Tu Lili (Orcid ID: 0000-0003-4294-1928)
Qin Yuan (Orcid ID: 0000-0003-4887-8878)
Xu Zhongping (Orcid ID: 0000-0003-2559-9091)
Yang Xiyan (Orcid ID: 0000-0003-1305-8677)
Wang Maojun (Orcid ID: 0000-0002-4791-3742)
Zhang Xianlong (Orcid ID: 0000-0002-7703-524X)

Single-cell RNA-seq reveals fate determination control of an individual fiber cell initiation in cotton (*Gossypium hirsutum*)

Yuan Qin[1,3], Mengling Sun[1,3], Weiwen Li[1], Mingqi Xu[1], Lei Shao[1], Yuqi Liu[1], Guannan Zhao[1], Zhenping Liu[1], Zhongping Xu[1], Jiaqi You[1], Zhengxiu Ye[1], Jiawen Xu[1], Xiyan Yang[1], Maojun Wang[1], Keith Lindsey[2], Xianlong Zhang[1] and Lili Tu[1,*]

[1] National Key Laboratory of Crop Genetic Improvement, Hubei Hongshan Laboratory, Huazhong Agricultural University, Wuhan, Hubei Province 430070, China

[2] Department of Biosciences, Durham University, Durham, UK

[3] These authors contributed equally.

[*] Correspondence: Lili Tu (lilitu@mail.hzau.edu.cn)

Fax: +86 027 87283955

Tel: +86 027 87283955

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/pbi.13918

Summary

Cotton fiber is a unicellular seed trichome, and lint fiber initials per seed as a factor determines fiber yield. However, the mechanisms controlling fiber initiation from ovule epidermis are not understood well enough. Here, with single-cell RNA sequencing (scRNA-seq), a total of 14,535 cells were identified from cotton ovule outer integument of Xu142_LF line at four developmental stages (1.5, 1, 0.5 days before anthesis and the day of anthesis). Three major cell types, fiber, non-fiber epidermis and outer pigment layer were identified and then verified by RNA *in situ* hybridization. A comparative analysis on scRNA-seq data between Xu142 and its fiberless mutant Xu142 *fl* further confirmed fiber cluster definition. The developmental trajectory of fiber cell was reconstructed, and fiber cell was identified differentiated at 1 day before anthesis. Gene regulatory networks at four stages revealed the spatiotemporal pattern of core transcription factors, and *MYB25-like* and *HOX3* were demonstrated played key roles as commanders in fiber differentiation and tip-biased diffuse growth, respectively. A model for early development of a single fiber cell was proposed here, which sheds light on further deciphering mechanism of plant trichome and the improvement of cotton fiber yield.

Key words: cotton fiber initiation, single cell transcriptomic atlas, *Gossypium hirsutum*, regulatory network, cell fate determination

Introduction

Cotton is an important cash crop worldwide, and supplies the largest proportion of natural fiber to textile industry. Cotton fibers are unicellular, and initiate from ovule epidermis, making it a good model for studying mechanisms of cell fate determination. There are two types of cotton fiber according to mature fiber length: lint and fuzz. The initiation of lint (long) fiber proceeds from the day post anthesis (0 DPA) to 3 DPA, with the initiation of fuzz (short) fiber typically commencing afterwards. The number of lint fiber initials per seed as a factor determines fiber yield, while lint initiation is typically from 25% of all epidermal cells (Stewart, 1975). Therefore, exploring the mechanisms and analyzing key factors and networks that regulate lint fiber cell fate, can provide a theoretical basis for the genetic improvement of fiber yield, so as to help improve the economic benefits of cotton planting.

Numerous studies have investigated the mechanisms of fiber development by means of expression profiling. Since early twenty-first century, 14 highly expressed cDNAs were identified in cotton fiber using cDNA arrays (Li *et al.*, 2002), after that several expression profiles have been performed on various stages of fiber development (Gou *et al.*, 2007; Shi *et al.*, 2006), between different cotton species (Tu *et al.*, 2007), during domestication (Hovav *et al.*, 2008; Rapp *et al.*, 2010), between normal cotton and fiber-related mutants (Wu *et al.*, 2006). A few transcription factor (TF) genes, such as R2R3-type MYB TFs *GhMYB109* (Pu *et al.*, 2008), *GhMYB25* (Machado *et al.*, 2009), *GhMYB25-like* (Walford *et al.*, 2011; Wan *et al.*, 2016), and HD-ZIP family TFs *GhHD-1* (Walford *et al.*, 2012), *GhHOX3* (Shan *et al.*, 2014) have been verified positively regulating lint fiber initiation. A model has been proposed in which no lint fiber will initiate if the combined expression levels of *MYB25-like_At* and *MYB25-like_Dt* are below a critical threshold level at 0 days post anthesis (DPA) (Zhu *et al.*, 2018). In addition, suppression of sucrose synthase activity by at least 70% in the ovule epidermis (Ruan *et al.*, 2003), or of a vacuolar invertase gene *GhVIN1* (Wang *et al.*, 2014), led to a fiberless phenotype. Furthermore, naked seeds were produced when ovules were cultured with adding no indoleacetic acid (IAA) (Zeng *et al.*, 2019) or excess IAA

transport inhibitor (Zhang *et al.*, 2017; Zhang *et al.*, 2011), or with high concentration of zeatin (ZT), a kind of cytokinin (Zeng *et al.*, 2019). Recently, some review papers have summarized that fiber initiation was affected by complex cross-talk among MYB MIXTA-like TFs, sugar signals and plant hormones (Huang *et al.*, 2021; Tian and Zhang, 2021; Wang *et al.*, 2021a). Despite these studies, when did the members in the complex start work, whether they take action only in fiber cells or also in adjacent cells during the continuous fiber initiation process, were partly ambiguous and needs more details.

Now, single-cell RNA sequencing (scRNA-seq) has been developed and brings unprecedented opportunities to the field of plant research (Denyer and Timmermans, 2022; Mo and Jiao, 2022; Ryu *et al.*, 2021; Seyfferth *et al.*, 2021; Shaw *et al.*, 2021). The first effective high-throughput scRNA-seq in plants exploited single-cell transcriptome sequencing of *Arabidopsis* root tissue protoplasts (Ryu *et al.*, 2019). Subsequently, a series of developmental processes in *Arabidopsis* had been explored at single cell resolution, such as the development of root tips (Denyer *et al.*, 2019; Jean-Baptiste *et al.*, 2019; Shahan *et al.*, 2022; Wendrich *et al.*, 2020; Zhang *et al.*, 2019), lateral root (Gala *et al.*, 2021), vegetative shoot apex (Zhang *et al.*, 2021b), stomatal cell lineage (Liu *et al.*, 2020), developing leaf (Kim *et al.*, 2021; Liu *et al.*, 2022a; Lopez-Anido *et al.*, 2021; Tenorio Berrio *et al.*, 2022) and vein pattern in the cotyledons (Liu *et al.*, 2022b). At the same time, scRNA-seq has also been widely used in other plants, including rice (Liu *et al.*, 2021b; Wang *et al.*, 2021b; Zhang *et al.*, 2021a; Zong *et al.*, 2022), corn (Li *et al.*, 2022; Ortiz-Ramirez *et al.*, 2021; Satterlee *et al.*, 2020; Sun *et al.*, 2022; Xu *et al.*, 2021), peanut (Liu *et al.*, 2021a), tea plant (Wang *et al.*, 2022), tomato (Omary *et al.*, 2022) and poplar (Chen *et al.*, 2021; Li *et al.*, 2021; Xie *et al.*, 2022). These studies provide new insights into heterogeneity of gene expression between different cell types, and molecular trajectory of cell differentiation during development.

To further decipher the detailed gene regulatory network in fiber initiation, we performed scRNA-seq on cotton ovules. The developmental trajectory starting from

early differentiated fiber cell was reconstructed. All the results can interactively be mined on the web, which is freely available at http://cotton.hzau.edu.cn/CLC/. The single cell resolution transcriptomes provide a valuable resource, and give us a deep understanding on the elaborate cotton fiber initiation process.

Results

Fiber cells begin to protrude at 0.5 days before anthesis

The phenotype of cotton Xu142_LF line (seed index 10-12 g, lint index 6-7 g and lint percentage 35%-38%) has been described in our previous work (Hu *et al.*, 2018). A more detailed morphological observations on Xu142_LF ovules were performed at -1.5, -1, -0.5, 0 DPA, respectively (Fig. 1). By SEM, no fiber cell protrusion detected on ovule epidermis at -1.5 DPA (Fig. 1a) and -1 DPA (Fig. 1b), while at -0.5 DPA can be detected (Fig. 1c), and protrusions number increased at 0 DPA (Fig. 1d,e).

To further investigate the dynamic process of fiber initiation, samples were taken every 4 hours (h) before and after fiber cell protrusion (from -1 to 0 DPA, Fig. S1). No epidermal cell protruded at 08:00, 12:00, 16:00 of -1 DPA. At 20:00, fiber cells began to protrude. At 24:00, the number of fiber initials increased. At 4:00 of 0 DPA, fiber initials grew larger, similar in size to those at 8:00 of 0 DPA (Fig. S1). From 20:00 (-0.5 DPA), the number of fiber initials shows an increasing trend (Fig. 1f).

Ovule epidermis at those four stages were also observed by TEM. Consistent with SEM observations, epidermal cells didn't protrude at -1.5 and -1 DPA (Fig. 1g,h), but began to protrude at -0.5 DPA (Fig. 1i) and continued to expand at 0 DPA (Fig. 1j). Considering lint fiber cell start protruding at -0.5 DPA, their fate is likely to be determined as early as -1 DPA, or -1.5 DPA, or possibly earlier.

Fiber cell cluster is identified from cotton ovule outer integument by scRNA-seq

To explore the molecular mechanism determining fiber cell fate, we performed scRNA-seq on Xu142_LF ovules. Ovule samples collected at -1.5, -1, -0.5 and 0 DPA were used for protoplast isolating (Fig. S2a,b). About 20,000 protoplasts were initially

loaded onto the 10x Genomics platform. After separation, RNA from individual protoplasts were extracted for library construction followed by high-throughput sequencing (Fig. S2c). As shown in longitudinal section of ovules after enzymolysis, only cells from ovule outer integument were released (Fig. S2d-g).

In Xu142_LF 0 DPA sample (LF_0d), a total of 3,679 cells with 50,753 genes were detected (Table S1). After a strict gene/cell filtering process (Document S1), LF_0d sample obtained 35,169 gene transcripts with high reliability across 1,703 cells. This filtration was performed on other samples one by one, then high-quality gene-cell matrices were obtained. Overall, 738-2045 filtered cells per sample were obtained (Table S1).

To examine the robustness of the scRNA-seq results, LF_0d scRNA-seq data were compared with 0 DPA ovule outer integument bulk RNA-seq data (Hu *et al.*, 2018). The correlation coefficient (R) was 0.63 with $p < 2.2e-16$, showing a very significant correlation between them (Fig. S3a). There were 3,974 protoplasting-induced differentially expressed genes (DEGs) identified, with 2,233 DEGs up-regulated and 1,741 down-regulated (Table S2). GO analysis suggested that the up-regulated genes were involved in 'response to stress and stimulus' and 'regulation of cell death', amongst others (Fig. S3b); the down-regulated genes were mainly involved in 'primary metabolic process' and 'biosynthetic process' (Fig. S3c). Next, UMAP and *t*-SNE algorithm were used to visualize and explore LF_0d dataset after linear dimensional reduction. Unsupervised analyses grouped 1,703 cells into 9 clusters (Fig. 2a,b). Similarly, 9 clusters were also observed after 3,974 DEGs in response to protoplasting were removed (Fig. S3d), and clustered cells were almost kept in the same cell types as before (Fig. S3e). This suggested that cell wall enzymolysis had only a minor effect on cell clustering, the same as reported in rice root (Liu *et al.*, 2021b).

For cell cluster definition, accumulation of reported fiber gene transcripts in single-cell populations were analyzed (Fig. 2c, Fig. S4a). The fiber genes included *MYB25-like* (Walford *et al.*, 2011; Wan *et al.*, 2016), *MYB25* (Machado *et al.*, 2009), *MML4* (Wu *et al.*, 2018), *MML9* (Bedon *et al.*, 2014), *HD1* (Walford *et al.*, 2012) and *HOX3*

(Shan et al., 2014). All these genes tended to be highly or preferentially expressed in cluster 5 (Fig. 2c, Fig. S4b), indicated that cluster 5 may be fiber cell. To further enable cell type assignment to particular clusters, a series of enriched genes for each cluster were identified (Fig. 2d, Fig. S4c, Table S3). With RNA *in situ* hybridization of the representative genes, all cell clusters from cotton ovule outer integument can be defined into three major types: fiber cell, non-fiber epidermis and outer pigment layer (Fig. 2e, Fig. 3). For example, *DUF* (*Ghir_D07G016770*, gene function unknown) was preferentially expressed in fiber cells, *Erg6* (*Ghir_A04G010380,* a methyltransferase encoding gene) in non-fiber epidermis and *HbdA* (*Ghir_D01G005520*, 3-hydroxyacyl-CoA dehydrogenase) in outer pigment layer (Fig. 3, Fig. S5).

Fiber cell definition is further proved by scRNA-seq of *Xu142 fl* fiberless mutant

Single-cell transcriptomes were generated from 0 DPA ovule protoplasts of wild type cultivar Xu142 and its fiberless mutant Xu142 *fl*. Clustering these two samples together generated 5 clusters (Fig. 4a, Table S3), which were assigned to three major cell types using *in situ* hybridization verified genes: fiber (cluster 4), non-fiber epidermis (cluster 0, 2, 3), outer pigment layer (cluster 1, Fig. 4a, Fig. S6a). When comparing the clustering results between Xu142 and Xu142 *fl*, the fiber cell cluster (cluster 4) was found to exist only in Xu142, not in Xu142 *fl* (Fig. 4b, Table S4). The expression distribution of three fiber marker genes (*MYB25*, *MML9* and *HOX3*) also showed expressing in Xu142, not in Xu142 *fl* (Fig. S6b). These again proved that the fiber cell definition was reliable.

'Ribosome' pathway is enriched in 0 DPA fiber cells

To compare the fiber cluster enriched genes between LF_0d and Xu142_0d, a clustering analysis on Xu142_0d sample was performed. The 738 high-quality cells from Xu142_0d clustered into 6 clusters. Cluster 2 was identified as fiber cell with expression pattern analysis of fiber markers (*MYB25* and *MML9*, Fig. S6c,d), and 984 genes were identified enriched in fibers (Table S3). When comparing these 984 genes with the 517 genes enriched in LF_0d fibers (Table S3), it was found that 474 genes overlapped, which means a similarity of gene expression between them (Fig. S6e).

To explore the pathways that are active in fiber cells, KEGG enrichment analysis was performed on the 517 genes enriched in LF_0d fibers. These genes were involved in 38 different signaling pathways (Table S5). Among them, 312 genes were involved in the most significant pathway 'Ribosome' (Table S5), which indicated that a major activity in 0 DPA fibers was 'peptide biosynthetic process'. It was consistent with the characteristics of expanding fiber initial cells, i.e., more abundant transcripts involved in protein synthesis to meet the high demand for new cell wall and membrane components (Wu *et al.*, 2007). The remaining genes were enriched in several pathways such as 'Fatty acid', 'Flavonoid', 'Oxidative phosphorylation', 'Sucrose' and 'Phagosome' (Fig. 4c, Table S5). Besides, some TFs were identified from the 517 genes, including *MML9*, *MYB25*, *HD1*, *PDF2*, *HOX3*, *HDA3* and *VFP5* (Fig. 4c). These TFs should play positive roles in 0 DPA fiber development, as some of them have been experimentally verified (Machado *et al.*, 2009; Shan *et al.*, 2014; Walford *et al.*, 2012).

Fiber cell was differentiated at -1 DPA

In order to clarify the dynamic changes of gene expression accompanying fiber cell initiation, the single-cell transcriptomes in Xu142_LF across 4 stages (-1.5, -1, -0.5, 0 DPA) were combined. After unsupervised clustering, totally 5,137 cells were clustered into 6 clusters (Fig. 5a), and genes enriched in each cluster were identified (Table S3). Based on the well-defined cell types in LF_0d and cell barcode mapping (Table S6), the combined sample can be defined as fiber (cluster 5), epidermis (cluster 1, 3) and outer pigment layer (cluster 0, 2, 4; Fig. 5a). The expression pattern of fiber marker genes (*MYB25*, *MML9* and *HD1*) further verified that cluster 5 was fiber cell cluster (Fig. S7a).

When displaying the clustering result separately at four stages (Fig. S7b-e), no fiber cell was identified at -1.5 DPA (Fig. S7b), while the next three stages contained fiber cluster (Fig. S7c-e). The expression pattern of *MYB25* and *MML9* also showed that they didn't express at -1.5 DPA, but expressed at next three stages (Fig. S7f,g). According to previous phenotypic observation, -0.5 DPA was the time point showing visible fiber initials (Fig. 1i). Here, the transcriptome data suggested that fiber cells were

differentiated at -1 DPA (Fig. S7c), although they were morphologically undistinguishable to its neighboring cells at this stage (Fig. 1h). Since then, the number of fiber cell (cluster 5) increases stepwise along development (Fig. 5b).

Next, the developmental trajectory of fiber cells was explored. In total, 226 fiber cells (16 cells from LF_-1d, 30 cells from LF_-0.5d and 180 cells from LF_0d, Fig. 5b) were selected for re-clustering. There were 5 sub-clusters identified (Fig. 5c). When displaying the re-clustering result separately (Fig. S7h), it was found that -1 DPA fiber cells were mainly located to sub-cluster 5.2. Then, pseudotime analysis on these fiber cells built a developmental trajectory (Fig. 5d). The starting of the trajectory was specified at the left branch, as sub-cluster 5.2 cells (early fiber cells) were located on this branch. There was a node in the trajectory, with branching into two directions (Fig. 5d). The MYB-MIXTA-like TFs, such as *MYB25-like*, *MML4, MYB25*, were prominently highly expressed at the beginning branch (Fig. 5e), in agreement with their key roles in fiber cell differentiation (Machado *et al.*, 2009; Walford *et al.*, 2011; Wu *et al.*, 2018). The homeodomain leucine zipper transcription factor, *HD1* (Walford *et al.*, 2012), was expressed almost at all the branches (Fig. 5e), suggesting that *HD1* gene has a broader temporal pattern during fiber initiation, distinct from MYB-MIXTA-like TFs.

It is known that once early fiber cell differentiated, they will undergo a process of diffuse growth then transformed into tip-biased diffuse growth (Qin and Zhu, 2011; Yu *et al.*, 2019). To further define the cell types on branch 1 and branch 2, a total of 167 genes were firstly identified as most significantly related to branching (qval < 0.001, Table S7). A heat map containing these 167 genes was produced and they were sorted into 5 modules (Fig. 5f). Among them, 59 genes in gene module 1 (M1) are predominantly expressed in branch 1 cells. GO results show that they are mainly enriched in 'regulation of ethylene-activated signaling pathway', 'sucrose alpha-glucosidase activity', 'cuticle hydrocarbon biosynthetic process' and so on (Fig. 5f, Table S8). These GO terms were reported mainly enriched in fiber elongation (Shi *et al.*, 2006). In details, multiple genes involved in ethylene signaling, such as ethylene responsive element binding factor (*ERF4, ERF9* and *ERF11*) , and 1-

aminocyclopropane-1-carboxylic acid (ACC) synthase gene (*ACS6*) (Shi *et al.*, 2006) were all prominently expressed in branch 1 cells (Fig. 5g, Table S7). There are 79 genes in M5, which are predominantly expressed in branch 2 cells. They are mainly enriched in 'ribosome', 'translation', 'lipid transport' and 'ATP transmembrane transporter activity' (Fig. 5f, Table S8), these GO terms were reported highly enriched in diffuse growing fiber cells (Wu *et al.*, 2007). The genes encoding large and small ribosomal subunit protein (*RPL24, RPS5B, RPL3A, RPS6B*) and lipid transfer protein (*LTP1*) (Wu *et al.*, 2006) were highly enriched in branch 2 (Fig. 5g, Table S7). Moreover, an E6 protein encoding gene (*Ghir_D05G016260*), which plays a role in diffuse growing fiber cell (Ji *et al.*, 2003), was identified and verified by *in situ* hybridization (Fig. 5g,h).

As described above, fiber initiation can be subdivided into several processes. Precursor fiber cell have not differentiated at -1.5 DPA (Fig. S7b), and then differentiated into fiber cell at -1 DPA (Fig. S7c and Fig. 5b) and start protruding by diffuse growth at -0.5 DPA (Fig. 1i), then transformed into tip-biased diffuse growth from 0 DPA (Fig. 1j).

The highly interconnected gene regulatory networks coordinate fiber cell initiation

The results above have shown that these four stages (-1.5, -1, -0.5, 0 DPA) represented distinct processes of fiber initiation. To explore gene regulatory networks in each process, clustering analysis were performed on four samples individually (Fig. S8a). By analyzing transcript accumulation of fiber marker gene *MML4* and the *in situ* hybridization verified genes (*DUF, Erg6* and *HbdA*, Fig. S8b), the clusters in each sample can be defined into fiber, epidermis and outer pigment layer (Fig. S8a). Then, genes enriched in each cluster were identified (Table S9), and WGCNA was performed on four samples, respectively. Several distinct gene modules (labeled by different colors) were identified as shown in the dendrogram (Fig. S9, Table S10). The modules which containing genes enriched in fiber cells (-1, -0.5, 0 DPA) or cells that may differentiate into fiber (-1.5 DPA) were selected for co-expression network construction (Fig. S9-S10, Table S11).

Among these four networks, some core components shared between adjacent stages, while some were specific at one stage. For example, the PROTODERMAL FACTOR1 (*PDF1*) gene, whose silencing caused retardation of fiber initiation (Deng *et al.*, 2012), was identified as a core network component both at -1.5, -0.5 and 0 DPA (Fig. S10). Two genes coding lipid transfer protein (*LTP3* and *LTP6*), which can transport metabolites around the cell for their membrane biosynthesis (Wu *et al.*, 2007; Wu *et al.*, 2006), were the core components at both -1.5 and -1 DPA network. At -0.5 DPA network, AMP-dependent synthetase and ligase (*LACS1*), ABC-2 type transporter (*ABCG13*), Ribosomal protein (*RPS29A*) and a transcription factor *MYB25* were located in the core (Fig. S10). At 0 DPA, the core network components including some function unknown genes, such as *Ghir_D08G021100*, *HUTL*, *Era* and others.

Filtering the network down by TFs, four new TF regulatory networks were obtained (Fig. 6). The core network components in latter three stages (-1, -0.5, 0 DPA) were *MYB25-like*, *HD1*, and *HOX3*, respectively. These three genes are superstars in fiber cell development (Cao *et al.*, 2020; Shan et al., 2014; Walford et al., 2011, 2012; Wan et al., 2016), but the detailed temporal pattern of how they work in individual fiber cell has not yet been answered. After comparative analysis, *MYB25-like_At* (Ghir_A12G017450) and *MYB25-like_Dt* (Ghir_D12G017660) were only identified at -1 and -0.5 DPA network (Fig. 6), as both of them have been reported associated with lint fiber development (Zhu *et al.*, 2018). While *HD1* gene, who expressed predominantly in epidermal tissues during early fiber development (Walford *et al.*, 2012), was identified in all these four stages. Meanwhile, the expression pattern of PROTODERMAL FACTOR 2 (*PDF2*) gene (Abe *et al.*, 2003) was consistent with *HD1* (Fig. 6). Another HD-ZIP TF *HOX3*, its silencing greatly reduces fiber length whereas overexpression leads to longer fiber (Shan *et al.*, 2014), was only identified at 0 DPA network (Fig. 6). And also, a R2R3 MYB TF *MYB109*, who plays a role in fiber elongation (Pu *et al.*, 2008), showed the same pattern as *HOX3*.

As core network components, MIXTA-like MYB genes *MYB25* and *MML9* were only identified at -0.5 and 0 DPA networks (Fig. 6). *MYB25* played a role in regulating

specialized outgrowth of fiber cell (Machado *et al.*, 2009). *MML9* was reported preferentially expressed during fiber initiation (Zhang *et al.*, 2015). At -1.5 DPA, the core gene *WRKY44* (Ghir_A04G008530), also named *TTG2*, was reported in *Arabidopsis* involved in a regulatory module for regulation of seed coat mucilage synthesis (Xu *et al.*, 2022).

MYB25-like determines fiber differentiation at -1 DPA and _HOX3_ determines fiber tip-biased diffuse growth at 0 DPA

MYB25-like defunction was previously reported to be the determinant of cotton fiberless mutant (Walford *et al.*, 2011; Wan *et al.*, 2016). Histochemical localization of *MYB25-like*-GUS expression was localized in the epidermal layer of 0 DPA ovules (Walford *et al.*, 2011). Here, compared to sense probe hybrid result, *MYB25-like* expressed in ovule outer integument layer, including outer pigment layer and epidermis, and more highly in fiber cells (Fig. 7a). Two independent lines with different editing types resulting both *MYB25-like_At* and *MYB25-like_Dt* knocked out were obtained (Fig. S11a) and they showed totally fiberless phenotype (Fig. S11b). Distinct from *GhMYB25-like* RNA interference suppression lines, which still has few lint fibers attached on mature seeds (Walford *et al.*, 2011), the seeds of our CRISPR mutant lines were totally glabrous. When observing 0 DPA ovules with SEM, no fiber initials was found in both *MYB25-like_CR* lines as compared to WT (Jin668, the transgenic receptor material, Fig. 7b,c).

Cotton *PDF2* was a member of HD-ZIP class IV homeodomain protein family, which shares the highest homology with *AtPDF2* in *Arabidopsis*. *AtPDF2* and its paralogue *ATML1* are functionally interchangeable and act on *Arabidopsis* shoot epidermal cell differentiation (Abe *et al.*, 2003; Rombola-Caldentey *et al.*, 2014). Cotton *PDF2* was validated highly expressed in ovule epidermis and fiber cells by *in situ* hybridization assays (Fig. 7d). Two types of *PDF2* CRISPR mutant lines (*PDF2_CR*, both *PDF2* homologs from At and Dt subgenomes knocked out) were created (Fig. S12a) and their fiber initials were significantly decreased on 0 DPA ovules as compared to WT (Fig. 7e,f). However, the mature fiber phenotype and fiber quality of *PDF2_CR* lines were

nearly the same as WT, except for a decrease in fuzz fiber density (Fig. S12b-c).

Cotton *HOX3* was highly expressed in 0 DPA fiber cells validated by *in situ* hybridization assays (Fig. 7g). *HOX3* transcript level was sharply decreased in homozygous transgene co-suppression lines, and resulting in retarded fiber elongation (Shan *et al.*, 2014). Here, through CRISPR-Cas9 based gene editing technology, the *HOX3* CRISPR mutant line (*HOX3_CR*, both *HOX3* homologs from At and Dt subgenomes knocked out) was obtained (Fig. S13a). Compared to WT, the phenotype of *HOX3_CR* line was seemingly naked seeds (Fig. S13b), different from the *HOX3* transgene co-suppression lines which attached with very short fibers (Shan *et al.*, 2014). When observed with SEM, normal fiber initials can be seen on 0 DPA ovules both in WT and *HOX3_CR* line (Fig. 7h). At 1 DPA, fiber initials normally elongated on WT ovules but not on *HOX3_CR* line (Fig. 7h).

Above all, the specific function of *MYB25-like* and *HOX3* were more clearly clarified using CRISPR gene editing technology. *MYB25-like* determines fiber differentiation, so defunctionalization of *MYB25-like* produces no fiber initials; *HOX3* determines fiber tip-biased diffuse growth, so cotton plants with *HOX3* defunctionalization can perform normal fiber differentiation and diffuse growth, but cannot transform into tip-biased diffuse growth. Besides, *PDF2* play a role in fiber initiation, while it's not as critical as *MYB25-like* and *HOX3*.

Discussion

The challenges of scRNA-seq in non-model crops

scRNA-seq has flourished in plants (Ryu *et al.*, 2021), but challenges remain, especially for non-model crops. The biggest obstacle is definition for each cell type. In model plants *Arabidopsis* and poplars, a large number of known markers had been reported, which was enough to define almost all the cell types (Chen *et al.*, 2021; Zhang *et al.*, 2019). While in rice, corn, and cotton (in this study), cell type definition still relies on RNA *in situ* hybridization (Liu *et al.*, 2021b; Satterlee *et al.*, 2020; Wang *et al.*, 2021b; Xu *et al.*, 2021), which is still a technologically dependent and time-

consuming challenge. Spatial transcriptome technology overcame this disadvantage and has been successfully applied in *Arabidopsis* leaves, which showed the *bona fide* single-cell spatial transcriptome profiles (Xia *et al.*, 2022). However, this technology still has many technical barriers in plant, for example tissue optimization, due to the existence of cell wall.

The mutants are very effective to verify cell type definition. After scRNA-seq on root tips of *rhd6* (lack root hair) and *gl2* (lack non-hair cells) mutant, the cluster of root hair cells and non-hair cells were respectively reduced compared to wild type (Ryu *et al.*, 2019). The similar trend was observed in *Arabidopsis* root tips before and after heat stress treatment. For example, root hair cluster cell number was decreased in heat shock sample (Jean-Baptiste *et al.*, 2019). Here, we sequenced cotton Xu142 *fl* fiberless mutant ovules, and no fiber cell cluster was identified in Xu142 *fl* (Fig. 4b). This not only proves the reliability of scRNA-seq in cotton ovules but also confirms the accuracy of our cell type definition.

Fiber cell initiation successively experiences differentiation, diffuse growth and tip-biased diffuse growth

The study on fiber initiation started from last century (Stewart, 1975). The morning of anthesis (0 DPA) was always selected as a representative stage for fiber initiation (Haigler *et al.*, 2012; Lee *et al.*, 2007). And a lot of SEM observation and comparative transcriptome profiling was performed on 0 DPA ovules (Hu *et al.*, 2018; Qin *et al.*, 2019; Walford *et al.*, 2012; Zhang *et al.*, 2011). In this study, fiber initials can be observed on Xu142_LF line ovule epidermis at -0.5 DPA (Fig. 1, Fig. S1). Further, Xu142_LF fiber cell was identified differentiated at -1 DPA (Fig. 5b). Therefore, with Xu142_LF line growing at Wuhan city, its -1 DPA samples should be collected for fiber differentiation study, and -0.5 DPA samples for fiber diffuse growth study.

scRNA-seq has been widely used in plants (Ryu *et al.*, 2021; Seyfferth *et al.*, 2021; Shaw *et al.*, 2021), but reported studies were limited to one specific developmental stage, such as primary root tips of 5 days after germination in *Arabidopsis* (Ryu *et al.*,

2019), root tips of 5 days rice seedlings (Zhang *et al.*, 2021a), leaf blades of 7 days seedlings in peanut (Liu *et al.*, 2021a), 5-10 mm developing ears of corn (Xu *et al.*, 2021) and stem below the third internode of 4-month-old poplar (Chen *et al.*, 2021). Here, our sampling strategy included 4 developmental stages during fiber initiation. In single sample, cell heterogeneity can be identified (Fig. 2). In multiple samples, the elaborate developmental dynamics of fiber cell differentiation, diffuse growth and tip-biased diffuse growth could be recognized (Fig. 5).

Integrating fiber cell developmental trajectory, TF regulatory networks, and core network components functional validation, a model was proposed for fiber initiation focusing on a single cell (Fig. 8a). A fiber cell successively experiences the process of differentiation, diffuse growth and tip-biased diffuse growth, with different key regulators involved in each process. *HD1* and *PDF2* expressed at all four stages (Fig. 6 and Fig. 8a). They play roles in ovule epidermis before fiber differentiation, and also in fiber cell after then. *MYB25-like* was identified preferentially expressed at -1 DPA (Fig. 5e and Fig. 6), promoting fiber cell differentiation. After fiber cell differentiated, *MYB25* and *MML9* start to express and gradually increased (Fig. S7f,g), accompanying with fiber cell diffuse growth (Fig. 8a). *MYB25* was likely regulated by MYB25-like (Walford *et al.*, 2011) then play a role in promoting fiber diffuse growth. At 0 DPA, a new core network component *HOX3* appeared (Fig. 6) and promoted fiber tip-biased diffuse growth (Fig. 8a).

On whole ovule, fiber initiation firstly occurred on chalazal end and progressed gradually towards micropylar end (Fig.7 and Fig. 8b), according to earlier morphological observations (Stewart, 1975). Once fiber cell differentiated at -1 DPA, the continuous development leads to multiple processes occurring simultaneously on ovule epidermis (Fig. 8b). As shown in re-clustering result (Fig. S7h), -1 DPA cells located only at the start of trajectory. At -0.5 DPA, cell distribution expanded (Fig. S7h), means these cells including both early fiber cell and diffuse growing cells (Fig. 8b). The 0 DPA cells were located at all the five sub-clusters (Fig. S7h), including cells in differentiation, diffuse growth and tip-biased diffuse growth (Fig. 8b). With scRNA-

seq, gene expression pattern in cells at different developmental stages can be more finely characterized, which is an important reason why this study can subdivide fiber initiation into three processes and identify the regulatory network of each process.

MYB25-like and _HOX3_ play most important roles as commanders in fiber differentiation and fiber tip-biased diffuse growth

Through cotton transgenic verification, *PDF2* gene mutation only decreased fiber initials at 0 DPA (Fig. 7e). In *Arabidopsis*, single mutant of *AtPDF2* or its paralogue *ATML1* display normal shoot development, while double mutant results in severe defects in shoot epidermal cell differentiation (Abe *et al.*, 2003; Rombola-Caldentey *et al.*, 2014). In cotton, *PDF2* homologs *GhHOX1* (Wang *et al.*, 2004) and *GhHD1* (Walford *et al.*, 2012) had been cloned and verified promoting fiber development. *PDF2* might also functionally interchangeable to *GhHOX1* or *GhHD1*, or both of them. So, double mutants on *PDF2* and its paralogue might display obvious phenotype in cotton.

There is no report to clarify the spatiotemporal pattern of *MYB25-like* gene at single cell resolution. Here, expression pattern on fiber developmental trajectory showed that *MYB25-like* was preferentially expressed in early fiber cells (Fig. 5e). With qRT-PCR measurements, *MYB25-like* showed higher expression in -1 to 3 DPA ovules (Walford *et al.*, 2011), in can be inferred that the expression of *MYB25-like* mainly contributed by the newly differentiated fiber cells. On one hand, *MYB25-like* gene starts to be highly expressed at -1 DPA (Walford et al., 2011), which is consistent with our result (Fig. 6). On the other hand, its high expression lasts until 3 DPA, a stage when fuzz fiber begins to differentiate, which indicated that *MYB25-like* is also participated in fuzz initiation.

Hormone response and sugar signal genes were reported involved in fiber initiation (Huang *et al.*, 2021; Tian and Zhang, 2021; Wang *et al.*, 2021a). Interestingly, *MYB25-like* was down-regulated in fiber-deficient mutants of hormone- or sugar-signaling genes. For example, *MYB25-like* activity was suppressed in *GhJAZ2* overexpressing lines (Hu *et al.*, 2016); the transcripts of *MYB25-like* was dramatically reduced in

GhVIN1-RNAi lines (Wang *et al.*, 2014) and down-regulated in *GhPIN1a*-RNAi ovules (Zhang *et al.*, 2017); *MYB25-like* was severely down-regulated in ovules cultured with excess ZT (*trans*-zeatin, a kind of cytokinin) (Zeng *et al.*, 2019). These suggested that hormone and sugar signaling may act on the upstream of *MYB25-like* and through this gene to exert their effects on fiber development. At least among currently known fiberless mutants (such as Xu142 *fl*, *GhVIN1*-RNAi lines and *GhPIN1a*-RNAi lines) (Walford et al., 2011; Wang et al., 2014; Zhang et al., 2017), *MYB25-like* gene was the direct commander controlling fiber cell differentiation.

HOX3 gene highly expressed at 0 DPA (Fig. 6) and mainly in fiber cells (Fig. 2c and Fig. 7g). Fiber cell can normally differentiate and diffuse grow but cannot transform into tip-biased diffuse growth in *HOX3_CR* line (Fig. 7h). Therefore, *HOX3* seems to act as a commander controlling fiber cell development transformation. As shown in Fig. 8a, two commanders (*MYB25-like* and *HOX3*) positively regulating early fiber development are functioning like a 'relay race' model. The RNA-seq, chromatin immunoprecipitation sequencing (ChIP-seq) with these cotton mutants and identification of TF regulatory multiple complexes have been ongoing in our lab, the molecular mechanism hidden below 'relay race' model will be revealed in the near future.

Core regulators controlling the differentiation and tip-biased diffuse growth of *Arabidopsis* trichomes and cotton fibers may be conserved

In *Arabidopsis*, mutant of two genes, *GL1* (R2R3-type MYB transcription factor GLABRA1) and *GL2* (homeobox transcription factor GLABRA2), produced no leaf trichomes (Hulskamp *et al.*, 1994; Szymanski *et al.*, 1998), but they were functioning in different ways. For *gl2* leaves some large cells bigger than normal epidermal cells appeared and their nuclei were also in the same size as wild-type trichomes, which means trichome cells can differentiate and enlarge, but no local outgrowth (Hulskamp *et al.*, 1994). However, all the epidermal cells of *gl1* leaves were uniform in size and shape, and no specialized cells was found, which suggested that *GL1* is required for trichome cell differentiation (Hulskamp *et al.*, 1994). In this aspect, *MYB25-like* gene

functions similarly to *GL1*, and *HOX3* to *GL2*. Because they were in same gene family respectively, but also the similar ways they acting on fiber cell or trichomes. In addition, other signals affecting trichome development such as phytohormones (Qi *et al.*, 2014; Qi *et al.*, 2011; Zhou *et al.*, 2013), miRNAs (Xue *et al.*, 2014; Yu *et al.*, 2010), R3 MYBs (Gan *et al.*, 2011; Vadde *et al.*, 2019; Wang *et al.*, 2007; Zhao *et al.*, 2008) or histone demethylase (Hung *et al.*, 2020), were always acting on the upstream of *GL1*. This also showed a similar pattern as cotton *MYB25-like*. Therefore, the core regulators controlling differentiation and tip-biased diffuse growth of *Arabidopsis* leaf trichomes and cotton fibers appear to be conserved. This also provides some inspiration for exploring other types of epidermal tissue, such as tomato trichomes, kapok fiber, tea trichomes, and okra seed coat mucilage.

Conclusion

In summary, with the help of scRNA-seq technology and reasonable multi-stage sampling strategy, fiber cell differentiation, diffuse growth and tip-biased diffuse growth process during fiber initiation were more finely delineated. Through gene regulatory network analysis and CRISPR-Cas9 based gene function verification, the *MYB25-like* gene was newly defined as a commander acting at -1 DPA on fiber cell differentiation. In addition, *HOX3* was proved to be another commander controlling fiber development transformation into tip-biased diffuse growth. It's the first report by applying scRNA-seq technology in cotton fiber cell, and our result provide a more refined and detailed stage definition of fiber initiation. The valuable resource provided here will help to further explore the mechanism of fiber development, plant trichomes differentiation and single cell fate determination.

Materials and Methods

Plant growth and sample collection

The cotton Lint-Fuzz (Xu142_LF) line, derived from recombinant inbred lines of Xu142 × Xu142 *fl* (Hu et al., 2018), was planted in the experimental field at Huazhong Agriculture University, Wuhan, China. Cotton plants were grown under conventional

field management. Samples were collected when cotton plants began flowering. -1 DPA flower buds, 0 DPA flowers were collected in the morning (8:00 a.m.), and -1.5 DPA and -0.5 DPA buds were collected in the evening (8:00 p.m.).

Scanning electron microscopy (SEM) and transmission electron microscopy (TEM)

Cotton bolls of Xu142_LF were collected at -1.5 DPA and every four hours from -1 DPA to 0 DPA (i.e., 8:00, 12:00, 16:00, 20:00, 24:00 at -1 DPA and 4:00, 8:00 at 0 DPA). Ovules were fixed for further SEM observation. The detailed processes were performed as reported (Hu *et al.*, 2018).

Ovules at -1.5, -1, -0.5, 0 DPA were collected and fixed with 2.5% glutaraldehyde, followed by fixing with 1% OsO_4, acetone gradient dehydration, resin infiltration and embedment. Sections of 90 nm were cut, counterstained, then visualized using a TEM (New Bio-TEM H-7500, HITACHI, Japan), according to the methods reported (Liu *et al.*, 2015; Min *et al.*, 2013).

Tissue digestion and scRNA-seq library preparation

Ovules collected from Xu142_LF at -1.5, -1, -0.5 and 0 DPA, and from 0 DPA of Xu142 and Xu142 *fl* were respectively placed into a 30 mm-diameter Petri dish containing 3 mL enzyme solution (1.5% [w/v] cellulose ["ONOZUKA" R-10, Yakult], 1% [w/v] hemicellulose [Sigma-Aldrich], 0.75% [w/v] Macerozyme [R-10, Solarbio], 0.4 M Mannitol, 20 mM MES [pH 5.7], 20 mM KCl, 10 mM $CaCl_2$, 0.1% [w/v] bovine serum albumin (BSA)). Four cotton balls were harvested and all ovules in these balls were collected together as one sample. Petri dishes were placed in a vacuum drying oven and kept at 0.1 atmospheric pressure for 5 minutes to promote the removal of gases in the ovule to facilitate the full submersion of the ovule in the enzyme solution. The dish was rotated at 60 rpm for 4 h at 25°C.

The enzyme solution was filtered with a 40 μm cell strainer, 1× PBS solution containing 0.04% BSA was added to the filtrate, gently inverted and mixed, centrifuged at 100 rcf for 2 min, and the supernatant was carefully discarded. Filtering, rinsing, and centrifuging steps was repeated twice to obtain a single cell suspension. The cell suspension was kept on ice to prevent cell death. Protoplast concentration was

determined using a hemocytometer, and the ideal concentration is 700-1200 cells/μl. Protoplasts were stained with 0.4% trypan blue solution for detecting viability.

A commercially available droplet-based system from 10× Genomics Inc. (Zheng *et al.*, 2017) was used to isolate protoplasts. The protoplast suspension was loaded into Chromium microfluidic chips with 3' v2 chemistry. RNA from the barcoded cells was subsequently reverse-transcribed, then sequencing libraries were constructed and sequenced on NovaSeq (Illumina) platform using Hiseq PE150 strategy.

Processing of scRNA-seq data

The detailed calculation methods and parameter settings involved in this process are described in Document S1. In brief, the raw data in FASTQ format were first processed to obtain clean reads. *Gossypium hirsutum* TM-1 genome (Wang *et al.*, 2019) was used as reference genome. Only reads that were uniquely mapped were used for UMI counting. The Seurat package (v. 4.0.1) implemented in R (v. 4.0.0) was used for gene-cell matrices analysis, including doublets, no-load cells and dead cells filtration. Principal Components Analysis (PCA), Uniform Manifold Approximation and Projection (UMAP) and *t*-Stochastic Neighbor Embedding (*t*-SNE) analyses were performed for visualizing data in 2-d space. Cluster-enriched genes were identified with Seurat function 'FindAllMarkers'. The subset of related clusters was extracted and processed for pseudo-time analysis. The weighted gene co-expression network analysis (WGCNA) was performed in R software (v. 4.0.0) following the official process. The gene regulatory network displaying, personalized Gene Ontology (GO) and pathway enrichment analysis were performed using OmicShare tools, a free online platform for data analysis (www.omicshare.com/tools).

Analysis of RNA-seq data

Bulk RNA-seq data of Xu142_LF 0 DPA ovule outer integument from our previous study (Hu *et al.*, 2018) was used here. Clean reads were newly mapped to the updated TM-1 reference genome (Wang *et al.*, 2019) using Hisat2 (v2.1.0) software (Kim *et al.*, 2015). The mapping reads were sorted to filter those reads representing PCR duplicates. Sequencing reads with mapping quality of less than 25 were filtered using SAMTOOLS

(v.0.1.19) (Li *et al.*, 2009), the remaining were used to calculate gene expression levels using STRINGTIE software (v.1.3.4) with default settings (Pertea *et al.*, 2015).

RNA *in situ* hybridization

In situ hybridization was carried out as described in cotton research previously (Zhang *et al.*, 2017). Briefly, 0 DPA ovaries of Xu142_LF were collected and embedded in paraffin. 10 μm paraffin sections were de-paraffinized, rehydrated and incubated overnight with the Dig-labelled RNA probe (Roche). Sections were then incubated with alkaline phosphatase-conjugated anti-digoxigenin (anti-Dig-AP, Roche) and the signal was detected by nitro-blue tetrazolium/5-bromo-4-chloro-3-inodyl-phosphate (NBT/BCIP) color substrate solution (Roche). Sections incubated with sense RNA probe were used as negative control. Images were captured using fully motorized upright microscope (Leica DM6B) in bright-field mode. Primers are listed in Table S12.

Vector construction and plant transformation

CRISPR technology was employed to create the *GhMYB25-like-CR* (Ghir_A12G017450/Ghir_D12G017660*)*, *GhHOX3-CR* (Ghir_A12G028530/ Ghir_D12G028680) and *GhPDF2-CR* (Ghir_A10G001030/Ghir_D10G001810) mutants in cotton. For targeting *MYB25-like* genes, two sgRNAs (CTCCATGTAGCGACAAGGTG; CGCCCTTCTTGGAAACAGGT) were designed for targeting MYB25-like. The primers listed in the Table S12 were used to amplify tRNA and gRNA from the template pGTR vector. Two different fragments containing tRNA-sgRNA1 and gRNA-tRNA-sgRNA2 were integrated together by an overlapping extension PCR. Finally, the PCR products were purified and inserted the Bsa I-digested pRGEB32-GhU6.9 vector using ClonExpress II One Step Cloning Kit (Vazyme) (Wang *et al.*, 2018). For targeting *GhHOX3* (ACCGGTACAACCTGTTCATA) and *GhPDF2* (CAACTGTGTCTCCTTACTTA) genes, single sgRNA was used in the *GhHOX3-CR* and *GhPDF2-CR* vector. The positive vectors were transformed into *Agrobacterium tumefaciens* strain EHA105 for cotton transformation. JIN668 was the transgenic receptor (Li *et al.*, 2019).

On-target analysis of gene-edited plants

T_0 positive transgenic plants were screened by PCR analysis using Cas9 forward and reverse primers (Table S12). For identification of mutated alleles in T_1 transgenic lines, high-throughput (Hi-Tom) sequencing was adopted (Liu *et al.*, 2019). First, the targeted regions were amplified by PCR using site-specific primers. Second, barcode primers were used to add barcodes to the first-round PCR products. After the second-round PCR amplification, the products of all samples were mixed in equal amounts and purified to perform next-generation sequencing (NGS). Finally, NGS data was analyzed by Hi-Tom platform (http://hi-tom.net/hi-tom/).

Acknowledgements

This work was supported by grants from the National Natural Science Foundation of China (31830062 and 31901576), and the funds from the Interdisciplinary Sciences Research Institute of Huazhong Agricultural University (2662021JC005). We thank Novogene Corporation (http://www.novogene.com/) for assistance with scRNA-seq library preparation and sequencing.

Conflicts of interest

The authors declared that they have no conflict of interest.

Author contributions

L.T. and X.Z. conceived the project and designed the experiments. Y.Q. performed protoplast isolating, M.S. performed the transgenic experiments. Y.Q., M.S. and L.S. performed *in situ* hybridization experiment. Y.Q., M.S. and J.Y. performed bioinformatics analysis. Z.Y., W.L., J.X., M.X., L.S., Y.L. and G.Z. contributed to the vector construction. Y.Q., Z.L. and Z.X. built a website for data sharing. Y.Q., M.S. and L.T. wrote the manuscript, X.Y., M.W., K.L. and X.Z. revised the manuscript with feedback from all other authors.

Data availability

The scRNA-seq data have been deposited in the NCBI SRA database

(https://www.ncbi.nlm.nih.gov/bioproject/) with BioProject number PRJNA600131.

References

Abe, M., Katsumata, H., Komeda, Y., and Takahashi, T. 2003. Regulation of shoot epidermal cell differentiation by a pair of homeodomain proteins in Arabidopsis. *Development* **130**: 635-643.

Bedon, F., Ziolkowski, L., Walford, S.A., Dennis, E.S., and Llewellyn, D.J. 2014. Members of the MYBMIXTA-like transcription factors may orchestrate the initiation of fiber development in cotton seeds. *Front Plant Sci* **5**: 179.

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

Chen, Y., Tong, S., Jiang, Y., Ai, F., Feng, Y., Zhang, J., Gong, J., Qin, J., Zhang, Y., Zhu, Y., *et al.* 2021. Transcriptional landscape of highly lignified poplar stems at single-cell resolution. *Genome Biol* **22**: 319.

Deng, F., Tu, L., Tan, J., Li, Y., Nie, Y., and Zhang, X. 2012. GbPDF1 is involved in cotton fiber initiation via the core cis-element HDZIP2ATATHB2. *Plant Physiol* **158**: 890-904.

Denyer, T., Ma, X., Klesen, S., Scacchi, E., Nieselt, K., and Timmermans, M.C.P. 2019. Spatiotemporal Developmental Trajectories in the Arabidopsis Root Revealed Using High-Throughput Single-Cell RNA Sequencing. *Dev Cell* **48**: 840-852.

Denyer, T., and Timmermans, M.C.P. 2022. Crafting a blueprint for single-cell RNA sequencing. *Trends Plant Sci* **27**: 92-103.

Gala, H.P., Lanctot, A., Jean-Baptiste, K., Guiziou, S., Chu, J.C., Zemke, J.E., George, W., Queitsch, C., Cuperus, J.T., and Nemhauser, J.L. 2021. A single-cell view of the transcriptome during lateral root initiation in Arabidopsis thaliana. *Plant Cell* **33**: 2197-2220.

Gan, L., Xia, K., Chen, J.G., and Wang, S. 2011. Functional characterization of TRICHOMELESS2, a new single-repeat R3 MYB transcription factor in the regulation of trichome patterning in Arabidopsis. *BMC Plant Biol* **11**: 176.

Gou, J.Y., Wang, L.J., Chen, S.P., Hu, W.L., and Chen, X.Y. 2007. Gene expression and metabolite profiles of cotton fiber during cell elongation and secondary cell wall synthesis. *Cell Res* **17**: 422-434.

Haigler, C.H., Betancur, L., Stiff, M.R., and Tuttle, J.R. 2012. Cotton fiber: a powerful single-cell model for cell wall and cellulose research. *Front Plant Sci* **3**: 104.

Hovav, R., Udall, J.A., Chaudhary, B., Hovav, E., Flagel, L., Hu, G., and Wendel, J.F. 2008. The evolution of spinnable cotton fiber entailed prolonged development and a novel metabolism. *PLoS Genet* **4**: e25.

Hu, H., He, X., Tu, L., Zhu, L., Zhu, S., Ge, Z., and Zhang, X. 2016. GhJAZ2 negatively regulates cotton fiber initiation by interacting with the R2R3-MYB transcription factor GhMYB25-like. *Plant J* **88**: 921-935.

Hu, H., Wang, M., Ding, Y., Zhu, S., Zhao, G., Tu, L., and Zhang, X. 2018. Transcriptomic repertoires depict the initiation of lint and fuzz fibres in cotton (Gossypium hirsutum L.). *Plant Biotechnol J* **16**: 1002-1012.

Huang, G., Huang, J.Q., Chen, X.Y., and Zhu, Y.X. 2021. Recent Advances and Future Perspectives in Cotton Research. *Annu Rev Plant Biol* **72**: 437-462.

Hulskamp, M., Misra, S., and Jurgens, G. 1994. Genetic dissection of trichome cell development in Arabidopsis. *Cell* **76**: 555-566.

Hung, F.Y., Chen, J.H., Feng, Y.R., Lai, Y.C., Yang, S., and Wu, K. 2020. Arabidopsis JMJ29 is involved in trichome development by regulating the core trichome initiation gene GLABRA3. *Plant J* **103**: 1735-1743.

Jean-Baptiste, K., McFaline-Figueroa, J.L., Alexandre, C.M., Dorrity, M.W., Saunders, L., Bubb, K.L., Trapnell, C., Fields, S., Queitsch, C., and Cuperus, J.T. 2019. Dynamics of Gene Expression in Single Root

Cells of Arabidopsis thaliana. *Plant Cell* **31**: 993-1011.

Ji, S.J., Lu, Y.C., Feng, J.X., Wei, G., Li, J., Shi, Y.H., Fu, Q., Liu, D., Luo, J.C., and Zhu, Y.X. 2003. Isolation and analyses of genes preferentially expressed during early cotton fiber development by subtractive PCR and cDNA array. *Nucleic Acids Res* **31**: 2534-2543.

Kim, D., Langmead, B., and Salzberg, S.L. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12**: 357-360.

Kim, J.Y., Symeonidi, E., Pang, T.Y., Denyer, T., Weidauer, D., Bezrutczyk, M., Miras, M., Zollner, N., Hartwig, T., Wudick, M.M., *et al.* 2021. Distinct identities of leaf phloem cells revealed by single cell transcriptomics. *Plant Cell* **33**: 511-530.

Lee, J.J., Woodward, A.W., and Chen, Z.J. 2007. Gene expression changes and early events in cotton fibre development. *Ann Bot* **100**: 1391-1401.

Li, C.H., Zhu, Y.Q., Meng, Y.L., Wang, J.W., Xu, K.X., Zhang, T.Z., and Chen, X.Y. 2002. Isolation of genes preferentially expressed in cotton fibers by cDNA filter arrays and RT-PCR. *Plant Sci* **163**: 1113-1120.

Li, H., Dai, X., Huang, X., Xu, M., Wang, Q., Yan, X., Sederoff, R.R., and Li, Q. 2021. Single-cell RNA sequencing reveals a high-resolution cell atlas of xylem in Populus. *J Integr Plant Biol* **63**: 1906-1921.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078-2079.

Li, J., Wang, M., Li, Y., Zhang, Q., Lindsey, K., Daniell, H., Jin, S., and Zhang, X. 2019. Multi-omics analyses reveal epigenomics basis for cotton somatic embryogenesis through successive regeneration acclimation process. *Plant Biotechnol J* **17**: 435-450.

Li, X., Zhang, X., Gao, S., Cui, F., Chen, W., Fan, L., and Qi, Y. 2022. Single-cell RNA sequencing reveals the landscape of maize root tips and assists in identification of cell type-specific nitrate-response genes. *The Crop Journal* , https://doi.org/10.1016/j.cj.2022.02.004.

Liu, H., Hu, D., Du, P., Wang, L., Liang, X., Li, H., Lu, Q., Li, S., Liu, H., Chen, X., *et al.* 2021a. Single-cell RNA-seq describes the transcriptome landscape and identifies critical transcription factors in the leaf blade of the allotetraploid peanut (Arachis hypogaea L.). *Plant Biotechnol J* **19**: 2261-2276.

Liu, J., Pang, C., Wei, H., Song, M., Meng, Y., Ma, J., Fan, S., and Yu, S. 2015. iTRAQ-facilitated proteomic profiling of anthers from a photosensitive male sterile mutant and wild-type cotton (Gossypium hirsutum L.). *J Proteomics* **126**: 68-81.

Liu, Q., Liang, Z., Feng, D., Jiang, S.J., Wang, Y.F., Du, Z.Y., Li, R.X., Hu, G.H., Zhang, P.X., Ma, Y.F., *et al.* 2021b. Transcriptional landscape of rice roots at the single-cell resolution. *Mol Plant* **14**: 384-394.

Liu, Q., Wang, C., Jiao, X., Zhang, H., Song, L., Li, Y., Gao, C., and Wang, K. 2019. Hi-TOM: a platform for high-throughput tracking of mutations induced by CRISPR/Cas systems. *Sci China Life Sci* **62**: 1-7.

Liu, W., Zhang, Y., Fang, X., Tran, S., Zhai, N., Yang, Z., Guo, F., Chen, L., Yu, J., Ison, M.S., *et al.* 2022a. Transcriptional landscapes of de novo root regeneration from detached Arabidopsis leaves revealed by time-lapse and single-cell RNA sequencing analyses. *Plant Commun* **3**: 100306.

Liu, Z., Wang, J., Zhou, Y., Zhang, Y., Qin, A., Yu, X., Zhao, Z., Wu, R., Guo, C., Bawa, G., *et al.* 2022b. Identification of novel regulators required for early development of vein pattern in the cotyledons by single-cell RNA-sequencing. *Plant J* **110**: 7-22.

Liu, Z., Zhou, Y., Guo, J., Li, J., Tian, Z., Zhu, Z., Wang, J., Wu, R., Zhang, B., Hu, Y., *et al.* 2020. Global Dynamic Molecular Profiling of Stomatal Lineage Cell Development by Single-Cell RNA Sequencing. *Mol Plant* **13**: 1178-1193.

Lopez-Anido, C.B., Vaten, A., Smoot, N.K., Sharma, N., Guo, V., Gong, Y., Anleu Gil, M.X., Weimer, A.K.,

and Bergmann, D.C. 2021. Single-cell resolution of lineage trajectories in the Arabidopsis stomatal lineage and developing leaf. *Dev Cell* **56**: 1043-1055.

Machado, A., Wu, Y., Yang, Y., Llewellyn, D.J., and Dennis, E.S. 2009. The MYB transcription factor GhMYB25 regulates early fibre and trichome development. *Plant J* **59**: 52-62.

Min, L., Zhu, L., Tu, L., Deng, F., Yuan, D., and Zhang, X. 2013. Cotton GhCKI disrupts normal male reproduction by delaying tapetum programmed cell death via inactivating starch synthase. *Plant J* **75**: 823-835.

Mo, Y., and Jiao, Y. 2022. Advances and applications of single-cell omics technologies in plant research. *Plant J* **110**: 1551-1563.

Omary, M., Gil-Yarom, N., Yahav, C., Steiner, E., Hendelman, A., and Efroni, I. 2022. A conserved superlocus regulates above- and belowground root initiation. *Science* **375**: eabf4368.

Ortiz-Ramirez, C., Guillotin, B., Xu, X., Rahni, R., Zhang, S., Yan, Z., Coqueiro Dias Araujo, P., Demesa-Arevalo, E., Lee, L., Van Eck, J., *et al.* 2021. Ground tissue circuitry regulates organ complexity in maize and Setaria. *Science* **374**: 1247-1252.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., and Salzberg, S.L. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33**: 290-295.

Pu, L., Li, Q., Fan, X., Yang, W., and Xue, Y. 2008. The R2R3 MYB transcription factor GhMYB109 is required for cotton fiber development. *Genetics* **180**: 811-820.

Qi, T., Huang, H., Wu, D., Yan, J., Qi, Y., Song, S., and Xie, D. 2014. Arabidopsis DELLA and JAZ proteins bind the WD-repeat/bHLH/MYB complex to modulate gibberellin and jasmonate signaling synergy. *Plant Cell* **26**: 1118-1133.

Qi, T.C., Song, S.S., Ren, Q.C., Wu, D.W., Huang, H., Chen, Y., Fan, M., Peng, W., Ren, C.M., and Xie, D.X. 2011. The Jasmonate-ZIM-Domain Proteins Interact with the WD-Repeat/bHLH/MYB Complexes to Regulate Jasmonate-Mediated Anthocyanin Accumulation and Trichome Initiation in Arabidopsis thaliana. *Plant Cell* **23**: 1795-1814.

Qin, Y., Sun, H., Hao, P., Wang, H., Wang, C., Ma, L., Wei, H., and Yu, S. 2019. Transcriptome analysis reveals differences in the mechanisms of fiber initiation and elongation between long- and short-fiber cotton (Gossypium hirsutum L.) lines. *BMC Genom* **20**: 633.

Qin, Y.M., and Zhu, Y.X. 2011. How cotton fibers elongate: a tale of linear cell-growth mode. *Curr Opin Plant Biol* **14**: 106-111.

Rapp, R.A., Haigler, C.H., Flagel, L., Hovav, R.H., Udall, J.A., and Wendel, J.F. 2010. Gene expression in developing fibres of Upland cotton (Gossypium hirsutum L.) was massively altered by domestication. *BMC Biol* **8**: 139.

Rombola-Caldentey, B., Rueda-Romero, P., Iglesias-Fernandez, R., Carbonero, P., and Onate-Sanchez, L. 2014. Arabidopsis DELLA and two HD-ZIP transcription factors regulate GA signaling in the epidermis through the L1 box cis-element. *Plant Cell* **26**: 2905-2919.

Ruan, Y.L., Llewellyn, D.J., and Furbank, R.T. 2003. Suppression of sucrose synthase gene expression represses cotton fiber cell initiation, elongation, and seed development. *Plant Cell* **15**: 952-964.

Ryu, K.H., Huang, L., Kang, H.M., and Schiefelbein, J. 2019. Single-Cell RNA Sequencing Resolves Molecular Relationships Among Individual Plant Cells. *Plant Physiol* **179**: 1444-1456.

Ryu, K.H., Zhu, Y., and Schiefelbein, J. 2021. Plant Cell Identity in the Era of Single-Cell Transcriptomics. *Annu Rev Genet* **55**: 479-496.

Satterlee, J.W., Strable, J., and Scanlon, M.J. 2020. Plant stem-cell organization and differentiation at single-cell resolution. *P Natl Acad Sci USA* **117**: 33689-33699.

Seyfferth, C., Renema, J., Wendrich, J.R., Eekhout, T., Seurinck, R., Vandamme, N., Blob, B., Saeys, Y.,

Helariutta, Y., Birnbaum, K.D., *et al.* 2021. Advances and Opportunities in Single-Cell Transcriptomics for Plant Research. *Annu Rev Plant Biol* **72**: 847-866.

Shahan, R., Hsu, C.W., Nolan, T.M., Cole, B.J., Taylor, I.W., Greenstreet, L., Zhang, S., Afanassiev, A., Vlot, A.H.C., Schiebinger, G., *et al.* 2022. A single-cell Arabidopsis root atlas reveals developmental trajectories in wild-type and cell identity mutants. *Dev Cell* **57**: 543-560.

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

Shaw, R., Tian, X., and Xu, J. 2021. Single-Cell Transcriptome Analysis in Plants: Advances and Challenges. *Mol Plant* **14**: 115-126.

Shi, Y.H., Zhu, S.W., Mao, X.Z., Feng, J.X., Qin, Y.M., Zhang, L., Cheng, J., Wei, L.P., Wang, Z.Y., and Zhu, Y.X. 2006. Transcriptome profiling, molecular biological, and physiological studies reveal a major role for ethylene in cotton fiber cell elongation. *Plant Cell* **18**: 651-664.

Stewart, J.M. 1975. Fiber initiation on the cotton ovule (Gossypium hirsutum). *Amer. J. Bot.* **62**: 723-730.

Sun, G., Xia, M., Li, J., Ma, W., Li, Q., Xie, J., Bai, S., Fang, S., Sun, T., Feng, X., *et al.* 2022. The maize single-nucleus transcriptome comprehensively describes signaling networks governing movement and development of grass stomata. *Plant Cell* **34**: 1890-1911.

Szymanski, D.B., Jilk, R.A., Pollock, S.M., and Marks, M.D. 1998. Control of GL2 expression in Arabidopsis leaves and trichomes. *Development* **125**: 1161-1171.

Tenorio Berrio, R., Verstaen, K., Vandamme, N., Pevernagie, J., Achon, I., Van Duyse, J., Van Isterdael, G., Saeys, Y., De Veylder, L., Inze, D., *et al.* 2022. Single-cell transcriptomics sheds light on the identity and metabolism of developing leaf cells. *Plant Physiol* **188**: 898-918.

Tian, Y., and Zhang, T. 2021. MIXTAs and phytohormones orchestrate cotton fiber development. *Curr Opin Plant Biol* **59**: 101975.

Tu, L.L., Zhang, X.L., Liang, S.G., Liu, D.Q., Zhu, L.F., Zeng, F.C., Nie, Y.C., Guo, X.P., Deng, F.L., Tan, J.F., *et al.* 2007. Genes expression analyses of sea-island cotton (Gossypium barbadense L.) during fiber development. *Plant Cell Rep* **26**: 1309-1320.

Vadde, B.V.L., Challa, K.R., Sunkara, P., Hegde, A.S., and Nath, U. 2019. The TCP4 Transcription Factor Directly Activates TRICHOMELESS1 and 2 and Suppresses Trichome Initiation. *Plant Physiol* **181**: 1587-1599.

Walford, S.A., Wu, Y., Llewellyn, D.J., and Dennis, E.S. 2011. GhMYB25-like: a key factor in early cotton fibre development. *Plant J* **65**: 785-797.

Walford, S.A., Wu, Y., Llewellyn, D.J., and Dennis, E.S. 2012. Epidermal cell differentiation in cotton mediated by the homeodomain leucine zipper gene, GhHD-1. *Plant J* **71**: 464-478.

Wan, Q., Guan, X., Yang, N., Wu, H., Pan, M., Liu, B., Fang, L., Yang, S., Hu, Y., Ye, W., *et al.* 2016. Small interfering RNAs from bidirectional transcripts of GhMML3_A12 regulate cotton fiber development. *New Phytol* **210**: 1298-1310.

Wang, L., Cook, A., Patrick, J.W., Chen, X.Y., and Ruan, Y.L. 2014. Silencing the vacuolar invertase gene GhVIN1 blocks cotton fiber initiation from the ovule epidermis, probably by suppressing a cohort of regulatory genes via sugar signaling. *Plant J* **78**: 686-696.

Wang, L., Kartika, D., and Ruan, Y.L. 2021a. Looking into 'hair tonics' for cotton fiber initiation. *New Phytol* **229**: 1844-1851.

Wang, M., Tu, L., Yuan, D., Zhu, D., Shen, C., Li, J., Liu, F., Pei, L., Wang, P., Zhao, G., *et al.* 2019. Reference genome sequences of two cultivated allotetraploid cottons, Gossypium hirsutum and Gossypium barbadense. *Nat Genet* **51**: 224-229.

Wang, P.C., Zhang, J., Sun, L., Ma, Y.Z., Xu, J., Liang, S.J., Deng, J.W., Tan, J.F., Zhang, Q.H., Tu, L.L., *et al.* 2018. High efficient multisites genome editing in allotetraploid cotton (Gossypium hirsutum) using CRISPR/Cas9 system. *Plant Biotechnol J* **16**: 137-150.

Wang, Q., Wu, Y., Peng, A., Cui, J., Zhao, M., Pan, Y., Zhang, M., Tian, K., Schwab, W., and Song, C. 2022. Single-cell transcriptome atlas reveals developmental trajectories and a novel metabolic pathway of catechin esters in tea leaves. *Plant Biotechnol J*, https://doi.org/10.1111/pbi.13891.

Wang, S., Kwak, S.H., Zeng, Q., Ellis, B.E., Chen, X.Y., Schiefelbein, J., and Chen, J.G. 2007. TRICHOMELESS1 regulates trichome patterning by suppressing GLABRA1 in Arabidopsis. *Development* **134**: 3873-3882.

Wang, S., Wang, J.W., Yu, N., Li, C.H., Luo, B., Gou, J.Y., Wang, L.J., and Chen, X.Y. 2004. Control of plant trichome development by a cotton fiber MYB gene. *Plant Cell* **16**: 2323-2334.

Wang, Y., Huan, Q., Li, K., and Qian, W. 2021b. Single-cell transcriptome atlas of the leaf and root of rice seedlings. *J Genet Genomics* **48**: 881-898.

Wendrich, J.R., Yang, B., Vandamme, N., Verstaen, K., Smet, W., Van de Velde, C., Minne, M., Wybouw, B., Mor, E., Arents, H.E., *et al.* 2020. Vascular transcription factors guide plant epidermal responses to limiting phosphate conditions. *Science* **370**: 810.

Wu, H., Tian, Y., Wan, Q., Fang, L., Guan, X., Chen, J., Hu, Y., Ye, W., Zhang, H., Guo, W., *et al.* 2018. Genetics and evolution of MIXTA genes regulating cotton lint fiber development. *New Phytol* **217**: 883-895.

Wu, Y., Llewellyn, D.J., White, R., Ruggiero, K., Al-Ghazi, Y., and Dennis, E.S. 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.

Wu, Y., Machado, A.C., White, R.G., Llewellyn, D.J., and Dennis, E.S. 2006. Expression profiling identifies genes expressed early during lint fibre initiation in cotton. *Plant Cell Physiol* **47**: 107-127.

Xia, K., Sun, H.X., Li, J., Li, J., Zhao, Y., Chen, L., Qin, C., Chen, R., Chen, Z., Liu, G., *et al.* 2022. The single-cell stereo-seq reveals region-specific cell subtypes and transcriptome profiling in Arabidopsis leaves. *Dev Cell* **57**: 1299-1310.

Xie, J., Li, M., Zeng, J., Li, X., and Zhang, D. 2022. Single-cell RNA sequencing profiles of stem-differentiating xylem in poplar. *Plant Biotechnol J* **20**: 417-419.

Xu, X., Crow, M., Rice, B.R., Li, F., Harris, B., Liu, L., Demesa-Arevalo, E., Lu, Z., Wang, L., Fox, N., *et al.* 2021. Single-cell RNA sequencing of developing maize ears facilitates functional analysis and trait candidate gene discovery. *Dev Cell* **56**: 557-568.

Xu, Y., Wang, Y., Du, J., Pei, S., Guo, S., Hao, R., Wang, D., Zhou, G., Li, S., O'Neill, M., *et al.* 2022. A DE1 BINDING FACTOR 1-GLABRA2 module regulates rhamnogalacturonan I biosynthesis in Arabidopsis seed coat mucilage. *Plant Cell* **34**: 1396-1414.

Xue, X.Y., Zhao, B., Chao, L.M., Chen, D.Y., Cui, W.R., Mao, Y.B., Wang, L.J., and Chen, X.Y. 2014. Interaction between Two Timing MicroRNAs Controls Trichome Distribution in Arabidopsis. *Plos Genet* **10**: e1004266.

Yu, N., Cai, W.J., Wang, S., Shan, C.M., Wang, L.J., and Chen, X.Y. 2010. Temporal control of trichome distribution by microRNA156-targeted SPL genes in Arabidopsis thaliana. *Plant Cell* **22**: 2322-2335.

Yu, Y., Wu, S., Nowak, J., Wang, G., Han, L., Feng, Z., Mendrinna, A., Ma, Y., Wang, H., Zhang, X., *et al.* 2019. Live-cell imaging of the cytoskeleton in elongating cotton fibres. *Nat Plants* **5**: 498-504.

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

Zhang, M., Zeng, J.Y., Long, H., Xiao, Y.H., Yan, X.Y., and Pei, Y. 2017. Auxin Regulates Cotton Fiber Initiation via GhPIN-Mediated Auxin Transport. *Plant Cell Physiol* **58**: 385-397.

Zhang, M., Zheng, X., Song, S., Zeng, Q., Hou, L., Li, D., Zhao, J., Wei, Y., Li, X., Luo, M., *et al.* 2011. Spatiotemporal manipulation of auxin biosynthesis in cotton ovule epidermal cells enhances fiber yield and quality. *Nat Biotechnol* **29**: 453-458.

Zhang, T., Hu, Y., Jiang, W., Fang, L., Guan, X., Chen, J., Zhang, J., Saski, C.A., Scheffler, B.E., Stelly, D.M., *et al.* 2015. Sequencing of allotetraploid cotton (Gossypium hirsutum L. acc. TM-1) provides a resource for fiber improvement. *Nat Biotechnol* **33**: 531-537.

Zhang, T.Q., Chen, Y., Liu, Y., Lin, W.H., and Wang, J.W. 2021a. Single-cell transcriptome atlas and chromatin accessibility landscape reveal differentiation trajectories in the rice root. *Nat Commun* **12**: 2053.

Zhang, T.Q., Chen, Y., and Wang, J.W. 2021b. A single-cell analysis of the Arabidopsis vegetative shoot apex. *Dev Cell* **56**: 1056-1074.

Zhang, T.Q., Xu, Z.G., Shang, G.D., and Wang, J.W. 2019. A Single-Cell RNA Sequencing Profiles the Developmental Landscape of Arabidopsis Root. *Mol Plant* **12**: 648-660.

Zhao, M., Morohashi, K., Hatlestad, G., Grotewold, E., and Lloyd, A. 2008. The TTG1-bHLH-MYB complex controls trichome cell fate and patterning through direct targeting of regulatory loci. *Development* **135**: 1991-1999.

Zheng, G.X., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., *et al.* 2017. Massively parallel digital transcriptional profiling of single cells. *Nat Commun* **8**: 14049.

Zhou, Z., Sun, L., Zhao, Y., An, L., Yan, A., Meng, X., and Gan, Y. 2013. Zinc Finger Protein 6 (ZFP6) regulates trichome initiation by integrating gibberellin and cytokinin signaling in Arabidopsis thaliana. *New Phytol* **198**: 699-708.

Zhu, Q.H., Yuan, Y., Stiller, W., Jia, Y., Wang, P., Pan, Z., Du, X., Llewellyn, D., and Wilson, I. 2018. Genetic dissection of the fuzzless seed trait in Gossypium barbadense. *J Exp Bot* **69**: 997-1009.

Zong, J., Wang, L., Zhu, L., Bian, L., Zhang, B., Chen, X., Huang, G., Zhang, X., Fan, J., Cao, L., *et al.* 2022. A rice single cell transcriptomic atlas defines the developmental trajectories of rice floret and inflorescence meristems. *New Phytol* **234**: 494-512.

Supporting information

Fig. S1 Dynamic changes on Xu142_LF ovule epidermis during fiber initiation.

Fig. S2 Tissue digestion and protoplast isolation.

Fig. S3 Comparation between single-cell RNA-seq and bulk RNA-seq of Xu142_LF.

Fig. S4 Expression and identification of cluster enriched genes.

Fig. S5 RNA *in situ* hybridization of cell-type representative genes.

Fig. S6 Single-cell RNA-seq and clusters identification on Xu142 and Xu142 *fl*.

Fig. S7 Fiber cell cluster identification on the LF combined sample.

Fig. S8 Clustering and annotation of four LF samples.

Fig. S9 Hierarchical cluster tree showing co-expression modules identified by WGCNA.

Fig. S10 Networks related to fiber cell initiation.

Fig. S11 Characterization of cotton CRISPR editing lines on *MYB25-like* genes.

Fig. S12 Characterization of cotton CRISPR editing lines on *PDF2* genes.

Fig. S13 Characterization of cotton CRISPR editing lines on *HOX3* genes.

Table S1 Summary of single-cell RNA sequencing results for each sample.

Table S2 Summary of protoplasting induced differentially expressed genes.

Table S3 List of enriched genes for each cluster.

Table S4 Cell number of each cluster identified from Xu142-vs-Xu142 *fl*.

Table S5 KEGG analysis of LF_0d fiber cell enriched genes.

Table S6 Cell barcode mapping between single sample and combined sample.

Table S7 Summary of 167 branch-dependent genes.

Table S8 GO enrichment analysis of 167 branch-dependent genes.

Table S9 Summary of cluster enriched genes.

Table S10 WGCNA on four LF samples.

Table S11 Weight coefficient of fiber development related genes.

Table S12 The primers used in this study.

Document S1 Details of bioinformatics analysis.

Figure legends

Fig. 1 The phenotype of ovule epidermis of Xu142_LF line during fiber initiation.

(a-d) Xu142_LF ovule epidermis observed by scanning electron microscopy (SEM) at -1.5, -1, -0.5 and 0 DPA, respectively. **(e)** Statistics of fiber initials number on ovule epidermis per unit area (100 * 100 μm^2) from **(a-d)**. **(f)** Statistics of fiber initials number on ovule epidermis per unit area (100 * 100 μm^2) from Fig. S1. The blue solid triangles and hollow triangles marked stages in (**e**) and (**f**) mean the same developmental stage, respectively. **(g-j)** Ovule epidermal cells observed by transmission electron microscope (TEM) at -1.5, -1, -0.5 and 0 DPA, respectively. Red arrows in (**i**) and (**j**) point to fiber cells. Bars: 100 μm (**a-d**); 10 μm (**g-j**).

Fig. 2 Cluster annotation of single-cell transcriptomes from cotton ovule outer integument.

(a) UMAP visualization of putative clusters from 1,703 cells in cotton ovule outer integument of LF_0d sample. Each dot denotes a single cell. Colors denote corresponding cell clusters. Resolution was 0.8. **(b)** *t*-SNE projection plot showing major clusters of the 1,703 individual cell transcriptomes of LF_0d sample. **(c)** UMAP projection plots showing transcript accumulation for known fiber markers in individual cells. Color intensity indicates the relative transcript level for the indicated gene in each cell. **(d)** Expression pattern of top 4 genes enriched in each cluster of LF_0d. Dot diameter, proportion of cluster cells expressing a given gene; color, average expression across cells in that cluster. **(e)** Schematic diagram of longitudinal section of 0 DPA cotton ovule showing spatial distribution of cell clusters in ovule outer integument. The left part is a magnified view of the shaded part on the right. F, fiber; Epi, epidermis; OPL, outer pigment layer; OI, outer integument layer; II, inner integument layer; N, nucellus.

Fig. 3 RNA *in situ* hybridization of cell-type representative marker genes.

(a) UMAP projection plots showing transcript accumulation for *DUF* (*Ghir_D07G016770*, function unknown), a novel fiber marker gene from cluster 5. **(b)** RNA *in situ* hybridization of *DUF* with the sense probe as a negative control. **(c)** Expression of non-fiber epidermis novel marker gene *Erg6* (*Ghir_A04G010380*, a methyltransferase encoding gene) in cluster 2 and 3. **(d)** RNA *in situ* hybridization of epidermis marker (*Erg6*) with the sense probe as a negative control. **(e)** UMAP projection plots showing transcript accumulation for outer pigment layer novel marker gene *HbdA* (*Ghir_D01G005520*, 3-hydroxyacyl-CoA dehydrogenase family protein) in cluster 0, 1, 4, 6, 7, 8. **(f)** RNA *in situ* hybridization of *HbdA* with the sense probe as a negative control. The hybridization signals of these marker genes in whole cotton ovules were shown in Fig. S5. Sections (10 μm) from Xu142_LF 0 DPA ovules were used for *in situ* hybridization. OI, outer integument; II, inner integument. Red arrows indicate fiber cells, and red arrowheads indicate non-fiber epidermal cells. Yellow arrows indicate outer pigment layer. Scale bars, 50 μm.

Fig. 4 Clarifying of fiber cell cluster and its enriched pathways.

(a) UMAP visualization of putative clusters in cotton ovule outer integument of Xu142 vs Xu142 *fl* sample. Each dot denotes a single cell. Colors denote corresponding cell clusters. Resolution was 0.2. **(b)** UMAP visualization of putative clusters which shown separately according to Xu142 and

Xu142 *fl*. **(c)** Schematic diagram showing the main pathways and regulatory factors active in 0 DPA fiber cells. The red protrusions on the ovule epidermis represent protruded fiber cells.

Fig. 5 Developmental trajectory of fiber cells.

(a) UMAP visualization of putative clusters of the combined sample. The scRNA-seq data of -1.5, -1, -0.5 and 0 DPA samples were combined for clustering. Each dot denotes a single cell. Colors denote corresponding cell clusters. Resolution was 0.2. **(b)** Cell number of each type identified from the combined sample, which shown at -1.5, -1, -0.5 and 0 DPA, respectively. **(c)** Re-clustering of all fiber cells identified in the combined sample. **(d)** A pseudotime trajectory showing fiber cells development. The developmental branch locations of 5 sub-clusters. Different colors represent the cells from each sub-cluster. **(e)** Expression patterns of fiber marker genes (*MYB25-like*, *MML4*, *MYB25* and *HD1*). The colors represent expression levels of these genes in individual cells. **(f)** Heat map displays the 167 branch-dependent genes with qval < 0.001. Each row represents one gene. These genes were clustered into 5 modules with distinct expression patterns. Different colors represent the gene expression level. The representative genes of each module are shown in the middle panel. The gene ontology (GO) terms for each module are shown on the right panel. **(g)** Representative marker genes (*ERF4*, *CINV2*, *ACS6* and *BZIP53*) expressed in branch 1 (upper panel), and expression of representative marker genes (*RPL24*, *RPS5B*, *LTP1* and *E6*) in branch 2 (lower panel). The color bar indicates relative expression levels. **(h)** RNA *in situ* hybridization of *E6* gene (*Ghir_D05G016260*) in LF_0d ovules with the sense probe as a negative control. OI, outer integument; II, inner integument. Red arrows indicate fiber cells, and red arrowheads indicate non-fiber epidermal cells. Scale bars, 50 μm.

Fig. 6 Transcription factors (TFs) regulatory network predicts key regulators in the four distinctive developmental processes during fiber initiation.

A total of 15, 15, 16, 9 TFs were identified related to fiber cell development at -1.5, -1, -0.5 and 0 DPA, respectively. Each circle represents a gene, and triangle represents transcription factor. Node size is equivalent to the number of predicted connections. Node color represents the weight abundance of predicted connections. Lines indicate edge weight (> 0.01) for each pair of genes. Edge width represents the strength of the predicted connection.

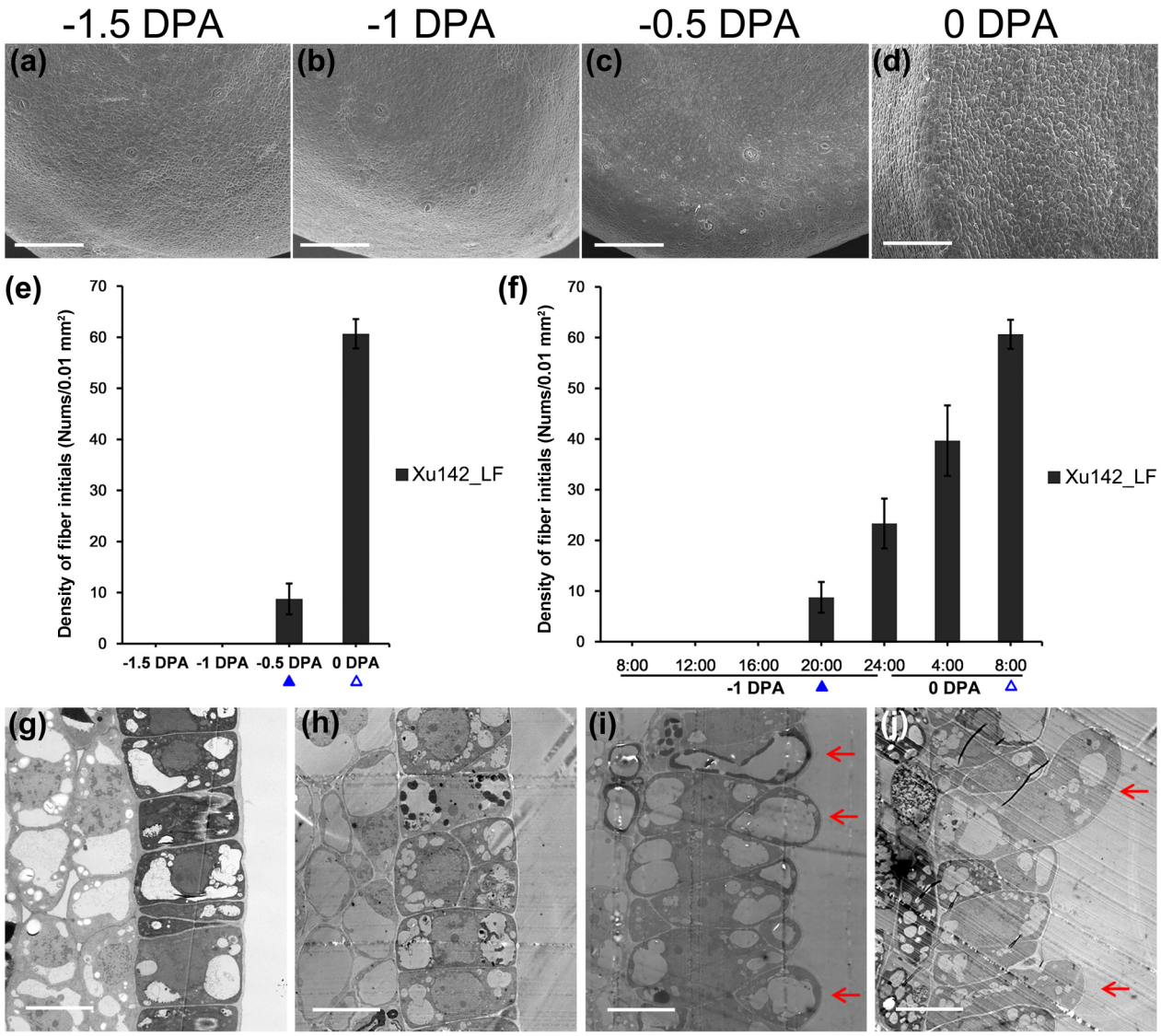
Fig. 7. Function verification of cotton *MYB25-like*, *PDF2* and *HOX3* genes.

(a) *MYB25-like* RNA *in situ* hybridization in LF_0d ovules. The sense probe was used as the negative control. OI, outer integument; II, inner integument. Red arrows indicate fiber cells, and red arrowheads indicate non-fiber epidermal cells. Yellow arrows indicate outer pigment layer. Scale bars, 50 μm. **(b)** Fiber initials morphology between WT (Jin668, the transgenic receptor material) and two types of *MYB25-like* CRISPR mutant (*MYB25-like _CR*, both *MYB25-like* homologs from At and Dt subgenomes knocked out) plants at 0 DPA. The upper panel shows the whole ovule, the lower panel is a magnified view of the ovule middle region. Scale bars, 500 μm (upper panel) and 50 μm (lower panel). **(c)** Statistics analysis of initial fiber densities (numbers per 0.01 mm^2) on 0 DPA ovules between WT and *MYB25-like_CR* plants. **(d)** RNA *in situ* hybridization of *PDF2* in LF_0d ovules with the sense probe as a negative control. OI, II, red arrows and red arrowheads have the same meaning as described in **(a)**. Scale bars, 50 μm. **(e)** Fiber initials morphology in two types of *PDF2_CR* (both *PDF2* homologs from At and Dt subgenomes knocked out) lines and WT plants at 0 DPA. The lower panel is a magnified view of the middle region of whole ovule (upper panel). Scale bars, 200 μm (upper panel) and 50 μm (lower panel). **(f)** Statistics analysis of initial fiber densities (numbers per 0.01 mm^2) on 0 DPA ovules between WT and *PDF2_CR* plants. All asterisks indicate significant differences when compared with the WT plants (T-test, ** $P < 0.01$). **(g)** RNA *in situ* hybridization of *HOX3* in LF_0d ovules with the sense probe as a negative control. OI, II, red arrows and red arrowheads have the same meaning as described in **(a)**. Scale bars, 50 μm. **(h)** Fiber initials morphology in WT and *HOX3_CR* (both *HOX3* homologs from At and Dt subgenomes knocked out) line at 0 DPA and 1 DPA, respectively. The lower panel is a magnified view of the middle region of whole ovule (upper panel). Scale bars, 200 μm (upper) and 50 μm (lower).

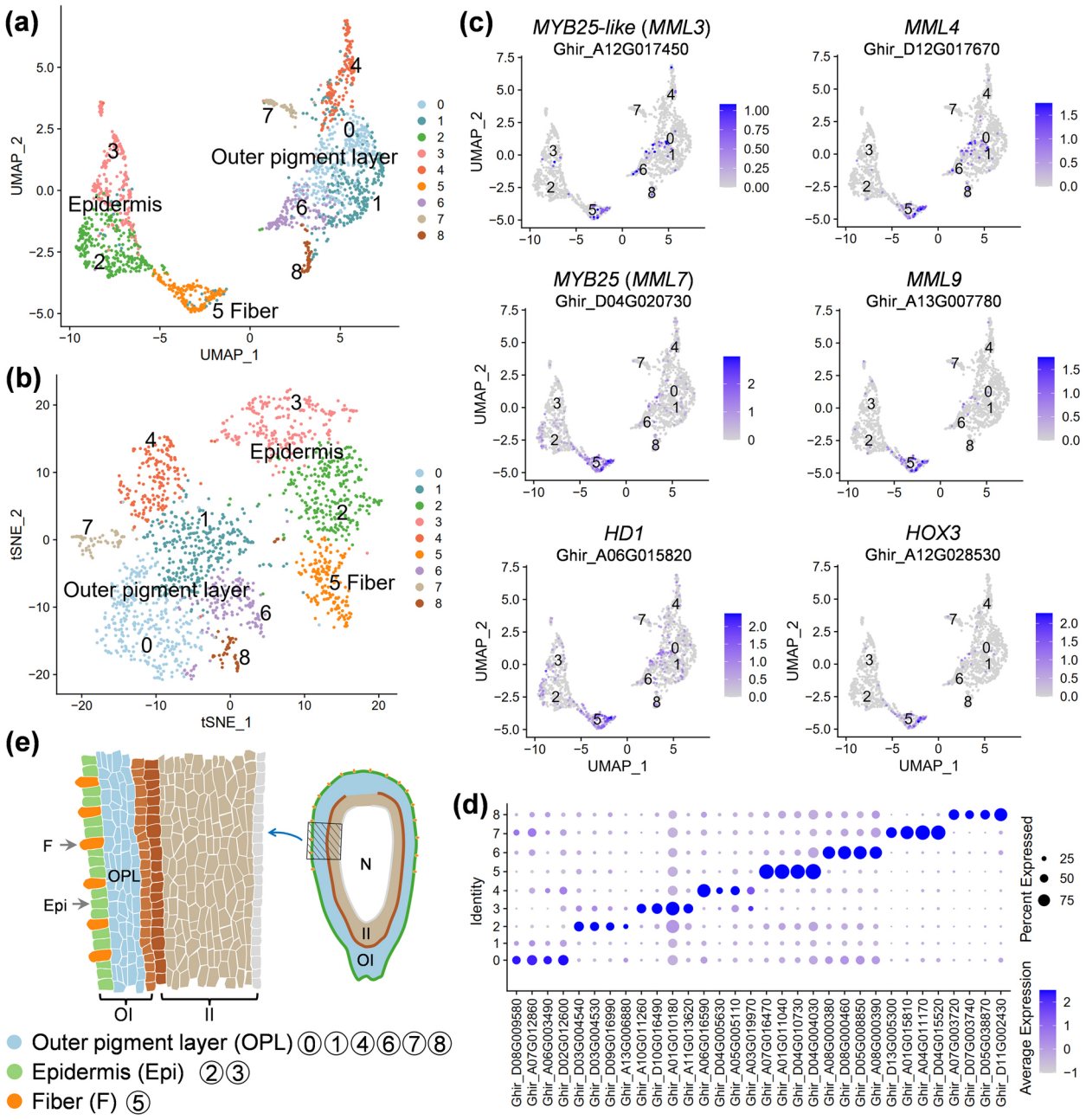
Fig. 8. The proposed model for fiber initiation focusing on a single cell.

(a) On cotton ovule epidermis, a precursor fiber cell successively experienced the process of differentiation, diffuse growth and tip-biased diffuse growth during its initiation (above), as represented by four stages (-1.5, -1, -0.5 and 0 DPA). Different hub transcription factors that regulating each process are shown (below). **(b)** On the whole ovule, lint fiber differentiation firstly occurred on chalazal end at -1 DPA (marked by light sky blue). The differentiation progress

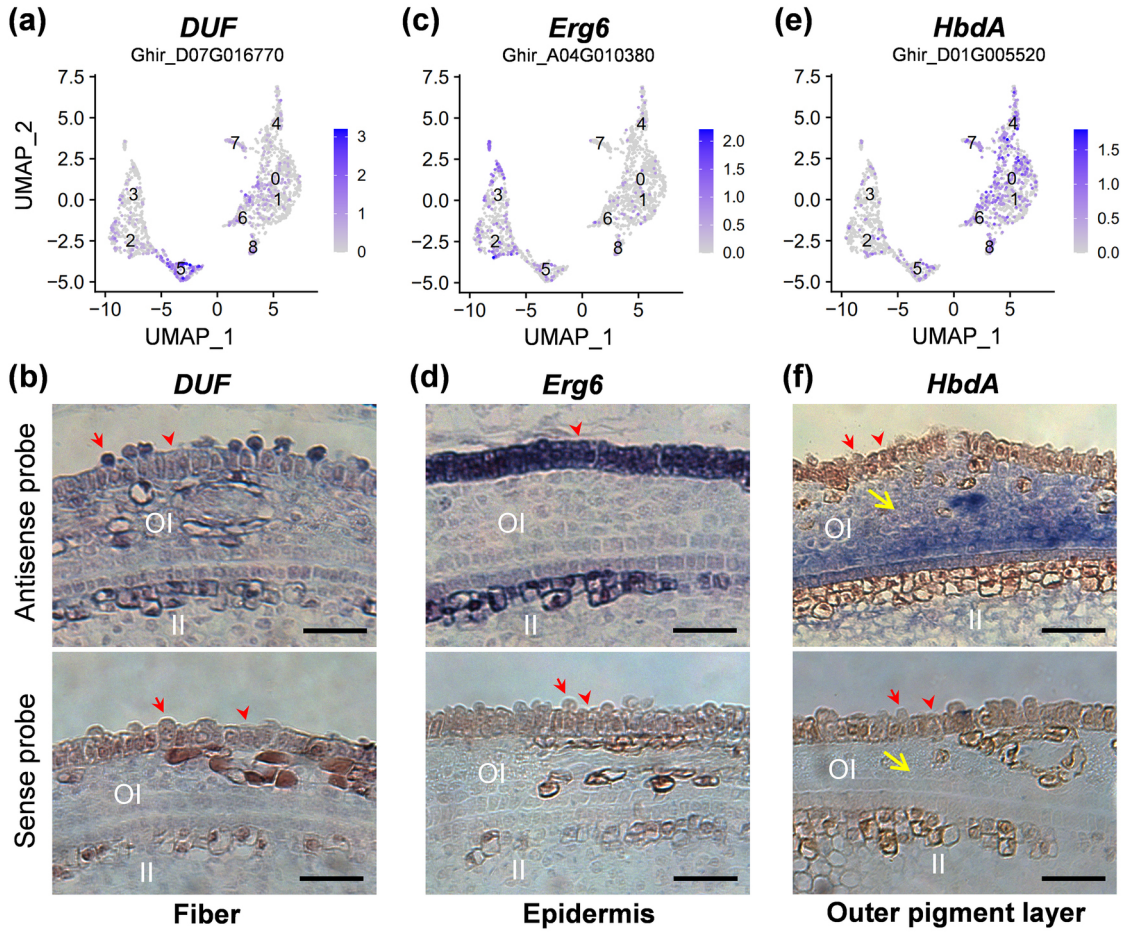
gradually towards micropylar end. At -0.5 DPA, fiber initials could be seen at chalazal end (marked by cyan) due to its diffuse growth. At the same time, fiber differentiation has occurred in the middle of the ovule (light sky blue). At 0 DPA, the previously protruded fiber initials transformed into tip-biased diffuse growth (dark blue), and the previously differentiated fibers began to protrude (cyan) due to diffuse growth. At the same time, fiber differentiation occurred at ovule micropylar end (light sky blue). To sum up, there were no fiber cell differentiated at -1.5 DPA; fiber differentiation occurred on chalazal end at -1 DPA; both fiber diffuse growth and differentiation occurred at -0.5 DPA; at 0 DPA, fiber tip-biased diffuse growth, diffuse growth and differentiation occurred simultaneously.



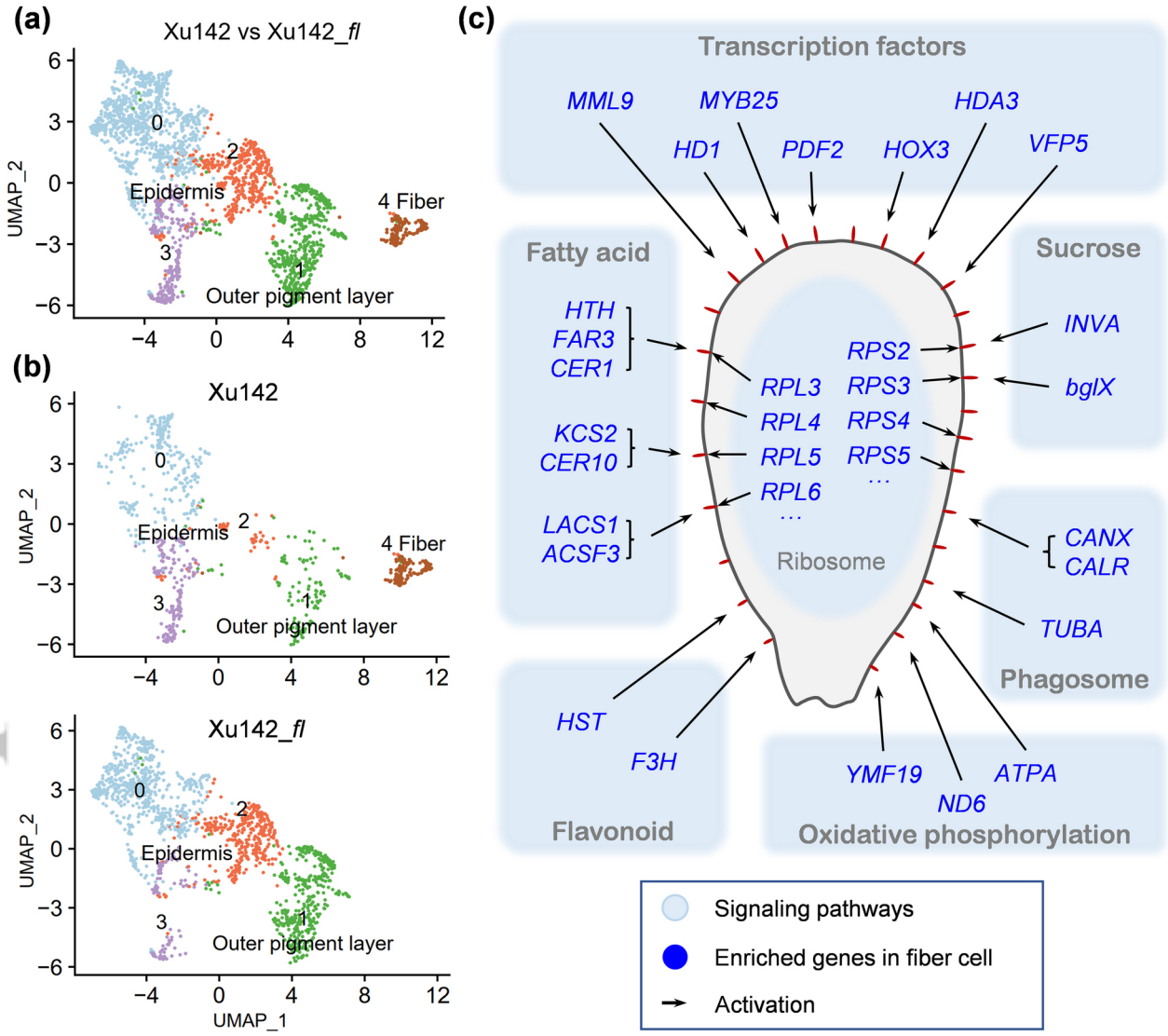
PBI_13918_Fig.1.jpg



PBI_13918_Fig.2.jpg

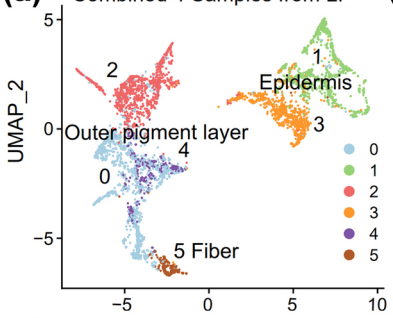


PBI_13918_Fig.3.jpg



PBI_13918_Fig.4.jpg

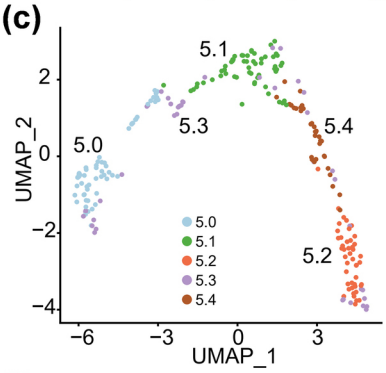
(a)



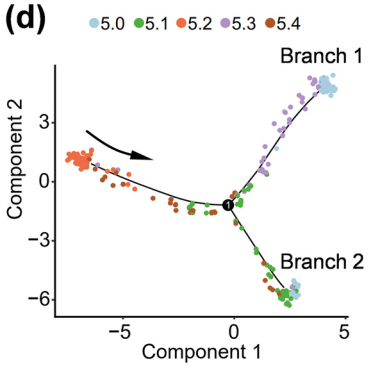
(b)

Cell type	Cluster	LF_-1.5d (1917 cells)	LF_-1d (696 cells)	LF_-0.5d (821 cells)	LF_0d (1703 cells)
Outer pigment layer	0, 2, 4	1106 (57.7%)	254 (36.5%)	542 (66.0%)	1043 (61.2%)
Epidermis	1, 3	811 (42.3%)	426 (61.2%)	249 (30.3%)	480 (28.2%)
Fiber	5	0	16 (2.3%)	30 (3.7%)	180 (10.6%)

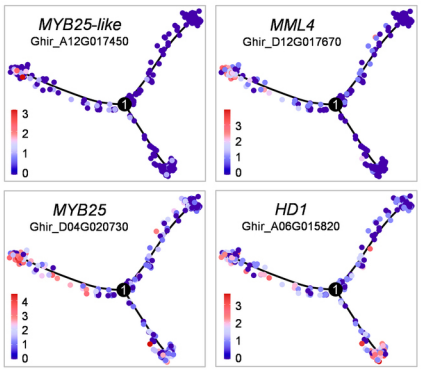
(c)



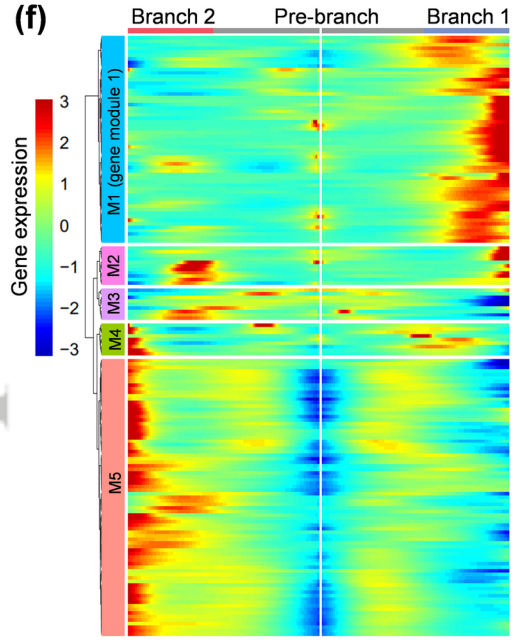
(d)



(e)

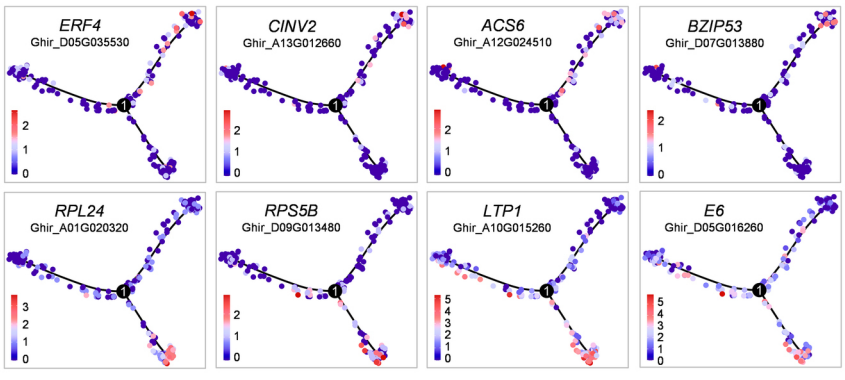


(f)

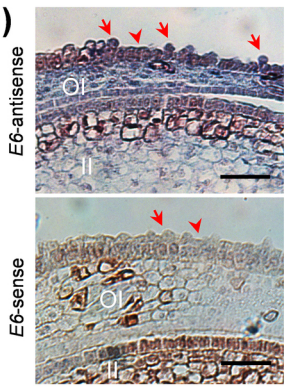


highlight genes		GO terms	
Ghir_A05G005610	*EBF1*	GO:0009723	response to ethylene
Ghir_D05G035530	*ERF4*	GO:0010104	regulation of ethylene-activated signaling pathway
Ghir_A03G019350	*GCH*	GO:0003935	GTP cyclohydrolase II activity
Ghir_A13G012660	*CINV2*	GO:0004575	sucrose alpha-glucosidase activity
Ghir_A12G024510	*ACS6*	GO:0033993	response to lipid
Ghir_A08G026490	*WAX2*	GO:0006723	cuticle hydrocarbon biosynthetic process
Ghir_D07G013880	*BZIP53*	GO:0003677	DNA binding
Ghir_A08G009020	*NAD6*	GO:0016655	oxidoreductase activity, acting on NAD(P)H
Ghir_D12G025340	*MGL*	GO:0001104	RNA polymerase II transcription cofactor activity
Ghir_A12G015580	*CRK3*	GO:0006468	protein phosphorylation
Ghir_D10G012310	*LTP3*	GO:0006869	lipid transport
Ghir_D10G012300	*LTP1*	GO:0006869	lipid transport
Ghir_D11G029310	*RPL18*	GO:0003735	structural constituent of ribosome
Ghir_A01G020320	*RPL24*	GO:0015934	large ribosomal subunit
Ghir_D09G013480	*RPS5B*	GO:0015935	small ribosomal subunit
Ghir_D05G027660	*RPL3A*	GO:0043043	peptide biosynthetic process
Ghir_A02G014340	*RPS6B*	GO:0006412	translation
Ghir_A10G015260	*LTP1*	GO:0006869	lipid transport
Ghir_A12G026040	*UBQ1*	GO:0016567	protein ubiquitination
Ghir_D12G015220	*AAC3*	GO:0005347	ATP transmembrane transporter activity
Ghir_A13G018840	*DUF538*	GO:0010090	trichome morphogenesis
Ghir_D05G016260	*E6*	GO:0044237	cellular metabolic process

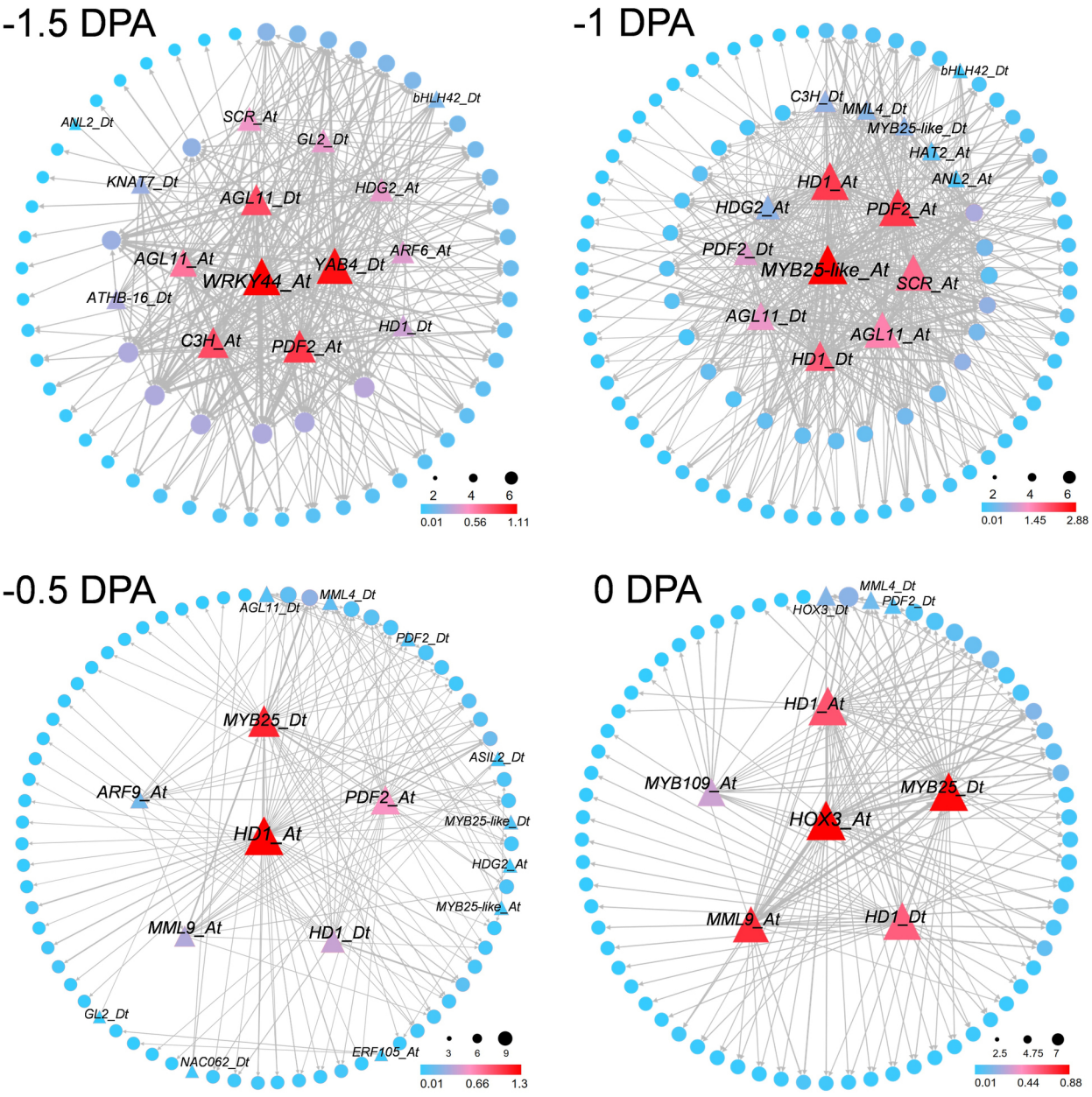
(g)



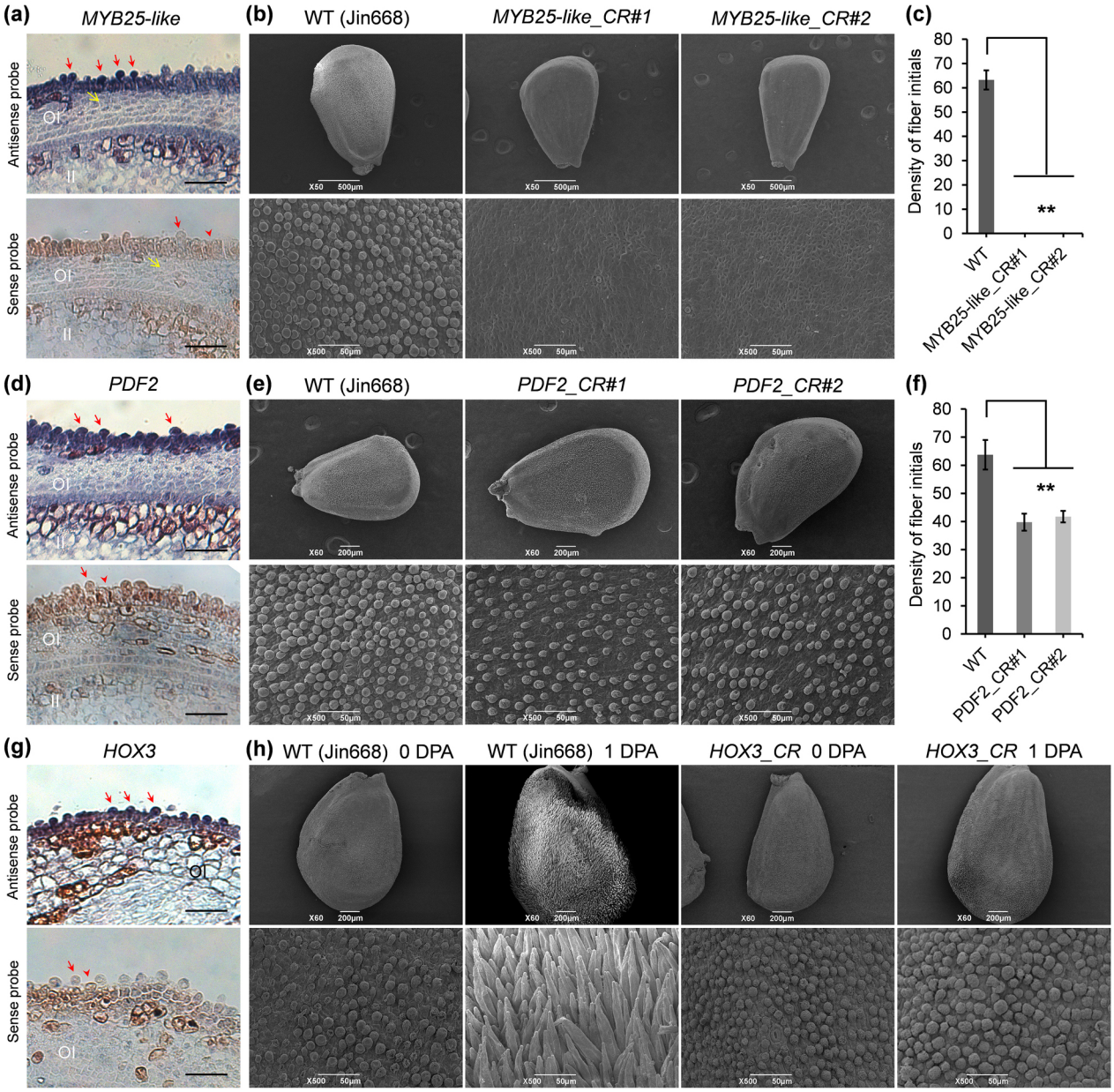
(h)



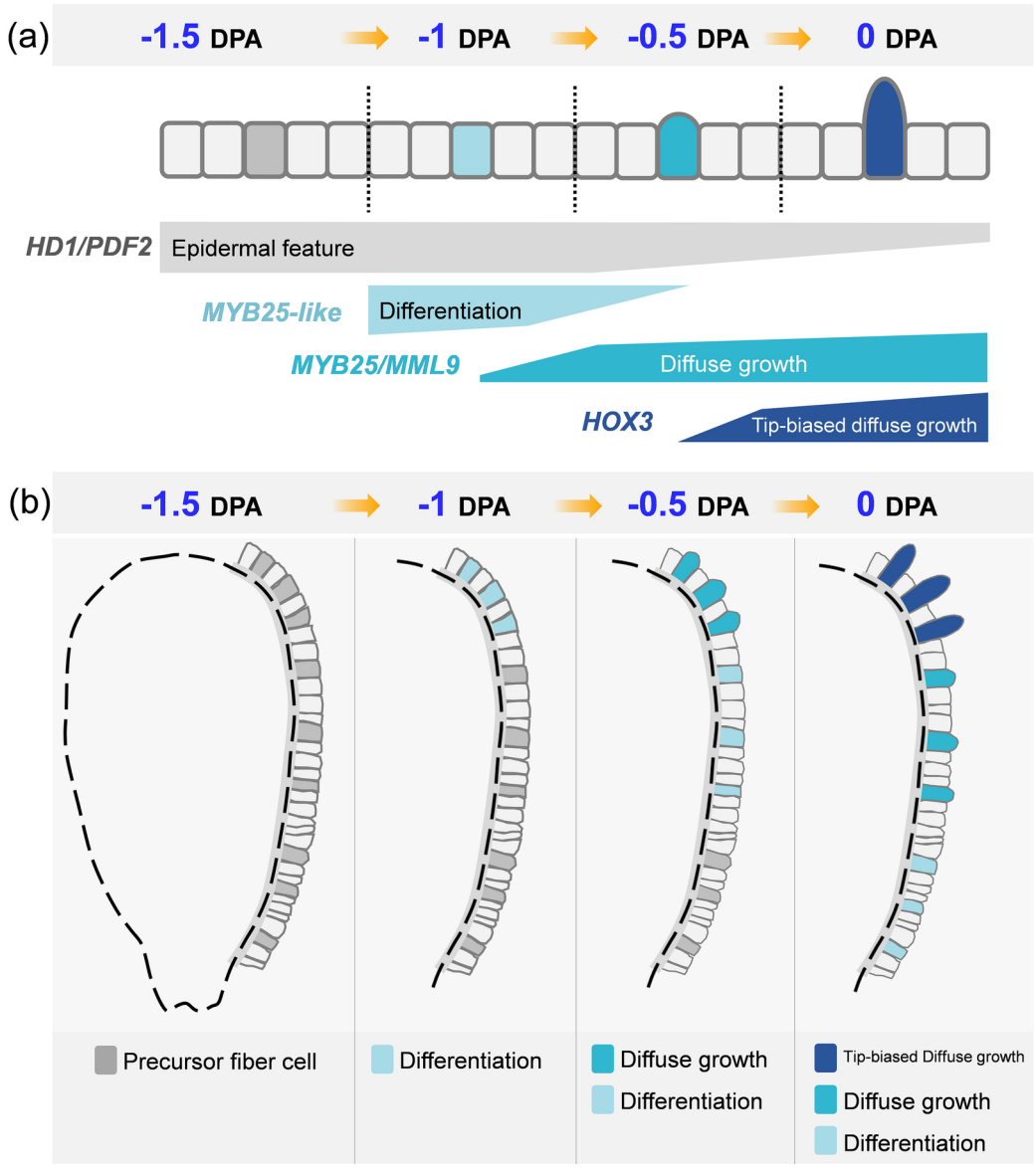
PBI_13918_Fig.5.jpg



PBI_13918_Fig.6.jpg



PBI_13918_Fig.7.jpg



PBI_13918_Fig.8.jpg