CellPress
Partner Journal

Molecular Plant
Research article

Single-cell transcriptomic analysis reveals the developmental trajectory and transcriptional regulatory networks of pigment glands in *Gossypium bickii*

Yue Sun[1,5], Yifei Han[1,5], Kuang Sheng[1], Ping Yang[2], Yuefen Cao[1], Huazu Li[1], Qian-Hao Zhu[3], Jinhong Chen[1,4], Shuijin Zhu[1,4,*] and Tianlun Zhao[1,4,*]

[1]College of Agriculture and Biotechnology, Zhejiang University, Hangzhou 310058, China

[2]Agricultural Experiment Station, Zhejiang University, Hangzhou 310058, China

[3]CSIRO Agriculture and Food, GPO Box 1700, Canberra, ACT 2601, Australia

[4]Institute of Hainan, Zhejiang University, Hangzhou 310058, China

[5]These authors contributed equally to this article.

***Correspondence: Shuijin Zhu (shjzhu@zju.edu.cn), Tianlun Zhao (tlzhao@zju.edu.cn)**

https://doi.org/10.1016/j.molp.2023.02.005

ABSTRACT

Comprehensive utilization of cottonseeds is limited by the presence of pigment glands and its inclusion gossypol. The ideal cotton has glandless seeds but a glanded plant, a trait found in only a few Australian wild cotton species, including *Gossypium bickii*. Introgression of this trait into cultivated species has proved to be difficult. Understanding the biological processes toward pigment gland morphogenesis and the associated underlying molecular mechanisms will facilitate breeding of cultivated cotton varieties with the trait of glandless seeds and glanded plant. In this study, single-cell RNA sequencing (scRNA-seq) was performed on 12 222 protoplasts isolated from cotyledons of germinating *G. bickii* seeds 48 h after imbibition. Clustered into 14 distinct clusters unsupervisedly, these cells could be grouped into eight cell populations with the assistance of known cell marker genes. The pigment gland cells were well separated from others and could be separated into pigment gland parenchyma cells, secretory cells, and apoptotic cells. By integrating the pigment gland cell developmental trajectory, transcription factor regulatory networks, and core transcription factor functional validation, we established a model for pigment gland formation. In this model, light and gibberellin were verified to promote the formation of pigment glands. In addition, three novel genes, *GbiERF114* (*ETHYLENE RESPONSE FACTOR 114*), *GbiZAT11* (*ZINC FINGER OF ARABIDOPSIS THALIANA 11*), and *GbiNTL9* (*NAC TRANSCRIPTION FACTOR-LIKE 9*), were found to affect pigment gland formation. Collectively, these findings provide new insights into pigment gland morphogenesis and lay the cornerstone for future cotton scRNA-seq investigations.

Key words: scRNA-seq, pigment gland morphogenesis, cell development, cotton

Sun Y., Han Y., Sheng K., Yang P., Cao Y., Li H., Zhu Q.-H., Chen J., Zhu S., and Zhao T. (2023). Single-cell transcriptomic analysis reveals the developmental trajectory and transcriptional regulatory networks of pigment glands in *Gossypium bickii*. Mol. Plant. **16**, 694–708.

INTRODUCTION

Cotton (*Gossypium* spp) is the leading natural fiber crop worldwide, and cottonseed is an excellent source of high-quality protein (23%) and oil (21%) (Sunilkumar et al., 2006; Rathore et al., 2020; Huang et al., 2021). However, the utilization of cottonseed is currently hampered by the presence of pigment glands containing gossypol deposits, which are toxic to humans and other monogastric animals (Sunilkumar et al., 2006; Rathore et al., 2020). Gossypol is a yellowish terpenoid that contributes to the natural resistance of cotton to insects and pathogens such as *Helicoverpa armigera* and *Verticillium*

Published by the Molecular Plant Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and CEMPS, CAS.

dahliae (Williams et al., 2011; Gao et al., 2013; Tian et al., 2018). Therefore, the ideal cotton should be glandless in the cottonseed, but glanded in other tissues of the plant (Sunilkumar et al., 2006; Ma et al., 2016). To this end, it is essential to understand the developmental processes involved in pigment gland formation and gossypol biosynthesis and their molecular mechanisms (Rathore et al., 2020; Zhang and Wedegaertner, 2021).

Pigment glands are distributed in most tissues and organs as dark opaque dots, which are one of the major characteristics of the *Gossypium* genus and its relatives (McMichael, 1960; Sunilkumar et al., 2006). The gland cavity is surrounded by flattened cells when mature, and there have been two diverse views interpreting the morphogenesis of pigment glands, i.e., the schizogenous process and the lysigenous process (Liu et al., 2010). Given the observation of the release of gossypol by autolysis, the lysigenous process is generally considered to be involved in programmed cell death (Dangl et al., 2000; Liu et al., 2010). Genetically, pigment gland formation in cotton is complex, with numerous regulatory factors. As of this writing, six independent genetic loci, gl_1, gl_2, gl_3, gl_4, gl_5, and gl_6, controlling pigment gland formation have been reported, with gl_2 and gl_3 being the major genetic contributors (McMichael, 1960). The dominant glandless gene, *GoPGF* (*Gossypium Pigment Gland Formation*), or Gl_2^e, which originated from a mutant of *Gossypium barbadense*, has been identified by fine genetic mapping, and low expression of this gene by silencing or knockout resulted in the glandless phenotype in cotton (Cheng et al., 2016; Ma et al., 2016; Li et al., 2021). Comparative transcriptome analysis on glandless near-isogenic cotton lines identified three *cotton gland formation* (*CGF*) genes, *CGF1*, *CGF2*, and *CGF3* (synonym of *GoPGF*) (Janga et al., 2019). Moreover, comparative transcriptome analysis of several glanded and glandless cultivars identified a *MYB* transcription factor named *Cotton Gland Pigmentation 1* (*CGP1*) as a regulator of gland pigmentation, which controlled the synthesis of gossypol by interacting with GoPGF to form a heterodimer in the nucleus (Gao et al., 2020). Another newly identified gland-associated gene, *GauGRAS1* (synonym of *Gossypium Stem Pigment Gland Forming Gene*, *GoSPGF*), was confirmed to be responsible for the formation of glands on some cotton plant tissues (Cai et al., 2020; Zang et al., 2021). Although several cotton genes have been identified as being associated with pigment gland formation, the developmental processes and regulatory networks involved in pigment gland formation remain largely unclear.

Single-cell RNA sequencing (scRNA-seq) revolutionized the details of whole transcriptomes from tissue to cell (Shaw et al., 2021). In plants, scRNA-seq has been used to dissect the transcriptome profile of roots from three aspects: the developmental regulatory network, the molecular mechanism of mutant strains, and the stress-response mechanism (Denyer et al., 2019; Jean-Baptiste et al., 2019; Ryu et al., 2019; Zhang et al., 2019; Wendrich et al., 2020; Wang et al., 2021b; Roszak et al., 2021; Shahan et al., 2022). The transcriptional landscapes of the shoot apical meristem of *Arabidopsis* (*Arabidopsis thaliana*) and maize (*Zea mays*) have been deciphered to understand the basic principles underlying cell division and differentiation at single-cell resolution (Satterlee et al., 2020; Tian et al., 2020; Zhang et al., 2021a). Furthermore, scRNA-seq technology was utilized to study specific lineage cell development, such as stomata in leaves and the vasculature in stems (Liu et al., 2020; Chen et al., 2021b; Bezrutczyk et al., 2021; Kim et al., 2021; Lopez-Anido et al., 2021). Single-cell multiomics technologies can provide further insights into the transcriptional regulation of cellular states and interactions. For example, integration of scRNA-seq/single-nucleus RNA-seq and the single-cell assay for transposase-accessible chromatin sequencing revealed the association of chromatin accessibility with gene expression at the single-cell level in individual cell types (Dorrity et al., 2021; Farmer et al., 2021; Marand et al., 2021). However, due to the diversity of plant species, the difficulty involved in preparation of plant protoplasts, and the fewer available cell-type marker genes in most plant species, plant single-cell transcriptome profiling is still in its early stage, particularly for non-model plants and non-meristematic tissues (Chen et al., 2021b; Liu et al., 2021). Nevertheless, scRNA-seq offers an unprecedented opportunity to study the morphogenesis of cotton pigment glands.

Gossypium bickii and several other Australian wild cotton species are of great breeding value in both cottonseed utilization and stress-resistance improvement for the characteristic of delayed pigment gland morphogenesis in the cotyledon (Fryxell, 1965; Zhu et al., 1999). The cotton species with this unique trait are the ideal material for investigating the morphogenesis of pigment glands. Here, we performed scRNA-seq using cotyledons from germinating *G. bickii* seeds at 48 h after imbibing. Integrating bulk and single-cell sequencing analysis and virus-induced gene silencing (VIGS) verification enabled us to understand the biological processes underlying the morphogenesis of cotton pigment glands and the molecular mechanisms regulating the trait of delayed pigment gland morphogenesis.

RESULTS

Construction of a single-cell transcriptome atlas of the cotton cotyledon

The dormant cottonseeds of *G. bickii* are glandless and gossypol-free (Sheng et al., 2022). At approximately 36 h after seed imbibition, the pigment glands become sparsely visible in the cotyledons. Subsequently, more pigment glands are clearly presented in the cotyledons and the young seedlings are normal glanded (Supplemental Figure 1A–1C) (Sheng et al., 2022). Similarly, gossypol is not detected until seed germination for 12 h and accumulates gradually in cotyledons during germination (Supplemental Figure 1D). This characteristic of gradual formation of pigment glands during seed germination is termed "delayed pigment gland morphogenesis" (Fryxell, 1965; Zhu et al., 1999; Sheng et al., 2022). Therefore, *G. bickii* is an ideal material for investigation of the molecular mechanisms responsible for pigment gland morphogenesis compared with normal pigment gland cotton strains.

To construct the trajectory of pigment gland development using scRNA-seq, cotyledons from germinating *G. bickii* seeds at 48 h after imbibition were sampled, because this time point is the active phase of pigment gland morphogenesis, with pigment glands at different developmental stages (Supplemental Figure 1A and Figure 1A). Protoplasts were prepared from the

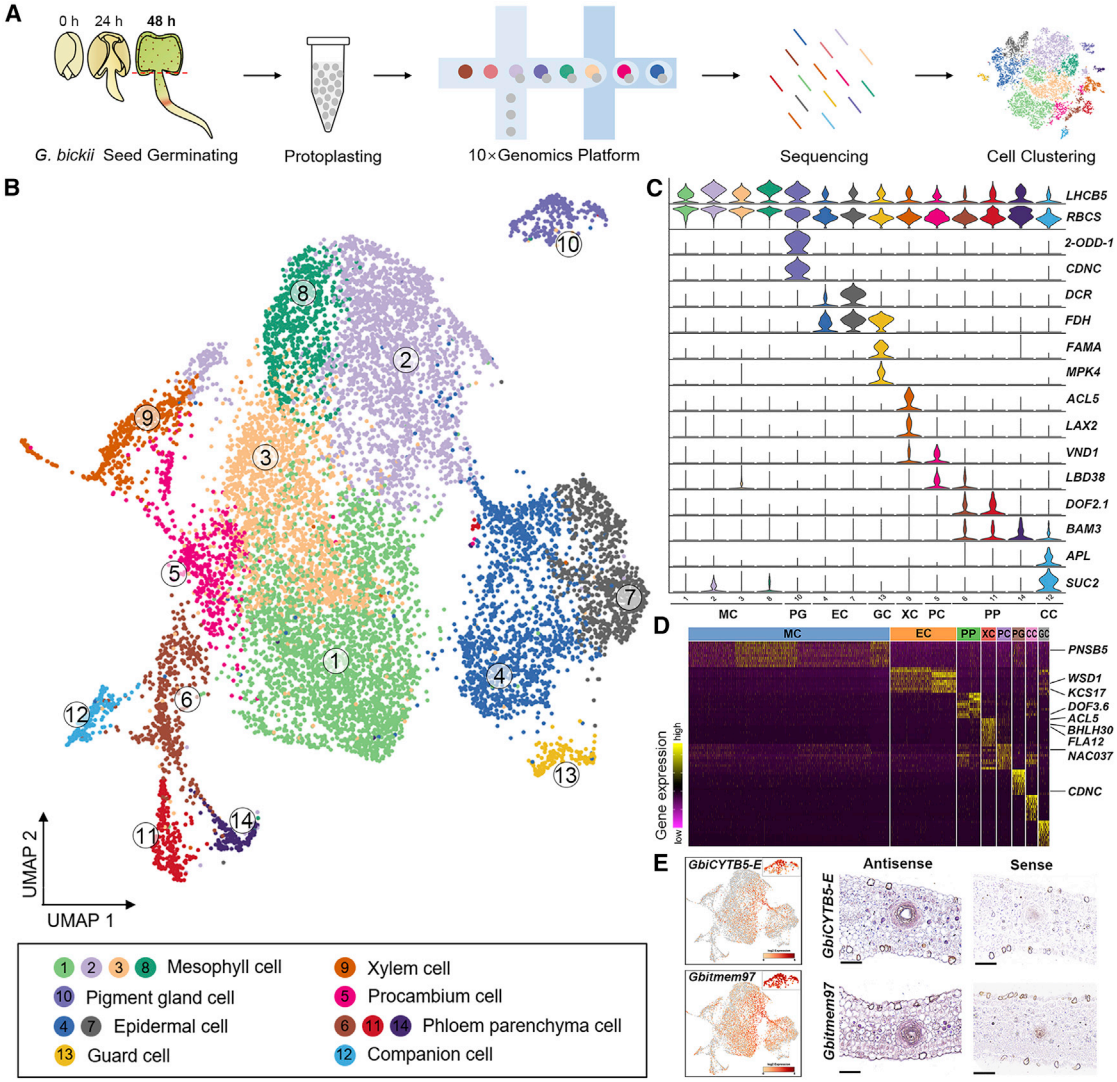

Figure 1. Construction of a single-cell transcriptome atlas of cotton cotyledon.

(A) The workflow for the cotton scRNA-seq using cotyledons from germinating *G. bickii* seeds at 48 h after imbibition.

(B) UMAP visualization of 14 distinct clusters derived from 12 222 cells. Each dot represents a single cell that is color coded to show cell clusters.

(C) Violin plots showing the expression pattern of representative cell-specific marker genes used to assign cell types.

(D) Heatmap showing the expression levels of the top 10 significantly upregulated genes in each cell population. The known marker genes are shown on the right. Color bar indicates the relative expression level. The full names of the selected genes are summarized in Supplemental Table 5.

(E) Expression of new marker genes *GbiCYTB5-E* and *Gbitmem97* and results of RNA *in situ* hybridization of them with the sense probe as a negative control. Scale bars, 50 μm.

48-h-old cotyledons by enzymatic hydrolysis, with a concentration of more than 2700 protoplast cells/μl. According to trypan blue staining, the percentage of live cells was over 90%. The protoplast cells were subjected to droplet-based scRNA-seq using the 10X Genomics scRNA-seq platform (Figure 1A). Transcriptome data of 13 495 cells were successfully obtained, with a mean number of 30 174 reads per cell and a median of 1915 expressed genes (Supplemental Table 1). After quality control, transcriptomes of 12 222 single cells were retained and subjected to downstream analyses (Supplemental Table 2). Principal-component analysis and graph-based unsupervised clustering, which did not rely on known markers, were performed to uncover 14 distinct cell clusters (Figure 1A and 1B). Clusters of cells were visualized in two dimensions using the uniform manifold approximation and projection (UMAP) algorithm and the *t*-distributed stochastic neighborhood embedding tool (Figure 1B, Supplemental Figure 2A, and Supplemental Table 3).

Reported marker genes in the *Arabidopsis* and plant scRNA-seq databases were used to identify the identity of each cluster (Figure 1C) (Chen et al., 2021a; Kim et al., 2021; Liu et al., 2020; Lopez-Anido et al., 2021; Xu et al., 2021b; Zhang et al., 2021a). As a result, 14 distinct cell clusters were grouped into eight cell populations, including mesophyll cell, pigment gland cell, epidermal cell (EC), guard cell, xylem cell, procambium cell, phloem parenchyma cell, and companion cell (Supplemental Figure 2B).

The cell populations were dominated by mesophyll cells (~61% of all cells), which comprised clusters 1, 2, 3, and 8. The known marker genes of mesophyll cells, such as those involved in photosynthesis, including *CHLOROPHYLL A/B BINDING PROTEIN* (*CAB*), *RUBISCO SMALL SUBUNIT* (*RBCS*), and *LIGHT HARVESTING COMPLEX OF PHOTOSYSTEM II* (*LHCB*), were predominantly expressed (Liu et al., 2020; Lopez-Anido et al., 2021; Zhang et al., 2021a). *YABBY* (*YAB*) transcripts, markers for spongy parenchyma, were enriched in cluster 2, indicating that the cluster may represent spongy mesophyll (Kim et al., 2021; Lopez-Anido et al., 2021). The epidermal-specific genes, including *3-KETOACYL-COA SYNTHASE* (*KCS*) and *DEFECTIVE IN CUTICULAR RIDGES* (*DCR*), were highly expressed in the EC population (~19%; clusters 4 and 7) (Zhang et al., 2021a; Kim et al., 2021). Guard cells (cluster 13) were identified based on enrichment of the specifically expressed guard cell marker genes *FAMA* (*FMA*) and *MITOGEN-ACTIVATED PROTEIN KINASE* (*MPK*) (Zhang et al., 2021a; Kim et al., 2021; Liu et al., 2020; Lopez-Anido et al., 2021). Using genes reported to be related to gland formation and gossypol biosynthesis in upland cotton (*Gossypium hirsutum*), such as *CDNC* (*(+)-δ-cadinene synthase subfamily C*), *CYP706B1* (*cytochrome P450 monooxygenase family 706 subfamily B polypeptide 1*), and *DH1* (*short-chain alcohol dehydrogenase 1*), as markers, cluster 10 was identified as the pigment gland cells, which were obviously separated from other clusters on the UMAP plot, consistent with their specific physiological functionalities and unique expression profiles (Supplemental Table 4 and Supplemental Figure 3) (Ma et al., 2016; Huang et al., 2021). Moreover, RNA *in situ* hybridization of *GbiGoPGF* (homologous gene of *GoPGF*) further supported the assignment of pigment gland cell identity (Supplemental Figure 4).

Vasculature cell populations were specifically located on the left side of the UMAP. Transcripts of the xylem gene *ACAULIS 5* (*ACL5*) and *LIKE AUXIN RESISTANT 2* (*LAX2*) were markedly overrepresented in cluster 9 cells, while the phloem gene *DOF ZINC FINGER PROTEIN 2* (*DOF2*) and *BETA-AMYLASE 3* (*BAM3*) were found in cells of clusters 6, 11, and 14 (Zhang et al., 2021a; Kim et al., 2021; Lopez-Anido et al., 2021). Procambium cells were identified by the dominant expression of *VASCULAR RELATED NAC-DOMAIN PROTEIN 1* (*VND1*) and *LOB DOMAIN-CONTAINING PROTEIN 38* (*LBD38*) in cluster 5 cells (Zhang et al., 2021a; Kim et al., 2021; Lopez-Anido et al., 2021), whereas transcripts of the companion cell marker genes, such as *ALTERED PHLOEM DEVELOPMENT* (*APL*) and *SUCROSE-PROTON SYMPORTER* (*SUC2*), were exclusively expressed in cluster 12 cells (Zhang et al., 2021a; Kim et al., 2021; Lopez-Anido et al., 2021). According to the predicted direction of cell differentiation using scVelo, procambium cells clearly bifurcated into phloem and xylem lineages as expected, while the companion cells were differentiated at an early stage of phloem development (Supplemental Figure 2C) (Chen et al., 2021b).

To characterize and verify new cell marker genes, the expression profiles of the top 10 significantly upregulated genes that were enriched in each cell population were visualized and compared (Figure 1D and Supplemental Table 5). These genes with similar expression profiles in each population were proposed to be the marker genes of the corresponding cell population, and these included some known marker genes, such as *PNSB5* (*PHOTOSYNTHETIC NDH SUBCOMPLEX B 5*, *Gbi10G1461*) in mesophyll cells and *WSD1* (*WAX ESTER SYNTHASE/DIACYLGLYCEROL ACYLTRANSFERASE 1*, *Gbi07G1673*) and *KCS17* (*3-KETOACYL-COA SYNTHASE 17*, *Gbi09G3067*) in ECs (Figure 1D and Supplemental Table 4) (Zhang et al., 2021a; Kim et al., 2021; Lopez-Anido et al., 2021). RNA *in situ* hybridization of the new cell marker genes *GbiCYTB5-E* (*Cytochrome b5 isoform E*, *Gbi02G1455*) and *Gbitmem97* (*Transmembrane protein 97, Gbi12G2048*) further verified the reliability of the pigment gland cell population definition and the availability of new cell marker genes (Figure 1E). In addition, the expression patterns of the potential top 10 marker genes of the pigment gland cell population from bulk RNA-seq datasets were further supported by the power of scRNA-seq to extract specific pigment gland cells (Supplemental Table 6 and Supplemental Figure 5). These new marker genes will be useful tools for distinguishing cell types in scRNA-seq studies of plants, particularly for cotton.

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were performed on all cell populations to evaluate the method used in identifying specifically expressed cell-type genes and the ability to correctly attribute biological processes to a cell type. As expected, a majority of the enriched terms or pathways were associated with the cell type of interest (Supplemental Figure 6). For instance, the mesophyll cells, which play a pivotal role in light capture, were found to be mainly associated with photosynthesis (Loudya et al., 2021). ECs were enriched in terms related to lipid biosynthesis and mitochondrial components (Lopez-Anido et al., 2021). The pigment gland cells were enriched in pathways related to terpenoid biosynthesis (Ma et al., 2016).

The developmental trajectory of pigment gland cells

The pigment gland cell cluster was well separated from other cell populations (Figure 1B). Given the complexity of cotton pigment glands, pigment gland cells were further unsupervisedly divided into three sub-cell clusters, with 144, 88, and 56 cells, respectively (Figure 2A) (Liu et al., 2010). To determine the identity of each sub-cell cluster, the expression pattern of the aforementioned marker genes in each sub-cluster was analyzed (Supplemental Table 7). In contrast to the undistinguishable and universal expression of known pigment gland formation genes in all sub-clusters, many gossypol biosynthesis genes were highly enriched in sub-cluster 2, indicating the secretory cell identity of the sub-cluster 2 cells (Figure 2B) (Liu et al., 2010). Based on the expression patterns of the parenchyma-marker genes in *Arabidopsis*, *ACTIN DEPOLYMERIZING FACTOR 6* (*ADF6*) and *AZELAIC ACID INDUCED 1* (*AZI1*), sub-cluster 1 was determined to be pigment gland parenchyma cells (Figure 2B) (Kim et al., 2021). While the genes involved in transportation of proanthocyanidin precursors into vacuoles in *Arabidopsis*, such as *DETOXIFICATION 41*(*DTX41*), and the cuticle-marker gene *ATP-BINDING CASSETTE G11* (*ABCG11*) were enriched in sub-cluster 3 (Figure 2B), the cell type of sub-cluster 3 could not be unambiguously determined (Liu et al., 2020; Kim et al., 2021; Lopez-Anido et al., 2021). We further did GO analysis on the sub-cell clusters to verify cell type by identifying the key biological functions of differentially expressed genes (DEGs) (Supplemental Figure 7). Consistent with the reported characteristics of pigment gland

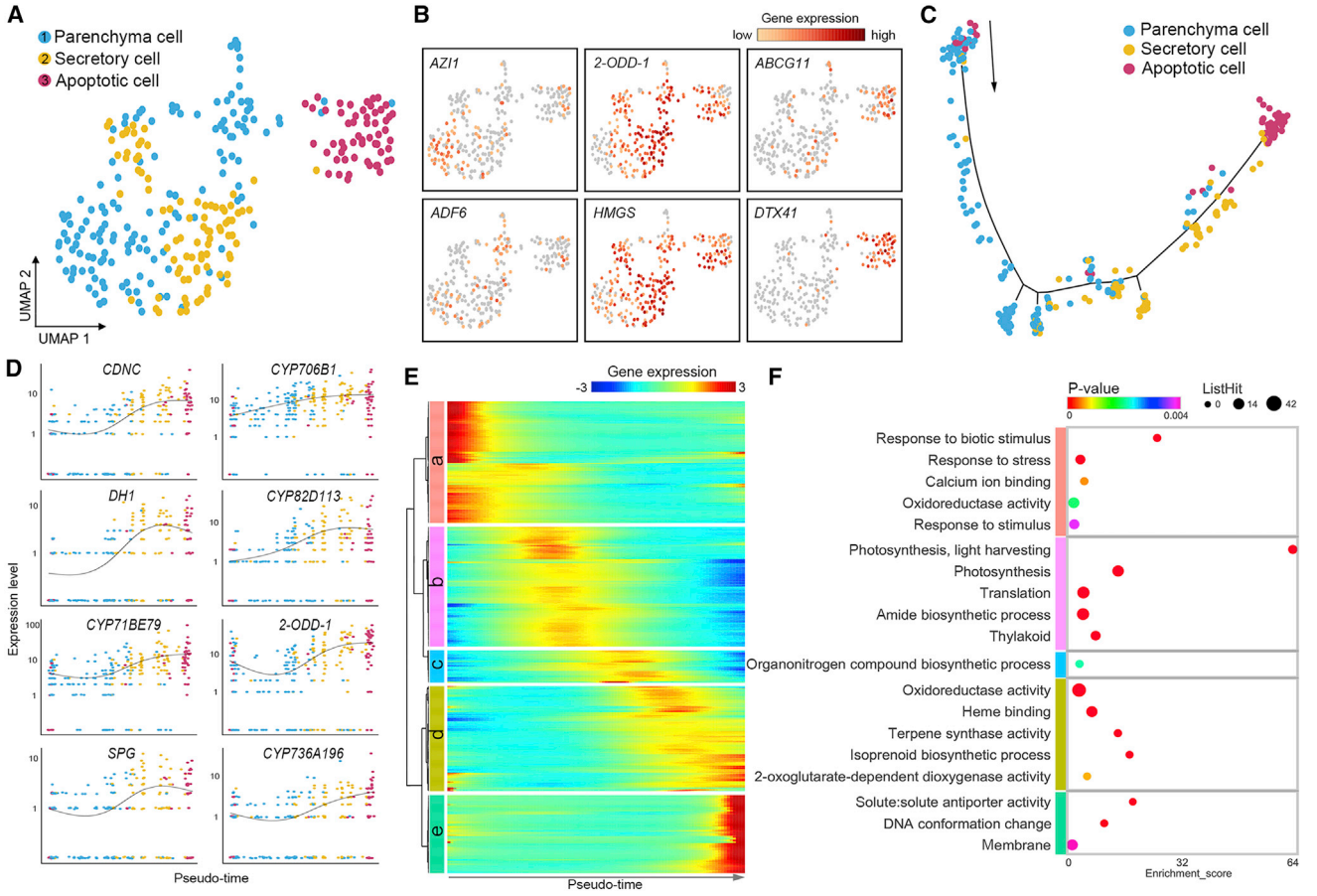

Figure 2. The developmental trajectory of pigment gland cells.

(A) UMAP plot showing three sub-cell clusters of the pigment gland cell population. Each dot represents a single cell that is color coded based on sub-clusters.

(B) Expression patterns of representative genes of sub-cell clusters on the UMAP. Clustering is the same as that shown in **(A)**. The color bar indicates relative expression levels.

(C) The successive differentiation trajectory of pigment gland cells. Each dot represents a single cell. The black arrow indicates the start and direction of the trajectory.

(D) Expression profile of the genes involved in gossypol biosynthesis along pseudo-time. Black line indicates expression tendency. Dots represent individual cells color coded based on sub-clusters.

(E) Heatmap showing the significantly differentially expressed genes of the five gene modules along pseudo-time. The arrow on the bottom of the heatmap indicates the trajectory order.

(F) Dot plot showing the GO enrichment analysis result of the DEGs of the five identified gene modules.

cells (Liu et al., 2010), the terms related to ribosome and translation were enriched in gland parenchyma cells, and the term related to terpenoid biosynthesis was enriched in secretory cells. Due to the enrichment of oxidoreductase activity but the low expression of mitochondrial and ribosome genes, it was conjectured that sub-cluster 3 may be pre-apoptotic cells that are programmed for autolysis (Supplemental Figure 7) (Reape et al., 2008; Liu et al., 2010; Daneva et al., 2016). In addition, in bulk RNA-seq data of *G. bickii* cotyledons (0, 12, 24, 36, and 48 h after seed imbibition), the expression patterns of the top five marker genes of sub-cell clusters were increased along with germination and most marker genes of secretory cells and apoptotic cells were not expressed at 0 and 12 h after seed imbibition (Supplemental Table 8), which were consistent with the development of pigment glands in the cotyledons and further confirmed sub-cell cluster assignments (Supplemental Figure 1B).

To further clarify the developmental pattern of pigment glands, the continuous differentiation trajectories of three types of pigment gland cell were analyzed. The developmental trajectory started with cells from the sub-cluster of parenchyma cells; after a bifurcation point, which marks developmental state transition, a portion of the parenchyma cells gradually differentiated into secretory cells (Figure 2C). Following further development of secretory cells, the trajectory ended with the pre-apoptotic cells. Consistent with the previously reported results, the key enzymes for gossypol biosynthesis, including *CDNC*, *CYP706B1*, *DH1*, *CYP82D113*, *CYP71BE79*, *2-ODD-1*, and *SPG*, were expressed in secretory cells along the cell differentiation trajectory (Huang et al., 2020) (Figure 2D). Some gossypol biosynthesis genes, including *CYP71BE79* and *CYP736A196*, were also highly expressed in pre-apoptotic cells at the axis end, suggesting the capability of apoptotic cells for gossypol biosynthesis in addition to release of gossypol by cell apoptosis. This is consistent with

the published cytomorphology reports that abundant black gossypol was first observed to be enveloped by autophagosomes and released after cell disintegration (Liu et al., 2010).

To reveal the genes involved in spatial-temporal regulation of pigment gland development, a total of 649 genes were first identified as most significantly related to pseudo-time order and sorted into five modules (Figure 2E). The module (module a) at the beginning of pseudo-time was enriched with genes involved in the response to biotic and abiotic stresses, including insects, heat, and hydrogen peroxide, suggesting that activation of pigment gland cell development might depend on an external stimulus and that the pigment gland formation might be one of the outcomes related to regulation of stress response (Huang et al., 2020) (Supplemental Table 9). One of the later modules (module d) showed signatures of terpenoid biosynthesis due to high enrichment of terpene synthase and dioxygenase activity, consistent with gossypol synthesis in mature secretory cells (Liu et al., 2010; Huang et al., 2020). The last module (module e) involved DNA conformation change, solute antiporter activity, and membrane, all of which are typical characteristics of apoptotic cells (Figure 2F) (Liu et al., 2010).

Construction of a transcription factor regulatory network of pigment gland morphogenesis

Transcription factors (TFs) are key regulators of all biological processes in plants, including formation of pigment glands (Huang et al., 2020). Of the TFs identified in pigment gland cells, most were found to be related to the differentiation of pigment gland parenchyma, including *bHLH* (*Basic Helix-Loop-Helix*), *AP2/ERF* (*APETALA2/Ethylene Responsive Factor*), *C2H2* (*C2H2 zinc-finger*), *GRAS* (*GAI-RGA- and -SCR*), *NAC* (*NAM, ATAF1, 2, and CUC2*), and *WRKY* (Janga et al., 2019). The transition from parenchyma to secretory cells was regulated by *AP2/ERF*s and *WRKY*s, while that from secretory cells to apoptotic cells was regulated by *MYB*s (Cai et al., 2020) (Supplemental Figure 8 and Figure 3D).

GoPGF, a *bHLH* TF, has been proven to play a significant role in the initiation of pigment gland parenchyma, but the transcriptional regulatory network regulating the morphogenesis of pigment gland structure still remains poorly understood (Cheng et al., 2016; Ma et al., 2016; Janga et al., 2019; Li et al., 2021). DNA-affinity purification sequencing (DAP-seq) analysis for *GoPGF* was performed to look for its downstream genes (Figure 3A and Supplemental Table 10). Given that directly modulated targets of a TF should be co-expressed in the same cell types, the genes co-expressed with *GoPGF* were excavated to narrow down the list of potential *GoPGF* targets involved in the regulation of pigment gland development. And by integrating the results of DAP-seq and scRNA-seq, a TF network with *GoPGF* as the hub gene was established (Figure 3B). *AP2/ERF*s, *bHLH*s, *C2H2*s, *MYB*s, *NAC*s, and *WRKY*s were shown to be the major interactors of *GoPGF* involved in the regulation of the formation of pigment gland parenchymal cells or release of gossypol from apoptotic cells (Figure 3B and 3C) (Gao et al., 2020).

Accordingly, a working model for the development of cotton pigment glands was proposed (Figure 3D). The pigment gland parenchyma matures to become the outer layer of the pigment gland. A portion of parenchyma maintains the ability to transform into secretory cells, which further develop into apoptotic cells. The mature cotton pigment glands consist of multiple layers of mature parenchyma, a monolayer of secretory cells in the outer cavity, and the residue of apoptotic cells in the inner cavity (Figure 3D) (Liu et al., 2010). The entire process of cotton pigment gland development involves a large number of regulating TFs such as *bHLH*, *AP2/ERF*, and *MYB* at different stages (Figure 3D).

To test the model for the formation and developmental progression of cotton pigment glands that emerged from scRNA-seq datasets, the expression pattern of the above representative TFs was analyzed from bulk RNA-seq data of *G. bickii* cotyledons (0, 12, 24, 36, and 48 h after seed imbibition), and most of them were increased along with the germination (Figure 4A). VIGS assays were applied to verify the functions of the eight TFs. Among them, silencing *GbiGoPGF* (*bHLH*), identified as the homolog of *GoPGF* in *G. hirsutum* (Ma et al., 2016), led to a completely glandless phenotype (Figure 4B). Silencing all eight TFs, including the novel genes *GbiERF114* (*AP2/ERF*), *GbiZAT11* (*C2H2*), and *GbiNTL9* (*NAC*) and the known homologous genes *GbiCGF1* (*bHLH*), *GbiGoPGF* (*bHLH*), *GbiGoSPGF* (*GRAS*), *GbiCGP1* (*MYB*), and *GbiCGF2* (*NAC*) (Janga et al., 2019; Cai et al., 2020; Gao et al., 2020; Zang et al., 2021), was shown to have a significant effect on pigment gland development (Figure 4B and Supplemental Figure 9). The pigment gland density and gossypol content in both leaves and stems of these silenced seedlings were all significantly lower than that in the negative control TRV:00 seedlings (Figure 4C and 4D). The result suggested that these types of TFs may play a major role in cotton pigment gland formation.

Involvement of light and gibberellic acid in the formation of pigment glands

Guided by the bulk samples of cotyledons (0, 12, 24, 36, and 48 h after seed imbibition) in *G. bickii* with pigment gland density information, 45 Scissor^{+} cells and 57 Scissor^{-} cells were selected in pigment gland cells, which were associated with the pigment gland density phenotypes (Supplemental Figures 1C and 10A). To uncover the factors affecting the process of pigment gland formation, functional enrichment analysis was performed among these cells. As a result, the photosynthesis-related pathways, such as response to light stimulus and light harvesting, were activated in Scissor^{+} cells (Supplemental Figure 10B). In addition, by predicting the function of *cis*-regulatory elements in the promoter sequence (2000 bp) of the above pigment gland development TFs with PlantCARE (Lescot et al., 2002), multiple elements functioning in responses to light and gibberellin were also found (Figure 5A).

To investigate the role of light and gibberellin in pigment gland formation, the development of pigment glands on cotyledons of germinating *G. bickii* seeds treated with dark, light, gibberellic acid (GA_3), and paclobutrazol (a synthetic inhibitor of gibberellins) was investigated. Under the dark, dark + paclobutrazol, and light + paclobutrazol conditions, no pigment gland was

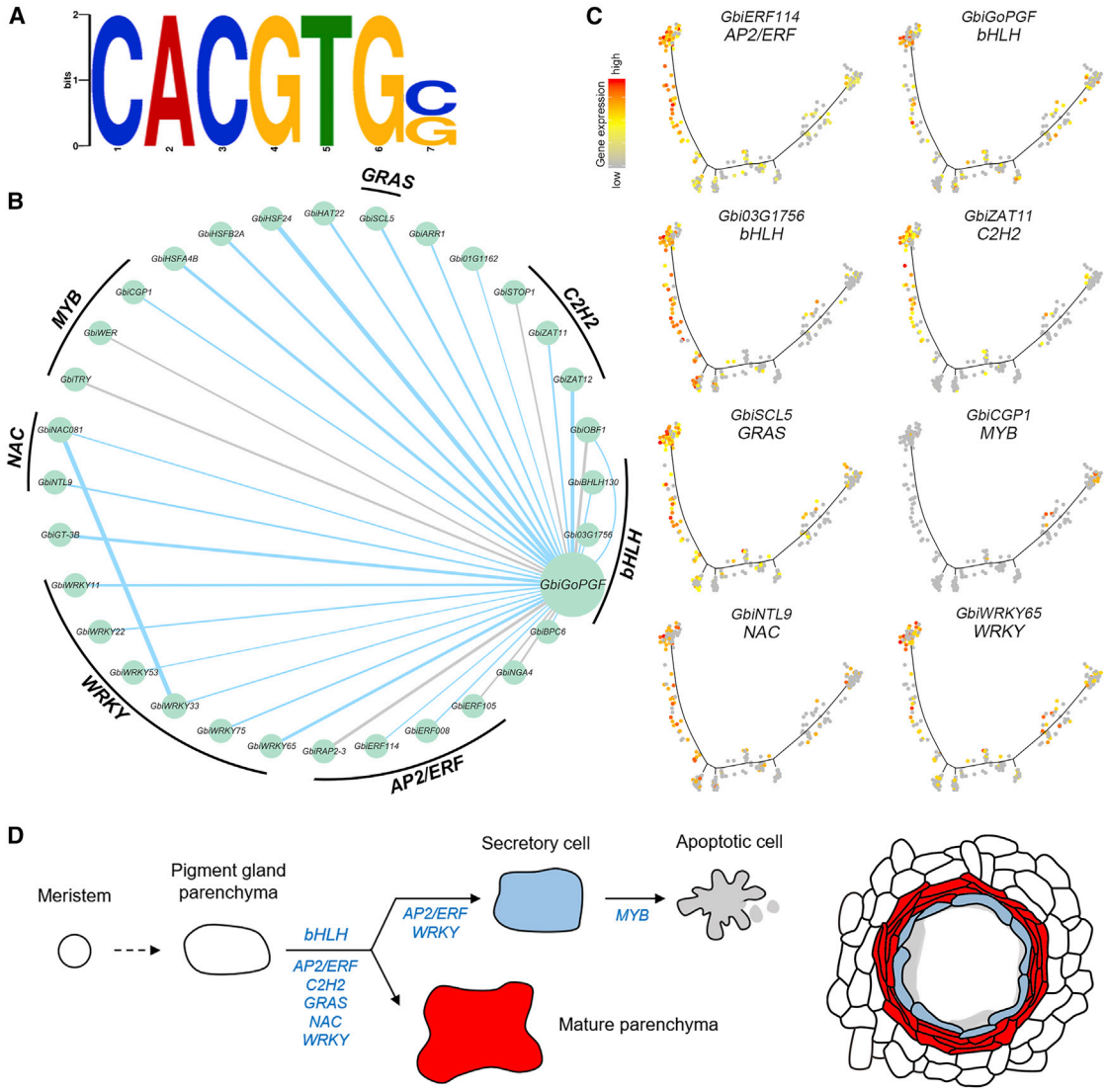

Figure 3. Transcription factor regulatory network of pigment gland cell development.

(A) Cotton consensus DNA motif (G-box: CACGTG) bound by GoPGF, as determined by DAP-seq.

(B) Gene-regulatory network inferred by transcription factors that are dynamically expressed across pigment gland differentiation based on DAP-seq and scRNA-seq data. Blue and gray, positive and negative regulation. Node size is equivalent to the number of predicted connections.

(C) The expression patterns of representative interactors in the differentiation process along pseudo-time. Each dot represents a single cell. The colors represent expression levels of the genes in individual cells.

(D) A working model for the formation and developmental progression of cotton pigment glands. The right shows a schematic of the anatomy and corresponding cell types of a mature cotton pigment gland, with mature parenchyma in red, secretory cells in blue, and residue of apoptotic cells in gray.

observed on cotyledons until 60 h after seed imbibition. The pigment gland density of the dark + paclobutrazol condition was significantly lower than that of the dark and light + paclobutrazol conditions at 60 and 84 h after seed imbibition (Figure 5C). In contrast, pigment glands were seen on cotyledons at 36 h after seed imbibition under the dark + GA_3, light, and light + GA_3 conditions (Figure 5B). Among them, the pigment gland density of the light + GA_3 condition was greater than that of the light condition and significantly greater than that of the dark + GA_3 condition at 36, 48, 60, and 84 h after seed imbibition (Figure 5C). These findings demonstrated that light and GA_3 are able to promote the formation of pigment glands.

The positive results of GA_3 were consistent with the recent findings that GA signal was transferred to *GoSPGF* to enhance the expression of *GoPGF* (Zang et al., 2021). To understand the transcriptional patterns under changes in light, the expression levels of pigment gland development TFs in cotyledons at 24 h after *G. bickii* seed imbibition under the treatments of dark and light were compared. Among them, the expression levels of *GbiCGF1* (*bHLH*), *GbiGoPGF* (*bHLH*), *GbiZAT11* (*C2H2*), *GbiGoSPGF* (*GRAS*), and *GbiCGP1* (*MYB*) were downregulated in the absence of light, while the expression levels of *GbiERF114* (*AP2/ERF*), *GbiNTL9* (*NAC*), and *GbiCGF2* (*NAC*) were upregulated (Figure 5D and Supplemental Table 11). In addition, the expression levels of most gossypol biosynthesis

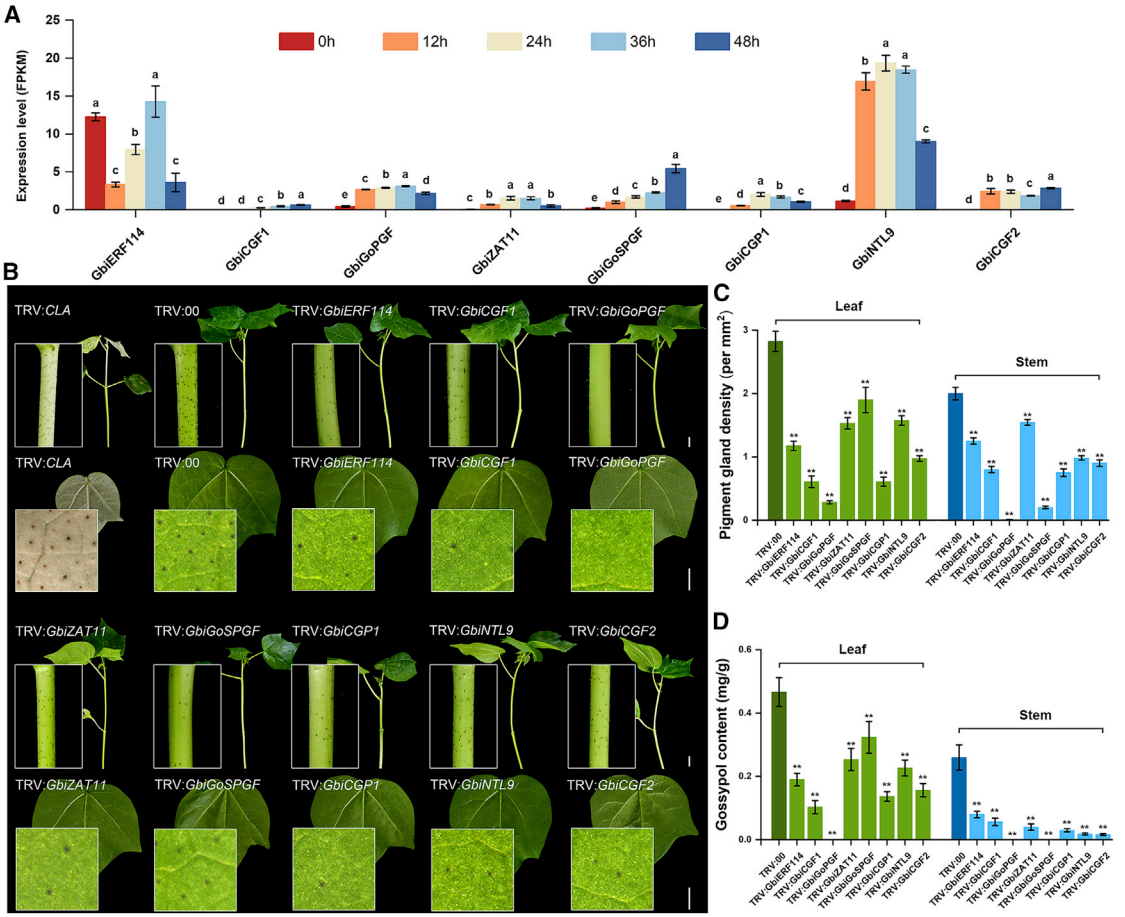

Figure 4. Functional verification of transcription factors involved in pigment gland development.

(A) Fragments per kilobase of transcript per million mapped fragments (FPKM) expression values of the representative transcription factors in *G. bickii* cotyledons at 0, 12, 24, 36, and 48 h after seed imbibition by bulk RNA-seq. Values are the means ± SD, different lowercase letters mean significant difference ($p < 0.05$), n = 3, *t*-test.

(B) Pigment gland phenotypes in leaves and stems of TM-1, TRV:*CLA* (*cloroplastos alterados 1*, positive control), TRV:00 (empty, negative vector control), TRV:*GbiERF114* silenced, TRV:*GbiCGF1* silenced, TRV:*GbiGoPGF* silenced, TRV:*GbiZAT11* silenced, TRV:*GbiGoSPGF* silenced, TRV:*GbiCGP1* silenced, TRV:*GbiNTL9* silenced, and TRV:*GbiCGF2* silenced cotton plants. The detail views of the leaf (20×) and stem (15×) images are shown on the left. Scale bars (1 cm) are shown on the right of each line.

(C) Pigment gland density on leaves and stems of the negative control and silenced seedlings (n = 15, $^{**}p < 0.01$, *t*-test).

(D) Gossypol content in leaves and stems of the negative control and silenced seedlings (n = 15, $^{**}p < 0.01$, *t*-test).

genes were upregulated under dark conditions, suggesting that light may inhibit gossypol synthesis (Figure 5E and Supplemental Table 11).

DISCUSSION

Pigment glands are regarded as defensive structures with bioactive and specialized metabolites for natural resistance to pests and pathogens in cotton. A fundamental biology topic in cotton and other glanded plants such as tomato is understanding the formation and development of glands (Huang et al., 2021). To address this issue, it is necessary to define the cell types within the gland-containing tissue and gain insight into plant life activities from a more microscopic perspective. In the present scRNA-seq study, the clear separation between pigment gland cells and other cell types is very significant for research on pigment glands, which make up only a small proportion of plant tissues (Xu et al., 2021a; Zhang et al., 2021a, 2021b; Chen et al., 2021b; Liu et al., 2021; Meir et al., 2021; Ryu et al., 2021; Kang et al., 2022; Shahan et al., 2022; Zong et al., 2022).

The study of pigment gland development started in the past century (McMichael, 1960), and the present results support the hypothesis of a lysigenous process, a typical form of autophagic programmed cell death (Dangl et al., 2000; Liu et al., 2010). Guided by morphological observations, the pigment gland development continuum of the 288 cells derived solely from pigment gland lineages was refigured into two vital stages (Liu et al., 2010). First, young pigment gland parenchyma, derived from the meristem, is constantly developing under substantial TF regulation. Two major differentiation trajectories co-exist in parenchyma, one toward central secretory cells and another toward mature gland parenchyma distributed around the cavity. Then,

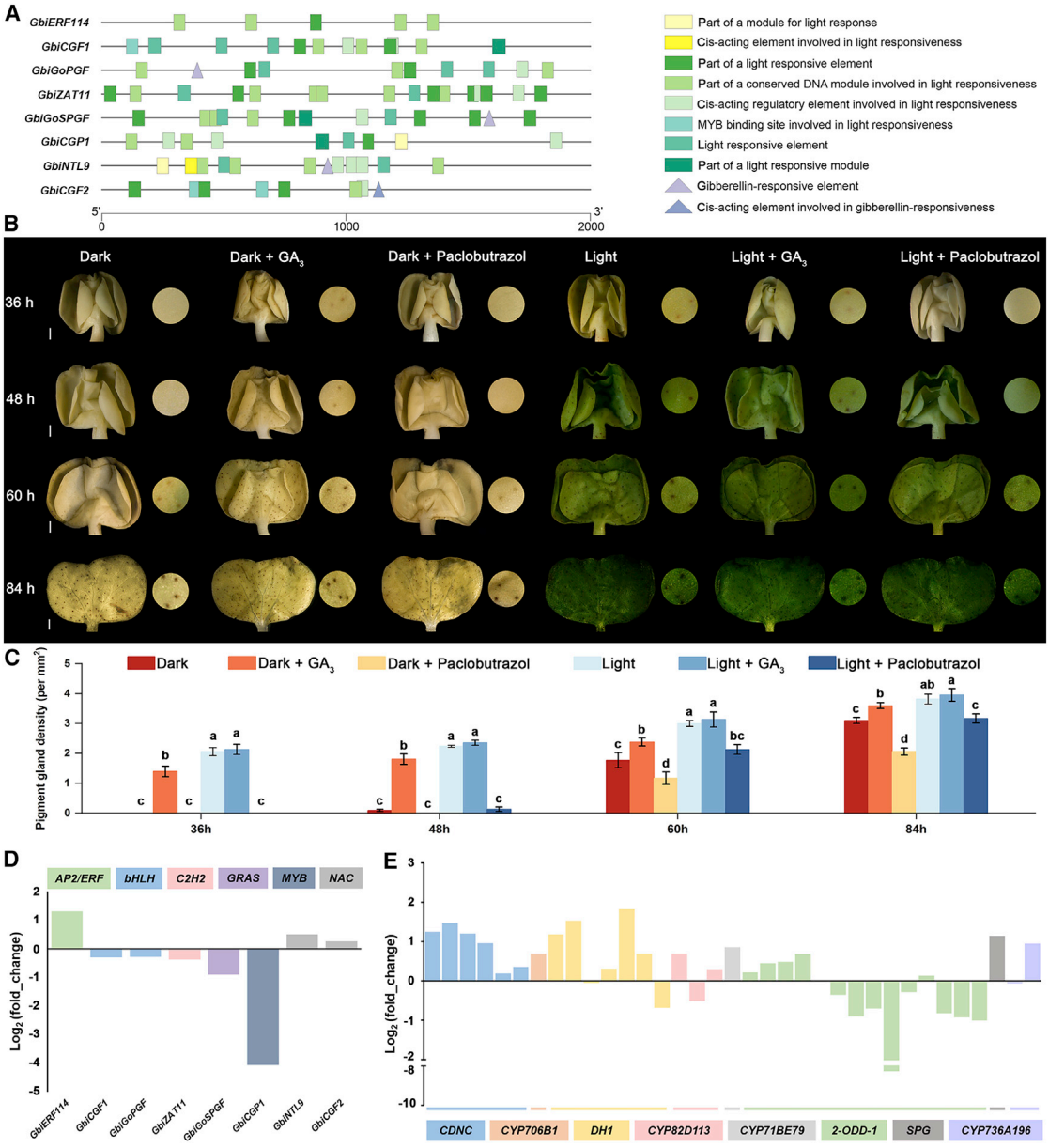

Figure 5. Light and GAs are required for pigment gland cell development.

(A) *cis*-acting elements in the 2000-bp promoter upstream of the coding region of transcription factors that regulate pigment gland development.

(B) The pigment gland phenotypic differences among cotyledons at 36, 48, 60, and 84 h after *G. bickii* seed imbibition under the treatments of dark, light, GA_3, and paclobutrazol. Scale bars, 1 mm. The circles next to the cotyledons show part of the corresponding cotyledon enlarged by four times.

(C) The statistical comparison of pigment gland density among cotyledons at 36, 48, 60, and 84 h after *G. bickii* seed imbibition under the treatments of dark, light, GA_3, and paclobutrazol. The results are derived from three biological repetitions. Values are the means ± SD, different lowercase letters mean significant difference ($p < 0.05$), *t*-test.

(D) The expression fold change of the pigment gland development transcription factors between the treatments of dark and light. The bars represent the expression fold change of cotyledons at 24 h after *G. bickii* seed imbibition under the treatment of dark versus that under the treatment of light. Different types of TFs are shown in different colors. The statistical analyses of gene expression are summarized in Supplemental Table 11.

(E) The expression fold change of the gossypol biosynthesis genes between the treatments of dark and light. Genes are arranged from left to right based on their order in the gossypol synthesis pathway. Genes of the same family are indicated by the same color. The gene IDs are shown in Supplemental Table 4. The statistical analyses of gene expression are summarized in Supplemental Table 11.

autolysis of secretory cells by programmed cell death releases secretion products, including gossypol, into the cavity.

Transcriptional heterogeneity, both in germplasms with different pigment gland traits and among plant tissues at different growth stages, has been used to identify key genes for pigment gland formation. Recently, *GoPGF*, several *CGF* genes, *GoSPGF*, and *CGP1* have been identified by bulk RNA-seq analysis and have provided critical insights into pigment gland development (Ma et al., 2016; Janga et al., 2019; Cai et al., 2020; Gao et al.,

2020; Zang et al., 2021). However, the bulk RNA-seq datasets were insufficient for exploring all the pigment gland formation signals due to the omission of low-expression genes in whole tissues (Figure 4A). In this study integrating pigment gland cells' developmental trajectory, TF regulatory networks, and core TFs' functional validation, a relatively complete model was proposed for pigment gland formation (Figure 3D). *GoPGF*, a *bHLH* TF, was shown as the master regulator in driving pigment gland cell development, and the cooperative regulatory roles of *AP2/ERF*s, *bHLH*s, *C2H2*s, *GRAS*s, *MYB*s, *NAC*s, and *WRKY*s with *GoPGF* were verified (Ma et al., 2016; Janga et al., 2019; Cai et al., 2020; Gao et al., 2020; Wang et al., 2021a; Yi et al., 2021; Zang et al., 2021). Furthermore, the novel genes *GbiERF114* (*AP2/ERF*), *GbiZAT11* (*C2H2*), and *GbiNTL9* (*NAC*) were identified and shown to reduce the pigment gland density and gossypol content after VIGS. These findings provide insights into the mechanisms of *GoPGF*'s and other TFs' functions in pigment gland formation.

The light-related pathways were identified as highly associated with pigment gland density by integrating bulk and single-cell sequencing data. Light- and GA-mediated promotion of pigment gland development was experimentally confirmed. The experiment also suggested a superimposed effect of light and GA on the promotion (Figure 5B), which may be due to the possible cross talk between light and GA in the promotion of pigment gland development (Fernández-Calvo et al., 2011; Gangappa et al., 2010; Gao et al., 2020; Hong et al., 2012; Qi et al., 2015). Further experiments are needed to explore this cross talk in pigment gland development and know whether the delayed pigment gland morphogenesis trait is related to lack of light signal transduction during ovule development.

Gossypol is stored in pigment glands as the main inclusion, and the relationship between pigment glands and gossypol has long been controversial. According to the physiological and biochemical phenotypes observed in glanded cotton cultivars, gossypol content is highly correlated with the density of pigment glands (Singh and Weaver, 1972; Wilson and Smith, 1976), but *Gossypium stocksii* and *Gossypium somalense* are exceptions that have normal pigment gland structure but no gossypol in seed (Xiang et al., 1993). Restraining pigment gland formation by gene silencing or knockout leads to significant reduction in gossypol content (Ma et al., 2016; Janga et al., 2019; Cai et al., 2020), while silencing of key genes in gossypol biosynthesis has less effect on pigment gland formation (Sunilkumar et al., 2006; Ma et al., 2016). The development process of pigment gland cells revealed by the present scRNA-seq data can provide insights into these conflicting observations. For cell development, gossypol biosynthesis occurs primarily in secretory cells, which are developed from parenchyma cells of the pigment gland (Figure 3D). Therefore, once pigment gland formation is restrained, biosynthesis and accumulation of gossypol will also be constrained (Ma et al., 2016; Janga et al., 2019). However, gossypol biosynthesis and pigment gland formation are controlled by different molecular mechanisms without direct connection (Ma et al., 2016; Gao et al., 2020; Zhao et al., 2020) (Supplemental Figure 11). It was also shown in comparative transcriptome analysis between treatments of dark and light that light may promote pigment gland development but inhibits gossypol biosynthesis. It was speculated that *GoPGF* and its downstream genes might be regulated by light, which is in accordance with the previous study that *ATMYC2* (homologous gene of *GoPGF*) responds to light signaling (Gangappa et al., 2010), while some light-induced TFs might be negative regulators of gossypol biosynthesis in secretory cells (Jing and Lin, 2020). Further investigations are necessary to confirm the hypothesis and to elucidate this complicated relationship.

With the help of scRNA-seq technology and the unique traits of *G. bickii*, new insights are provided into pigment gland morphogenesis, and the valuable resource provided here will help us further explore the mechanisms of delayed pigment gland morphogenesis. Some of these genes could be used in the production of cotton with ideal pigment gland traits in the future through further methods such as transgenic technology. scRNA-seq technology is constantly advancing in plants, but challenges remain, such as the protoplasting method and RNA *in situ* hybridization in non-model crops. Looking forward to the future, the application of scRNA-seq/single-nucleus RNA-seq together with spatial transcriptomics, chromatin accessibility, DNA methylation, and proteomics can improve and obtain a broader view of individual plant cell characteristics (Macaulay et al., 2017; Stuart and Satija, 2019; Ma et al., 2020; Dorrity et al., 2021; Farmer et al., 2021; Marand et al., 2021). It is anticipated that multiple single-cell analyses will be used to address many aspects of cotton biology, especially the development of cotton fibers, which are extensively elongated and unbranched single cells derived from the seed epidermis (Pang et al., 2009; Haigler et al., 2012; Huang et al., 2021; Pei et al., 2022).

METHODS

Cotyledon collection and protoplast preparation

Mature seeds of *G. bickii* were sterilized and sown on filter paper soaked in water in Petri dishes at 28°C with light. Cotyledons of the germinating seeds were carefully harvested at 48 h after seed sowing and moved to a sterile water-moistened Petri dish. The harvested cotyledons were cut sufficiently into thin slices with a sharp blade and transferred to a small Petri dish. Three milliliters of freshly prepared enzyme solution (20 mM KCl, 10 mM $CaCl_2$, 0.1% BSA [w/v], 20 mM MES, 0.4 M mannitol, 1.5% Cellulase R-10 [Yacult, Japan], and Macerozyme R-10 [Yacult, Japan] [pH 5.7], 0.22-μm filtration sterilization) was added to the Petri dish in advance. The cut tissue slices were immersed in the enzymatic digestion solution, placed in a 23°C constant temperature shaker, and digested at 180 rpm in the dark for 3–4 h. After digestion, the protoplasts were filtered with a 40-μm cell sieve and centrifuged at 300 *g* at room temperature for 5 min to collect the protoplasts. The protoplasts were washed two or three times with WI solution (0.5 M mannitol, 20 mM KCl, 4 mM MES [pH 5.7], 0.22-μm filtration sterilization), centrifuged at 300 *g* for 5 min, and resuspended to 50 μl. Cell activity was examined by 0.04% trypan blue staining and cell concentration was measured using a hemocytometer and a light microscope. Ideal samples contained more than 5×10^4 cells with viability over 80%.

scRNA-seq library construction and scRNA-seq data preprocessing

The scRNA-seq libraries were constructed with the Chromium Next GEM Single Cell 3' GEM, Library & Gel Bead Kit v.3.1 according to the user manual supplied with the kit. The libraries were sequenced by an Illumina sequencer NovaSeq6000 with 150-bp paired-end reads.

The Cell Ranger software pipeline (version 3.1.0) provided by 10× Genomics was used to de-multiplex cellular barcodes, map reads to the *G. bickii* reference genome and transcriptome using the STAR aligner (Sheng et al., 2022), and down-sample reads as required to generate normalized aggregate data across samples, producing a matrix of gene counts versus cells. We processed the unique molecular identifier (UMI) count matrix using the R package Seurat (Butler et al., 2018) (version 3.1.1). To remove low-quality cells and likely multiplet captures, we applied a criterion to filter out cells with UMI/gene numbers out of the limit of the mean value ±2-fold of the standard deviation assuming a Gaussian distribution of each cell's UMI/gene numbers. Following visual inspection of the distribution of cells by the fraction of expressed mitochondrial genes, we further discarded low-quality cells. After applying these quality control criteria, we obtained 12 222 single cells for the downstream analyses. Library size normalization was performed with the Normalize Data function in Seurat (Butler et al., 2018) to obtain the normalized count.

Top variable genes across single cells were identified using the method described by Macosko et al. (Macosko et al., 2015). Principal-component analysis was performed to reduce the dimensionality. Graph-based clustering was performed to cluster cells according to their gene expression profile, and cells were visualized using a UMAP algorithm and *t*-stochastic neighbor embedding analyses (Becht et al., 2019). DEGs were identified using the FindMarkers. A $p < 0.05$ and |log2 fold change| > 0.58 were set as the thresholds for significantly differential expression. Enrichment analyses of GO terms and KEGG pathways of DEGs were performed using R based on the hypergeometric distribution.

RNA velocity analysis

To perform the RNA velocity analysis, the spliced reads and unspliced reads were recounted using the Python script velocyto.py (La Manno et al., 2018) (https://github.com/velocyto-team/velocyto.py) on the Cell Ranger output folder. The calculation of RNA velocity values for each gene in each cell and the embedding RNA velocity vector to low-dimension space were performed with the R package velocyto.R v.0.6. Velocity fields were projected onto the UMAP embedding obtained in Seurat.

Pseudo-time analysis

Mesophyll cells (clusters 1, 2, 3, and 8) were selected for reconstruction of mesophyll differentiation trajectory, and pigment gland cells (sub-clusters 1, 2, and 3) were selected for reconstruction of the developmental trajectory of pigment gland cells. We determined the developmental pseudo-time with the Monocle2 package (version 2.9.0) (Trapnell et al., 2014). The raw count was first converted from Seurat object into CellDataSet object with the importCDS function in Monocle. We used the differentialGeneTest function of the Monocle2 package to select ordering genes (qval < 0.01) that were likely to be informative in the ordering of cells along the pseudo-time trajectory. The dimensional reduction clustering analysis was performed with the reduceDimension function, followed by trajectory inference with the orderCells function using the default parameters. Gene expression was plotted with the plot_genes_in_pseudotime function to track changes over pseudo-time.

scRNA-seq co-expression analysis

The scRNA-seq datasets of pigment gland cells (sub-clusters 1, 2, and 3) were used for calculating co-expression values. Pearson's correlation coefficient between the expression levels of each gene pair was calculated using the "cor.test" function in R to identify genes with similar expression patterns. Considering that the features of single-cell sequencing data are different from those bulk RNA-seq (e.g., dropout events), it may not be strict to designate a threshold rashly. *GbiGoPGF* was then extracted from the datasets with |co-expression value| > 0.1. The gene-regulatory network constructed from the target TFs of the *GbiGoPGF* obtained through DAP-seq, which was further screened by the scRNA-seq co-expression analysis, and the gene correlation network were displayed using Cytoscape (v.3.9.0) (Shannon et al., 2003).

RNA *in situ* hybridization

This was carried out as described in scRNA-seq samples collection. Cotyledons of the 48-h-germinating seeds were collected. Probe sequences for all genes are listed in Supplemental Table 12. Tissue sections were incubated overnight with RNA digoxigenin-labeled mature probe (Servicebio) at 42°C. The next day, slides were incubated with anti-digoxigenin alkaline phosphatase-conjugated antibody (Roche, 1:2000). Alkaline phosphatase substrate deposition using NBT/BCIP reaction was performed and sections were imaged using an Olympus BX50 microscope with a Spot Insight 6 camera. Slides were then washed in PBS and incubated with rabbit anti-GFAP antibody (Dako, 1:1000), diluted in 1% BSA PBS + 0.1% Triton X-100 overnight. Slides were then washed 3× in PBS 0.1% Triton X-100 and stained.

DNA-affinity purification sequencing

DAP-seq binding assays were performed as described previously with modification (O'Malley et al., 2016; Bartlett et al., 2017) and using a DNA Affinity Purification Sequencing Kit (cat. no. D202009, Bluescape) according to the manufacturer's instructions. Fresh *G. hirsutum* leaves were used for genomic DNA exaction. Fragmented gDNA was constructed into libraries using the NEXTFLEX Rapid DNA Seq Kit (PerkinElmer, Austin, TX, USA). The coding sequence of *GoPGF* was cloned into a pFN19K HaloTag T7 SP6 Flexi expression vector and was expressed using the TNT SP6 Coupled Wheat Germ Extract System (Promega). Expressed proteins were directly captured using Magne Halo Tag Beads (Promega). The GoPGF-bound beads were incubated with adapter-ligated gDNA libraries. Eluted DNAs were sequenced on an Illumina NovaSeq6000 with two technical duplicates. Beads without the addition of protein were taken as the input negative control DAP libraries. The fastp software default parameters were used to filter the raw data to obtain high-quality sequencing data/clean data for downstream analysis (Chen et al., 2018). Clean reads were mapped to the *G. hirsutum* genome sequence (http://www.cottongen.org/; Hu et al., 2019) to get unique mapped reads using BOWTIE2 (Langmead and Salzberg, 2012). MACS2 callpeak (Zhang et al., 2008) and IDR software (Li et al., 2011) were used to merge the peaks of the two biological duplicates with Q < 0.05. Motif discovery was performed using the MEME-ChIP software (Machanick and Bailey, 2011). The bound peaks were annotated using Homer software (Heinz et al., 2010). The promoter regions were determined as the binding peaks within 2000 bp upstream of the transcription start site. Finally, genes were aligned to *G. bickii* by blast with the criterion of identity over 90%. The summary statistics of DAP-seq are shown in Supplemental Figure 12 and Supplemental Tables 13 and 14.

Analysis of bulk RNA-seq data

After soaking in water for 6 h, the de-hulled *G. bickii* seeds were germinated on filter paper soaked with water in Petri dishes at 28°C under continuous light or dark conditions. Bulk RNA-seq samples were extracted from *G. bickii* cotyledons at 0, 12, 24, 36, and 48 h after seed imbibition and 24 h under light and 24 h in the dark. Three biological replicates were obtained for each sample. Isolated RNA was used for cDNA library construction with the NEBNext Ultra RNA Library Prep Kit following the manufacturer's instructions (Illumina, San Diego, CA, USA). All libraries were subjected to paired-end 150-bp sequencing on the Illumina HiSeq 4000 platform. Illumina short paired-end reads produced previously were trimmed using Trimmomatic (v.0.36) with default parameters. The high-quality reads were aligned to the *G. bickii* genome using HISAT2 (v.2.0.4) (Kim et al., 2019). The counts of reads aligned to each transcript were calculated with HTSeq (v.0.6.0) software (Anders et al., 2015), and fragments per kilobase of transcript per million mapped fragments were used to estimate the expression level of each gene.

Virus-induced gene silencing system

A VIGS assay was used to knock down the target genes amplified from *G. bickii* cDNA. The primers used are listed in Supplemental Table 12. TRV-based vectors (*pTRV1* and *pTRV2*) were used. The PCR fragment was ligated into *Bam*HI–*Kpn*I-digested *pTRV2*. The *pTRV1* and recombinant *pTRV2* vectors were introduced into *Agrobacterium* strain GV3101 and the *Agrobacterium* suspension was injected into cotyledons of 7-day-old *G. hirsutum* seedlings at a ratio of 1:1. The instructions and protocols for vector construction and virus transformation were described in a previous study (Li et al., 2019). All VIGS experiments were performed three times. Fifteen replicates were obtained for each sample to measure the gossypol content and pigment gland density. The protocols for sample treatment method and gossypol content determination are described in a previous study (Tian et al., 2018). Total RNA was extracted from different tissues using the RNAprep Pure Plant Kit (Tiangen). Total RNA (500 ng) was reverse transcribed using the ReverTra Ace qPCR RT Kit (with gDNA remover) (Toyobo, Osaka, Japan). qPCR analysis was performed using a LightCycler FastStart DNA Master SYBR Green I kit (Roche, Basel, Switzerland) on a LightCycler96 real-time PCR detection system (Roche). Three independent biological replicates were used for each analysis with at least three technical replicates each. The relative expression levels were evaluated using the comparative cycle threshold method according to Livak and Schmittgen (2001). GhUBQ was used as the internal control, which was stably expressed in cotton plants and not affected by treatments and genotypes. The expression of genes in the A sub-genome and D sub-genome was measured. All primers are listed in Supplemental Table 12.

Integrating bulk and single-cell sequencing data

To link pigment gland cells with specific phenotypes, we used the Scissor algorithm, a novel R package (version 2.1.0), which is a novel single-cell data analysis approach that utilizes bulk phenotypes to identify the most highly phenotype-associated cell subpopulations from single-cell sequencing data (Wang et al., 2022). Its usage is detailed on the website https://sunduanchen.github.io/Scissor/vignettes/Scissor_Tutorial.html#execute-scissor-to-select-the-informative-cells-1. Scissor integrates phenotype-associated bulk expression data and single-cell data by quantifying the similarity between every single cell and bulk sample. To identify relevant subpopulations, it then optimizes a regression model on the correlation matrix with the sample phenotype. The Scissors algorithm incorporates a reliability test to rule out false associations between identified cell subsets and bulk phenotypes. This statistical test can determine whether the inferred phenotype-cell association is reliable ($p < 0.05$) or a false positive. In actual operation, we set the family = "gaussian", alpha = "0.5", cutoff = "0.2" by a Gaussian linear regression model. In the results, both Scissor$^+$ cells and Scissor$^-$ cells are the Scissor selected cells that are most highly associated with the specified phenotypes. Among them, Scissor$^+$ indicates that the selected cells are positively correlated with the phenotype of interest, and Scissor$^-$ indicates that the selected cells are negatively correlated with the phenotype of interest.

Effect of light and GA on pigment gland morphogenesis in cotyledons

G. bickii seeds were soaked in 100 μmol/l GA_3 and 100 μmol/l paclobutrazol (a synthetic inhibitor of gibberellins) for 6 h, and then the de-hulled seeds were germinated on filter paper soaked with water in Petri dishes at 28°C to investigate the effect of gibberellin on pigment gland formation. For investigation of the effect of light, after soaking in water for 6 h, the de-hulled *G. bickii* seeds were germinated on filter paper soaked with water in Petri dishes at 28°C under light or dark conditions. Three biological replicates were obtained for each condition, and each replicate contained 50 seedlings.

ACCESSION NUMBERS

All sequencing data generated in this study have been submitted to Gene Expression Omnibus GEO (www.ncbi.nlm.nih.gov/geo/) under accession nos. GSE224635 and GSE224574.

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Molecular Plant Online*.

FUNDING

This work was supported by the Genetically Modified Organisms Breeding Major Project of China (2016ZX08005005), the National Key Technology R&D Program of China (2016YFD0101404), the China Agricultural Research System (CARS-15-27), the Jiangsu Collaborative Innovation Center for Modern Crop Production, and the National Science Foundation of China (32101764).

AUTHOR CONTRIBUTIONS

S.Z. and T.Z. conceived and designed the project. Y.S. and Y.H. performed the experiments. K.S., P.Y., Y.C., H.L., and J.C. contributed to project discussion. Y.S. and T.Z. analyzed the data and wrote the manuscript draft, and S.Z. and Q.-H.Z. revised it. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

We thank OE Biotech Co. Ltd. (Shanghai, China) for providing scRNA-seq and Dr. Yongbing Ba and Qiuyan Shi for assistance with bioinformatics analysis. We thank Bluescape Hebei Biotech Co. Ltd. (Baoding, PR China) for their support in the DAP-seq service. We thank Deli Sun (Agricultural Experiment Station, Zhejiang University) for the kind help with the materials planting. No conflict of interest is declared.

Received: November 11, 2022
Revised: December 31, 2022
Accepted: February 7, 2023
Published: February 10, 2023

REFERENCES

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics **31**:166–169. https://doi.org/10.1093/bioinformatics/btu638.

Bartlett, A., O'Malley, R.C., Huang, S.-s.C., Galli, M., Nery, J.R., Gallavotti, A., and Ecker, J.R. (2017). Mapping genome-wide transcription-factor binding sites using DAP-seq. Nat. Protoc. **12**:1659–1672. https://doi.org/10.1038/nprot.2017.055.

Becht, E., McInnes, L., Healy, J., Dutertre, C.-A., Kwok, I.W.H., Ng, L.G., Ginhoux, F., and Newell, E.W. (2018). Dimensionality reduction for visualizing single-cell data using UMAP. Nat. Biotechnol. **37**:38–44. https://doi.org/10.1038/nbt.4314.

Bezrutczyk, M., Zöllner, N.R., Kruse, C.P.S., Hartwig, T., Lautwein, T., Köhrer, K., Frommer, W.B., and Kim, J.Y. (2021). Evidence for phloem loading via the abaxial bundle sheath cells in maize leaves. Plant Cell **33**:531–547. https://doi.org/10.1093/plcell/koaa055.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. **36**:411–420. https://doi.org/10.1038/nbt.4096.

Cai, Y., Cai, X., Wang, Q., Wang, P., Zhang, Y., Cai, C., Xu, Y., Wang, K., Zhou, Z., Wang, C., et al. (2020). Genome sequencing of the Australian wild diploid species *Gossypium australe* highlights disease resistance and delayed gland morphogenesis. Plant Biotechnol. J. **18**:814–828. https://doi.org/10.1111/pbi.13249.

Chen, H., Yin, X., Guo, L., Yao, J., Ding, Y., Xu, X., Liu, L., Zhu, Q.H., Chu, Q., and Fan, L. (2021a). PlantscRNAdb: a database for plant single-cell RNA analysis. Mol. Plant **14**:855–857. https://doi.org/10.1016/j.molp.2021.05.002.

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics **34**:i884–i890. https://doi.org/10.1093/bioinformatics/bty560.

Chen, Y., Tong, S., Jiang, Y., Ai, F., Feng, Y., Zhang, J., Gong, J., Qin, J., Zhang, Y., Zhu, Y., et al. (2021b). Transcriptional landscape of highly lignified poplar stems at single-cell resolution. Genome Biol. **22**:319. https://doi.org/10.1186/s13059-021-02537-2.

Cheng, H., Lu, C., Yu, J.Z., Zou, C., Zhang, Y., Wang, Q., Huang, J., Feng, X., Jiang, P., Yang, W., et al. (2016). Fine mapping and candidate gene analysis of the dominant glandless gene Gl_2^e in cotton (*Gossypium* spp.). Theor. Appl. Genet. **129**:1347–1355. https://doi.org/10.1007/s00122-016-2707-1.

Daneva, A., Gao, Z., Van Durme, M., and Nowack, M.K. (2016). Functions and regulation of programmed cell death in plant development. Annu. Rev. Cell Dev. Biol. **32**:441–468. https://doi.org/10.1146/annurev-cellbio-111315-124915.

Dangl, J.L., Eietrich, R.A., and Thomas, H. (2000). Senescence and programmed cell death. Biochemistry and Molecular Biology of Plants. In American Society of Plant Physiologists, Rockville, pp. 1044–1100.

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.e5. https://doi.org/10.1016/j.devcel.2019.02.022.

Dorrity, M.W., Alexandre, C.M., Hamm, M.O., Vigil, A.-L., Fields, S., Queitsch, C., and Cuperus, J.T. (2021). The regulatory landscape of *Arabidopsis thaliana* roots at single-cell resolution. Nat. Commun. **12**:3334. https://doi.org/10.1038/s41467-021-23675-y.

Farmer, A., Thibivilliers, S., Ryu, K.H., Schiefelbein, J., and Libault, M. (2021). Single-nucleus RNA and ATAC sequencing reveals the impact of chromatin accessibility on gene expression in *Arabidopsis* roots at the single-cell level. Mol. Plant **14**:372–383. https://doi.org/10.1016/j.molp.2021.01.001.

Fernández-Calvo, P., Chini, A., Fernández-Barbero, G., Chico, J.M., Gimenez-Ibanez, S., Geerinck, J., Eeckhout, D., Schweizer, F., Godoy, M., Franco-Zorrilla, J.M., et al. (2011). The *Arabidopsis* bHLH transcription factors MYC3 and MYC4 are targets of JAZ repressors and act additively with MYC2 in the activation of jasmonate responses. Plant Cell **23**:701–715. https://doi.org/10.1105/tpc.110.080788.

Fryxell, P. (1965). A revision of the Australian species of *Gossypium* with observations on the occurrence of *Thespesia* in Australia (Malvaceae). Aust. J. Bot. **13**:71–102. https://doi.org/10.1071/BT9650071.

Gangappa, S.N., Prasad, V.B.R., and Chattopadhyay, S. (2010). Functional interconnection of MYC2 and SPA1 in the photomorphogenic seedling development of Arabidopsis. Plant Physiol. **154**:1210–1219. https://doi.org/10.1104/pp.110.163717.

Gao, W., Long, L., Zhu, L.-F., Xu, L., Gao, W.-H., Sun, L.-Q., Liu, L.-L., and Zhang, X.-L. (2013). Proteomic and virus-induced gene silencing (VIGS) analyses reveal that gossypol, brassinosteroids, and jasmonic acid contribute to the resistance of cotton to *Verticillium dahliae*. Mol. Cell. Proteomics **12**:3690–3703. https://doi.org/10.1074/mcp.M113.031013.

Gao, W., Xu, F.C., Long, L., Li, Y., Zhang, J.L., Chong, L., Botella, J.R., and Song, C.P. (2020). The gland localized *CGP1* controls gland pigmentation and gossypol accumulation in cotton. Plant Biotechnol. J. **18**:1573–1584. https://doi.org/10.1111/pbi.13323.

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. https://doi.org/10.3389/fpls.2012.00104.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell **38**:576–589. https://doi.org/10.1016/j.molcel.2010.05.004.

Hong, G.J., Xue, X.Y., Mao, Y.B., Wang, L.J., and Chen, X.Y. (2012). *Arabidopsis* MYC2 interacts with DELLA proteins in regulating sesquiterpene synthase gene expression. Plant Cell **24**:2635–2648. https://doi.org/10.1105/tpc.112.098749.

Hu, Y., Chen, J., Fang, L., Zhang, Z., Ma, W., Niu, Y., Ju, L., Deng, J., Zhao, T., Lian, J., et al. (2019). *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. Nat. Genet. **51**:739–748. https://doi.org/10.1038/s41588-019-0371-5.

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. https://doi.org/10.1146/annurev-arplant-080720-113241.

Huang, J.Q., Fang, X., Tian, X., Chen, P., Lin, J.L., Guo, X.X., Li, J.X., Fan, Z., Song, W.M., Chen, F.Y., et al. (2020). Aromatization of natural products by a specialized detoxification enzyme. Nat. Chem. Biol. **16**:250–256. https://doi.org/10.1038/s41589-019-0446-8.

Janga, M.R., Pandeya, D., Campbell, L.M., Konganti, K., Villafuerte, S.T., Puckhaber, L., Pepper, A., Stipanovic, R.D., Scheffler, J.A., and Rathore, K.S. (2019). Genes regulating gland development in the cotton plant. Plant Biotechnol. J. **17**:1142–1153. https://doi.org/10.1111/pbi.13044.

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. https://doi.org/10.1105/tpc.18.00785.

Jing, Y., and Lin, R. (2020). Transcriptional regulatory network of the light signaling pathways. New Phytol. **227**:683–697. https://doi.org/10.1111/nph.16602.

Kang, M., Choi, Y., Kim, H., and Kim, S.G. (2022). Single-cell RNA-sequencing of Nicotiana attenuata corolla cells reveals the biosynthetic pathway of a floral scent. New Phytol. **234**:527–544. https://doi.org/10.1111/nph.17992.

Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. **37**:907–915. https://doi.org/10.1038/s41587-019-0201-4.

Kim, J.Y., Symeonidi, E., Pang, T.Y., Denyer, T., Weidauer, D., Bezrutczyk, M., Miras, M., Zöllner, 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. https://doi.org/10.1093/plcell/koaa060.

La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lönnerberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. Nature **560**:494–498. https://doi.org/10.1038/s41586-018-0414-6.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods **9**:357–359. https://doi.org/10.1038/nmeth.1923.

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van De Peer, Y., Rouzé, P., and Rombauts, S. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res. **30**:325–327. https://doi.org/10.1093/nar/30.1.325.

Li, C., He, Q., Zhang, F., Yu, J., Li, C., Zhao, T., Zhang, Y., Xie, Q., Su, B., Mei, L., et al. (2019). Melatonin enhances cotton immunity to *Verticillium* wilt via manipulating lignin and gossypol biosynthesis. Plant J. **100**:784–800. https://doi.org/10.1111/tpj.14477.

Li, B., Liang, S., Alariqi, M., Wang, F., Wang, G., Wang, Q., Xu, Z., Yu, L., Naeem Zafar, M., Sun, L., et al. (2021). The application of temperature sensitivity CRISPR/LbCpf1 (LbCas12a) mediated genome editing in allotetraploid cotton (*G. hirsutum*) and creation of nontransgenic, gossypol-free cotton. Plant Biotechnol. J. **19**:221–223. https://doi.org/10.1111/pbi.13470.

Li, Q., Brown, J.B., Huang, H., and Bickel, P.J. (2011). Measuring reproducibility of high-throughput experiments. Ann. Appl. Stat. **5**:1752–1779. https://doi.org/10.1214/11-AOAS466.

Liu, H., Hu, D., Du, P., Wang, L., Liang, X., Li, H., Lu, Q., Li, S., Liu, H., Chen, X., et al. (2021). 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. https://doi.org/10.1111/pbi.13656.

Liu, W.Z., Zhou, Y.F., Wang, X., and Jiao, Z.J. (2010). Programmed cell death during pigment gland formation in *Gossypium hirsutum* leaves. Plant Biol. **12**:895–902. https://doi.org/10.1111/j.1438-8677.2009.00291.x.

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. https://doi.org/10.1016/j.molp.2020.06.010.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ Method. Methods **25**:402–408. https://doi.org/10.1006/meth.2001.1262.

Lopez-Anido, C.B., Vatén, 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.e4. https://doi.org/10.1016/j.devcel.2021.03.014.

Loudya, N., Mishra, P., Takahagi, K., Uehara-Yamaguchi, Y., Inoue, K., Bogre, L., Mochida, K., and López-Juez, E. (2021). Cellular and transcriptomic analyses reveal two-staged chloroplast biogenesis underpinning photosynthesis build-up in the wheat leaf. Genome Biol. **22**:151. https://doi.org/10.1186/s13059-021-02366-3.

Ma, A., McDermaid, A., Xu, J., Chang, Y., and Ma, Q. (2020). Integrative methods and practical challenges for single-cell multi-omics. Trends Biotechnol. **38**:1007–1022. https://doi.org/10.1016/j.tibtech.2020.02.013.

Ma, D., Hu, Y., Yang, C., Liu, B., Fang, L., Wan, Q., Liang, W., Mei, G., Wang, L., Wang, H., et al. (2016). Genetic basis for glandular trichome formation in cotton. Nat. Commun. **7**:10456. https://doi.org/10.1038/ncomms10456.

Macaulay, I.C., Ponting, C.P., and Voet, T. (2017). Single-cell multiomics: multiple measurements from single cells. Trends Genet. **33**:155–168. https://doi.org/10.1016/j.tig.2016.12.003.

Machanick, P., and Bailey, T.L. (2011). MEME-ChIP: motif analysis of large DNA datasets. Bioinformatics **27**:1696–1697. https://doi.org/10.1093/bioinformatics/btr189.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell **161**:1202–1214. https://doi.org/10.1016/j.cell.2015.05.002.

Marand, A.P., Chen, Z., Gallavotti, A., and Schmitz, R.J. (2021). A *cis*-regulatory atlas in maize at single-cell resolution. Cell **184**:3041–3055.e21. https://doi.org/10.1016/j.cell.2021.04.014.

McMichael, S.C. (1960). Combined effects of glandless genes gl_2 and gl_3 on pigment glands in the cotton plant. Agron. J. **52**:385–386. https://doi.org/10.2134/agronj1960.00021962005200070005x.

Meir, Z., Aviezer, I., Chongloi, G.L., Ben-Kiki, O., Bronstein, R., Mukamel, Z., Keren-Shaul, H., Jaitin, D., Tal, L., Shalev-Schlosser, G., et al. (2021). Dissection of floral transition by single-meristem transcriptomes at high temporal resolution. Nat. Plants **7**:800–813. https://doi.org/10.1038/s41477-021-00936-8.

O'Malley, R.C., Huang, S.-s.C., Song, L., Lewsey, M.G., Bartlett, A., Nery, J.R., Galli, M., Gallavotti, A., and Ecker, J.R. (2016). Cistrome and epicistrome features shape the regulatory DNA landscape. Cell **166**:1292–1598. https://doi.org/10.1016/j.cell.2016.08.063.

Pang, M., Woodward, A.W., Agarwal, V., Guan, X., Ha, M., Ramachandran, V., Chen, X., Triplett, B.A., Stelly, D.M., and Chen, Z.J. (2009). Genome-wide analysis reveals rapid and dynamic changes in miRNA and siRNA sequence and expression during ovule and fiber development in allotetraploid cotton (*Gossypium hirsutum* L.). Genome Biol. **10**:R122. https://doi.org/10.1186/gb-2009-10-11-r122.

Pei, L., Huang, X., Liu, Z., Tian, X., You, J., Li, J., Fang, D.D., Lindsey, K., Zhu, L., Zhang, X., et al. (2022). Dynamic 3D genome architecture of cotton fiber reveals subgenome-coordinated chromatin topology for 4-staged single-cell differentiation. Genome Biol. **23**:45. https://doi.org/10.1186/s13059-022-02616-y.

Qi, T., Wang, J., Huang, H., Liu, B., Gao, H., Liu, Y., Song, S., and Xie, D. (2015). Regulation of jasmonate-induced leaf senescence by antagonism between bHLH subgroup IIIe and IIId factors in Arabidopsis. Plant Cell **27**:1634–1649. https://doi.org/10.1105/tpc.15.00110.

Rathore, K.S., Pandeya, D., Campbell, L.M., Wedegaertner, T.C., Puckhaber, L., Stipanovic, R.D., Thenell, J.S., Hague, S., and Hake, K. (2020). Ultra-low gossypol cottonseed: selective gene silencing opens up a vast resource of plant-based protein to improve human nutrition. CRC Crit. Rev. Plant Sci. **39**:1–29. https://doi.org/10.1080/07352689.2020.1724433.

Reape, T.J., Molony, E.M., and McCabe, P.F. (2008). Programmed cell death in plants: distinguishing between different modes. J. Exp. Bot. **59**:435–444. https://doi.org/10.1093/jxb/erm258.

Roszak, P., Heo, J.o., Blob, B., Toyokura, K., Sugiyama, Y., de Luis Balaguer, M.A., Lau, W.W.Y., Hamey, F., Cirrone, J., Madej, E., et al. (2021). Cell-by-cell dissection of phloem development links a maturation gradient to cell specialization. Science **374**:eaba5531. https://doi.org/10.1126/science.aba5531.

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. https://doi.org/10.1146/annurev-genet-071719-020453.

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. https://doi.org/10.1104/pp.18.01482.

Satterlee, J.W., Strable, J., and Scanlon, M.J. (2020). Plant stem-cell organization and differentiation at single-cell resolution. Proc. Natl. Acad. Sci. USA **117**:33689–33699. https://doi.org/10.1073/pnas.2018788117.

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.e9. https://doi.org/10.1016/j.devcel.2022.01.008.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. **13**:2498–2504. https://doi.org/10.1101/gr.1239303.

Shaw, R., Tian, X., and Xu, J. (2021). Single-cell transcriptome analysis in plants: advances and challenges. Mol. Plant **14**:115–126. https://doi.org/10.1016/j.molp.2020.10.012.

Xiang, S., and Yang, W. (1993). Studies on gossypol and its enantiomers in the seeds of cotton Gossypium. Sci. Agric. Sin. **26**:31–35.

Sheng, K., Sun, Y., Liu, M., Cao, Y., Han, Y., Li, C., Muhammad, U., Daud, M.K., Wang, W., Li, H., et al. (2023). A reference-grade genome assembly for *Gossypium bickii* and insights into its genome evolution and formation of pigment glands and gossypol. Plant Commun. **4**:100421. https://doi.org/10.1016/j.xplc.2022.100421.

Singh, I.D., and Weaver, J.B., Jr. (1972). Growth and infestation of boll weevils on normal-glanded, glandless, and high-gossypol strains of cotton. J. Econ. Entomol. **65**:821–824. https://doi.org/10.1093/jee/65.3.821.

Stuart, T., and Satija, R. (2019). Integrative single-cell analysis. Nat. Rev. Genet. **20**:257–272. https://doi.org/10.1038/s41576-019-0093-7.

Sunilkumar, G., Campbell, L.M., Puckhaber, L., Stipanovic, R.D., and Rathore, K.S. (2006). Engineering cottonseed for use in human nutrition by tissue-specific reduction of toxic gossypol. Proc. Natl. Acad. Sci. USA **103**:18054–18059. https://doi.org/10.1073/pnas.0605389103.

Tian, C., Du, Q., Xu, M., Du, F., and Jiao, Y. (2020). Single-nucleus RNA-seq resolves spatiotemporal developmental trajectories in the tomato shoot apex. Preprint at bioRxiv. https://doi.org/10.1101/2020.09.20.305029.

Tian, X., Ruan, J.X., Huang, J.Q., Yang, C.Q., Fang, X., Chen, Z.W., Hong, H., Wang, L.J., Mao, Y.B., Lu, S., et al. (2018). Characterization of gossypol biosynthetic pathway. Proc. Natl. Acad. Sci. USA **115**:E5410–E5418. https://doi.org/10.1073/pnas.1805085115.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. **32**:381–386. https://doi.org/10.1038/nbt.2859.

Wang, F., Chen, M., Ma, J., Wang, C., Wang, J., Xia, H., Zhang, D., and Yao, S. (2022). Integrating bulk and single-cell sequencing reveals the phenotype-associated cell subpopulations in sepsis-induced acute lung injury. Front. Immunol. **13**:981784. https://doi.org/10.3389/fimmu.2022.981784.

Wang, M., Qiu, X., Pan, X., and Li, C. (2021a). Transcriptional factor-mediated regulation of active component biosynthesis in medicinal plants. Curr. Pharm. Biotechnol. **22**:848–866. https://doi.org/10.2174/1389201021666200622121809.

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. https://doi.org/10.1016/j.jgg.2021.06.001.

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**:eaay4970. https://doi.org/10.1126/science.aay4970.

Wilson, F.D., and Smith, J.N. (1976). Some genetic relationships between gland density and gossypol content in *Gossypium hirsutum* L. Crop Sci. **16**:830–832. https://doi.org/10.2135/cropsci1976.0011183X001600060023x.

Williams, J.L., Ellers-Kirk, C., Orth, R.G., et al. (2011). Fitness cost of resistance to Bt cotton linked with increased gossypol content in pink bollworm larvae. PLoS One **6**, e21863. https://doi.org/10.1371/journal.pone.0021863.

Xu, X., Crow, M., Rice, B.R., Li, F., Harris, B., Liu, L., Demesa-Arevalo, E., Lu, Z., Wang, L., Fox, N., et al. (2021a). Single-cell RNA sequencing of developing maize ears facilitates functional analysis and trait candidate gene discovery. Dev. Cell **56**:557–568.e6. https://doi.org/10.1016/j.devcel.2020.12.015.

Xu, Z., Wang, Q., Zhu, X., Wang, G., Qin, Y., Ding, F., Tu, L., Daniell, H., Zhang, X., and Jin, S. (2022). Plant Single Cell Transcriptome Hub (PsctH): an integrated online tool to explore the plant single-cell transcriptome landscape. Plant Biotechnol. J. **20**:10–12. https://doi.org/10.1111/pbi.13725.

Yi, X., Wang, X., Wu, L., Wang, M., Yang, L., Liu, X., Chen, S., and Shi, Y. (2021). Integrated analysis of basic Helix loop Helix transcription factor family and targeted terpenoids reveals candidate *AarbHLH* genes involved in terpenoid biosynthesis in *artemisia* argyi. Front. Plant Sci. **12**:811166. https://doi.org/10.3389/fpls.2021.811166.

Zang, Y., Xu, C., Xuan, L., Ding, L., Zhu, J., Si, Z., Zhang, T., and Hu, Y. (2021). Identification and characteristics of a novel gland-forming gene in cotton. Plant J. **108**:781–792. https://doi.org/10.1111/tpj.15477.

Zhang, J., and Wedegaertner, T. (2021). Genetics and breeding for glandless upland cotton with improved yield potential and disease resistance: a review. Front. Plant Sci. **12**:753426. https://doi.org/10.3389/fpls.2021.753426.

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. https://doi.org/10.1016/j.molp.2019.04.004.

Zhang, T.Q., Chen, Y., and Wang, J.W. (2021a). A single-cell analysis of the *Arabidopsis* vegetative shoot apex. Dev. Cell **56**:1056–1074.e8. https://doi.org/10.1016/j.devcel.2021.02.021.

Zhang, T.Q., Chen, Y., Liu, Y., Lin, W.H., and Wang, J.W. (2021b). Single-cell transcriptome atlas and chromatin accessibility landscape reveal differentiation trajectories in the rice root. Nat. Commun. **12**:2053. https://doi.org/10.1038/s41467-021-22352-4.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-seq (MACS). Genome Biol. **9**:R137. https://doi.org/10.1186/gb-2008-9-9-r137.

Zhao, T., Xie, Q., Li, C., Li, C., Mei, L., Yu, J.Z., Chen, J., and Zhu, S. (2020). Cotton roots are the major source of gossypol biosynthesis and accumulation. BMC Plant Biol. **20**:88. https://doi.org/10.1186/s12870-020-2294-9.

Zhu, S., Ji, D., Wang, R., and Wang, H. (1999). Studies on the gossypol trend of the cotyledon during seed germination and the relationship between gossypol and gland formation in the wild species of *Gossypium* in Australia. Cotton Science **11**:169–173.

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. https://doi.org/10.1111/nph.18008.