on the microRNAs ligated with 3' and 5' adapters. Thirdly, small cDNA fractions of length ranging between 22 nt and 30 nt were isolated by using 6% denaturing polyacrylamide gel electrophoresis. Finally, the cDNA construct was purified, and the library was validated. The libraries sequencing was done using the Illumina HiSeq 2500 platform.

### Identification of conserved and novel microRNAs

Clean reads were run through miRPlant software for identification of plant microRNAs from microRNA sequencing data [58]. *Gossypium hirsutum* TM-1 genome [34] was used for mapping reads with software's default parameters. The miRPlant, predicted miRNAs were queried against the miRBase database (version 21, <http://www.mirbase.org/>) to identify conserved and reported earlier microRNAs. Matched sequences with no more than two mismatches were considered as candidate conserved or formerly reported miRNAs and were assigned to the corresponding miRBase family. Predicted microRNAs with more than 2 mismatches were considered as potential novel miRNAs; they lack sufficient similarity to assign to a microRNA family [76]. Statistical importance of differential expression levels of microRNAs in the RNA-seq data was calculated with the Audic and Claverie statistic using IDEG6 software [77].

### Interaction determination between miRNAs and lncRNAs

The interaction between miRNAs and lncRNAs and their potential roles in regulating gene expression were previously described in higher plants. Upland cotton lncRNAs were predicted as miRNA targets using the psRNATarget [78].

### RT-qPCR analysis

To examine the lncRNA expression patterns, cotton tissues from Ligon-lintless-1 and its wild-type were harvested at different stages of cotton fiber development such as 0, 5, 8, 10, and 15 DPA. High quality of RNA was treated by DNase I (TaKaRa, Japan) to eliminate contaminating DNA. The cDNA was generated from

2 µg of RNA in reaction volume of 20 µL using ReverTra Ace PCR-qRT kit based on the manufacturer's requirement. RT-qPCR analysis was applied to test lncRNAs expression levels during developmental stages of cotton fiber. RT-qPCR experiment analysis was performed using the SYBER premix ExTaq kit (TaKaRa, Japan) and the Applied Biosystems 7500 Real-Time PCR system. SYBR Green was estimated the amplification targeted lncRNA. The housekeeping β-actin gene of the cotton was applied as a reference and primers were used for RT-qPCR. The following thermal cycler settings and lncRNA expression analysis were done according to salih et al. [79].

### Additional files

**Additional file 1: Table S1.** The numbers of annotated genes (mRNAs) and other transcript reads mapped to the TM-1 reference genome (XLSX 13 kb)

**Additional file 2: Table S2.** Differentially expressed lncRNAs in Ligon-lintless-1 and wild-type during cotton fiber development were expressed at 0 and 8 DPA. (XLSX 2671 kb)

**Additional file 3: Table S3.** Differentially expressed lncRNAs during cotton fiber development at 0 DPA in Ligon-lintless-1 and wild-type (XLSX 10464 kb)

**Additional file 4: Table S4.** lncRNAs were predicted as miRNA targets using the psRNATarget. (XLSX 27 kb)

**Additional file 5: Table S5.** Differentially expressed miRNAs in Ligon-lintless-1 mutant as compared to wild-type were expressed during cotton fiber development at 0 and 8 DPA. (XLSX 11 kb)

### Abbreviations

bp: Bias pair; DPA: Day post anthesis; Li-1: Ligon-lintless-1; lncRNAs: Long non-coding RNAs; WT: Wild-type

### Acknowledgments

The authors appreciate the National Science and Technology of China and China Scholarship Council (CSC) all contributed with funding towards the research.

### Authors' contributions

HS, SPH and XMD designed the experiment. WFG and WX grew the cotton seedlings and performed the experiments. HS analyzed the results and prepared the manuscript. MRO and XMD and revised the manuscript. All authors reviewed and approved the final manuscript.

### Funding

This study was supported by the National Natural Science foundation of china (No. 31601353) and the National Key Project of Research and development Plan (No. 2016YFD0100203).

### Availability of data and materials

All related data are available within the manuscript and its additional files.

### 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**

<sup>1</sup>Institute of Cotton Research, Chinese Academy of Agricultural Science (ICR, CAAS)/State Key Laboratory of Cotton Biology, Anyang 455000, China.

<sup>2</sup>Zalingei University, Central Darfur, Sudan.

Received: 16 November 2018 Accepted: 16 July 2019

Published online: 19 August 2019

**References**

- Tiwari SC, Wilkins TA. Cotton (*Gossypium hirsutum*) seed Trichomes expand via diffuse growing mechanism. *Can J Bot Can Bot*. 1995;73:746–57.
- Lee JJ, Woodward AW, Chen ZJ. Gene expression changes and early events in cotton fibre development. *Ann Bot*. 2007;100:1391–401 doi:mcm232 [pii]v10.1093/aob/mcm232.
- Basra AS, Malik CP. Development of the cotton Fiber. In: International Review of Cytology; 1984. p. 65–113. [https://doi.org/10.1016/S0074-7696\(08\)61300-5](https://doi.org/10.1016/S0074-7696(08)61300-5).
- Taliercio EW, Boykin D. Analysis of gene expression in cotton fiber initials. *BMC Plant Biol*. 2007;7:22.
- Meinert MC, Delmer DP. Changes in biochemical composition of the cell wall of the cotton fiber during development. *Plant Physiol*. 1977;59:1088–97.
- Liu B, Zhu Y, Zhang T. The R3-MYB gene *GhCPC* negatively regulates cotton Fiber elongation. *PLoS One*. 2015;1:1–17.
- Bolton JJ, Soliman KM, Wilkins TA, Jenkins JN. Aberrant expression of critical genes during secondary Cell Wall biogenesis in a cotton mutant, Ligon Lintless-1 (Li-1). *Comp Funct Genomics*. 2009;2009:659301. <https://doi.org/10.1155/2009/659301>.
- Ding M, Jiang Y, Cao Y, Lin L, He S, Zhou W, et al. Gene expression profile analysis of Ligon lintless-1 (Li1) mutant reveals important genes and pathways in cotton leaf and fiber development. *Gene*. 2014;535:273–85.
- Kwak P, Wang Q, Chen X, Qiu C, Yang Z. Enrichment of a set of microRNAs during the cotton fiber development. *BMC Genomics*. 2009;10:457. <https://doi.org/10.1186/1471-2164-10-457>.
- Wang QQ, Liu F, Chen XS, Ma XJ, Zeng HQ, Yang ZM. Transcriptome profiling of early developing cotton fiber by deep-sequencing reveals significantly differential expression of genes in a fuzzless/lintless mutant. *Genomics*. 2010;96:369–76.
- Ji SJ, Lu YC, Feng JX, Wei G, Li J, Shi Yong-Hui YH, et al. Isolation and analyses of genes preferentially expressed during early cotton fiber development by subtractive PCR and cDNA array. *Nucleic Acids Res*. 2003; 31:2534–43.
- Kohel RJ. (1978) Linkage tests in upland cotton. III. 1978;10–3.
- Karaca M, Saha S, Jenkins JN, Zipf A, Kohel R, Stelly DM. Simple sequence repeat (SSR) markers linked to the Ligon lintless (Li(1)) mutant in cotton. *J Hered*. 1999;93:221–4.
- Rong J, Pierce GJ, Waghmare VN, Rogers CJ, Desai A, Chee PW, et al. Genetic mapping and comparative analysis of seven mutants related to seed fiber development in cotton. *Theor Appl Genet*. 2005;111:1137–46.
- Gilbert MK, Turley RB, Kim HJ, Li P, Thyssen G, Tang Y, et al. Transcript profiling by microarray and marker analysis of the short cotton (*Gossypium hirsutum* L.) fiber mutant Ligon lintless-1 (Li1). *BMC Genomics*. 2013;14:403.
- Salih H, Leng X, He S-P, Jia Y, Gong W, Du X-M. Characterization of the early fiber development gene, Ligon-lintless 1 (Li1), using microarray. *Plant Gene*. 2016;6:59–66. <https://doi.org/10.1016/j.plgene.2016.03.006>.
- Liu K, Sun J, Yao L, Yuan Y. Transcriptome analysis reveals critical genes and key pathways for early cotton fiber elongation in Ligon lintless-1 mutant. *Genomics*. 2012;100:42–50.
- Yao Y, Zhang B, Dong CJ, Du Y, Jiang L, Liu JY. Comparative proteomic and biochemical analyses reveal different molecular events occurring in the process of fiber initiation between wild-type allotetraploid cotton and its fuzzless-lintless mutant. *PLoS One*. 2015
- Gilbert MK, Kim HJ, Tang Y, Naoumkina M, Fang DD. 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*. 2014;9:23–5.
- Naoumkina M, Thyssen GN, Fang DD. RNA-seq analysis of short fiber mutants Ligon-lintless-1 (Li 1) and - 2 (Li 2) revealed important role of aquaporins in cotton (*Gossypium hirsutum* L.) fiber elongation. *BMC Plant Biol*. 2015;15:65.
- Zhu Q, Wang M. Molecular Functions of Long Non-Coding RNAs in Plants. 2012;176–190.
- Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, et al. The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. 2012;1775–1789.
- Creville P. Regulation of the floral repressor gene FLC : the complexity of transcription in a chromatin context' n and Caroline Dean. 2011;38–44.
- Sha S, Li J, Sun Q. *Biochimica et Biophysica Acta Functions of plants long non-coding RNAs* ☆. 2016;1859:155–62.
- Bardou F, Ariel F, Simpson CG, Romero-barríos N, Laporte P, Balzergue S, et al. Long noncoding RNA modulates alternative splicing regulators in Arabidopsis. *Dev Cell*. 2014;30:166–76. <https://doi.org/10.1016/j.devcel.2014.06.017>.
- Kornblihtt AR. A long noncoding way to alternative splicing in plant development. *Dev Cell*. 2014;30:117–9.
- Liu X. Long non-coding RNAs and their biological roles in plants. *Genomics Proteomics Bioinformatics*. 2015;13:137–47. <https://doi.org/10.1016/j.gpb.2015.02.003>.
- 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.). 2015;1181–97.
- Heo JB, Sung S, Kim DH, Doyle MR, Sung S, Amasino RM, et al. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science*. 2011;331:76–9. <https://doi.org/10.1126/science.1197349>.
- Ding J, Shen J, Mao H, Xie W, Li X, Zhang Q. RNA-directed DNA methylation is involved in regulating photoperiod-sensitive male sterility in rice. *Mol Plant*. 2012;5:1210–6.
- Naoumkina M, Thyssen GN, Fang DD, Hinchliffe DJ, Florane CB. Small RNA sequencing and degradome analysis of developing fibers of short fiber revealed a role for miRNAs and their targets in cotton fiber elongation. *BMC Genomics*. 2016;17:1–15. <https://doi.org/10.1186/s12864-016-2715-1>.
- Mujahid H, Pendarvis K, Reddy J, Nallamilli B, Reddy K, Nanduri B, et al. Comparative proteomic analysis of cotton Fiber development and protein extraction method comparison in late stage fibers. *Proteomes*. 2016;4:7. <https://doi.org/10.3390/proteomes4010007>.
- Zhao P-M, Wang L-L, Han L-B, Wang J, Yao Y, Wang H-Y, et al. Proteomic identification of differentially expressed proteins in the Ligon lintless mutant of upland cotton (*Gossypium hirsutum* L.). *J Proteome Res*. 2010;9:1076–87. <https://doi.org/10.1021/pr900975t>.
- Zhang T, Hu Y, Jiang W, Fang L, Guan X, Chen J, et al. Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nat Biotechnol*. 2015;33:531–7.
- Kim D, Perteza G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. 2013.
- Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. *Nat Methods*. 2012;9:357–9.
- 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;28:511–5. <https://doi.org/10.1038/nbt.1621>.
- Guttman M, Garber M, Levin JZ, Donaghey J, Robinson J, Adiconis X, et al. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat Biotechnol*. 2010;28:503–10. <https://doi.org/10.1038/nbt.1633>.
- 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 Res*. 2007;35(SUPPL):2.
- 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 Res*. 2013;41:e166.
- Punta M, Coghill P, Eberhardt R, Mistry J, Tate J, Boursnell C, et al. The Pfam protein families databases. *Nucleic Acids Res*. 2012;30:1–12 40 D290-D301. <https://doi.org/10.1093/nar/gkp985>.
- Thyssen GN, Fang DD, Turley RB, Florane CB, Li P, Mattison CP, et al. A Gly65Val substitution in an actin, GHACTION1, disrupts cell polarity and F-actin organization resulting in dwarf, lintless cotton plants. *Plant J*. 2017;90:111–21.
- Ma Q, Wu M, Pei W, Li H, Li X, Zhang J, et al. Quantitative phosphoproteomic profiling of fiber differentiation and initiation in a fiberless mutant of cotton. *BMC Genomics*. 2014;15:466. <https://doi.org/10.1186/1471-2164-15-466>.
- Chaudhary B, Hovav R, Rapp R, Verma N, Udall JA, Wendel JF. Global analysis of gene expression in cotton fibers from wild and domesticated *Gossypium barbadense*. *Evol Dev*. 2008;10:567–82.

45. Dasani SH, Thaker VS. Role of abscisic acid in cotton fiber development. *Russ J Plant Physiol.* 2006;53:62–7. <https://doi.org/10.1134/S1021443706010080>.
46. Yoon J-H, Abdelmohsen K, Gorospe M. Functional interactions among microRNAs and long noncoding RNAs. *Semin Cell Dev Biol.* 2014;09:1–14. <https://doi.org/10.1016/j.semcdb.2014.05.015>.
47. 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 Exp Bot.* 2014;65:4975–83.
48. Rhoades MW, Reinhart BJ, Lim LP, Burge CB, Bartel B, Bartel DP. Prediction of plant microRNA targets. *Cell.* 2002;110:513–20.
49. Wang M, Sun R, Li C, Wang Q, Zhang B. MicroRNA expression profiles during cotton (*Gossypium hirsutum* L) fiber early development. *Sci Rep.* 2017;7:44454. <https://doi.org/10.1038/srep44454>.
50. Xie F, Jones DC, Wang Q, Sun R, Zhang B. Small RNA sequencing identifies miRNA roles in ovule and fibre development. *Plant Biotechnol J.* 2015;13:355–69.
51. 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. *Sci Rep.* 2017;7:3.
52. Wang HY, Wang J, Gao P, Jiao GL, Zhao PM, Li Y, et al. Down-regulation of GhADF1 gene expression affects cotton fibre properties. *Plant Biotechnol J.* 2009;7:13–23.
53. Liang W, Fang L, Xiang D, Hu Y, Feng H, Chang L, et al. Transcriptome Analysis of Short Fiber Mutant Ligon lintless-1 ( Li 1 ) Reveals Critical Genes and Key Pathways in Cotton Fiber Elongation and Leaf Development. 2015; 1:1–18.
54. Li XB, Cai L, Cheng NH, Liu JW. Molecular characterization of the cotton GhTUB1 gene that is preferentially expressed in fiber. *Plant Physiol.* 2002; 130:666–74.
55. Khemka N, Singh VK, Garg R, Jain M, Djebali S, Guttman M, et al. Genome-wide analysis of long intergenic non-coding RNAs in chickpea and their potential role in flower development. *Sci Rep.* 2016;6(August):33297. <https://doi.org/10.1038/srep33297>.
56. Zhang Y-C, Liao J-Y, Li Z-Y, Yu Y, Zhang J-P, Li Q-F, et al. Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice. *Genome Biol.* 2014;15:512. <https://doi.org/10.1186/s13059-014-0512-1>.
57. Zhu B, Yang Y, Li R, Fu D, Wen L, Luo Y, et al. RESEARCH PAPER RNA sequencing and functional analysis implicate the regulatory role of long non-coding RNAs in tomato fruit ripening. 2015; Pol II.
58. Cui J, Luan Y, Jiang N, Bao H, Meng J, Science L, et al. Comparative transcriptome analysis between resistant and susceptible tomato allows the identification of lincRNA16397 conferring resistance to. *Plant J.* 2016.
59. Lu X, Chen X, Mu M, Wang J, Wang X, Wang D, et al. Genome-wide analysis of long noncoding rnas and their responses to drought stress in cotton (*Gossypium hirsutum* L.). *PLoS One.* 2016;11:1–18.
60. An W, Gong W, He S, Pan Z, Sun J, Du X. MicroRNA and mRNA expression profiling analysis revealed the regulation of plant height in *Gossypium hirsutum*. *BMC Genomics.* 2015;16:886. <https://doi.org/10.1186/s12864-015-2071-6>.
61. Yoo MJ, Wendel JF. Comparative evolutionary and developmental dynamics of the cotton (*Gossypium hirsutum*) Fiber transcriptome. *PLoS Genet.* 2014; 10:e1004073.
62. Ho MH, Saha S, Jenkins JN, Ma DP. Characterization and promoter analysis of a cotton RING-type ubiquitin ligase (E3) gene. *Mol Biotechnol.* 2010;46:140–8.
63. Mei W, Qin Y, Song W, Li J, Zhu Y. Cotton GhPOX1 encoding plant class III peroxidase may be responsible for the high level of reactive oxygen species production that is related to cotton fiber elongation. *J Genet Genomics.* 2009;36:141–50.
64. Zou C, Wang Q, Lu C, Yang W, Zhang Y, Cheng H, et al. Transcriptome analysis reveals long noncoding RNAs involved in fiber development in cotton (*Gossypium arboreum*). *Sci China Life Sci.* 2016;59:164–71.
65. Necsulea A, Soumillon M, Warnefors M, Liechti A, Daish T, Zeller U, et al. The evolution of lincRNA repertoires and expression patterns in tetrapods. *Nature.* 2014;505:635–40.
66. Shi YH, Zhu SW, Mao XZ, Feng JX, Qin YM, Zhang L, et al. Transcriptome profiling, molecular biological, and physiological studies reveal a major role for ethylene in cotton fiber cell elongation. *Plant Cell.* 2006;18:651–64. <https://doi.org/10.1105/tpc.105.040303>.
67. Qin Y-M, Hu C-Y, Pang Y, Kastaniotis AJ, Hiltunen JK, Zhu Y-X. Saturated very-long-chain fatty acids promote cotton Fiber and Arabidopsis cell elongation by activating ethylene biosynthesis. *Plant Cell Online.* 2007;19:3692–704. <https://doi.org/10.1105/tpc.107.054437>.
68. Giovane A, Servillo L, Balestrieri C, Raiola A, D'Avino R, Tamburrini M, et al. Pectin methylesterase inhibitor. *Biochimica et Biophysica Acta - Proteins and Proteomics.* 1696;2004:245–52.
69. Li W, Shang H, Ge Q, Zou C, Cai J, Wang D, et al. Genome-wide identification, phylogeny, and expression analysis of pectin methylesterases reveal their major role in cotton fiber development. *BMC Genomics.* 2016;17.
70. Deng T, Yao H, Wang J, Wang J, Xue H, Zuo K. GhLTPG1, a cotton GPI-anchored lipid transfer protein, regulates the transport of phosphatidylinositol monophosphates and cotton fiber elongation. *Sci Rep.* 2016;6:26829.
71. Kumar S, Pandey P, Kumar K, Rajamani V, Padmalatha KV, Dhandapani G, et al. Delineating the glycoproteome of elongating cotton fiber cells. *Data Br.* 2015;5:717–25.
72. Salih H, Gong W, He S, Mustafa NS, Du X. Comparative transcriptome analysis of TUCPs in *Gossypium hirsutum* Ligon-lintless-1 mutant and their proposed functions in cotton fiber development. *Mol Gen Genomics.* 2018; 0:0. <https://doi.org/10.1007/s00438-018-1482-x>.
73. Ghanbarian AT, Hurst LD. Neighboring genes show correlated evolution in gene expression. *Mol Biol Evol.* 2015;32:1748–66.
74. Hou M, Tang X, Tian F, Shi F, Liu F, Gao G. AnnoLnc: a web server for systematically annotating novel human lincRNAs. *BMC Genomics.* 2016;17:931.
75. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008;9:559.
76. Meyers BC, Axtell MJ, Bartel B, Bartel DP, Baulcombe D, Bowman JL, et al. Criteria for Annotation of Plant MicroRNAs. 2008;20 December:3186–90.
77. Audic S, Claverie JM. The significance of digital gene expression profiles. *Genome Res.* 1997;7:986–95.
78. Wu HJ, Ma YK, Chen T, Wang M, Wang XJ. PsRobot: a web-based plant small RNA meta-analysis toolbox. *Nucleic Acids Res.* 2012;40.
79. Salih H, Gong W, Mkulama M, Du X. Analysis of the WD40 protein family in cotton. 2018;547 May:539–47.

## 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](https://biomedcentral.com/submissions)

