



# Genome-Wide Transcriptome Profiling Revealed Cotton Fuzz Fiber Development Having a Similar Molecular Model as *Arabidopsis* Trichome

Qun Wan, Hua Zhang, Wenxue Ye, Huaitong Wu, Tianzhen Zhang\*

National Key Laboratory of Crop Genetics and Germplasm Enhancement, Cotton Research Institute, Nanjing Agricultural University, Nanjing, China

## Abstract

The cotton fiber, as a single-celled trichome, is a biological model system for studying cell differentiation and elongation. However, the complexity of gene expression and regulation in the fiber complicates genetic research. In this study, we investigated the genome-wide transcriptome profiling in Texas Marker-1 (TM-1) and five naked seed or fuzzless mutants (three dominant and two recessive) during the fuzz initial development stage. More than three million clean tags were generated from each sample representing the expression data for 27,325 genes, which account for 72.8% of the annotated *Gossypium raimondii* primary transcript genes. Thousands of differentially expressed genes (DEGs) were identified between TM-1 and the mutants. Based on functional enrichment analysis, the DEGs downregulated in the mutants were enriched in protein synthesis-related genes and transcription factors, while DEGs upregulated in the mutants were enriched in DNA/chromatin structure-related genes and transcription factors. Pathway analysis showed that ATP synthesis, and sugar and lipid metabolism-related pathways play important roles in fuzz initial development. Also, we identified a large number of transcription factors such as MYB, bHLH, HB, WRKY, AP2/EREBP, bZIP and C2H2 zinc finger families that were differently expressed between TM-1 and the mutants, and were also related to trichome development in *Arabidopsis*.

**Citation:** Wan Q, Zhang H, Ye W, Wu H, Zhang T (2014) Genome-Wide Transcriptome Profiling Revealed Cotton Fuzz Fiber Development Having a Similar Molecular Model as *Arabidopsis* Trichome. PLoS ONE 9(5): e97313. doi:10.1371/journal.pone.0097313

**Editor:** Samuel P. Hazen, University of Massachusetts Amherst, United States of America

**Received:** December 12, 2013; **Accepted:** April 16, 2014; **Published:** May 13, 2014

**Copyright:** © 2014 Wan et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The State Basic Research Development Program of China (973 Program, 2011CB109300), the National High Technology Research and Development Program of China (863 Program) (2011AA10A102), and the Priority Academic Program Development of Jiangsu Higher Education Institutions. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist. As a PLOS ONE Editorial Board member, this does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

\* E-mail: cotton@njau.edu.cn

## Introduction

Cotton (*Gossypium* spp.) is an important commercial crop and the largest source of natural textile fibers grown throughout the world. Cotton fibers used in textiles originate from the outer epidermal layer of the maturing seed, and are classified into two types: lint and fuzz. Initiation of lint fibers is a quasi-synchronous process that occurs in developing ovules during anthesis. The fuzz fibers initiate growth at 4 DPA (days post anthesis) and elongate to approximately 0.5 cm, much shorter than lint fibers [1].

Many genes from *Arabidopsis* have been identified that control the initiation and morphogenesis of trichomes, and most of them encode transcription factors including MYB (*GLABROUS1*, *TRIPTYCHON*, *CAPRICE*, *WEREWOLF*) [2–5], WD-40 type (*TRANSPARENT TESTA GLABRA1*) [6,7], bHLH (*GLABROUS3*) [8,9], HD-ZIP (*GLABROUS2*) [10] and a WRKY-related transcription factor (*TRANSPARENT TESTA GLABRA2*) [11]. *TRANSPARENT TESTA GLABRA 1* (*TTG1*) encodes a small protein with WD-repeats, although no WD-repeat protein has either enzymatic activity or a DNA binding domain [12]. The identification of TTG1 as a WD40 repeat-containing protein suggests that TTG1 regulates MYC transcription factors or pathways in which MYC factors are involved [7]. *TTG2* encodes a WRKY transcription factor and acts downstream of the trichome initiation genes, *TTG1* and *GL1* [11]. bHLH family members have a basic helix-loop-

helix domain [13]. Mutant analyses have identified several plant bHLH proteins involved in anthocyanin biosynthesis, such as GL3, EGL3 and TT8 [8,9,14–16].

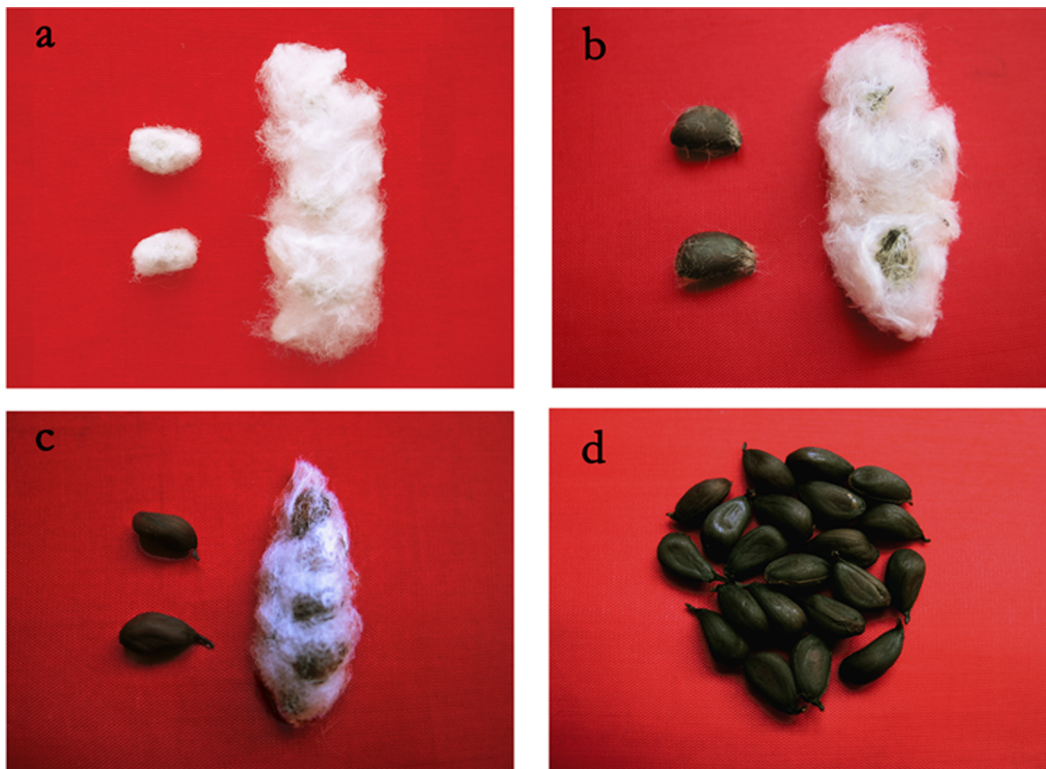
Cotton fibers share many similarities with *A. thaliana* leaves trichome development, and several studies have demonstrated a close relationship between these two types of cells using cotton fiber-related genes (Table S1). Six putative cotton MYB genes (*GhMYB1-6*) have been isolated, and these DNA-binding factors were shown to be involved in the differentiation and expansion of cotton seed trichomes [17]. *GhMYB109*, which encodes a R2R3 MYB transcription factor, was shown to be expressed specifically in fiber cell initials and expanding fibers [18]. Another R2R3 MYB gene, *GaMYB2*, which is homologous to *AtGL1*, was predominantly expressed early in cotton fibers and complemented *gl1* phenotypes in *Arabidopsis* [19]. Overexpressing *GhMYB2* or its downstream gene *GhRDL1* in *Arabidopsis* activates fiber-like hair production on 4–6% of the seed coats and has no obvious effect on trichome development in leaves or siliques [20]. In addition, overexpression of *GbMYB2* in *Arabidopsis* caused thicker leaf trichomes and longer roots to develop due to the activation of trichome development-related genes such as *GL2* [21]. *GhMYB25* encodes a homolog of *AmMIXTA/AmMYBML1* which involved in epidermal cell differentiation, is highly expressed in ovules, fiber cell initials and trichomes on leaf. Silencing of *GhMYB25* in cotton showed fiber and trichome development were suppressed, while

overexpression of *GhMYB25* increased cotton fibre initiation and leaf trichome number [22–26]. *GhMYB25-like* had a similar expression pattern with *GhMYB25* which significantly higher expression during fiber cell initiation (–3~3 DPA). Transgenic plants showed *GhMYB25-like* had significant regulatory roles in cotton fiber development. RNA interference suppression of *GhMYB25-like* resulted in cotton plants with fibreless seeds, but normal trichomes elsewhere implying *GhMYB25-like* playing a crucial role in the very early stages of fiber cell differentiation [26,27]. A cotton gene encoding an *Arabidopsis CPC* ortholog (*R3 MYB* gene) was identified and downregulated in fiber initials at 1 DPA [28]. In addition to the MYB genes, four putative homologues of *Arabidopsis TTG1* (*GhTTG1–GhTTG4*), have been isolated and were shown to form two groups, with *GhTTG1* and *GhTTG3* being closely related to each other, and *GhTTG2* and *GhTTG4* forming the second group, based on sequence comparisons of the four deduced proteins and *Arabidopsis TTG1* [29]. Three homeobox (HOX) genes, *GhHOX1*, *GhHOX2*, and *GhHOX3*, have been identified from cotton, showing 66%, 34%, and 37% protein sequence similarity to *Arabidopsis GL2*, respectively. *GhHOX1* was able to restore the glabrous phenotype of *gl2* mutant, indicating that this protein is a functional homologue of GL2 in controlling trichome development and may function in fiber development [30]. Two GL3-like bHLH cDNAs from cotton ovule, *GhDEL65* and *GhDEL61*, have been deposited in the Genbank [31,32]. It will be interesting to examine if they work like GL3 during cotton fiber development. Also, several ESTs (expressed sequence tags) from cotton have been published that share identity with *Arabidopsis* homologues in the NCBI database [33,34]. As many homologous genes have been isolated from

cotton and shown to play similar roles in trichome initiation in *Arabidopsis*, the GL1-GL3/EGL3-TTG1 protein complex may control fiber formation in cotton [33].

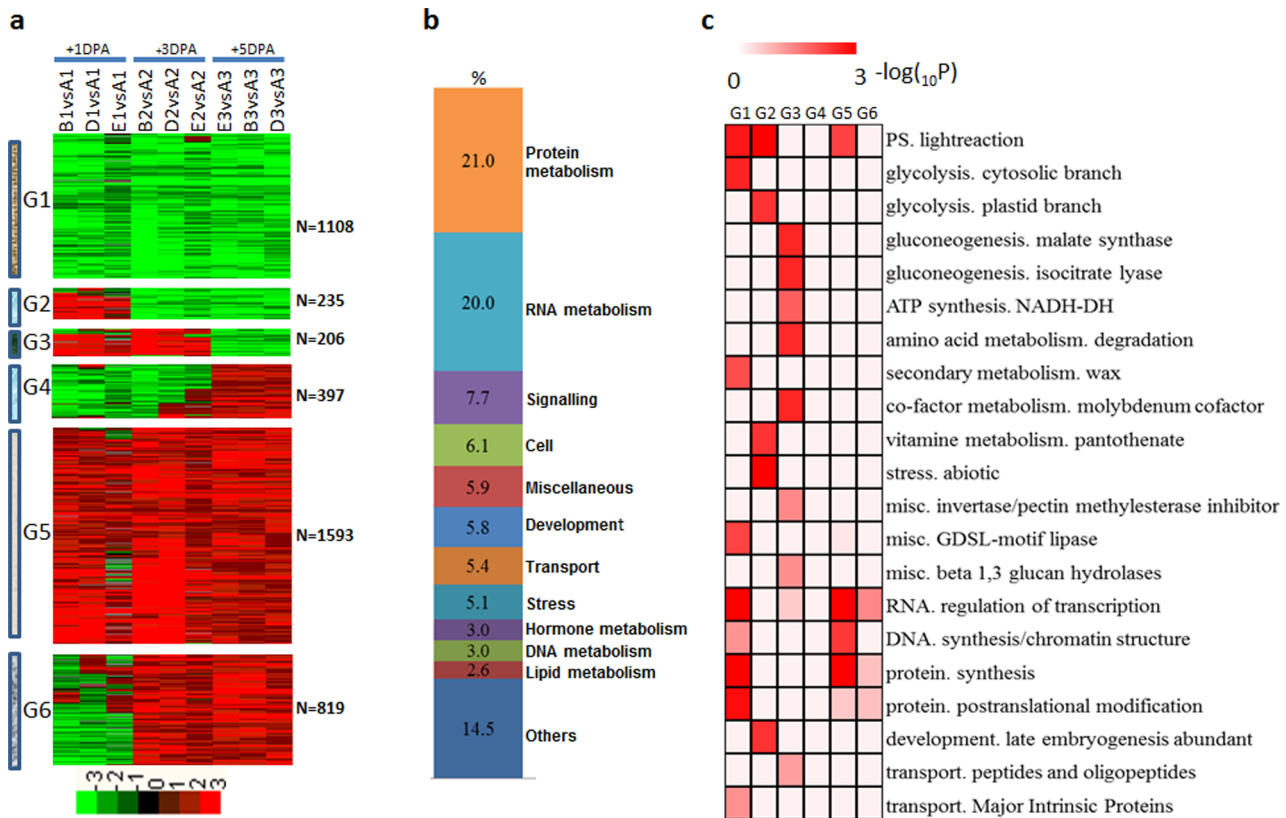
Several “qualitative” mutants in fiber development have been reported. The best characterized of these are the naked seed loci,  $N_1N_1$  and  $n_2n_2$ . The dominant naked seed mutant (NSM)  $N_1NSM$  is fuzzless but with a little lint on the seed [35]. The recessive naked seed mutant  $n_2NSM$  produces regular lint, but bears a naked seed phenotype with very little fuzz fibers present at the micropyle tips of the seed [36]. Fuzzless-lintless mutants (FLM) XZ142FLM, MD17FLM and SL1-7-1FLM are all completely without any fiber; SL1-7-1FLM possess the dominant naked seed gene  $N_1$  [37,38], XZ142FLM possess recessive naked seed gene  $n_2$  [39,40], while MD17FLM possess both  $N_1$  and  $n_2$  [37,41]. TM-1 with lint and fuzz fiber is upland genetic standard line, which widely used in research programs [42]. Although these six materials have different genetic background, critical genes or pathways can be identified by studying the common different expressed genes between WT and several same genotype mutants.

To gain a better understanding of gene regulation in the early stage of fuzz development, we present here the first genome-wide analysis of gene expression during cotton fuzz initial cell development using massively parallel deep-sequencing developed by Solexa/Illumina. As cell fate determination for fiber (lint and fuzz) must occur prior to the formation of fiber cell initials, we selected +1, +3 and +5 DPA ovules to analyze fuzz initial development. In this study, we annotated thousands of read signatures matching predicted genes, and quantified the transcript abundance in developing ovules and fibers. In addition, we have profiled gene expression in the mutants against fuzz-bearing ovules



**Figure 1. Cotton fiber morphology in the wild-type line and the mutant lines.** a: wild line: TM-1; b: recessive naked-seed mutant:  $n_2NSM$ ; c: dominant naked-seed mutant:  $N_1NSM$ ; d: fuzzless-lintless mutants: SL1-7-1FLM (with  $N_1$  gene), XZ142FLM (with  $n_2$  gene), MD17FLM (with  $N_1$  and  $n_2$ ). Matured seeds were separated from the opened bolls on the cotton plant. Ginned seeds (right) and matured seeds (left) showed on linted-fuzzy and linted-fuzzless panel.

doi:10.1371/journal.pone.0097313.g001



**Figure 2. Dynamic progression of common differentially expressed genes in the dominant mutants.** (a) Unsupervised hierarchical clustering of the 4,358 common DEGs in the dominant mutants. Common DEGs were clustered into six groups and the number of genes of each group was listed at right. Red region, genes upregulated in the mutants; green region, genes downregulated in the mutants. A, TM-1; B, SL1-7-1FLM; D, MD17FLM; E, N<sub>1</sub>NSM; 1, +1 DPA; 2, +3 DPA; 3, +5 DPA. (b) Functional distribution of common DEGs in the dominant mutants. (c) Functional category enrichment of common differentially expressed genes in the dominant mutants. doi:10.1371/journal.pone.0097313.g002

(wild type, WT), and found large changes in gene expression in the mutants.

**Materials and Methods**

**Plant Material Preparation and Total RNA Isolation**

*G. hirsutum* cv. Texas Marker-1 (TM-1) and five naked-seed or fuzzless mutants (XZ142FLM, MD17FLM, SL1-7-1FLM, N<sub>1</sub>NSM and n<sub>2</sub>NSM) were used in this study (Figure 1). SL1-7-1FLM, MD17FLM and N<sub>1</sub>NSM each possess the dominant naked seed gene *N<sub>1</sub>*, while XZ142FLM and n<sub>2</sub>NSM carry the recessive naked seed gene *n<sub>2</sub>*.

Plants were grown at Jiangpu Breeding Station, Nanjing (JBS/NAU) in 2010. All lines were self-pollinated, and the progeny were tested to verify the initial pattern. Buds were tied up the day before anthesis to ensure self-pollination. Bolls were harvested at +1, +3 and +5 DPA. Ovules were excised carefully from bolls, frozen in liquid nitrogen immediately, and stored at -70°C. Total RNA was extracted using the CTAB method [43].

**Sequencing and Digital Tag Profiling**

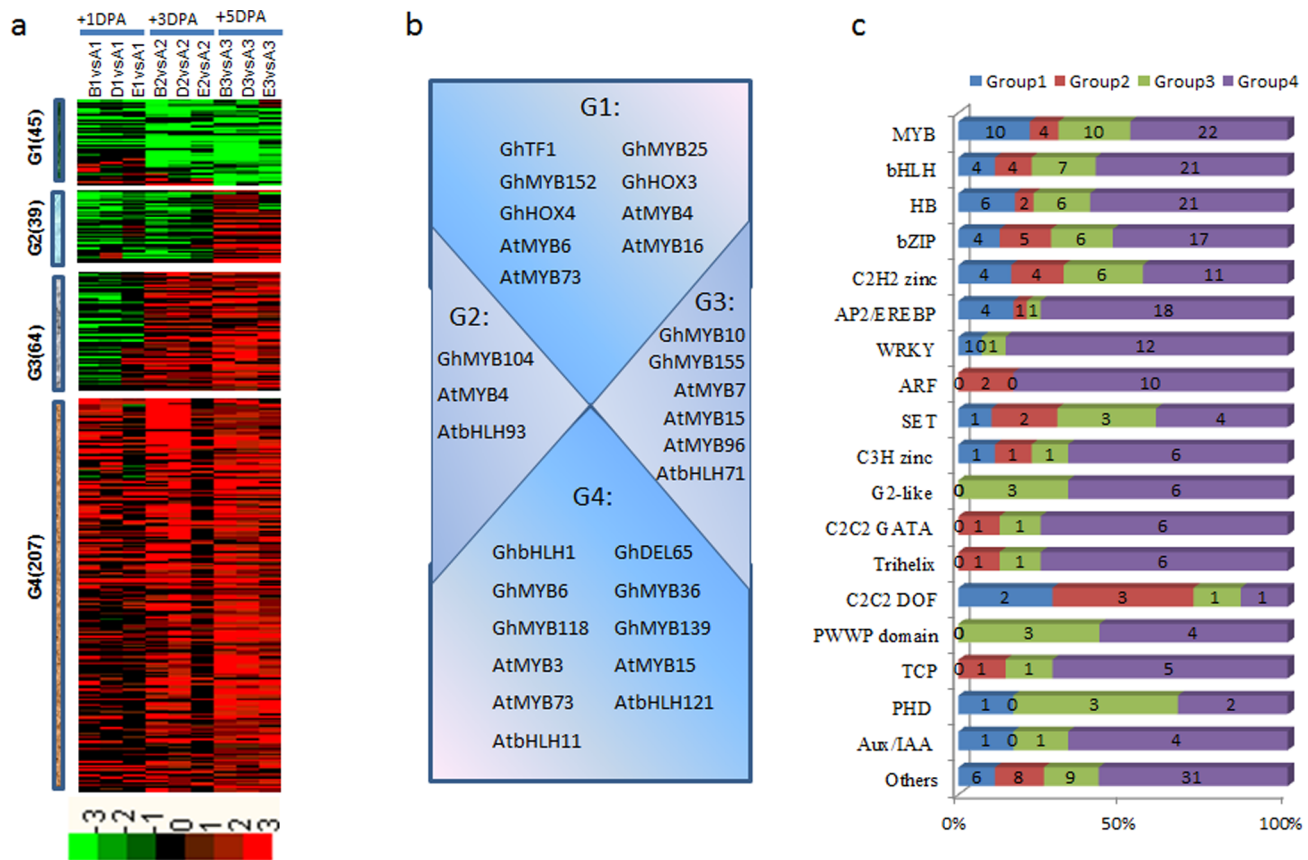
Library construction, sequencing and raw data processing were performed commercially by BGI (Beijing Genomics Institute at Shenzhen, China) via the sequencing by synthesis (SBS) on Illumina HiSeq 2000 System as described previously [44]. Digital tag profiling was performed as described by Wang et al [44] and

*Gossypium raimondii* primary transcript sequences (<http://www.phytozome.net>) was used as reference gene database.

**Defining Differentially Expressed Genes and Cluster Analysis**

Statistical analysis was performed to identify differentially expressed genes between the libraries using a rigorous algorithm described previously [45]. Gene expression was normalized to transcripts per million (TPM) clean tags. For gene expression variance, the statistical *t*-test was used to identify genes differently expressed between the libraries. *P* values were adjusted using the multiple testing procedures described by Benjamini and Yekutieli [46] for controlling the false discovery rate (FDR). In this study, we used a stringent value of FDR <0.001, and the absolute value of |log<sub>2</sub>Ratio| ≤ 1 as the threshold to judge the significant difference of gene expression.

K means clustering was performed with the open-source program Cluster3.0 (<http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm>). The genes in each cluster were then classified into Mapman functional categories [47]. Functional categories of the MapMan annotation were tested for significance of expression change by applying a two-sided Wilcoxon rank test with a Benjamini Yekutieli correction for multiple tests. Pathway analysis was mainly based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [48].



**Figure 3. Dynamics of transcription factor expression profiles in various dominant mutants.** (a) Unsupervised hierarchical clustering of 355 transcription factor genes included in the 4,358 common DEGs in the dominant mutants. Four groups were generated and the number of each group was in parentheses. Red region, genes upregulated in the mutants; green region, genes downregulated in the mutants. A, TM-1; B, SL1-7-1FLM; D, MD17FLM; E, N<sub>1</sub>NSM; 1, +1 DPA; 2, +3 DPA; 3, +5 DPA. (b) Representative functions and genes showing expression gradients. (c) Distribution of transcription factor families among G1, G2, G3 and G4. doi:10.1371/journal.pone.0097313.g003

### Quantitative Real Time RT-PCR (qPCR)

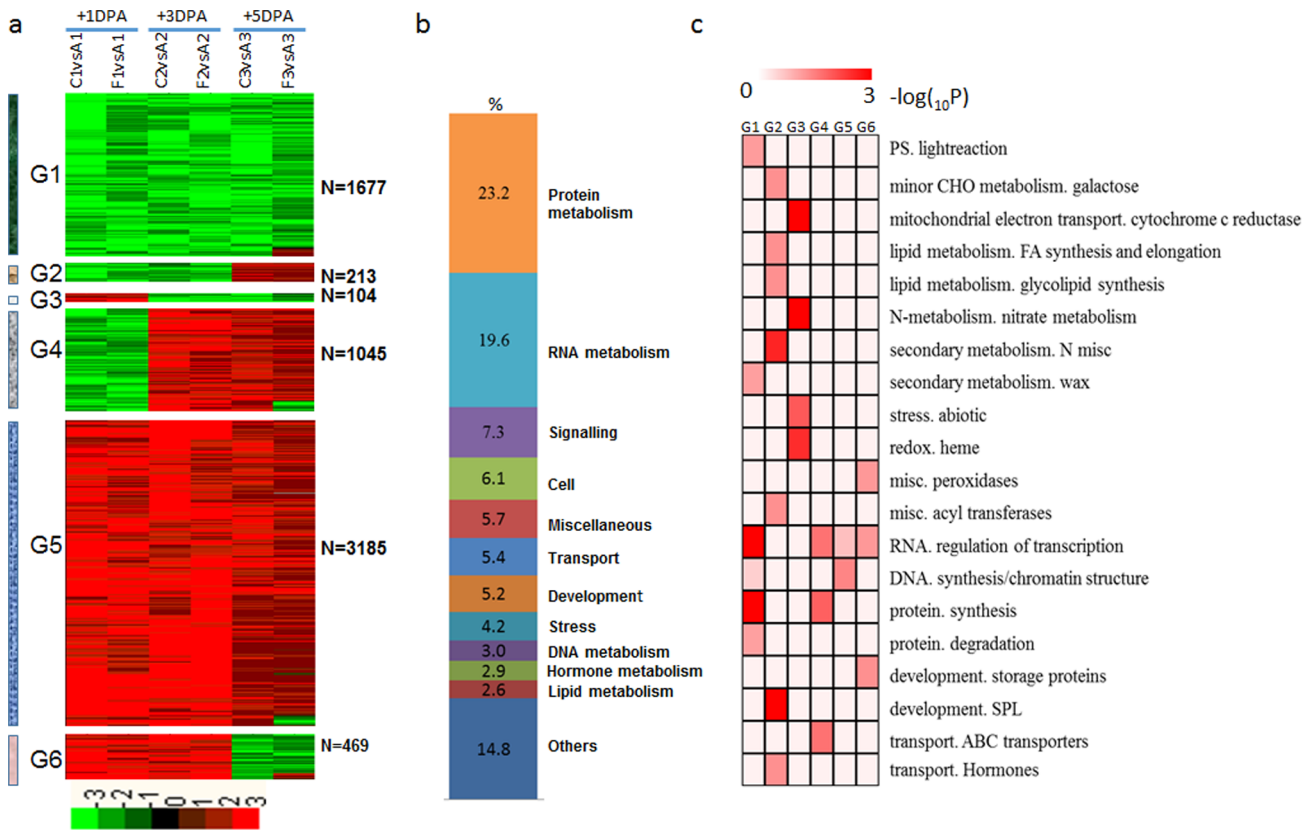
Verification of some differentially expressed genes (DEGs) was performed by real-time quantitative PCR (qPCR). The primers for the various genes were designed with Primer 3.0 (<http://frodo.wi.mit.edu/cgi-bin/primer3/primer3>), and synthesized commercially (Genscript, Nanjing, China); sequences are given in Table S10. Two microgramme total RNA was reversely transcribed using PrimeScript RT reagent Kit with gDNA Eraser (Perfect Real Time) (TaKaRa, Shiga, Japan). QPCR was performed using the LightCycler FastStart DNA Master SYBR Green I kit (Roche, Basel, Switzerland) in an ABI7500 Real-Time PCR detection system (Applied Biosystems, San Francisco, CA, USA). Each sample was PCR-amplified using 100ng cDNA template in triple reactions. The cotton *histone 3* gene [19] (ACC No. AF024716) was used as the positive control and amplified with the primer pair (F: 5'-GGTGGTGTGAAGAAGCCTCAT-3', and R: 5'-AAT-TTCACGAACAAGCCTCTGGAA-3'). The amplification efficiency of each gene was calculated. The qRT-PCR cycles were as follows: (1) 95°C, 10 min; (2) 40 cycles of 95°C for 15 s, ~60°C (temperature varied for different primers, Table S10) for 30 s and 72°C for 30 s; (3) a melting curve analysis from 65 to 95°C (1 s hold per 0.2°C increase) to check the specificity of the amplified product. Relative expression levels were determined by the  $2^{-\Delta Ct}$  method.

### Results

#### Sequencing Data Analysis

To obtain a global view of transcription relevant to cotton fuzz development, we used the Illumina HiSeq 2000 System to perform high-throughput tag-seq analysis on poly(A)-enriched RNAs from eighteen cotton ovule libraries including the cultivar TM-1 and five mutants during the fuzz initiation stage (+1 DPA, +3 DPA and +5 DPA). The total number of tags per library ranged from 3.5 to 4.7 million, and the number of tags with distinct sequences ranged from 0.27 to 0.44 million (Table S2). After removal of low quality tags, we obtained a total of 3.4 to 4.5 million clean tags that corresponded to about 0.15 million distinct tags (Table S2). The distribution of total and distinct tag counts over different tag abundance categories showed very similar profiles for all libraries (Figure S1). Among the distinct tags, less than 5% had a copy number higher than 100, whereas 38% of the tags were present between 5 and 50 copies, and more than 57% of the transcripts had 2–5 copies.

As there was no allotetraploid cotton genome sequence available, clean tags were mapped to *G. raimondii* genome sequence (<http://www.phytozome.net>). Approximately 73%–82% of the distinct tags (83–87% of the total tags) could be mapped to the reference genome (Table S2). All clean tags were aligned to the reference *G. raimondii* primary transcript sequences. Approximately 26%–35% of the distinct tags could be uniquely mapped to the



**Figure 4. Dynamic progression of common differentially expressed genes in the recessive mutants.** (a) Unsupervised hierarchical clustering of 6,693 common differentially expressed genes in the recessive mutants. Common DEGs were clustered into six groups and the number of genes of each group was listed at right. Red region, genes upregulated in the mutants; green region, genes downregulated in the mutants. A, TM-1; C, XZ142FLM; F, n<sub>2</sub>NSM; 1, +1 DPA; 2, +3 DPA; 3, +5 DPA. (b) Functional category distribution of common DEGs in the recessive mutants. (c) Functional category enrichment of common differentially expressed genes in the recessive mutants. doi:10.1371/journal.pone.0097313.g004

reference sequence. The tags that mapped to the database generated 19,829–22,213 tag-mapped transcripts for the libraries (Table S2).

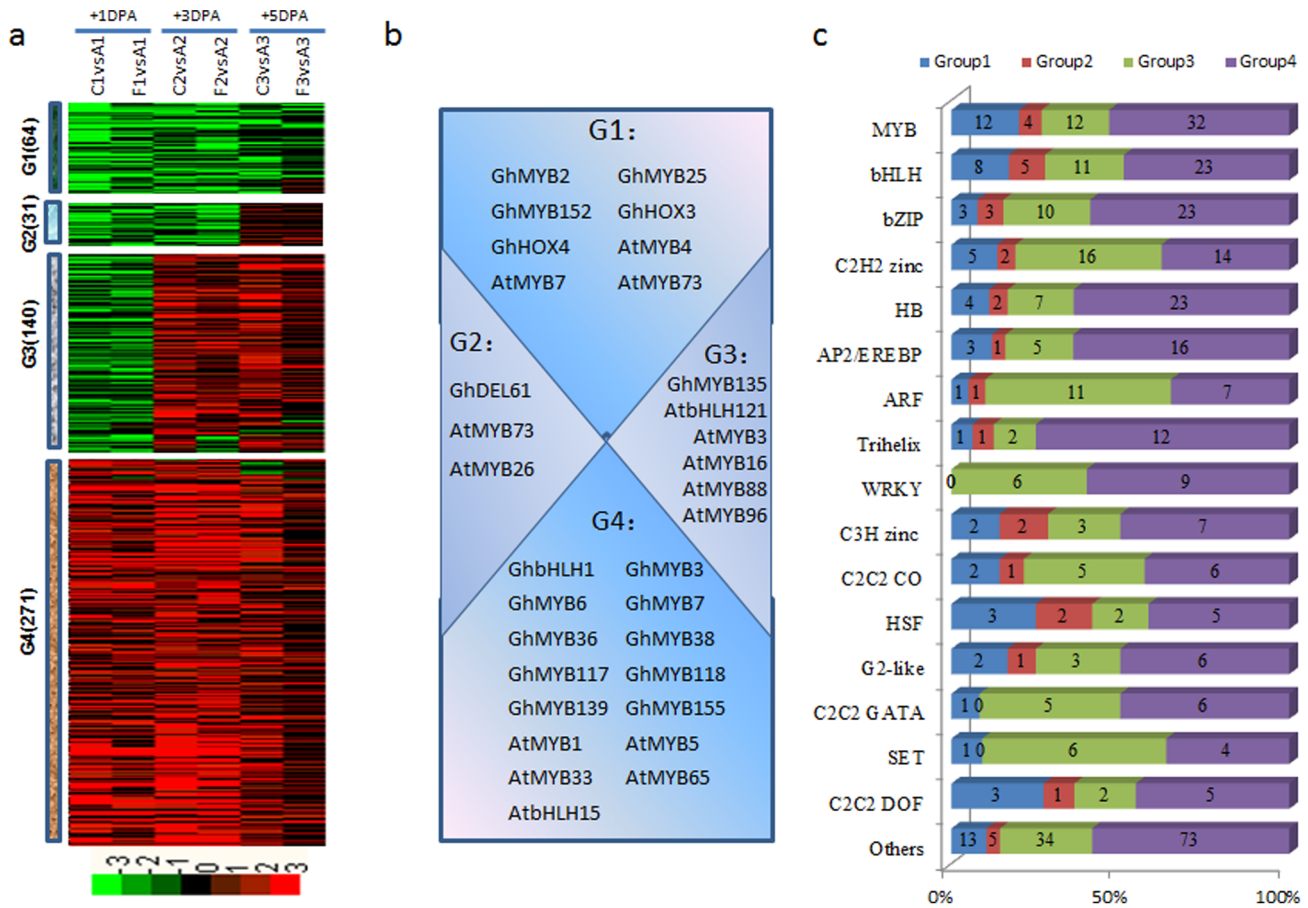
### Common DEGs between Dominant Naked-seed Mutants and TM-1 during Fuzz Development

To understand the molecular mechanisms of the dominant fuzzy phenotype, 4,358 common DEGs differentially expressed between the mutants MD17FLM, SL1-7-1FLM, N<sub>1</sub>NSM and the wild-type TM-1 were identified. Of these, 268 genes were up regulated and 557 genes down regulated at +1 DPA; 792 genes were up regulated and 699 genes down regulated at +3 DPA in the mutants; and 2,000 genes were up regulated with 957 genes down regulated at +5 DPA. Ten common differentially up regulated genes and 62 down regulated genes were identified at +1 DPA, +3 DPA and +5 DPA (Figure S2, Table S3).

We then used MapMan annotation to assign genes to functional categories and grouped the genes into six groups using the hierarchical clustering algorithm. Two main groups (Groups 1 and 5) accounted for ~62% of the DEGs at the three sampling times (Figure 2a). Excluding 866 genes belonging to the ‘not assigned or unknown’ categories, 3,492 genes had MapMan annotation assignments. Among these, 21.0% are related to protein metabolism, 20.0% to RNA metabolism, 7.7% to signaling, and the remaining genes to cell functions, development, transport, stress, hormone metabolism, DNA metabolism, or lipid metabolism

(Figure 2b). To further explore this dataset, we tested for enrichment by MapMan functional category using Fisher’s exact test ( $P < 0.01$ , FDR = 5%). Most of the MapMan bins showed enrichment for particular groups of expressed genes (Figure 2c); for example, genes that encode enzymes for protein synthesis and transcription factors in Group 1, light reaction and abiotic stress in Group 2, ATP synthase, amino acid metabolism, glyoxylate cycle in Group 3, and transcription factors and DNA synthesis in Group 5.

The dynamics of transcription factor accumulation during fuzz initiation were particularly well resolved in our data. Of the 1,596 differentially expressed transcription factor genes that we detected in the fuzz initial stage, 355 were common to the dominant mutants (Figure 3). Most of these genes (207) were upregulated in the dominant mutants (Group 4), including *GhbHLH1*, *GhDEL65*, *GhMYB6*, *GhMYB118*, *GhMYB139*, *AtMYB3*, *AtMYB73*, *AtbHLH121* and *AtbHLH11*. Only 45 transcription factors were downregulated (Group 1), including *GhTF1*, *GhMYB25*, *GhMYB152*, *GhHOX3*, *GhHOX4*, *AtMYB4*, *AtMYB6*, *AtMYB16* and *AtMYB73*. Thirty nine transcription factors showed downregulation at +1 DPA and +3 DPA, and upregulation at +5 DPA (Group 2); these included *GhMYB104*, *AtMYB4*, and *AtbHLH93*. An additional 64 transcription factors showed downregulation at +1 DPA, and upregulation at +3 DPA and +5 DPA (Group 3); these included *GhMYB10*, *GhMYB155*, *AtMYB7*, *AtMYB15*, *AtMYB96*, *AtbHLH71* (Figure 3a and 3b). We also identified family-specific expression trends (Figure 3c). Members of the C2C2 DOF



**Figure 5. Dynamics of transcription factor accumulation profiles in various recessive mutants.** (a) Unsupervised hierarchical clustering of 506 transcription factor genes included in the 6,693 common DEGs in the recessive mutants. Four groups were generated and the number of each group was in parentheses. Red region, genes upregulated in the mutants; green region, genes downregulated in the mutants. A, TM-1; C, XZ142FLM; F, n<sub>2</sub>NSM; 1, +1 DPA; 2, +3 DPA; 3, +5 DPA. (b) Representative functions and genes showing expression gradients. (c) Distribution of transcription factor families among G1, G2, G3 and G4. doi:10.1371/journal.pone.0097313.g005

zinc-finger families of transcriptional regulators are highly expressed in TM-1; Aux/IAA, WRKY, AP2/EREBP and G2-like families are highly expressed in the dominant mutants.

To identify the metabolic pathways that are active during fuzz initiation, we mapped the 4,358 commonly expressed genes in the dominant mutants to the reference KEGG canonical pathways. In total, we assigned 1448 genes to KEGG pathways. Most of these mapped to ATP synthesis, or sugar and lipid metabolism-related metabolic pathways such as starch and sucrose metabolism (82 members), oxidative phosphorylation (30 members), galactose metabolism (22 members), glycolysis/gluconeogenesis (31 members), fatty acid degradation (20 members; Table S4). These annotations provide a valuable resource for investigating the processes, functions, and pathways specific to the initiation of fuzz development.

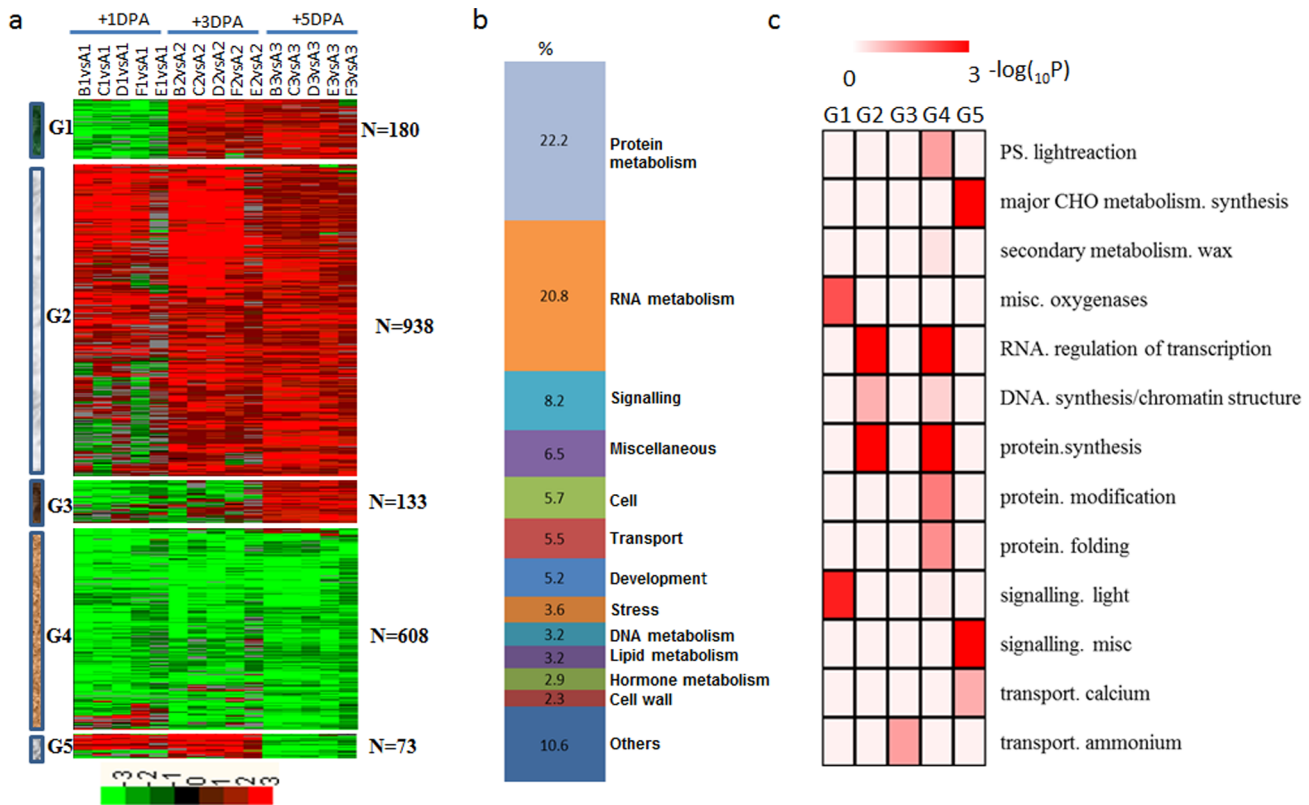
### Common DEGs between Recessive Naked-seed Mutants and TM-1 during Fuzz Development

To understand the molecular mechanisms underlying the recessive naked seed phenotype, 6,693 DEGs were identified that are common to mutants XZ142FLM and n<sub>2</sub>NSM compared with TM-1. Of these, 1,978 genes were up-regulated and 1,480 genes downregulated at +1 DPA; 2,971 genes were upregulated and 980 genes downregulated at +3 DPA; and 1,264 genes were

upregulated and 666 genes downregulated at +5 DPA. There were 192 upregulated genes and 120 downregulated genes common to the differentially expressed genes at +1 DPA, +3 DPA and +5 DPA (Figure S2, Table S5).

We identified six groups using the hierarchical clustering algorithm. Group 1 and Group 5 accounted for ~73% of the differentially expressed genes at the three sampled times (approx. 4,852 genes; Figure 4a). Five-thousand four-hundred twenty-one genes had MapMan annotations, excluding 19.0% belonging to the 'not assigned or unknown' categories. Among the annotated genes, 23.2% are related to protein metabolism, 19.6% to RNA metabolism, 7.3% to signaling, and the rest to cell functions, transport, development, stress, DNA metabolism, hormone metabolism or lipid metabolism (Figure 4b). Most of the MapMan bins showed enrichment for particular groups (Figure 4c); for example, genes that encode enzymes for protein synthesis and transcription factors in Group 1, SPL protein in Group 2, abiotic stress in Group 3, ABC transport in Group 4, DNA synthesis in Group 5, and peroxidases and storage protein in Group 6.

We identified 506 transcription factors that were expressed in common in the two recessive mutants (Figure 5). Among these, 64 transcription factors (Group 1) including *GhMYB2*, *GhMYB25*, *GhMYB152*, *GhHOX3*, *GhHOX4* were downregulated in the recessive mutants, and 271 transcription factors (Group4) were



**Figure 6. Dynamic progression of common differentially expressed genes in the dominant/recessive mutants.** (a) Hierarchical clustering of the 1,932 common DEGs in five mutants. Common DEGs were clustered into five groups and the number of genes of each group was listed at right. Red region, genes upregulated in the mutants; green region, genes downregulated in the mutants. A, TM-1; B, SL1-7-1FLM; C, XZ142FLM; D, MD17FLM; E, N<sub>1</sub>NSM; F, n<sub>2</sub>NSM; 1, +1 DPA; 2, +3 DPA; 3, +5 DPA. (b) Functional distribution of common DEGs in the dominant/recessive mutants. (c) Functional category enrichment of common DEGs in the dominant/recessive mutants. doi:10.1371/journal.pone.0097313.g006

upregulated including *GhbHLH1*, *GhMYB3*, *GhMYB36*, *GhMYB7*, *GhMYB36*, *GhMYB38*, *GhMYB117*, *GhMYB118*, *GhMYB139*, *GhMYB155*. Thirty-one transcription factors (Group 2), such as *GhDEL61*, *GhGL2-like1* were downregulated at +1 DPA and +3 DPA, but were upregulated at +5 DPA, 140 transcription factors (Group 3), such as *GhMYB135* were downregulated at +1 DPA, were upregulated at +5 DPA (Figure 5a and 5b). Members of the HSF families of transcriptional regulators were highly expressed in TM-1; the MYB, WRKY, AP2/EREBP, bHLH, ARF and C2C2(Zn) GATA families were highly expressed in the recessive mutants (Figure 5c).

Of 6,693 common DEGs in the recessive mutants, 2,356 were assigned to KEGG pathways. The pathways with the most representation for the unique sequences were ATP synthesis, or sugar and lipid metabolism-related metabolic pathways including starch and sucrose metabolism (123 members), oxidative phosphorylation (44 members), glycolysis/gluconeogenesis (43 members), galactose metabolism (40 members) and fatty acid degradation (28 members) (Table S6).

### Common DEGs between Naked-seed Mutants and TM-1

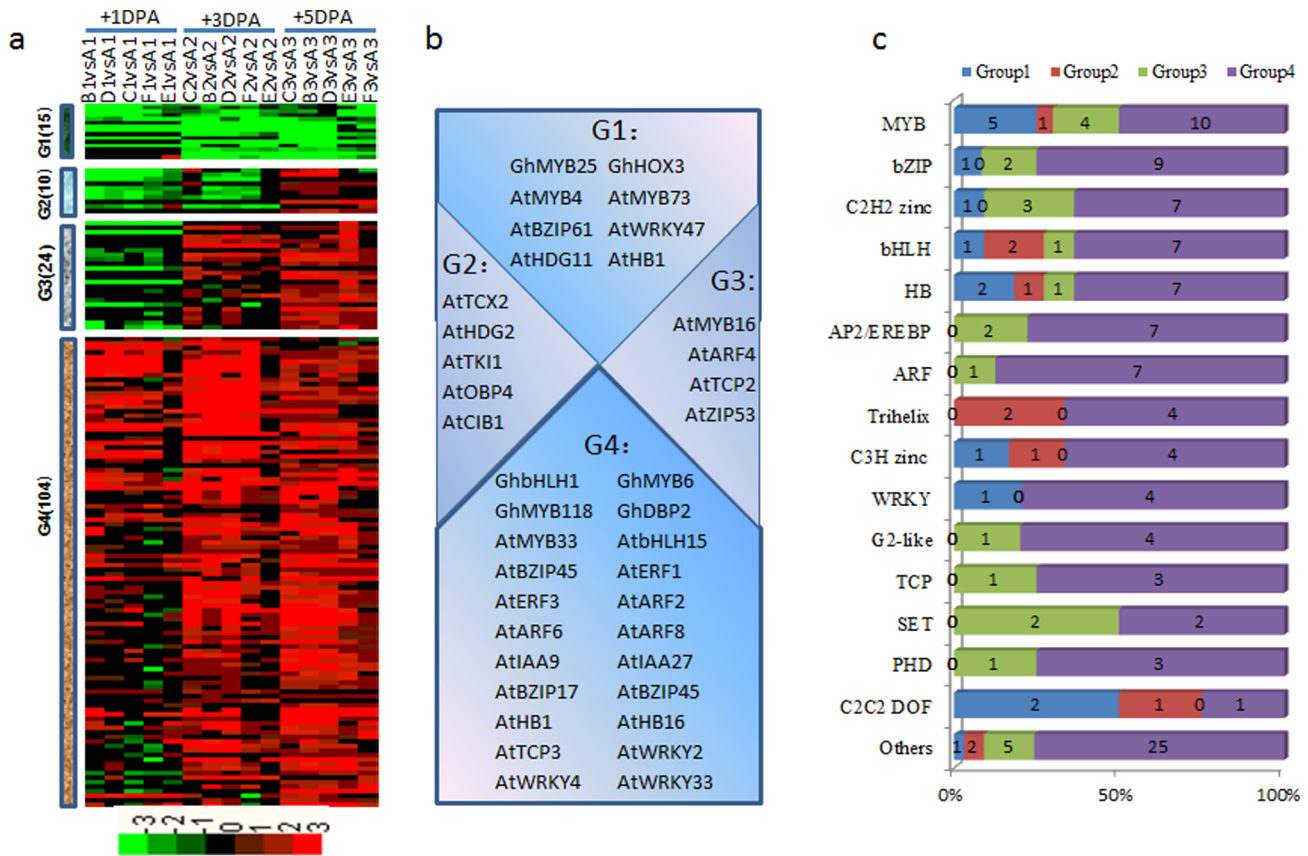
To uncover shared molecular mechanisms in the dominant and recessive fuzz-less mutants, we identified 1,932 DEGs that were common to the five mutants. Of these, 106 genes were upregulated and 314 downregulated at +1 DPA, 473 genes were upregulated and 215 downregulated at +3 DPA, and 737 genes were upregulated and 432 downregulated at +5 DPA (Figure S2, Table

S7). Four were three upregulated genes and 29 downregulated genes common to the +1 DPA, +3 DPA and +5 DPA samples (Table S8).

The 1,932 common DEGs were classified into four groups by hierarchical clustering. Nine-hundred and thirty-eight genes (Group 2) were upregulated at +1 DPA, +3 DPA and +5 DPA; 608 genes (Group 4) were downregulated at +1 DPA, +3 DPA and +5 DPA; 180 genes (Group 1) were downregulated at +1 DPA, and upregulated at +3 DPA and +5 DPA; 133 genes (Group 3) were downregulated at +1 DPA and +3 DPA, and upregulated at +5 DPA; 73 genes (Group 5) were upregulated at +1 DPA and +3 DPA, and downregulated at +5 DPA (Figure 6a).

One-thousand six-hundred and two genes were annotated by MapMan, excluding 17% belonging to the 'not assigned or unknown' categories. Among these genes, 22.2% were related to protein metabolism, 20.8% were related to RNA metabolism, 8.2% were related to signaling, with the remaining genes were related to cell functions, transport, development, stress, DNA metabolism, lipid metabolism, hormone metabolism, and cell wall (Figure 6b). Genes that encode oxygenases and light signaling were enriched in Group 1, encode enzymes for protein synthesis and regulation of transcription in Group 2 and 4, ammonium transport in Group 3, and major CHO synthesis in Group 5 (Figure 6c).

We found 153 differentially expressed transcription factor genes in common between the wild-type TM-1 and the naked mutants (Figure 7). Fifteen transcription factors were downregulated in the naked mutants (Group 1), including *GhMYB25* and *GhHOX3*. Ten



**Figure 7. Dynamics of transcription factor accumulation profiles in various dominant/recessive mutants.** (a) Unsupervised hierarchical clustering of 153 transcription factor genes included in the 1,932 common DEGs in the recessive mutants. Four groups were generated and the number of each group was in parentheses. Red region, genes upregulated in the mutants; green region, genes downregulated in the mutants. A, TM-1; B, SL1-7-1FLM; C, XZ142FLM; D, MD17FLM; E, N<sub>1</sub>NSM; F, n<sub>2</sub>NSM; 1, +1 DPA; 2, +3 DPA; 3, +5 DPA. (b) Representative functions and genes showing expression gradients. (c) Distribution of transcription factor families among G1, G2, G3 and G4. doi:10.1371/journal.pone.0097313.g007

transcription factors were downregulated at +1 DPA and +3 DPA, and upregulated at +5 DPA (Group 2), including *AtTCX2*, *AtHDG2*, *AtTKI1*, *AtOBP4* and *AtCIB1*. Forty transcription factors were downregulated at +1 DPA, and upregulated at +3 DPA and +5 DPA (Group 3), including *AtMYB16*, *AtARF4*, *AtTCP2*, *AtZIP53*. Another 271 transcription factors showed upregulation in the naked mutants (Group 4), including *GhbHLH1*, *GhMYB6*, *GhMYB118* and *GhDBP2* (Figure 7a and 7b). Members of the most families of transcriptional regulators such as MYB, bHLH, bZIP, C2C2(Zn), HB, AP2/EREBP, ARF and WRKY families were highly expressed in the recessive mutants (Figure 7c).

To identify the differential metabolic pathways active in fuzz initiation, we mapped the 1,932 common DEGs in the dominant and recessive mutants in the KEGG database. In total, we assigned 620 genes to KEGG pathways. Similar to our earlier results, most of these genes were related to ATP synthesis, and sugar and lipid metabolism pathways. For example, 38 genes were annotated to starch and sucrose metabolism 13 genes were annotated to galactose metabolism, and 11 genes to oxidative phosphorylation (Table S9).

#### Validation of Differentially Expressed Genes by qPCR

To determine whether the digital gene expression results were reliable, we conducted qPCR analysis of the expression levels of 21 representative differentially expressed genes, most of them

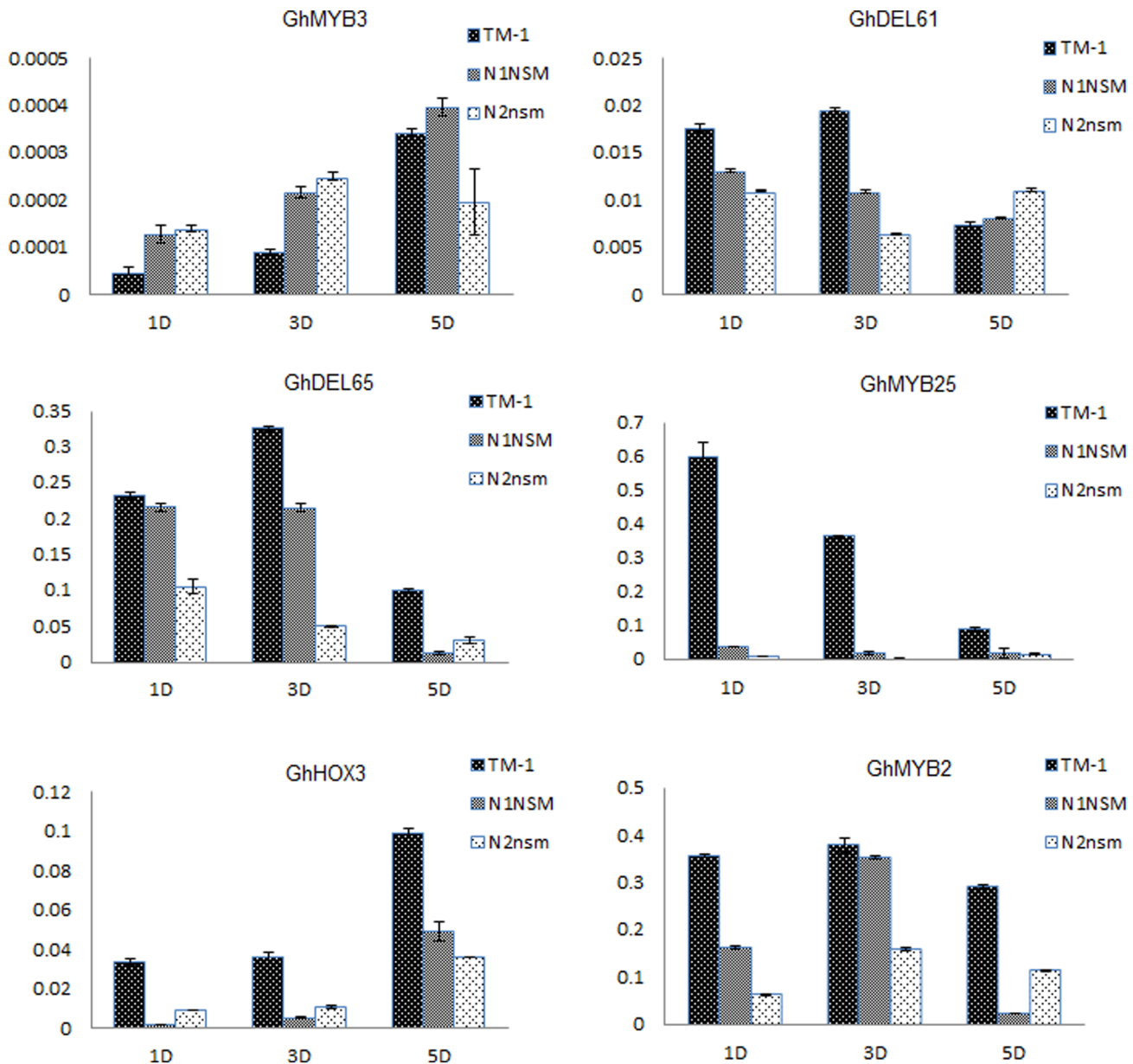
transcription factors. The qPCR results (Table S11) indicated that the expression levels estimated by DGE and qPCR were highly correlated ( $r^2 = 0.72-0.93$ ). The qPCR validation results confirmed the accuracy and reliability of the expression levels determined by digital gene expression analysis, which means that we can make reasonable deductions from the functional enrichment analysis of the DEGs. The qPCR results for expression of transcription factors indicated that *GhMYB3* had a high level of expression in fuzzy ovules at +1 DPA and +3 DPA, *GhDEL61* had a low expression level at +1 DPA and +3 DPA, and *GhDEL65* had a low expression level at +3 DPA and +5 DPA. Additionally, *GhMYB25*, *GhHOX3* and *GhMYB2* had low levels of expression in the fuzzless mutants ovules (Figure 8).

#### Discussion

##### Choice of Materials is Important for the Study of Fuzz Initial Cell Development

Cotton lint fibers are extremely long, single epidermal cells that develop on the outer surface of ovules, reaching upwards of 5 cm in some species [50]. Fibers initiate between -1 DPA and +1 DPA, and the fiber initials begin to elongate rapidly immediately after fertilization, extending out from the surface of the seed coat epidermis. Zhang et al. [1] showed that fuzz initiation begins at +4 DPA by scanning electron microscopy (SEM) examination of TM-





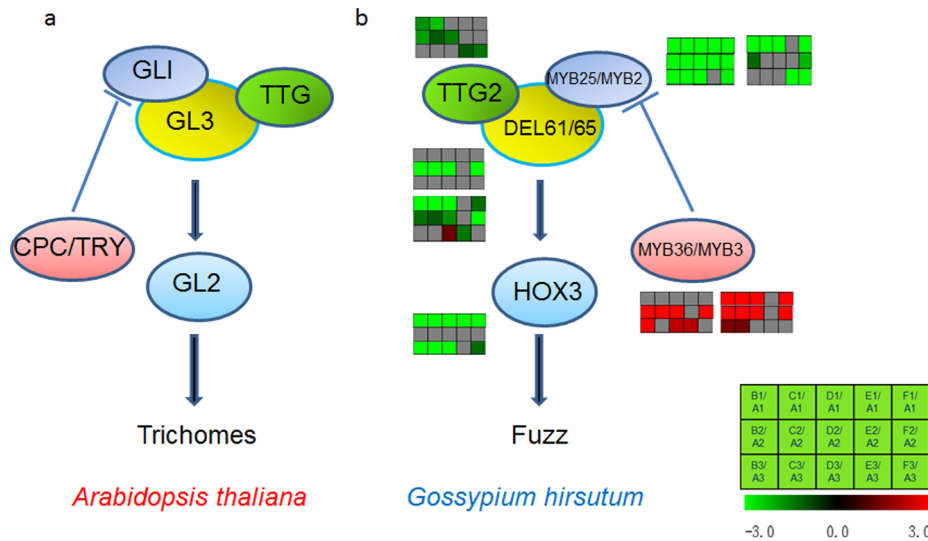
**Figure 8. QRT-PCR confirmation for the six selected differentially-expressed transcription factor genes.** The expression level of selected genes at +1 DPA, +3 DPA and +5 DPA in TM-1, N<sub>1</sub>NSM and n<sub>2</sub>nsm. Data shown are the means  $\pm$  SD of three biological replicates. doi:10.1371/journal.pone.0097313.g008

1 ovules, although the shape of fuzz protrusions differed from that of lint fibers. In our study, +1, +3 and +5 DPA ovules were employed for fuzz initial development. Two types of fibers, the long lint fibers and the short fuzz fibers, probably share common developmental pathways at least in early differentiation. However, the fuzz fiber appears to be under separate genetic control, as a number of genetic loci specifying absence of fuzz fiber, but with normal lint, have been identified [37]. Lintless mutants, however, only occur in conjunction with lack of fuzz fiber, so are essentially fiberless [49]. Cotton fiber mutants are invaluable for the investigation of genes that control fiber development at the molecular level. The natural fiber mutants are well suited for genetic, physiological, and molecular characterization of the mutant phenotype, avoiding the complex and time-consuming progress of inducing, screening, and verifying fiber mutants. In this

study, we selected five mutant lines, all of them fuzzless mutants that possessed different naked-seed genes. Thus, we can more clearly understand the mechanism of regulation of fuzz initial development by studying the five fuzzless mutants.

#### Many Specific Proteins that Relate to Fuzz Initial Development were Identified

A global analysis of the transcriptome will facilitate the characterization of gene expression and identification of regulatory mechanisms involved in fiber development [51,32]. In this study, we performed transcriptome profiling of fuzz-bearing and fuzzless ovules to identify genes that were differentially expressed during the fuzz initiation stage. Using a tag-based deep-sequencing approach [52], we could obtain a direct digital readout of cDNAs and achieve an essentially dynamic range of genes from the



**Figure 9. Model for the action of GL1-activating trichome development in *Arabidopsis thaliana* and fuzz development in *Gossypium hirsutum*.** a: Model for the action of GL1-activating trichome development in *Arabidopsis thaliana*. b: Model for the action of GL1-activating fuzz development in *Gossypium hirsutum* A: TM-1, B: SL1-7-1FLM, C: XZ142FLM, D: MD17FLM, E: N<sub>1</sub>NSM, F: n<sub>2</sub>NSM, 1: +1 DPA, 2: +3 DPA, 3: +5 DPA. Light red/green bars indicate cotton fiber gene expression in the upper/lower group. doi:10.1371/journal.pone.0097313.g009

libraries. Thus, the present study represents the most comprehensive analyses of the cotton fuzz transcriptome. Approximately 19,829–22,213 tag-mapped reference genes were identified for each library. Unfortunately, sequencing of the upland cotton genome is incomplete, so there are still a large proportion of unique tags mismatched. These unmatched unique sequences probably represent novel genes to be identified in future studies.

From +1 to +5 DPA, the cotton fibers and ovules are in a state of rapid development. Jensen (1968) observed the ultrastructure and composition of the cotton zygote and described a dramatic series of alterations in cell structure including zygote size, endoplasmic reticulum, microtubules, mitochondria, ribosomes and plastids [53]. During early development, fiber cells produced from the surface of the ovules and elongate quickly. We found that DEGs between TM-1 and fuzzless mutants involved in protein metabolism, RNA metabolism, and signaling categories were enriched significantly. The large number of RNA metabolism-related genes is consistent with the sharp increase in the total number of ribosomes observed in the zygote [53].

Based on the large number of genes with fiber-specific expression, the molecular dissection of cotton fiber initiation has provided new insights [33,34]. Lee et al. identified more than 20 genes that were greatly enriched at the fiber-bearing (+3 DPA) stage in TM-1 as compared to the N<sub>1</sub>NSM mutant [34,51]. Few studies have been performed to examine the initial pattern of cotton fuzz fiber development. In this study, we identified many DEGs between TM-1 and the fuzzless mutants. Protein synthesis-related genes had low levels of representation in both dominant and recessive mutants, while DNA and chromatin structure-related genes were highly represented. ATP synthesis, and sugar and lipid metabolism-related metabolic pathways play important roles in fuzz initial development. Recently, Du et al. [54] identified proteins related to fuzz fiber initiation in wild-type diploid cotton (*Gossypium arboreum* L.) and its fuzzless mutant by comparative proteomic analysis. They found 71 differentially expressed proteins between diploid Asiatic cotton DPL971 and the fuzzless mutant DPL972, mainly involved in cell response/signal transduction, redox homeostasis, protein metabolism, and energy/carbohydrate

metabolism [51]. The differential expression of these proteins demonstrated that rapidly differentiating and expanding fuzz fiber cells experience active protein metabolism [55,56].

#### Fuzz Development May Share Similar Molecular Mechanisms with Leaf Trichome Development in *Arabidopsis*

Illuminating the functions of key regulators in fuzz development could help explain the reasons for the delayed developmental and elongation steps of fuzz fiber development. Through molecular improvement of key transcriptional factors in cultivated varieties, cotton could produce longer fuzz fibers and have higher yields. Cell fate determination is a critical step in the developmental processes of plants, and involves the participation of a large number of transcription factors. The pattern of trichome development has been studied in depth in the model plant *Arabidopsis* [57]. We found many common differentially expressed transcription factors in the dominant and recessive mutants. Most of these were in the MYB, bHLH, HB and WRKY gene families. *GhTTG2*, a putative homolog of *Arabidopsis* *TTG2*, were downregulated in lintless-fuzzless mutants at +1 DPA and +3 DPA, and also in ovules of lint-fuzzless mutants in +5 DPA. *GhMYB25* and *GhMYB2*, putative homologues of *Arabidopsis* *GL1*, showed low expression levels in fuzzless mutants. The bHLH domain of *DEL61* and *DEL65*, which share a high degree of sequence identity with *Arabidopsis* *GL3* and *EGL3*, both had low expression levels in the fuzzless mutants. *HOX3*, a full-length coding sequence homolog of *AtGL2*, shares 72% identity with the homeobox conserved domain, and the expression level of *HOX3* in the fuzzless mutants was extremely low compared to that in TM-1. *GhMYB3/GhMYB36* pertaining to the MYB family had the high expression level in fuzzy ovules in dominant/recessive mutants (Figure 9). A model of the MYB25/MYB2-DEL61/65-TTG2 protein complex was described as the initial pattern of cotton fuzz, similar to the model of trichomes and root hairs in *Arabidopsis* [33,53]. Future studies including analyses of

protein function may shed light on the mechanism of cell initiation and formation of cotton fiber.

## Supporting Information

**Figure S1 Distribution of clean tag copy numbers for the 18 libraries.**

(TIF)

**Figure S2 Identity analysis of differentially expressed genes in various mutants.**

(TIF)

**Table S1 Genes involved in cotton fiber initial development and their *Arabidopsis* homologs.**

(XLSX)

**Table S2 Raw sequence data for the 18 transcriptome libraries.**

(XLSX)

**Table S3 List of 4,358 common DEGs between the mutants MD17FLM, SL1-7-1FLM, N1NSM and the wild-type TM-1.**

(XLS)

**Table S4 Pathway enrichment analysis for common differentially expressed genes in various dominant mutants.**

(XLS)

**Table S5 List of 6,693 common DEGs between the mutants XZ142FLM and n2NSM and the wild-type TM-1.**

(XLS)

## References

- Zhang DY, Zhang TZ, Sang ZQ, Guo WZ (2007) Comparative development of lint and fuzz using different cotton fiber-specific developmental mutants in *Gossypium hirsutum*. *J Integr Plant Bio* 49 (7): 975–983
- Larkin JC, Oppenheimer DG, Pollock S, Marks MD (1993) *Arabidopsis GLABROUS1* gene requires downstream sequences for function. *Plant Cell* 5: 1739–1748.
- Schellmann S, Schnitger A, Kirik V, Wada T, Okada K, et al. (2002) *TRIPTYCHON* and *CAPRICE* mediate lateral inhibition during trichome and root hair patterning in *Arabidopsis*. *EMBO J* 21: 5036–5046.
- Wada T, Tachibana T, Shimura Y, Okada K (1997) Epidermal cell differentiation in *Arabidopsis* determined by a Myb homolog, CPC. *Science* 277: 1113–1116.
- Lee MM, Schiefelbein J (2001) Developmentally distinct MYB genes encode functionally equivalent proteins in *Arabidopsis*. *Development* 128: 1539–1546.
- Larkin JC, Oppenheimer DG, Lloyd AM, Pappozzi ET, Marks MD (1994) Roles of the *GLABROUS1* and *TRANSPARENT TESTA GLABRA* genes in *Arabidopsis* trichome development. *Plant Cell* 6: 1065–1076.
- Walker AR, Davison PA, Bolognesi-Winfield AC, James CM, Srinivasan N, et al. (1999) The *TRANSPARENT TESTA GLABRA1* locus, which regulates trichome differentiation and anthocyanin biosynthesis in *Arabidopsis*, encodes a WD40 repeat protein. *Plant Cell* 11: 1337–1349.
- Payne CT, Zhang F, Lloyd AM (2000) *GL3* encodes a bHLH protein that regulates trichome development in *Arabidopsis* through interaction with *GL1* and *TTG1*. *Genetics* 156: 1349–1362.
- Zhang F, Gonzalez A, Zhao M, Payne CT, Lloyd A (2003) A network of redundant bHLH proteins functions in all TTG1-dependent pathways of *Arabidopsis*. *Development* 130: 4859–4869.
- Szymanski DB, Jilk RA, Pollock SM, Marks MD (1998) Control of *GL2* expression in *Arabidopsis* leaves and trichomes. *Development* 125: 1161–1171.
- Johnson CS, Kolevski B, Smyth DR (2002) *TRANSPARENT TESTA GLABRA2*, a trichome and seed coat development gene of *Arabidopsis*, encodes a WRKY transcription factor. *Plant Cell* 14: 1359–1375.
- Neer EJ, Schmidt CJ, Nambudripad R, Smith TF (1994) The ancient regulatory protein family of WD-repeat proteins. *Nature* 371: 297–300.
- Murre C, McCaw PS, Baltimore D (1989) A new DNA-binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MYOD and MYC proteins. *Cell* 56: 777–783.
- Bernhardt C, Lee MM, Gonzalez A, Zhang F, Lloyd A, et al. (2003) The *bHLH* genes *GLABRA3* (*GL3*) and *ENHANCER OF GLABRA3* (*EGL3*) specify epidermal cell fate in the *Arabidopsis* root. *Development* 130: 6431–6439.
- Ramsay NA, Walker AR, Mooney M, Gray JC (2003) Two basic-helix-loop-helix genes (*MYC146* and *GL3*) from *Arabidopsis* can activate anthocyanin biosynthesis in a white-flowered *Matthiola incana* mutant. *Plant Mol Biol* 52(3): 679–688.
- Nesi N, Debeaujon I, Jond C, Pelletier G, Caboche M, et al. (2000) The *TT8* gene encodes a basic helix-loop-helix domain protein required for expression of *DFR* and *BAN* genes in *Arabidopsis* siliques. *Plant Cell* 12: 1863–1878.
- Loguercio LL, Zhang JQ, Wilkins TA (1999) Differential regulation of six novel MYB-domain genes defines two distinct expression patterns in allotetraploid cotton (*Gossypium hirsutum* L.). *Mol Genet Genomics* 261: 660–671.
- Suo J, Liang X, Pu L, Zhang YS, Xue YB (2003) Identification of *GhMYB109* encoding a R2R3 MYB transcription factor that expressed specifically in fiber initials and elongating fibers of cotton (*Gossypium hirsutum* L.). *Biochim Biophys Acta* 1630: 25–34.
- Wang S, Wang JW, Yu N, Li CH, Luo B, et al. (2004) Control of plant trichome development by a cotton fiber *MYB* gene. *Plant Cell* 16: 2323–2334.
- Guan XY, Lee JJ, Pang MX, Shi XL, Stelly DM, et al. (2011) Activation of *Arabidopsis* seed hair development by cotton fiber-related genes. *PLoS One* 6(7): e21301
- Huang YQ, Liu X, Tang KX, Zuo KJ (2013) Functional analysis of the seed coat-specific gene *GhMYB2* from cotton. *Plant Physiol. Biochem.* 73: 16–22.
- Noda K, Glover BJ, Linstead P, Martin C (1994) Flower colour intensity depends on specialized cell shape controlled by a Myb-related transcription factor. *Nature* 369: 661–664.
- Perez-Rodriguez M, Jaffe FW, Butelli E, Glover BJ, Martin C (2005) Development of three different cell types is associated with the activity of a specific MYB transcription factor in the ventral petal of *Antirrhinum majus* flowers. *Development* 132: 359–370.
- Wu Y, Llewellyn DJ, White R, Ruggiero K, Al-Ghazi Y, et al. (2007) Laser capture microdissection and cDNA microarrays used to generate gene expression profiles of the rapidly expanding fibre initial cells on the surface of cotton ovules. *Planta* 226: 1475–1490.
- Machado A, Wu YR, Yang YM, Llewellyn DJ, Dennis ES (2009) The MYB transcription factor GhMYB25 regulates early fibre and trichome development. *Plant J* 59: 52–62
- Wu YR, Machado AC, White RG, Llewellyn DJ, Dennis ES (2006) Expression profiling identifies genes expressed early during lint fibre initiation in cotton. *Plant and Cell Physiology* 47 (1): 107–127
- Walford SA, Wu YR, Llewellyn DJ, Dennis ES (2011) GhMYB25-like: a key factor in early cotton fibre development. *Plant J* 65(5): 785–797

**Table S6 Pathway enrichment analysis for common differentially expressed genes in various recessive mutants.**  
(XLS)

**Table S7 List of 1,932 common DEGs between the mutants MD17FLM, SL1-7-1FLM, N1NSM, XZ142FLM, n2NSM and the wild-type TM-1.**  
(XLS)

**Table S8 The common differentially expressed genes in various mutants and times.**  
(XLS)

**Table S9 Pathway enrichment analysis for common differentially expressed genes in various dominant and recessive mutants.**  
(XLS)

**Table S10 Primers used for qRT-PCR analysis.**  
(XLS)

**Table S11 Correlation between qRT-PCR and DGE for 21 differentially expressed genes.**  
(XLS)

## Acknowledgments

We thank Prof. Dr. Wangzhen Guo for his support and comments during the preparation of this manuscript.

## Author Contributions

Conceived and designed the experiments: TZ. Performed the experiments: QW HZ WY HW. Analyzed the data: QW. Wrote the paper: QW TZ.

28. Taliervo EW, Boykin D (2007) Analysis of gene expression in cotton fiber initials. *BMC Plant Biol* 7: 22.
29. Humphries JA, Walker AR, Timmis JN, Orford SJ (2005) Two WD-repeat genes from cotton are functional homologues of the *Arabidopsis thaliana* *TRANSPARENT TESTA GLABRA1 (TTG1)* gene. *Plant Mol Biol* 57: 67–81.
30. Guan XX, Li QJ, Shan CM, Wang S, Mao YB, et al. (2008) The HD-Zip IV gene *GaHOX1* from cotton is a functional homologue of *Arabidopsis GLABRA2*. *Physiol Plant* 134: 174–182.
31. Mandaokar A, Kumar VD, Amway M, Browse J (2003) Microarray and differential display identify genes involved in jasmonate-dependent anther development. *Plant Mol Biol* 52: 775–786.
32. Shangguan XX, Xu B, Yu ZX, Wang LJ, Chen XY (2008) Promoter of a cotton fiber MYB gene functional in trichomes of *Arabidopsis* and glandular trichomes of tobacco. *J Exp Bot* 59(13): 3533–3542.
33. Guan XY, Yu N, Shangguan XX, Wang S, Lu S, et al. (2007) *Arabidopsis* trichome research sheds light on cotton fiber development mechanisms. *Chinese Sci Bull* 52: 1734–1741.
34. Lee JJ, Woodward AW, Chen ZJ (2007) Gene expression changes and early events in cotton fibre development. *Ann Bot* 100: 1391–1401.
35. Kearney TH, Harrison GJ (1927) The inheritance of smoothness seeds in cotton. *J Agric Res* 35: 193–217.
36. Turley JO, Benedict LI, Rolfe WH (1947) A recessive naked-seed character in upland cotton. *J Hered* 38: 313–320.
37. Turley RB, Kloth RH (2002) Identification of a third fuzzless seed locus in upland cotton (*Gossypium hirsutum* L.). *J Hered* 93(5): 359–364.
38. Turley RB, Kloth RH (2008) The inheritance model for the fiberless trait in upland cotton (*Gossypium hirsutum* L.) line SL1-7-1: variation on a theme. *Euphytica* 164: 123–132.
39. Zhang T, Pan J (1991) Genetic analysis of a fuzzless-lintless mutant in *Gossypium hirsutum*. *J Agric Sci* 7(3): 13–16.
40. Du XM, Pan JJ, Wang RH, Zhang TZ, Shi TZ (2001) Genetic analysis of presence and absence of lint and fuzz in cotton. *Plant Breed* 120: 519–522.
41. Turley RB (2002) Registration of MD 17 fiberless upland cotton as a genetic stock. *Crop Sci* 42: 994–995.
42. Kohel RJ (1973) Genetic nomenclature in cotton. *J Hered* 64: 291–295.
43. Jiang JX, Zhang TZ (2003) Extraction of total RNA in cotton tissues with CTAB-acidic phenolic method. *Cotton Sci* 15: 166–167.
44. Wang QQ, Liu F, Chen XS, Ma XJ, Zeng HQ, et al. (2010) Transcriptome profiling of early developing cotton fiber by deep-sequencing reveals significantly differential expression of genes in a fuzzless/lintless mutant. *Genomics* 96(6): 369–376.
45. Audic S, Claverie JM (1997) The significance of digital gene expression profiles. *Genome Res* 7: 986–995.
46. Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *Ann Stat* 29: 1165–1188.
47. Thimm O, Blasing O, Gibon Y, Nagel A, Meyer S, et al. (2004) MapMan: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J* 37: 914–939.
48. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, et al. (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Res* 36: D480–D484.
49. Hegedus Z, Zakrzewska A, Agoston VC, Ordas A, Racz P, et al. (2009) Deep sequencing of the zebrafish transcriptome response to mycobacterium infection. *Mol Immunol* 46: 2918–2930.
50. Stewart JM (1975) Fiber initiation on the cotton ovule (*Gossypium hirsutum*). *Am J Bot* 62: 723–730.
51. Lee JJ, Hassan OSS, Gao W, Wei NE, Kohel RJ, et al. (2006) Developmental and gene expression analyses of a cotton naked seed mutant. *Planta* 233: 418–432.
52. Morrissy AS, Morin RD, Delancy A, Zeng T, McDonald H, et al. (2010) Next-generation tag sequencing for cancer gene expression profiling. *Genome Res* 19: 1825–1835.
53. Jensen WA (1968) Cotton Embryogenesis: The Zygote. *Planta (Berl.)* 79: 346–366.
54. Du SJ, Dong CJ, Zhang B, Lai TF, Du XM (2013) Comparative proteomic analysis reveals differentially expressed proteins correlated with fuzz fiber initiation in diploid cotton (*Gossypium arboreum* L.). *J Proteomics* 82: 113–129.
55. Schaller A (2004) A cut above the rest: the regulatory function of plant proteases. *Planta* 220: 183–197.
56. Hovav R, Udall JA, Hovav E, Rapp R, Flagel L, et al. (2008) A majority of cotton genes are expressed in single-celled fiber. *Planta* 227: 319–329.
57. Ramsay NA, Glover BJ (2005) MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. *Trends Plant Sci* 10(2): 63–70.