Title

Identification of maize genes conditioning the early systemic infection of sugarcane mosaic virus by single-cell transcriptomics

Authors

Chen, Xi

Yao, Ru

Hua, Xia

et al.

Publication Date

2025-03-01

DOI

10.1016/j.xplc.2025.101297

Title page

A high-resolution single-cell atlas reveals cellular developmental trajectories and the novel transcript factors involving gossypol synthesis in roots of upland cotton (*Gossypium hirsutum* L.)

Lei Mei[1, 7, 9, †, *], Chengyu Ye[2, †], Jianyan Zeng[3, †], Peng Zheng[1, †], Lan Long[4], Hanyan Lai[3], Xianwen Zhang[5], Boxin Liu[6], Huimin Ye[7], Xuewei Wu[8], Yueyi Zhu[1], Simon C. Groen[10, 11, 12] and Sanjie Jiang[6,*]

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

[2] Department of Microbiology and Immunology, Emory University, Atlanta GA 30322, United States of America;

[3] College of Agronomy and Biotechnology, Southwest University, Chongqing 400715, China;

[4] Enshi Tujia & Miao Autonomous Prefecture Academy of Agricultural Sciences, Enshi 445000, China;

[5] Institute of Virology and Biotechnology, Zhejiang Academy of Agricultural Sciences, Hangzhou 310021, China;

[6] Beijing Genome Institute, Beijing 100108, China;

[7] Center of Research and Development, Selenium Science Development (Zhejiang) Company Limited, Hangzhou 311121, China;

[8] School of Agriculture, Yunnan University, Kunming 650091, China;

[9] Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, United Kingdom;

[10] Department of Nematology, University of California Riverside, Riverside CA 92521, United States of America;

[11] Center for Plant Cell Biology, Institute for Integrative Genome Biology, University of California Riverside, Riverside CA 92521, United States of America;

[12] Department of Botany and Plant Sciences, University of California Riverside, Riverside CA 92521, United States of America;

[†]Co-first authors: These authors contributed equally.

*Correspondence: Dr Lei Mei (meileihzruk@gmail.com; leimei@zju.edu.cn) and Dr Sanjie Jiang (jiangsanjie@genomics.cn)

Running title: scRNA-seq atlas and novel transcript factors of gossypol synthesis in upland cotton root

Short Summary: This study presents a high-resolution single-cell atlas of upland cotton roots, elucidating cellular locations and identifying key transcript factors associated with gossypol synthesis in the roots of upland cotton.

Abstract

The toxicity of gossypol limits the utilization of seeds from commercial cotton for human food and animal feed. Gossypol and its associated terpenoids can be synthesized across multiple organs in cotton, such as roots, stems, leaves, and seed kernels. In contrast to aerial parts, the synthesis of gossypol in distal cotton roots can eliminate interference from gland morphogenesis, providing an ideal organ for studying gossypol synthesis. Using single-cell RNA sequencing, we dissected the cell heterogeneity of upland cotton roots, releasing a high-resolution scRNA-seq atlas. The resulting single-cell transcriptome atlas partitioned the cotton root into 12 large cell populations of 16 transcriptionally distinct cell clusters, as defined through a series of novel gene markers. The scRNA-seq transcriptome profile of upland cotton roots was characterized through intra- and inter-species comparisons. Via multiple *in silico* approaches, we confirmed that the lateral root cap plays a crucial role in gossypol synthesis in upland cotton roots. Specifically, further developmental trajectory analysis revealed that cell type lateral root cap-1 (LRC-1) branched off and differentiated into two distinct cell types, LRC-2 and LRC-3. The development of LRC-2 from LRC-1 was crucial for gossypol production. In addition, our gene regulatory network analysis identified 53 novel core candidate transcription factors (TFs), such as *GhZAT12*, *GhbZIP53*, *GhERF14*, *GhNAC87*, and *GhMYB73*, in regulating gossypol synthesis. The top two TFs, *GhZAT12* and *GhbZIP53*, were selected for functional validation by virus-induced gene silencing (VIGS). This indicates their participation in gossypol synthesis in cotton roots without affecting gland morphogenesis in aerial parts. This suggests that gossypol synthesis is independent of gland development. This study is the first to release the single-cell transcriptional landscape of upland cotton roots and successfully identified TFs involved in root gossypol synthesis. Our findings offer valuable insights into

63 comprehending cell differentiation and cell type-specific metabolism, specifically
64 gossypol synthesis, at a single-cell resolution. Moreover, we establish a reference
65 model for identifying metabolism-related transcription factors in plants.

66 *Keywords:* single-cell RNA-seq; upland cotton (*Gossypium hirsutum L.*); gossypol
67 synthesis; differentiation trajectory; transcript factor

1. Introduction

Cotton is one of the most important economic crops globally since it not only produces natural fiber for the textile industry, but its seeds further provide plenty of nutrients, including proteins, oils as well as mineral elements, for human food and livestock feed (Gao et al., 2022; Lusas and Jividen, 1987). It had been estimated that global cottonseed production could meet the protein requirements of 500 million people annually (Sunilkumar et al., 2006). Gossypol exists as a sesquiterpene compound in cotton plants, conferring resistance to pests and pathogens, and this trait may be adaptive in hostile environments (Scheffler et al., 2012; Sun et al., 2010). However, gossypol is toxic to humans and other monogastric mammals, thereby limiting the direct utilization of cottonseed (Benedict et al., 2004).

As early as 1886, gossypol was isolated from a mixture with other pigments via the refinement of cottonseed oil (Dodou, 2005). The gossypol biosynthetic pathway was re-constructed in recent decades. ^{14}C tracing technology showed (+)-δ-cadinene as a pivotal precursor to all cadinene-type sesquiterpenoids involving 7- and 8-hydroxylated derivatives (Heinstein et al., 1970). Generally, (+)-δ-cadinene synthase, encoded by two genes in cotton, CDNC and CDNA (key abbreviations referring to Supplemental Table 1), catalyzes FPP (farnesyl diphosphate) and produces (+)-δ-cadinene. From this, furocalamen-2-one is generated by cytochrome P450 enzymes that act in tandem: CYP706B1, CYP82D113, and CYP71BE79 (Tian et al., 2018). Then, hemigossypol is created through the activity of 2-oxoglutarate/Fe(II)-dependent dioxygenase-1. Finally, hemigossypol dimerises, forming gossypol (Benedict et al., 2006; Effenberger et al., 2015). The mentioned gossypol synthesis pathway has been extensively validated (Gao et al., 2020; Maryam et al., 2022; Zang et al., 2021; Zhang et al., 2022). Additionally, Huang et al. recently explored the functional attributes of a

glyoxalase I variant (SPG) responsible for catalyzing the conversion of aldehydes to ketones, and they reported on the influence of dirigent proteins (DIR) in modulating the enantioselectivity of gossypol (Huang et al., 2020; Lin et al., 2023).

Gossypol occurs in the seeds and is also found in the roots, leaves, bolls, and stems (Stipanovic et al., 2006). In cotton, the gossypol gland, namely the pigment gland, is regarded as a storage organ for retaining gossypol, which manifests as black or brownish-red dots in the tissues of all cotton plants, excluding pollen, seed coat, the distal portion of the root, and other parts (Gao et al., 2022; Tian et al., 2018). The glands within the green organs of the cotton plant contain gossypol, hemi-gossypolone, and heliocides, while glands in the flower petals and seed kernels contain only gossypol (Pandeya et al., 2023). To date, several crucial genes that determine gland morphogenesis have been meticulously mapped out and functionally validated, such as *CGF1*, *CGF2*, *CGF3/PGF*, *SPGF*, *EF105*, and so on (Janga et al., 2019; Long et al., 2023; Zang et al., 2021). Furthermore, scRNA-seq analysis was employed to screen batches of genes associated with gland development in various organs. Notably, the identified genes, such as *GhJUB1*, *GbiERF114*, *GbiZAT11*, and *GbiNTL9*, were reported (Long et al., 2023; Sun et al., 2023). Importantly, all of these genes implicated in gland morphogenesis belong to the category of transcription factors (TFs).

In cotton breeding, the gland trait is a distinguishing marker for primarily discerning gossypol-producing cotton from gossypol-free varieties. This is since gossypol is typically detectable in the aerial parts with glands. Earlier, based on grafting and in vitro cultivation experiments, roots were considered the primary organ for synthesizing gossypol in cotton, and glands were regarded as the storage organ for gossypol, with gossypol synthesized in the roots ultimately transported to the above-ground parts of the cotton plant (Gao et al., 2020; Ye et al., 2023; Zhao et al., 2020).

However, previous studies could only indicate that the roots are one of the organs involved in gossypol synthesis, without directly proving that the aboveground part of gossypol is transferred from the underground part in a downward-to-upward manner. This transfer mechanism of gossypol has faced intense scrutiny in the academic community. The latest research suggests that the aboveground parts of cotton, especially the glands, also possess strong gossypol synthesis capabilities (Lin et al., 2023; Pandeya et al., 2023). The distal roots of cotton lack glandular structures, regardless of whether in glandular or glandless cultivars, but exhibit strong gossypol synthesis capabilities, making them ideal organ objects for studying gossypol synthesis by eliminating gland interference.

To date, there is limited knowledge regarding the cellular heterogeneity of gossypol synthesis in cotton roots. The regulatory mechanisms associated with gossypol synthesis in roots are poorly understood, primarily due to the absence of reported vital transcription factors related to this process. Earlier studies in the genetic model plant Arabidopsis revealed non-overlapping root cell types, which were identified by generating a series of cell-type-specific reporter lines and characterizing these using a combination of cell sorting, microarray, and RNA sequencing approaches (Birnbaum et al., 2003; Brady et al., 2007). However, it is still difficult to identify root cell types in cotton due to their transcriptional similarities and confusion on whether specific cell clusters show homogeneous or heterogeneous properties relative to a reference cell type (Zhang et al., 2019). As the complexity of the cotton genome with abundant gene copies, it is arduous to identify cotton cell types through conventional routes (Li et al., 2021; Rong et al., 2010). Single-cell RNA-sequencing (scRNA-seq) methodologies have overcome several of the limitations of genome-wide gene expression profiling of aggregate cell samples, allowing the categorization of individual

cells into coherent clusters and the discovery of novel cell types (Bawa et al., 2022). After pioneering studies established the application of scRNA-seq for describing gene expression patterns in Arabidopsis roots (Birnbaum et al., 2003; Serrano-Ron et al., 2021; Zhang et al., 2019), this novel methodology has further shown an excellent ability to compare and determine cell identity in other plants such as rice, tea and many others. In upland cotton, single-cell RNA sequencing (scRNA-seq) was initially employed to elucidate the initiation of individual fiber cells (Qin et al., 2022). Subsequently, it has been utilized to investigate other typical issues, such as glandular cell heterogeneity (Lin et al., 2023; Long et al., 2023). Nevertheless, the technology has yet used to systematically and comprehensively expand knowledge of heterogeneous gene expression patterns across different root cell types in upland cotton. Single-cell RNA sequencing (scRNA-seq) was employed in upland cotton roots to identify crucial transcription factors that govern gossypol synthesis.

2 Results

2.1 Gossypol was synthesized at a high level by roots in both glandular and glandless upland cotton

The upland cotton seedlings employed in this study exhibited the correct phenotypes on glands, involved in glandular and glandless cultivar CKG and CKN, respectively. As shown in Figure 1A, the glands were detected within the cotyledon leaves and stems of cultivar CKG, whereas none of the organs of cultivar CKN seedlings developed pigment glands. The average density of glands in CKG was 251.57 glands per square centimeter on cotyledon and 55.43 on stems, respectively (Figure 1B, I). This indicated that cotyledons might have a higher storage capacity for gossypol than stems.

To determine the gossypol contents, measurements of (+)-gossypol and (-)-gossypol standards were obtained (see Supplemental Figure 1) and utilized for the quantification of gossypol in various organs of CKG and CKN plants (Figure 1B, II-IV). Regarding the total gossypol content of each organ in CKG, the highest value was observed in cotyledons, constituting as much as 0.85 % of dry weight. In contrast, the content of stems was more than 20 times lower, with a value of 0.04%. Notably, the total gossypol content of roots was 0.28%, significantly lower than in cotyledons but higher than in stems, respectively ($P<0.05$). In cultivar CKN, the total gossypol content of roots closely resembled that of CKG roots, even though levels were too low to quantify in cotyledons and stems of this cultivar. It is important to note that the total gossypol content results from accumulating both (+)-gossypol and (-)-gossypol. The Contents of (+)-gossypol and (-)-gossypol exhibited remarkably similar trends across all cultivar-by-organ combinations, mirroring the patterns observed in the total gossypol content.

In the subsequent phase of our investigation, bulk tissue RNA-seq analyses were performed on the three organs, namely cotyledon leaves, stems, and roots, belonging to the two cultivars. The transcript abundance data was examined within the comprehensive framework of the fully reconstructed gossypol pathway, encompassing the upstream MVA pathway, as elucidated by Tian et al. in 2018 (refer to Figure 1C). To establish a direct correspondence of gene IDs between the current dataset and that generated by Tian and colleagues, we have presented a detailed mapping in Supplemental Table 2A. Further, the corroborative values corresponding to Figure 1C are provided in Supplemental Table 2B. We successfully quantified the expression levels of 99 genes implicated in gossypol biosynthesis, encompassing fourteen genes responsible for encoding enzymes crucial in biosynthetic pathways. Generally, the

genes demonstrated consistent expression patterns across CKG and CKN, showing significantly heightened expression in roots compared to cotyledons and stems.

Specially, Acyl CoA-cholesterol acyltransferase (ACAT) facilitates the conversion of Acetyl-CoA to acetoacetyl-CoA. Among the four genes coding for ACAT, *LOC107912464* and *LOC107924144* they have displayed heightened expression in roots. Eight genes encoding 3-hydroxy-3-methylglutaryl-coenzyme-A (HMG-CoA) synthase (HMGS) were identified, and half of those exhibited significantly higher expression in roots. Additionally, within the 3-hydroxy-3-methylglutaryl-coenzyme-A (HMG-CoA) reductase (HMGR) gene family, nine members, including *LOC107939227, LOC107908970, LOC107908973, LOC121203148, LOC107929696, LOC107908971, LOC107908972, LOC107939235,* and *LOC107942380,* displayed increased mRNA abundance in roots. Conversely, family members *LOC107928879, LOC107900974, LOC107920223,* and *LOC107894010* demonstrated broader and moderate expression across leaves, stems, and roots. Mevalonate kinase (MVK) and phosphomevalonate kinase (MVP), responsible for promoting mevalonate (MVA) conversion to mevalonate-5-diphosphate (MVA-5-PP), were encoded by two genes each, all of which exhibited higher mRNA abundances in roots. Diphosphomevalonate decarboxylase (PMD), involved in converting MVA-5-PP to isopentenyl diphosphate (IPP), was represented by nine gene family members. Among these, only *LOC107933302* displayed significant expression, particularly in roots. In the alternative pathway branch leading to IPP production, IPP isomerase (IPPI) they played a pivotal role in the reciprocal transformation between dimethylallyl diphosphate (DMAPP) and IPP. *LOC107909098* and *LOC107928371*, exhibited high expression in roots. Among the four genes identified encoding the essential protein FPS, responsible for converting IPP to FPP. *LOC107905701* and

LOC107905737 showed heightened expression in roots. Conversely, seven of the eleven genes belonging to the family encoding CDNC displayed similar expression levels in roots. The three genes encoding cytochrome P450 enzymes CYP706B1, CYP82D113, and CYP71BE79 also demonstrated higher expression in roots, further underscoring their significance in gossypol synthesis. Additionally, the alcohol dehydrogenase 1 (DH1) gene family tended to have elevated expression in roots. As the key rate-limiting enzyme in gossypol synthesis, 2-oxoglutarate/Fe (II)-dependent dioxygenase-1 (2-ODD-1) is encoded by a gene family comprising up to twenty members, the majority of which exhibited high expression levels in roots. Overall, the genes associated with gossypol synthesis showed comparable and elevated expression levels in the roots of both CKG and CKN, indicating the root (distal) was capable of gossypol synthesis.

2.2 A novel and fine single-cell atlas was reconstructed involving distal root

ScRNA-seq was conducted on the root tips of CKG and CKN to investigate the cell type composition of upland cotton roots. The root tip was selected due to the presence of cells with diverse developmental fates. The scRNA-seq analysis yielded profiles from 11,988 and 11,187 cells, with a mean read depth of 68,826 and 73,777 in CKG and CKN, respectively (Supplemental Table 3). Sequencing identified 55,126 and 55,117 genes across cells in CKG and CKN, respectively, with a median unique molecular identifier (UMI) count per cell of 1,604 in CKG and 1,744 in CKN. Notably, the proportions of Q_{30} bases in the UMIs were consistently above 97.30%, and more than 88.70% of reads were successfully mapped to the genome for both cultivars. Moreover, high correlations were observed between pseudo-bulked scRNA and bulk tissue RNA data for CKG ($\rho = 0.83$) and CKN ($\rho = 0.79$), with the latter correlation being slightly lower (refer to Supplemental Figure 2).

Following the removal of organellar genome-derived transcripts, 10,477 and 8,802 cells for CKG and CKN, respectively, remained for dimensionality reduction. Unsupervised clustering based on shared transcriptomic profiles resulted in 16 distinguishable clusters in planar space, as visualized on a UMAP plot (see Figure 2B). Specific marker genes from Arabidopsis were employed to annotate the cotton root cell clusters. These marker genes were subjected to BLAST using the online tool *Plantcellmarker* to identify their cotton homologs. Marker genes for various cell types, such as atrichoblast (*CYTOCHROME B5 ISOFORM D*, *AtCB5-D*), columella (*ROOT EXPRESSED TONOPLAST INTRINSIC PROTEIN*, *AtDELTA-TIP2*), cortex (*ENCODES A MEMBER OF THE AZI FAMILY OF LIPID TRANSFER PROTEINS*, *AtAZI5*), endodermis (*DIRIGENT-LIKE PROTEIN*), pericycle (*BIFUNCTIONAL INHIBITOR/LIPID-TRANSFER PROTEIN/SEED STORAGE 2S ALBUMIN SUPERFAMILY PROTEIN*), phloem (*EARLY NODULIN-LIKE PROTEIN 9*, *AtENODL9*), protophloem (*YCF1 PROTEIN*, *AtYCF1.1*), quiescent center (*HISTONE 3.1*, *AtH3.1*), stele (*HOMOLOG OF MEDICAGO TRUNCATULA NODULIN21*, *AtUMAMIT5*), trichoblast (*CYTOSOLIC THIOREDOXIN THAT REDUCES DISULFIDE BRIDGES OF TARGET PROTEINS*, *ATH2*), and xylem (*FASCICLIN-LIKE ARABINOGALACTAN PROTEINS 12*, *ATFLA12*) were homologous with *GhCYTb5*, *GhTIPrb75a*, *GhHPS*, *GhD25*, *GhDIR1*, *GhENODL3*, *GhTIC214*, *GhCSE4*, *GhWAT1*, *GhTRXH2*, and *GhFLA12* in upland cotton, respectively (refer to Supplemental Table 4A). These marker genes facilitated the attribution of clusters 1, 2, 3, 4, 8, 9, 10, 11, 12, 13 and 16 to the above cell types. The top ten marker genes for each cotton root cell cluster can be found in Supplemental Table 4B. Additionally, three Arabidopsis genes *EMBRYOGENESIS ABUNDANT (LEA) PROTEIN-LIKE PROTEIN* (*AtLEA*), *GLUTAMINE-DEPENDENT ASPARAGINE SYNTHETASE* (*AtASN1*), and

FIBRE EXPRESSED PROTEIN (*AtFEP*) explicitly expressed in lateral root cap cells (Zhang et al., 2019) were found to have cotton homologs (*GhRCP*, *GhAS1*, and *GhCFP*) belonging to clusters 5, 6, and 7, respectively. Clusters 5, 6, and 7 were identified as lateral root cap cell-2, -1, and -3 (LRC-2, -1, and -3), indicating their derivation from the same cell lineage/population. Similarly, Cluster 8 and 12 exhibited concurrent expression of stele-specific markers, suggesting their association with the stele lineage.

To validate the expression of selected markers, representative genes for endodermis (*LOC107930672*, *GhD25*) and lateral root cap-1 (*LOC107891486*, *GhRCP*) were employed for RNA fluorescence *in situ* hybridization (RNA-FISH). In crosscuts (the left upper panel of Figure 2E), abundant RNA is concentrated in the cell layer corresponding to the Casparian strip, a component of the endodermis near the stele. Overall, the enrichment of *LOC107930672* RNA in the Casparian strip was pronounced closer to the differentiation zone. A similar expression pattern for *LOC107930672* was observed in the roots of CKN (the right upper panel of Figure 2E), as depicted in the left upper panels of Figure 2E. Furthermore, *LOC107891486* was expressed in lateral root cap cells of both CKG (see the left lower panel of Figure 2E) and CKN (see the right lower panel of Figure 2E). As anticipated, these two genes displayed precise and targeted expression in their respective root parts, affirming the reliability of the *in silico* analysis of the scRNA-seq data.

2.3 Computational transcriptional comparison of root cell types between the upland cotton cultivars

Comparing divergent transcriptional patterns distinguishing cell clusters and cell features between the cotton cultivars CKG and CKN is crucial for understanding the cellular distinctions within their root systems. In this study, we assessed single-cell transcriptome profiles of various cell types to elucidate potential differences. Notably,

all 16 identified cell-type clusters exhibited a robust alignment between CKG and CKN, as illustrated in Figure 3A. Nevertheless, the ratios between cell types displayed some variability across the two cultivars, potentially attributed to technical artifacts arising during cell dissociation. Despite this, specific cell types such as atrichoblast, cortex, stele, pericycle, protophloem, lateral root cap-1, and unknown-1 were consistently represented by a substantial number of cells in both CKG and CKN, constituting more than half of the total cell population. Conversely, cell types with lower cell numbers, such as quiescent center cells, accounted for less than 4.5% of cells in both cultivars. These observed cell-type ratios were in concordance with anatomical root features previously measured. Furthermore, a high correlation of ratios between homologous cell types in CKG and CKN was noted, as depicted in Figure 3B, with exceptions observed in cell clusters lateral root cap-3, phloem, and unknown-2. We hypothesize that altered molecular activities within these specific cell types contribute to the observed discrepancies between cultivars. A Venn diagram analysis revealed that over 75% of expressed genes were shared between CKG and CKN in most cell clusters, as illustrated in Figure 3C. Remarkably, the stele and pericycle cell clusters exhibited exceptionally high percentages of shared gene expression, surpassing 80%. In contrast, lateral root cap-3 and phloem cell clusters displayed lower overlaps, falling below 60%, consistent with the observed correlation between homologous cell types (Figure 3B).

To explore the similarity in the expression of genes with molecular functions between the cultivars, we conducted KEGG term enrichment analysis for all cell types (Figure 3D & E). The top 11 enriched KEGG pathways that were enriched in shared expression patterns between cultivars were shown in Figure 3D, and the pathways related to the spliceosome, proteasome, and ubiquitin-mediated proteolysis were enriched for all cell types. These most strongly enriched pathways all take part in

processes that support cotton survival. Conversely, the top 14 pathways showing differentiated expression patterns between cultivars were associated with cysteine and methionine metabolism, α-linolenic acid metabolism, ascorbate and aldarate metabolism, phenylpropanoid biosynthesis, and biosynthesis of various plant secondary metabolites (see Figure 3E). Notably, these pathways tended to be involved in responses to environmental stimuli.

An in-depth exploration of expression divergence between CKG and CKN for core genes involved in phytohormone perception was conducted for each cell type, as illustrated in Supplemental Figure 3, with detailed information on the pertinent core genes provided in Supplemental Table 5. No discernible differential expression was observed in gene families associated with the perception of phytohormones across various root cell types between the two cultivars. Specifically, the gene families *GhGID1* and *GhPYR/GhPYL*, known for their pivotal roles in responding to gibberellin and abscisic acid, respectively, exhibited higher expression levels. Gene families GhNRP1, GhCRE1, and *GhETR*, responsive to salicylic acid, cytokinin, and ethylene, respectively, demonstrated a moderate expression level in most cell types of both cultivars. Notably, negligible gene expression for *GhBRI1* was detected in any cell type, suggesting a limited role for brassinosteroids at this developmental stage of young cotton roots. Notably, the *GhAUX1* gene family displayed relatively high expression in the stele and lateral root cap-1 cell types. While this implies cell type-specific responses to auxin in roots, no evident distinctions between the cultivars were detected. These findings collectively suggest a lack of regulatory distinctions in phytohormone responses in the roots of glandular and glandless cultivars. Combining the previous results, overall, the single-cell transcription profiles of cells from different cell types

between the two cultivars indicate a strong similarity in the transcription of roots at the single-cell level.

2.4 Evolutionarily conserved and divergent scRNA transcription patterns between cotton and Arabidopsis

To assess the conservation of cell type-dependent gene expression patterns between cotton and Arabidopsis, scRNA datasets from the two species were compared (Figure 4A). For this, a merged version of the cotton scRNA-seq datasets from CKG and CKN was considered, along with an Arabidopsis scRNA dataset derived by integrating four published root tip expression datasets (Supplemental Figure 4; Denyer et al., 2019; Jean-Baptiste et al., 2019; Wendrich et al., 2020; Zhang et al., 2019). The scRNA-seq datasets of cotton and Arabidopsis were combined and aligned based on the expression of orthologous genes. Subsequently, cells were clustered through dimension reduction, and homologous cell types were identified based on shared cluster membership. The integrated cell sets were divided into 15 main shared clusters (Figure 4B), and further grouped into 11 primary populations (Figure 4C). *Pearson*'s correlation coefficients indicated homology between cell populations of the two species (Figure 4D). Notably, gene expression in cell populations identified as meristem cells demonstrated a relatively high correlation between cotton and Arabidopsis. Similar patterns were observed for cell populations designated as stele (xylem). These cell types showing low transcriptional divergence appear to be more highly conserved between species. Another set of cell types, including stele (phloem), cortex, endodermis, and lateral root cap exhibited positive correlations between cotton and Arabidopsis. Given the role of stele cells in transporting water and nutrients in plants, the conserved expression patterns of genes in these cells likely confer widespread benefits for plant survival. Conversely, other cell types, such as atrichoblast and trichoblast, presented

much weaker positive correlations in expression patterns across species, indicating low expression conservation. Further analysis of the expression of orthologous marker genes for individual cell types revealed additional patterns of similarities and differences between cotton and Arabidopsis roots (Figure 4E). Features of the top 550 genes with divergent expression across species were depicted in a heat map (Figure 4F; Supplemental Table 6), and these differentially expressed genes (DEGs) were characterized by studying enrichment of GO (Figure 4G) and KEGG (Figure 4H) annotations. The most highly enriched GO terms were GO: 0005507 (copper ion binding), GO: 0016616 (oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor), GO:0016614 (oxidoreductase activity, working on CH-OH group of donors), and GO:0030170 (pyridoxal phosphate binding). The most highly enriched KEGG pathways included ath01200 (Carbon metabolism), ath00620 (Pyruvate metabolism), ath00010 (Glycolysis /Gluconeogenesis), and ath00350 (Tyrosine metabolism). The GO terms and KEGG pathways enriched among the DEGs suggest relatively strong between-species divergence in developmental plasticity and responses to environmental fluctuations.

2.5 ScRNA analysis of gossypol biosynthesis in cotton root

The expression of fourteen gene families, extensively studied for their involvement in gossypol synthesis, was evaluated in individual cell types and compared between CKG and CKN (Figure 5). Gene families responsible for the gossypol synthesis pathway (distinct from the MVA pathway) exhibited robust and stable expression, including genes implicated in *2-ODD-1*, *CYP71BE79*, *CYP82D113*, *DH1*, *CYP706B1*, and *CDNC*. Conversely, the gene families responsible for the MVA pathway, leading up to synthesizing the precursor FPP, exhibited variable expression patterns. Notably, the *HMGR* family showed relatively low expression across all cell

types, whereas the MVP and MVK gene families were expressed in fewer cells, possibly owing to their smaller membership. These findings were consistent with the bulk RNA-seq data (Figure 1C). The IPPI gene family demonstrated more widespread expression across all cell types. Overall, the gene families involved in gossypol synthesis showed distinct expression patterns in both two cultivars across individual cell types, as can be deduced from variations in color depth within the graphical display (Figure 5). To further identify if specific cell types contributed to gossypol production, we merged the scRNA-seq data for CKG and CKN (Supplemental Figure 5). The merged data revealed higher expression of the *2-ODD-1*, *DH1*, *CDNC*, *MVK*, *HMGR*, and *HMGS* gene families in LRC-1 and LRC-2 cell types. Additionally, three cytochrome gene families, *CYP71BE79*, *CYP82D113*, and *CYP706B1*, showed higher expression levels in all cell types, with slightly elevated expression in LRC-1 and LRC-2 cells. FPS, ACAT, and IPPI gene families displayed similar expression trends across all cell types, indicating indiscriminate expression. These results showed that LRC-1 and -2 were the most crucial cell types in cotton roots where gossypol production occurs.

Furthermore, the gene expression features in the lateral root cap cell types between cultivars were investigated. 1237, 2057, and 2020 genes were differently expressed between CKG and CKN in cell types LRC-1, LRC-2, and LRC-3, respectively (Supplemental Figure 6A). Consequently, the top divergent expressed genes were screened and mapped to the KEGG pathways. The top-ten up-regulated and down-regulated KEGG enrichments were presented as Supplemental Figure 6B. In all three cells types, the KEGG annotation of divergent gene expression ($P<0.05$). Between the two cultivated varieties, the differentially expressed genes (DEGs) of LRC-1, LRC-2, and LRC-3 were 1237, 2057, and 2020, respectively. Compared to the total number of protein-coding genes in the upland cotton genome, which exceeds 70,000, these

proportions are very small. It can be observed that there is a high degree of conservation in gene expression among the three cell types between the two cultivated varieties. It is worth noting that the number of differentially expressed genes in LRC-2 and LRC-3 was approximately 1.6 times that of LRC-1. This larger difference in gene expression in these two cell types compared to LRC-1 may suggest a higher level of functional differentiation in LRC-2 and LRC-3. Thus, merging the single-cell data for both cultivars in individual cell types is necessary and reliable, as mentioned above.

2.6 ScRNA-seq reveals distinct developmental trajectories of lateral root cap cells

The scRNA-seq technique facilitates the examination of the continuous trajectory of cell differentiation during root development, starting from cells in intermediate states that can be captured (Liu et al., 2021; Trapnell, 2015). Analogous to the ability of animal stem cells to further differentiate into diverse cell types, the shoot apical meristem (SAM) and root apical meristem (RAM) are acknowledged as two stem cell niches (SCNs) in plants, with the quiescent center (QC) constituting the domain of root stem cells. *Pseudo*-time analysis was conducted on all 16 cell clusters using *Monocle* (Figure 6A). In conjunction with the cell types illustrated in Figure 2B, QC cells were situated at the initial point in pseudo-time, from which the developmental trajectory of the cotton root emerged akin to a ‘radical tree’. Notably, the cell types pericycle, stele, xylem, protophloem, and phloem, formed a significant branch we define as ‘general stele’; columella and lateral root caps constituted a ‘top root tip’ branch. According to the *pseudo*-time results, the cell types atrichoblast and endodermis, and, similarly, trichoblast and cortex, appeared closer in ‘*pseudo*-time distance’, suggesting the differentiation of the epidermal cell layer into two primary cell types.

The findings above demonstrate that the cell types LRC-1 and LRC-2 play a crucial role in synthesizing gossypol. Consequently, we further investigated the

developmental trajectory of lateral root cap cell types utilizing the pseudo-time approach. The consistency in cell differentiation between the two cultivars was evident in this *pseudo*-time analysis, indicating robust conservation in cell developmental trajectories (Figure 6B). Specifically, the developmental trajectory of lateral root cap cell types commenced at LRC-1, traversed a bifurcation point indicating a developmental state transition, and then extended gradually along two divergent cell developmental paths encompassing LRC-2 (designated as branch 1) and LRC-3 (defined as branch 2). Subsequent projection of gene expression patterns of six core gene families in gossypol synthesis onto the *pseudo*-time developmental trajectories (Figure 6C and Supplemental Figure 7), revealed a pronounced expression of these genes along the pre-branching trajectory and branch 1. Moreover, expression patterns of these genes along the two cell differentiation trajectories exhibited a high degree of conservation between CKG and CKN. Following adjustment for pseudo-time order, 296 differentially expressed genes (DEGs) ($q<1e^{-4}$) were identified between cells post branches 1 and 2 and the pre-branching part of the trajectory. Visualization of these DEGs through a head map (Figure 6D, Supplemental Table 7), resulted in their categorization into three clusters based on distinct expression patterns among cells involved in lateral root cap development. Overall, gene expression patterns diverged significantly after cells traversed the branch point, with genes highly expressed within branch 1 being enriched in KEGG pathways such as terpenoid backbone biosynthesis (crucial for gossypol production), sesquiterpenoid and triterpenoid biosynthesis, lysine degradation, and glutathione metabolism. These pathways commonly play vital roles in plant responses to environmental stimuli. Conversely, the genes expressed prominently along branch 2 (representing cell type LRC-3), were enriched in pathways associated with energy metabolism, including pentose and glucoronate

interconversions, inositol phosphate metabolism, biosynthesis of nucleotide sugars, and starch and sucrose metabolism. The robust enrichment of these pathways implies active cell growth and development. Genes in cluster 3 exhibited high expression levels before the branching point and were associated with energy metabolism pathways such as starch and sucrose metabolism and alkaloid biosynthesis. These expression patterns suggested that the genes predominantly expressed before the branching point contributed to the involvement of cells of type LRC-1 in both growth and plant responses to the environment.

Subsequently, we further validated the developmental trajectories of the lateral root cap cells. We conducted dimensionality reduction on the scRNA data of the lateral root cap cells, and distinct separation of cell types LRC-1, LRC-2, and LRC-3 into consecutive bins along an arc was evident (Supplemental Figure 8A). The analysis captured a linear dependence relationship from LRC-1 to LRC-2 cells. scRNA velocity calculations and projections across the cells revealed that of these three cell types was calculated and projected across the cells revealed that LRC-3 cells exhibited a higher degree of differentiation and precipitated towards the end of the cell developmental trajectory of the lateral root cap cells (Supplemental Figure 8B). Finally, *CytoTrace* was employed to predict the differentiation ability of lateral root cap cell lineages. The results indicated a more profound differentiation in the forward developmental process from LRC-1 to LRC-2 than from LRC-1 to LRC-3 (Supplemental Figure 8C). Among the top twenty genes that most strongly correlated with cell differentiation according to *CytoTrace*, *LOC107891486 (GhRCP)* (Supplemental Figure 8C, the right panel), identified as the marker gene for cell type LRC-2, occupied the number one position.

2.7 Identification and verification of novel transcript factors governing the gossypol synthesis

The transcription factors (TFs) potentially influencing gossypol synthesis in upland cotton roots were identified systematically. Initially, 5022 TFs were retrieved from the complete upland cotton genome. The top 2000 genes were subsequently screened, adjusting for the pseudo-time trajectory of gossypol synthesis in lateral root cap cells, encompassing the three cell types. A list of 115 TFs was generated by intersecting the two gene sets above (Supplemental Table 8A). The 115 TFs were further categorized into 31 groups, encompassing major types such as MYB, ERF, bZIP, NAC, bHLH, LBD, WRKY, C2H2, and GRAS. Each of these major types comprises more than four type members (refer to Figure 7A). The core TF reciprocal interactions were subsequently analyzed and presented in Figure 7B, revealing the significant involvement of the C2H2 TF type in the TF-TF interaction network. Specifically, the C2H2 TF activated various TF types, including LBD, NAC, G2-like, bZIP, SBP, WRKY, HD-ZIP, MYB, bHLH, ERF, HSF, NF-YC, and ARF, while inhibiting others like ERF, MYB, and HB-other (see Figure 7B).

The 115 transcription factors (TFs) and 69 genes associated with gossypol synthesis mentioned above were analyzed. A comprehensive examination revealed 178 TF-gene interaction events (with a weight greater than 0.05, as detailed in Supplemental Table 8B). These interactions encompassed 53 TF gene families and 69 target genes directly implicated in the synthesis of gossypol (refer to Supplemental Table 8B for specific details). *GhZAT12*, a C2H2 TF type, emerged as a key player in these interactions, being involved in most events (up to 76 times), followed by *GhbZIP53* (bZIP, seven times), GhERF114 (ERF, six times), GhNAC87 (NAC, six times), and others (refer to Figure 7C). This highlights the central role of *GhZAT12* in the TF

regulation of gossypol synthesis. The TF-gene interaction network (see Figure 7D) demonstrated that MYB, bZIP, and ERF TF types had the highest members, with eight, seven, and seven, respectively. Additionally, apart from TF type C2H2 (such as *GhZAT12*), gene families *2-ODD-1, CYP82D113, DH1, HMGS,* and *MVP* were primarily regulated by bZIP, HD-bZIP, ERF, bZIP, and NAC, respectively. Furthermore, TF types bZIP, MYB, and NAC tended to handle the HMGR family, while the PMD family exhibited a greater propensity for interaction with TF types bZIP, ERF, and MYB. Notably, multiple TF types were involved in regulating the CDNC and FPS gene families.

Several genes, such as *GhCGF2, GhCGP1, GhCGF3 (GhPGF), GhCGF1*, and *GhSPGF*, along with TFs, were identified as contributors to the development of pigment glands in cotton (refer to Supplemental Table 8C) (Gao et al., 2020; Janga et al., 2019; Ma et al., 2016; Zang et al., 2021). Based on our single-cell data, it was observed that the TFs *GhCGF1* (bHLH, *LOC107953911*), *GhCGF2* (NAC, *LOC107936158* and *LOC107936796*), *GhCGF3/GhPGF* (bHLH, *LOC107942873* and *LOC107947206*), and *GhSPGF* (GRAS, *LOC107931299*) exhibited no expression in cell-types LRC-1, LRC-2, and LRC-3 (Figure 7E).

For the functional validation of putative transcription factors (TFs) implicated in cotton root gossypol synthesis, we selected the two highest-ranked TFs, *GhZAT12* and *GhbZIP53*, for a Virus-Induced Gene Silencing (VIGS) assay. Figure 8A illustrates that Seedling TRV2: GhCLA1 exhibited albino leaves, confirming the successful application of VIGS in this batch of plants. As depicted in Figure 8B, the relative expression of *GhZAT12* decreased in the roots, stems, and leaves of TRV2:GhZAT12 seedlings compared to those of the organs of TRV2:00 seedlings. Similarly, the relative expression of *GhbZIP53* showed a decrease in the same organs of TRV2:GhbZIP53

seedlings compared to TRV2:00 seedlings. These results demonstrate a consistent reciprocal trend in the replicates of both cultivars, CKG and CKN. Moreover, the gossypol contents of TRV2: GhZAT12 and TRV2:GhbZIP53 seedlings were reduced compared to those of TRV2:00 seedlings in the organs of cultivar CKG (see Figure 8C). The same trends were observed in the roots of CKN. These findings indicate that suppressing the expression of TFs *GhZAT12* and *GhbZIP53* can effectively inhibit gossypol synthesis in both glandular and glandless cotton. Once again, we examined the gland density changes in the aboveground sections of cotton seedlings TRV2:00, TRV2: GhZAT12, and TRV2:GhbZIP53. Compared to those of seedling TRV2:00, no statistically significant decrease ($P<0.05$) in gland density was observed in the leaves and stems of seedlings TRV2: GhZAT12 and TRV2:GhbZIP53 within the CKG group (Figure 8D). Apparently, the *GhZAT12* and *GhbZIP53* do not participate in the development process of glands in upland cotton seedlings.

3. Discussion

Single-cell sequencing provides a powerful tool for deciphering cellular heterogeneity in gene expression.

Roots are essential for the survival of cotton. In addition to their fundamental roles of providing a physical anchor and absorbing nutrients and water, cotton roots also play specific and crucial roles in synthesizing a remarkably diverse group of secondary metabolites, including gossypol (Smith, 1961; Zhang et al., 2021). Cells of eukaryotic multicellular organisms exhibit inherent heterogeneity (Zhang et al., 2019). The multifunction observed in cotton roots results from the heterogeneous cells and the intricate regulation of gene expression. Personalized proteins orchestrate various biological activities. Capturing gene expression at the individual cell or cell type level

563 is crucial for dissecting the intricate gene regularity networks involved in root
564 development and response to external environments (Liu et al., 2021).

565 The development of single-cell RNA sequencing (scRNA-seq) technology allows
566 for the dissection of cellular heterogeneity, the identification of new cell types, and the
567 revelation of developmental trajectories at a high-throughput level (Shulse et al., 2019).
568 Initially, the application of single-cell RNA sequencing (scRNA-seq) in plants was
569 confined to a few species, such as Arabidopsis (Denyer et al., 2019; Farmer et al., 2021;
570 Zhang et al., 2019), primarily due to technical barriers associated with the preparation
571 of high-quality protoplasts, which were hampered by the presence of a cell wall. Using
572 scRNA-seq tools to investigate plant-related issues has expanded to encompass a
573 broader range of species and organs. In the past two years, single-cell sequencing has
574 become a valuable tool for investigating critical aspects of cotton biology. This
575 application extends to the examination of diverse aspects such as the developmental
576 processes of fiber cells originating from ovules (Qin et al., 2022; Wang et al., 2023),
577 the progression of various cell types derived from anthers (Li et al., 2024), the
578 maturation of gland cells within cotyledons, true leaves, and stems, and the examination
579 of gland cell synthesis of gossypol (Lin et al., 2023; Long et al., 2023; Sun et al., 2023),
580 the fate of stem cell niche in cotton lateral meristem (Zhu et al., 2023), secondary
581 growth development of induced callus stem cells derived from stems, and to analyze
582 the response of Asian cotton (*Gossypol arboreum*) roots to salt stress (Li et al., 2023).

583 In this study, we obtained sufficient high-quality cotton root protoplasts by
584 improving the isolation method based on Arabidopsis and rice protocols. 23,175 single
585 cells were sequenced, and 19,279 credible cells were retained after multiple filtering
586 steps. The large number of cells with excellent scRNA-seq data enhances the reliability
587 of subsequent assays. Typically, a successful single-cell research study should include

at least 8,000 cells. The study met this criterion with the cultivar CKN, which had a cell count of 8,802. Based on this, nearly all major cell types were successfully identified through in silico analysis combined with novel cluster-specific marker genes. Using the defined cell types, the cotton root was reconstructed, and common taproot structures were adjusted accordingly. This is the first time a comprehensive single-cell RNA sequencing (scRNA-seq) atlas has been released for the root of upland cotton (*Gossypium hirsutum*) that contributes to major cotton production worldwide. Between the CKG and CKN cultivars, a notable inherent divergence was observed with respect to the presence of glands on the aerial parts. However, no significant differences were identified in root cell-type heterogeneity. Our scRNA-seq data confirmed a high positive correlation among homologous cell types, even when comparing non-homologous cell types between the two cultivars based on gene expression. This suggests that biological functions are conserved within individual homologous cell types, and multiple cell types are collectively involved in specific biological activities. A higher degree of divergence was observed across species compared to inter-cultivar differences. The shared gene expression displayed more outstanding distinctiveness in homologous cell types between cotton and Arabidopsis. This pattern is consistent with findings in other species, such as rice and Arabidopsis (Liu et al., 2021), as well as woodland strawberry and Arabidopsis (Bai et al., 2022).

The synthesis of gossypol exhibits cell-type specificity in cotton roots

Three sub-cell types, namely LRC-1, -2, and -3, were discovered in the lateral root cap. LRC-1 and columella shared specific marker genes, suggesting that LRC-1 is likely the part adjacent to the columella in the root tip. Through RNA-FISH assay, two representative cotton marker genes, *LOC107930672 (GhD25)* and *LOC107891486 (GhRCP)* were located in the endodermis and lateral root cap zone of the cotton root,

613 respectively, supporting the credibility of the predicted cell types in this study. In our
614 *in silico* analysis, *LOC107930672* was primarily expressed in the endodermis, as some
615 signal noise was observed in cell type unknown-2 (Figure 2C). The RNA fluorescence
616 signal was detected in a few of the cortex and stele cells (weaker signal) and the
617 endodermis (the upper panels of Figure 2E). Combining these findings, the cells
618 categorized into cell type unknown-2 likely involve heterogeneous kinds.
619 Simultaneously, the fluorescence signal captured the inner cells of the lateral root cap,
620 indicating that LRC-2 and LRC-3 are probably located in the outward direction from
621 LRC-1. In this study, we have identified distinct expression patterns among various
622 genes associated with the gossypol biosynthetic pathway across different cell types in
623 cotton roots. Specifically, compared to other cell types, all genes involved in the
624 gossypol biosynthetic pathway demonstrated significantly elevated expression levels in
625 the lateral root cap (LRC) cell type. Noteworthy is the observation that within the LRC
626 cell type, the subtypes LRC-1 and LRC-2 exhibited particularly heightened expression
627 levels. These results suggest a cell type-specific bias in gossypol synthesis within cotton
628 roots, with LRC-1 and LRC-2 playing pivotal roles in facilitating the gossypol
629 biosynthetic process. This was the first time the specific cell lines that contributed to
630 gossypol synthesis were identified successfully.

631 The concept of '*pseud*o-time' was introduced to align cells based on their
632 incremental changes in the transcriptome, establishing a temporal order along a
633 trajectory corresponding to a biological process. Our studies analyzed lateral root cap
634 cells using the *pseudo*-time inference tool *Monocle 2*, an unsupervised algorithm that
635 predicted the differentiation of cell types LRC-2 and LRC-3 from LRC-1. Examination
636 of representative gene expression associated with gossypol synthesis across the pseudo-
637 time order further indicated the successful transition of cell fate from LRC-1 to LRC-

2, leading to gossypol synthesis. The robustness of this prediction was validated using similar trajectory analysis tools, namely *CytoTrace* and *scVelo*. These findings support the accurate prediction of cellular differentiation and gene expression patterns along the identified trajectory.

The regulation of gossypol synthesis involves multiple transcription factors controlling multiple genes

Identifying transcription factors (TFs) can accelerate elucidating regulatory networks underlying plant metabolic responses to developmental processes or environmental adaptation (Franco-Zorrilla and Solano, 2017). Until now, only a limited number of TFs have been reported to play roles in regulating gossypol biosynthesis. In this study, we systematically identified the TF regulatory networks involved in gossypol biosynthesis through gene co-expression analysis in cell types LRC-1 and LRC-2. Our findings revealed that a single C2H2 zinc finger protein, ZINC FINGER PROTEIN 12 (*GhZAT12, LOC107931921*), participated in over half of the interplay events among the TF-gene interaction network, suggesting a central core role. Generally, C2H2 zinc finger proteins are known to play roles in plant development and growth (Wu, 2002; Yin et al., 2020). Furthermore, C2H2 zinc finger proteins have been documented to function in improving plant stress resistance by maintaining cellular ionic homeostasis and adjusting metabolism to stresses (Han et al., 2021). We also identified another potential hub C2H2 zinc finger protein, SENSITIVE TO PROTON RHIZOTOXICITY 1 (*GhSTOP1, LOC107899819*), which was reported to regulate various genes critical for adaptation to different stresses (Koyama et al., 2021). This indicates that gossypol synthesis is related to environmental cues, as gossypol exists as a phytoalexin and antioxidant metabolite in cotton. Additionally, we identified several other vital TFs involved in gossypol synthesis, including *GhbZIP53* (a member of the

family BASIC LEUCINE ZIPPER), *GhERF14* (a member of the family ETHYLENE RESPONSE FACTOR), *GhNAC87* (a member of the family NAM, ATAF1, ATAF2, CUC2), and *GhMYB73* (a member of the family MYELOBLASTOSIS). Notably, these TF types have commonly been shown to contribute to secondary metabolism and plant stress response (Liu et al., 2015; Nuruzzaman et al., 2013; Wang et al., 2022; Wang et al., 2021; Wu et al., 2022).

Inhibiting the expression of the *GhZAT12* and *GhbZIP53* genes significantly reduces the content of gossypol in VIGS plants, indicating the involvement of these two transcription factors in the process of gossypol synthesis. It is evident that by confirming the expression characteristics of genes in specific cell types along the target metabolic pathway, it is feasible to screen transcription factors coupled with the expression characteristics of genes related to the metabolic pathway from these specific cell types. Identifying functional transcription factors acting on the metabolic pathway from these transcription factors is achievable. By identifying transcription factors involved in gossypol synthesis, we provide a reference model for the identification of metabolism-related transcription factors in plants.

The gossypol biosynthesis is independent of gland morphogenesis in cotton

The ability to synthesize gossypol is inherent in the roots of both glandular and glandless cotton. This phenomenon was first reported in 1961 and has since been confirmed to occur in both types of cotton by culturing excised cotton roots. (Smith, 1961). In the early stages, roots were considered the primary organizing organ for synthesizing gossypol in cotton. One main reason was that genes related to gossypol synthesis were highly expressed in the roots, regardless of the presence of glands in glandular or glandless cotton. In the aboveground parts, the expression levels of these genes were extremely low, even though the expression in the aboveground parts of

glandular cotton was slightly higher than in glandless cotton. Glands were believed to only store gossypol, based on the detection of gossypol in the aboveground parts of glandular cotton but not in the aboveground parts of glandless cotton. However, until today, direct evidence of gossypol transport within the cotton plant is lacking, hindered by the lack of real-time tracking technology for gossypol inside cells. The organ and transport mode of systemic gossypol synthesis in cotton has been debated in the academic community. Recently, Pandeya (2023) demonstrated through grafting methods that the source of gossypol in seeds is not transported from the roots to the above-ground parts. Again, Lin (2023) identified gland-secreting cells in leaves directly through single-cell methods and proved their ability to synthesize gossypol. Specially, Zhang et al. (2024) explored the dual functionality of GhCGF3/GoPGF and determined that pigment glands serve as the primary synthetic site for gossypol in the aerial components of cotton. This indicates that glands are also essential organs for gossypol synthesis. Glands are distributed on the surface of the above-ground parts of the cotton plant, and the proportion of cells in above-ground tissues is tiny. This may be why it is challenging to detect expression in conventional RNA sequencing due to the very low content of the target tissue in overall sampling.

In the development of glands in cotton, crucial genes play a controlling role, and these typical key genes have been successively cloned and analysed. Essentially, suppressing the expression of these genes leads to a reduction or disappearance of glands, subsequently decreasing the content of gossypol in cotton (Janga et al., 2019; Ma et al., 2016; Sun et al., 2023). Currently, all cloned genes related to glands belong to transcription factors. We examined the expression of these typical transcription factors in cotton lateral root cap cells. We found that the transcription factors related to gland synthesis in cell types contributing primarily to gossypol synthesis were not

expressed. This suggests that the synthesis of gossypol is independent of gland development, and these transcription factors are not present in the identified set of transcription factors interacting with genes involved in gossypol synthesis in this study. Furthermore, there was no loss or reduction of glands in plants where we identified and functionally verified the re-expression inhibition of two transcription factors. This again demonstrates the independence of gossypol synthesis to gland development. At the level of root cell types, we found no significant differences in gossypol synthesis, overall gene expression levels based on cell types, and the expression of hormone-responsive genes between glandular and glandless cotton. This indirectly supports the notion that gossypol synthesis is independent of the gland development process. In summary, we speculate that the decrease in plant gossypol content due to the inhibition of gland development gene expression may not be because these genes affect the gossypol synthesis process, but rather due to the decrease or disappearance of glands, influencing the presence of gossypol in the plant.

5. Materials and methods

Plant materials, pigment gland phenotyping, and gossypol assay

Two upland cotton (*Gossypol hirsutum* L.) cultivars, CKG (glandular Coker-312) and CKN (glandless Coker-312) were employed in this study. Seed germination and seedling cultivation procedures followed Mei et al. (2022), and 10-day old seedlings were prepared for use after germination. The density of pigment glands involved in cotyledon leaves, stems, and roots, was assessed between the two cultivars using a stereo microscope and *Image Pro Plus 6.0* software. The pigment gland phenotypes in the two species were checked by evaluating their density in cotyledon leaves, stems, and roots. Quantification of gossypol was referred to Gao et al. (2022). Specifically, a standard gossypol solution was prepared by dissolving 0.01g of (+)/(-)-gossypol in 10

738 ml of acetonitrile. A gradient of 0.25, 0.50, 1.00, 2.50, 5.00, 10.00, 25.00, 40.00, 50.00,
739 and 100.00 mg/L (+)/(-)-gossypol was also prepared. Calibration curves for (+)/(-)-
740 gossypol were constructed according to the above preparations, as depicted in
741 Supplemental Figure 1. HPLC analysis was performed using an Agilent 1100 system
742 (Agilent, Santa Clara, USA) with an auto-sampler and UV detection. The HPLC
743 conditions included a C18 column (250 mm × 4.6 mm, 5 μm, Dikma, Richmond Hill,
744 USA) as the stationary phase, acetonitrile/0.2% phosphoric acid (80/20, v/v) as the
745 mobile phase, a 10 μL injection volume, and a flow rate of 1.0 ml/min. The UV
746 wavelength was 238 nm, and the retention times of (+)/(-)-gossypol were 7.9 and 14.1
747 min. The standard gossypol used was purchased from *Solarbio Science & Technology*
748 *Co., Ltd (Beijing, China)*. Furthermore, the quantification of gossypol in VIGS
749 seedlings was carried out using fresh frozen samples due to constraints on the
750 availability of mutant seedlings.

751 ***Bulk RNA-seq***

752 cotyledons, stems, and root tips from the 10-day seedlings of CKG and CKN were
753 harvested for bulk RNA-seq. Each sample contained three replicates. The total RNA
754 extraction referred to (Mei et al., 2022). Illumina's NEBNext®Ultra™ RNA Library
755 Prep Kit was employed to construct the library. The cDNA library was sequenced and
756 produced 150 bp paired-end reads by the Illumina NovaSeq Platform. The raw reads
757 were processed by *Perl* scripts, and the clean reads were mapped to the reference of the
758 upland cotton genome (version number: GCA_007990345.1 v2.1) using *HISAT2*.
759 DEGs were identified and screened out by *DESeq* with the threshold of $|\log_2 FC| \geqslant 1$,
760 as well as FDR $\leqslant$ 0.01. The lg (FPKM + 1) and Pearson correlation coefficients of bulk
761 and scRNA-seq data were calculated in *R*.

Protoplast isolation, library construction, and pre-processing of scRNA-seq data

The cotton root single cells were prepared based on the protoplast isolation. In detail, cotton root tips less than 0.5 cm (not longer than 1.0 cm exactly) were sampled and cut into 1-2 mm pieces. Next, the small pieces and enzyme cocktail solution were transferred to the 10 ml centrifuge tube and vacuumed to make the pieces drop. 10 ml enzyme cocktail was made from 3% cellulase RS, 1.5% macerozyme R10, 0.1% pectolyase Y-23, 10 mM MES (pH 5.7), 0.6 M mannitol, 0.1% BSA and 10 mM $CaCl_2$. After that, the incubating tubes were kept within the table concentrator at dark, 28 ℃ and 80 rpm for 4 h. The digesting solutions were microscopy checking at each hour. The well-digested protoplasts were filtered by cell strainers (40 μm in diameter, Falcon, Cat No./ID: 352340) four times. Next, the protoplasts were concentrated at 2000 rmp for 5 min and then washed with 8% mannitol for four times. The protoplast viability trypan blue staining (Supplemental Figure 9). The concentration of protoplasts was counted via a hemocytometer. The ratio of viable protoplast was more than 85%, and the concentration reached approximately 1800 cells/μL, allowing to next step operation. In our study, three biological replicates were conducted, and three were pooled before sorting. The scRNA-seq libraries were constructed on the 10x Chromium 3' Single Cell platform according to the manufacturer's instructions (10x Genomics, Pleasanton, CA, USA) (Liu et al., 2021). A batch of 16 000 cells was loaded on a microfluidic chip by Chromium Single Cell 30 v3 Reagent Kit. The structure of single-cell GEM (gel beads in emulsion) was formed after single-cell suspension and beads with a bell barcode. Agilent 2100 Bioanalyzer and Illumina Hiseq2000 Sequencer were used for quantification of the cDNA library and sequencing, respectively.

Cell Ranger 3.0.1 (10x Genomics) was used to process the raw scRNA-seq data. We first built the cotton genome reference (GCA_007990345.1 v2.1) using the

‘cellranger mkref’ function with genome and GFF annotation files obtained from NCBI as arguments. Sequencing data were aligned and quantified using the ‘cellranger count’ function against the reference genome, and the gene expression matrices were generated. The raw gene expression matrices were used for further downstream analyses (Farmer et al., 2021).

scRNA-seq data doublet detection, normalization, and clustering

DoubletFinder (v.2.0.3) was used to identify doublets (Mcginnis et al., 2019). The proportion of every cell's artificial nearest neighbors (pANN) was computed. The doublet threshold of pANN was defined according to the nExp value to generate final doublet predictions. The resultant cells only flagged as singlets were used for downstream analysis. The upcoming analysis was conducted via the Seurat package (v.4.0.3) in R (Mcginnis et al., 2019). We performed quality control procedures by filtering out low-quality cells and genes and normalizing data with the ‘Normalize Data’ function (LogNormalize method, scaling factor of 10,000). The low-quality cells and genes were filtered adjusting to the criteria: (1) The cells with the number of expressed genes fall into the interval between 400 and 5000; (2) the total number of detected mRNA (nCount_RNA) is less than 15000 (3) the percentage of mitochondrial UMIs less than 10%. The top 2000 highly variable genes were selected for further analysis. We scaled data with ‘Scale Data’ function, performed PCA analysis with ‘Run PCA’ function (50 principal components), and determined optimal PCs by ‘Elbow Plot’ function. Using Harmony, we removed batch effects between samples (Korsunsky et al., 2019). The further downstream analysis were based on harmony dimensionality reductions. We then applied ‘FindNeighbors’ and ‘FindClusters’ functions to construct the SNN graph and cluster cells based on Louvain. Cell clusters were visualized by ‘Run UMAP’ function. Mainly, Dimensionality reduction was also performed based on

Harmony embedding using PHATE. This dimensionality-reduction method is capable of capturing both local and global structures (Moon et al., 2019), for further analysis. The cluster-enriched genes (marker genes) were identified using 'Find All Markers' function using the default parameters in Seurat.

RNA florescent in situ hybridization analysis

For fluorescent *in situ* hybridization (FISH) analysis of *LOC107930672* and *LOC107891486*, probes of those were synthesized by Bersinbio Company (Guangzhou, China). Firstly, cell slides were fixed with 4% paraformaldehyde for 20 min. The slides were washed with PBS twice and dehydrated with ethanol. After that, slides were denatured, added with 20 μL hybridization reaction solution (2 μL probes + 18 μL hybridization reaction), and hybridized at 37 °C overnight. After that, the slides were washed with 25% formamide/2 × Saline Sodium Citrate (SSC) twice. The test on integrity and auto-fluorescence, as well as probes, refers to Supplemental Figure 10 and Supplemental Table 9. The antibodies were utilized for FISH analysis of these RNAs. Cells after fixation and permeabilization were incubated with the primary antibody overnight and then incubated with the secondary antibody for 1 h. Finally, the slides were stained with DAPI and subjected to fluorescent signal detection using Zeiss LSM800 confocal laser microscopy (Zeiss, Germany).

Comparison of scRNA-seq data in inter and intra-species

We first generated a comprehensive Arabidopsis cell atlas for a more scientific robust reference and further inter-species comparison. We collected Arabidopsis root scRNA-seq data from four labs (Denyer et al., 2019; Jean-Baptiste et al., 2019; Wendrich et al., 2020; Zhang et al., 2019). We performed robust QC and doublets removal for each dataset. We merged all the datasets and corrected the batch effects using the R package Harmony (Korsunsky et al., 2019). In general, we obtained a total

of 34747 cells after QC and doublets removal. The rest of the Seurat workflow with similar parameters was performed as mentioned above.

To perform the cross-species comparison, we first identified orthologous genes between cotton and Arabidopsis using OrthoFinder (Emms and Kelly, 2019). The protein sequences of cotton and Arabidopsis were downloaded from NCBI and Phytozome, which were used for the OrthoFinder was used to cluster similar protein sequences between rice and Arabidopsis by an all-against-all strategy with BLASTP (e-value: 1-e^5). Only the one-to-one orthologs were kept for further downstream analysis. We first integrated the Arabidopsis single-cell atlas we generated (excluding the dataset from Schiefelbein lab) with the cotton single-cell atlas using R package LIGER with UINMF (Kriebel and Welch, 2022). We selected shared and unshared orthologous using a variable gene threshold of 0.1 and a threshold of 0.3 for the Arabidopsis data and 0.3 for the cotton data. We then integrated by function 'optimize ALS()' with parameters set to K = 30, λ = 5. For quantile normalization, we used the Arabidopsis dataset as a reference and used knn_k = 50. For Louvain community detection, we used a resolution of 0.1. After the cluster, we obtained 15 clusters in total. We then calculated cross-species pairwise correlation between cell types following a previous study (Tosches et al., 2018). In detail, we first calculated the average gene expression level of each cell type using the 'Average Expression' function in Seurat. 'gene-specificity index' equation was used to transform the gene expression matrices into gene specificity matrices. Pairwise Spearman correlations between Arabidopsis and cotton were obtained based on the gene specificity matrices. For significance analysis of correlation coefficients, gene expression values were shuffled 1000 times across cell types. $P<0.05$ indicates significance.

Trajectory analysis and RNA velocity

We used R package *Monocle 2* (v.2.20.0) (Qiu et al., 2017) and *CytoTRACE* (v.0.3.3)(Gulati et al., 2020) to perform the trajectory and developmental inference. For the *Monocle 2* analysis, we first extracted only the cells from the LRC-1, LRC-2, and LRC-3 clusters. The single-cell data were converted to a Cell Data Set object first. *Monocle 2* was used to reconstruct the *pseudo*-time developmental trajectory in R. The 'estimate Size Factors' and 'estimate Dispersions' functions were used to standardize cell differences. Genes were selected using the function 'detect Genes' with the parameter 'min_expr' set to 0.1. Highly variable genes along the *pseudo*-time were identified using the 'differential Gene Test' function of *Monocle 2*. 'DDRTree' was used for the dimension reduction. Genes varying across branch points were analyzed using the BEAM function. *CytoTRACE* analysis was implemented with default parameters. PHATE embeddings were utilized in the *CytoTRACE* analysis.

Velocyto package was employed to infer the *RNA velocity* (La Manno et al., 2018). Spliced and unspliced mRNA counts were generated using *Velocyto* run10x (v.0.17.17) on scRNA-seq cell ranger outputs. The output loom files of each sample were combined using 'loompy'. The merged loom file was read into *Velocyto* via R (v.0.6). We then extracted the spliced and unspliced mRNA expression matrix of the LRC-1, LRC-2, and LRC-3 clusters. To filter out lowly variable genes, data normalization and variable gene detection were performed using function 'scv.pp.filter_and_normalize()' with parameters 'min_shared_counts' set to 30 and 'n_top_genes' set to 2000. We used stochastic modeling of RNA velocity in function scv.tl.velocity. Results were projected based on UMAP embedding.

KEGG enrichment analysis

Differentially expressed genes and marker genes in common between CKG and CKN were obtained first. *Cluster Profiler* was used for pathway enrichment analysis (Yu et al., 2012). The pathway enrichment with high statistical significance was annotated to a biological process. Significantly different values were determined with a False Discovery Rate-corrected P value < 0.05.

Analysis of transcription factors and Virus-induced gene silence (VIGS) assay

Completed upland cotton transcription facts (TFs) were retrieved from PlantTFDB (http://planttfdb.cbi.pku.edu.cn). The top 2000 genes were screened as adjusting the *pseudo*-time of lateral root cap cells. The 115 TFs, falling into the intersection between the above two sets, were employed for TF-TF and TF-gene interaction analysis, as the genes targeted by the TFs were involved in gossypol synthesis. The interaction analysis was performed via *GENIE3 (Gene Network Inference with Ensemble of trees)* (Huynh-Thu et al., 2010) and visualized via *Cytoscape (Vision 3.9.1)* (Shannon et al., 2003).

To suppress the expression of *GhZAT12* (*LOC107931921*) and *GhbZIP53* (*LOC107945994*) through Virus-Induced Gene Silencing (VIGS), about 300 bp fragments of these genes were amplified from cotton root cDNA. The sequences of the two cDNA and corresponding amplification primer pairs referred to Supplemental Table 10A. Subsequently, these fragments were incorporated into the pTRV2 vector using *BamHI* and *KpnI* restriction sites. The pTRV1, pTRV2, and the resulting pTRV2 VIGS vectors were then introduced into the Agrobacterium strain *GV3101* via electroporation (*MicroPulser, Bio-Rad, California, United States*). TRV2:GhCLA1 was utilized as a positive control. The VIGS assays followed established protocols (Ma et al., 2016). To examined the expression of silenced genes, RT-qPCR were employed

and the primer pairs referring to Supplemental Table 10B. The proceeds of RT-qPCR according to Mei et al. (2022).

Conflict of interest: All authors declare that they have no conflict of interest

Data availability

The high-throughput scRNA-seq data in this research have been deposited in the National Center for Biotechnology Information (NCBI) SRA database with Bioproject No. PRJNA852944 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA852944), and the bulk RNA-seq in GEO datasets with accession No. GSE206663 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE206663).

Supplemental Information

Supplemental information is available at Plant Communication Online.

Funding

This research was supported by grants from Genetically Modified Organisms Breeding Major Project of China (2020ZX08009-10B), Zhejiang Major Scientific and Technological Project of Agricultural (Upland crop) Breeding (2021C02064-6), Natural Science Foundation of Zhejiang Province (LQ22C02002), Natural Science Foundation of Hubei Province (2021CFB002), China Postdoctoral Science Foundation (2021M690633).

Author Contributions

LM and SJ conceived the study, designed the experiments, and prepared this manuscript. CY, BL and XZ performed the *in silico* analysis. JZ and HL performed the VIGS validation. PZ, XZ and YZ prepared protoplasts. LL, JZ and HY quantified the gossypol. HY and YZ performed RNA-FISH analysis. XW and SCG polished the writing. All authors have finally proofread the manuscript.

Acknowledgment

The authors declare no competing interests.

References

Bai, Y., Liu, H., Lyu, H., Su, L., Xiong, J., and Cheng, Z.M. (2022). Development of a single-cell atlas for woodland strawberry (*Fragaria vesca*) leaves during early *Botrytis cinerea* infection using single-cell RNA-seq. Hortic. Res. **9**: uhab055.

Bawa, G., Liu, Z., Yu, X., Qin, A., and Sun, X. (2022). Single-Cell RNA Sequencing for Plant Research: Insights and Possible Benefits. Int. J. Mol. Sci. **23**:4497.

Benedict, C.R., Liu, J., and Stipanovic, R.D. (2006). The peroxidative coupling of hemigossypol to (+)- and (−)-gossypol in cottonseed extracts. Phytochemistry **67**:356-361.

Benedict, C.R., Martin, G.S., Liu, J., Puckhaber, L., and Magill, C.W. (2004). Terpenoid aldehyde formation and lysigenous gland storage sites in cotton: variant with mature glands but suppressed levels of terpenoid aldehydes. Phytochemistry **65**:1351-1359.

Birnbaum, K., Shasha, D.E., Wang, J.Y., Jung, J.W., Lambert, G.M., Galbraith, D.W., and Benfey, P.N. (2003). A Gene Expression Map of the *Arabidopsis* Root. Science **302**:1956-1960.

Brady, S.M., Orlando, D.A., Lee, J., Wang, J.Y., Koch, J., Dinneny, J.R., Mace, D., Ohler, U., and Benfey, P.N. (2007). A High-Resolution Root Spatiotemporal Map Reveals Dominant Expression Patterns. Science **318**:801-806.

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.

Dodou, K. (2005). Investigations on gossypol: past and present developments. Expert Opin. Investig. Drugs **14**:1419-1434.

Effenberger, I., Zhang, B., Li, L., Wang, Q., Liu, Y., Klaiber, I., Pfannstiel, J., Wang, Q., and Schaller, A. (2015). Dirigent Proteins from Cotton (*gossypium* sp.) for the Atropselective Synthesis of Gossypol. Angew. Chem. Int. Ed. **54**:14660-14663.

Emms, D.M., and Kelly, S. (2019). Orthofinder: phylogenetic orthology inference for comparative genomics. Genome Biol. **20**.

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.

Franco-Zorrilla, J.M., and Solano, R. (2017). Identification of plant transcription factor target sequences. Biochim. Biophys. Acta Gene Regul. Mech. **1860**:21-30.

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.

Gao, W., Zhu, X., Ding, L., Xu, B., Gao, Y., Cheng, Y., Dai, F., Liu, B., Si, Z., and Fang, L., et al. (2022). Development of the engineered "glanded plant and glandless seed" cotton. Food Chem. (Oxf) **5**:100130.

Gulati, G.S., Sikandar, S.S., Wesche, D.J., Manjunath, A., Bharadwaj, A., Berger, M.J., Ilagan, F., Kuo, A.H., Hsieh, R.W., and Cai, S., et al. (2020). Single-cell transcriptional diversity is a hallmark of developmental potential. Science **367**:405-411.

Han, G., Li, Y., Qiao, Z., Wang, C., Zhao, Y., Guo, J., Chen, M., and Wang, B. (2021). Advances in the Regulation of Epidermal Cell Development by C2H2 Zinc Finger Proteins in Plants. Front. Plant Sci. **12**.

Heinstein, P.F., Herman, D.L., Tove, S.B., and Smith, F.H. (1970). Biosynthesis of gossypol. Incorporation of mevalonate-2-14C and isoprenyl pyrophosphates. J. Biol. Chem. **245**:4658-4665.

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

Huynh-Thu, V.A., Irrthum, A., Wehenkel, L., and Geurts, P. (2010). Inferring regulatory networks from expression data using tree-based methods. PloS One. **5**:e12776

Koyama, H., Wu, L., Agrahari, R.K., and Kobayashi, Y. (2021). STOP1 regulatory system: Centered on multiple stress tolerance and cellular nutrient management. Mol Plant **14**:1615-1617.

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.

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.

Kriebel, A.R., and Welch, J.D. (2022). UINMF performs mosaic integration of single-cell multi-omic datasets using nonnegative matrix factorization. Nat. Commun. **13**.

Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. Nat. Methods **16**:1289-1296.

La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lönnerberg, P., and Furlan, A., et al. (2018). RNA velocity of single cells. Nature **560**:494-498.

Li, J., Yuan, D., Wang, P., Wang, Q., Sun, M., Liu, Z., Si, H., Xu, Z., Ma, Y., and Zhang, B., et al. (2021). Cotton pan-genome retrieves the lost sequences and genes during domestication and selection. Genome Biol. **22**.

Li, P., Liu, Q., Wei, Y., Xing, C., Xu, Z., Ding, F., Liu, Y., Lu, Q., Hu, N., and Wang, T., et al. (2023). Transcriptional landscape of cotton roots in response to salt stress at single-cell resolution. Plant Commun. **5**: 100740.

Li, Y., Ma, H., Wu, Y., Ma, Y., Yang, J., Li, Y., Yue, D., Zhang, R., Kong, J., and Lindsey, K., et al. (2024). Single‐Cell Transcriptome Atlas and Regulatory Dynamics in Developing Cotton Anthers. Adv. Sci. (Weinh) **11**.

Lin, J., Chen, L., Wu, W., Guo, X., Yu, C., Xu, M., Nie, G., Dun, J., Li, Y., and Xu, B., et al. (2023). Single-cell RNA sequencing reveals a hierarchical transcriptional regulatory network of terpenoid biosynthesis in cotton secretory glandular cells. Mol Plant **16**:1990-2003.

Lin, J.L., Fang, X., Li, J.X., Chen, Z.W., Wu, W.K., Guo, X.X., Liu, N.J., Huang, J.F., Chen, F.Y., and Wang, L.J., et al. (2023). Dirigent gene editing of gossypol enantiomers for toxicity-depleted cotton seeds. Nat. Plants **9**:605-615.

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

Liu, J., Osbourn, A., and Ma, P. (2015). MYB Transcription Factors as Regulators of Phenylpropanoid Metabolism in Plants. Mol Plant **8**:689-708.

Long, L., Xu, F.C., Wang, C.H., Zhao, X.T., Yuan, M., Song, C.P., and Gao, W. (2023). Single-cell transcriptome atlas identified novel regulators for pigment gland morphogenesis in cotton. Plant Biotechnol. J. **21**:1100-1102.

Lusas, E.W., and Jividen, G.M. (1987). Glandless cottonseed: A review of the first 25 years of processing and utilization research. J. Am. Oil Chem. Soc. **64**:839-854.

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

Maryam, H., Ali, Z., Saddique, M.A.B., and Nawaz, F. (2022). Gh*CDNC* and Gh*CYP706B1* genes mediate gossypol biosynthesis in upland cotton. Mol. Biol. Rep. **49**:4919-4928.

McGinnis, C.S., Murrow, L.M., and Gartner, Z.J. (2019). Doubletfinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. Cell Syst. **8**:329-337.

Mei, L., Zhu, Y., Liu, H., Hui, Y., Xiang, J., Daud, M.K., Jiang, S., and Zhu, S. (2022). Genome-wide characterization on MT family and their expression in response to environmental cues in upland cotton (*Gossypium hirsutum* L.). Int. J. Biol. Macromol. **198**: 54-67.

Moon, K.R., van Dijk, D., Wang, Z., Gigante, S., Burkhardt, D.B., Chen, W.S., Yim, K., Elzen, A.V.D., Hirn, M.J., and Coifman, R.R., et al. (2019). Visualizing structure and transitions in high-dimensional biological data. Nat. Biotechnol. **37**:1482-1492.

Nuruzzaman, M., Sharoni, A.M., and Kikuchi, S. (2013). Roles of NAC transcription factors in the regulation of biotic and abiotic stress responses in plants. Front. Microbiol. **4**.

Pandeya, D., Campbell, L.M., Puckhaber, L., Suh, C., and Rathore, K.S. (2023). Gossypol and related compounds are produced and accumulate in the aboveground parts of the cotton plant, independent of roots as the source. Planta **257**.

Qin, Y., Sun, M., Li, W., Xu, M., Shao, L., Liu, Y., Zhao, G., Liu, Z., Xu, Z., and You, J., et al. (2022). Single-cell RNA-seq reveals fate determination control of an individual fibre cell initiation in cotton (*Gossypium hirsutum*). Plant. Biotechnol. J. **20**:2372-2388.

Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H.A., and Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell trajectories. Nat. Methods **14**:979-982.

Rong, J., Feltus, F.A., Liu, L., Lin, L., and Paterson, A.H. (2010). Gene copy number evolution during tetraploid cotton radiation. Heredity (Edinb) **105**:463-472.

Scheffler, J.A., Romano, G.B., and Blanco, C.A. (2012). Evaluating host plant resistance in cotton (*Gossypium hirsutum* L.) with varying gland densities to tobacco budworm (*Heliothis virescens* F.) and bollworm (*Helicoverpa zea* Boddie) in the field and laboratory. Agricultural Sciences **03**:14-23.

Serrano-Ron, L., Perez-Garcia, P., Sanchez-Corrionero, A., Gude, I., Cabrera, J., Ip, P., Birnbaum, K.D., and Moreno-Risueno, M.A. (2021). Reconstruction of lateral root formation through single-cell RNA sequencing reveals order of tissue initiation. Mol. Plant **14**:1362-1378.

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.

Shulse, C.N., Cole, B.J., Ciobanu, D., Lin, J., Yoshinaga, Y., Gouran, M., Turco, G.M., Zhu, Y., O Malley, R.C., and Brady, S.M., et al. (2019). High-Throughput Single-Cell Transcriptome Profiling of Plant Cell Types. Cell Rep. **27**:2241-2247.

Smith, F.H. (1961). Biosynthesis of Gossypol by Excised Cotton Roots. Nature **192**:888-889.

Stipanovic, R.D., Puckhaber, L.S., and Bell, A.A. (2006). Ratios of (+)- and (−)-Gossypol in Leaves, Stems, and Roots of Selected Accessions of *Gossypium hirsutum* Var.*marie galante* (Watt) Hutchinson. J. Agric. Food Chem. **54**:1633-1637.

Sun, Q., Cai, Y., Xie, Y., Mo, J., Yuan, Y., Shi, Y., Li, S., Jiang, H., Pan, Z., and Gao, Y., et al. (2010). Gene expression profiling during gland morphogenesis of a mutant and a glandless upland cotton. Mol. Biol. Rep. **37**:3319-3325.

Sun, Y., Han, Y., Sheng, K., Yang, P., Cao, Y., Li, H., Zhu, Q., 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.

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. **103**:18054-18059.

Tian, X., Ruan, J., Huang, J., Yang, C., Fang, X., Chen, Z., Hong, H., Wang, L., Mao, Y., and Lu, S., et al. (2018). Characterization of gossypol biosynthetic pathway. Proc. Natl. Acad. Sci. **115**.

Tosches, M.A., Yamawaki, T.M., Naumann, R.K., Jacobi, A.A., Tushev, G., and Laurent, G. (2018). Evolution of pallium, hippocampus, and cortical cell types revealed by single-cell transcriptomics in reptiles. Science **360**:881-888.

Trapnell, C. (2015). Defining cell types and states with single-cell genomics. Genome Res. **25**:1491-1498.

Wang, D., Hu, X., Ye, H., Wang, Y., Yang, Q., Liang, X., Wang, Z., Zhou, Y., Wen, M., and Yuan, X., et al. (2023). Cell-specific clock-controlled gene expression program regulates rhythmic fiber cell growth in cotton. Genome Biol. **24**:49.

Wang, H., Zhang, Y., Norris, A., and Jiang, C. (2022). S1-bZIP Transcription Factors Play Important Roles in the Regulation of Fruit Quality and Stress Response. Front Plant Sci. **12**.

Wang, X., Niu, Y., and Zheng, Y. (2021). Multiple Functions of MYB Transcription Factors in Abiotic Stress Responses. Int. J. Mol. Sci. **22**:6125.

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

Wu, L.C. (2002). ZAS: C2H2 zinc finger proteins involved in growth and development. Gene Expr. **10**: 137-152.

Wu, Y., Li, X., Zhang, J., Zhao, H., Tan, S., Xu, W., Pan, J., Yang, F., and Pi, E. (2022). ERF subfamily transcription factors and their function in plant responses to abiotic stresses. Front Plant Sci. **13**.

Ye, K., Teng, T., Yang, T., Zhao, D., and Zhao, Y. (2023). Transcriptome analysis reveals the effect of grafting on gossypol biosynthesis and gland formation in cotton. BMC Plant Biol. **23**.

Yin, J., Wang, L., Zhao, J., Li, Y., Huang, R., Jiang, X., Zhou, X., Zhu, X., He, Y., and He, Y., et al. (2020). Genome-wide characterization of the C2H2 zinc-finger genes in *Cucumis sativus* and functional analyses of four *CsZFPs* in response to stresses. Bmc Plant Biol. **20**.

Yu, G., Wang, L., Han, Y., and He, Q. (2012). Clusterprofiler: an R Package for Comparing Biological Themes among Gene Clusters. OMICS **16**:284-287.

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.

Zhang, C., Liu, X., Song, Y., Sun, Z., Zhang, J., Wu, H., Yang, Y., Wang, Z., and He, D. (2022). Comparative Transcriptome Analysis Reveals Genes Associated with the Gossypol Synthesis and Gland Morphogenesis in *Gossypium hirsutum*. Genes (Basel) **13**:1452.

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

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

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**.

Zhu, X., Xu, Z., Wang, G., Cong, Y., Yu, L., Jia, R., Qin, Y., Zhang, G., Li, B., and Yuan, D., et al. (2023). Single-cell resolution analysis reveals the preparation for reprogramming the fate of stem cell niche in cotton lateral meristem. Genome Biol. **24**:1-194

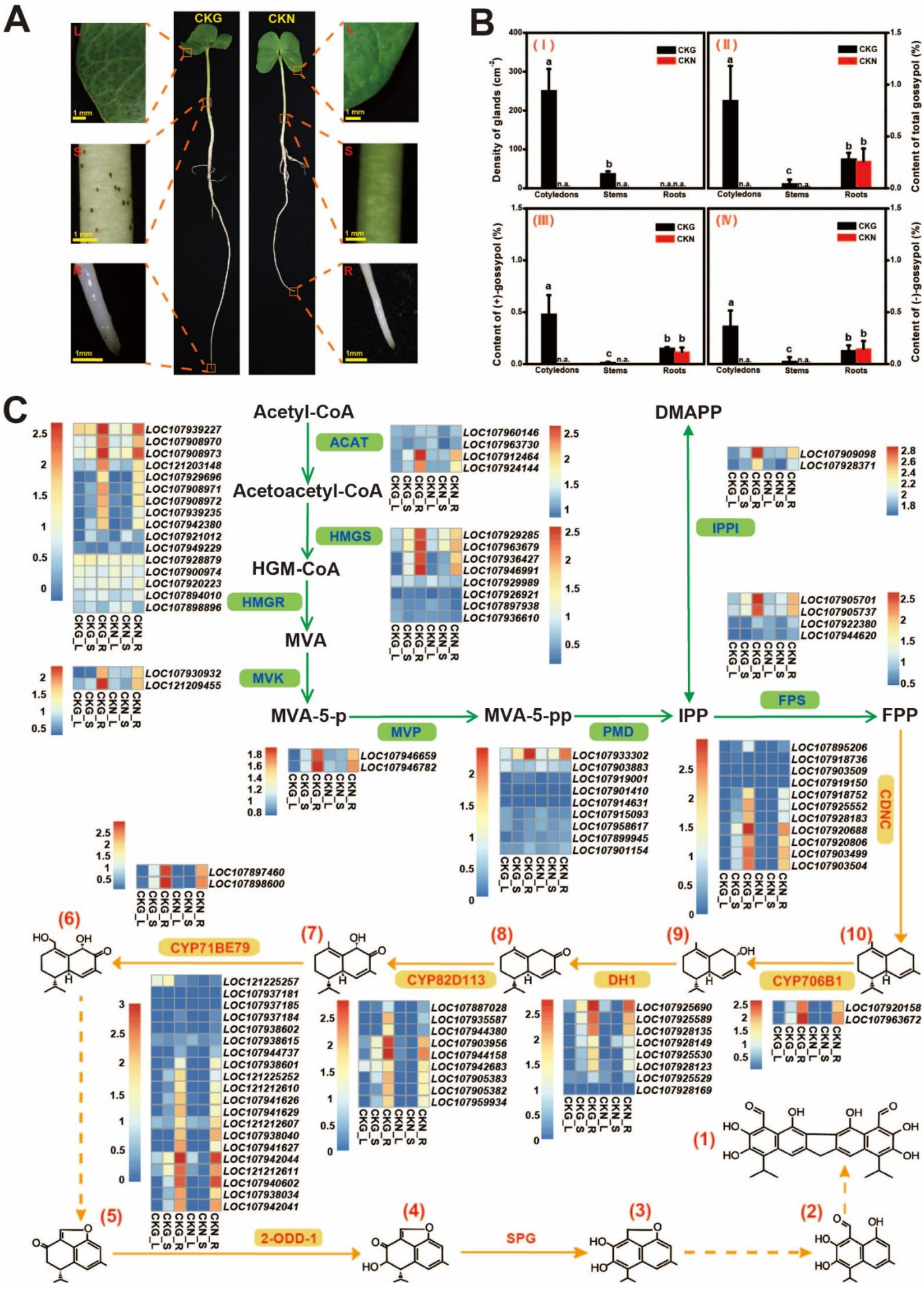

A
CKG
CKN
L
S
R
1 mm
1mm
B
(Ⅰ)
(Ⅱ)
(Ⅲ)
(Ⅳ)
CKG
CKN
Density of glands (cm-2)
Content of total gossypol (%)
Content of (+)-gossypol (%)
Content of (-)-gossypol (%)
Cotyledons
Stems
Roots
n.a.
C
Acetyl-CoA
ACAT
Acetoacetyl-CoA
HMGS
HGM-CoA
HMGR
MVA
MVK
MVA-5-p
MVP
MVA-5-pp
PMD
IPP
DMAPP
IPPI
FPS
FPP
CDNC
CYP706B1
DH1
CYP82D113
CYP71BE79
2-ODD-1
SPG
(1)
(2)
(3)
(4)
(5)
(6)
(7)
(8)
(9)
(10)
LOC107939227
LOC107908970
LOC107908973
LOC121203148
LOC107929696
LOC107908971
LOC107908972
LOC107939235
LOC107942380
LOC107921012
LOC107949229
LOC107928879
LOC107900974
LOC107920223
LOC107894010
LOC107898896
LOC107960146
LOC107963730
LOC107912464
LOC107924144
LOC107929285
LOC107963679
LOC107936427
LOC107946991
LOC107929989
LOC107926921
LOC107897938
LOC107936610
LOC107930932
LOC121209455
LOC107946659
LOC107946782
LOC107897460
LOC107898600
LOC107909098
LOC107928371
LOC107905701
LOC107905737
LOC107922380
LOC107944620
LOC107933302
LOC107903883
LOC107919001
LOC107901410
LOC107914631
LOC107915093
LOC107958617
LOC107899945
LOC107901154
LOC107895206
LOC107918736
LOC107903509
LOC107919150
LOC107918752
LOC107925552
LOC107928183
LOC107920688
LOC107920806
LOC107903499
LOC107903504
LOC107920158
LOC107963672
LOC107925690
LOC107925589
LOC107928135
LOC107928149
LOC107925530
LOC107928123
LOC107925529
LOC107928169
LOC107887028
LOC107935587
LOC107944380
LOC107903956
LOC107944158
LOC107942683
LOC107905383
LOC107905382
LOC107959934
LOC121225257
LOC107937181
LOC107937185
LOC107937184
LOC107938602
LOC107938615
LOC107944737
LOC107938601
LOC121225252
LOC121212610
LOC107941626
LOC107941629
LOC121212607
LOC107938040
LOC107941627
LOC107942044
LOC121212611
LOC107940602
LOC107938034
LOC107942041

Figure 1 Synthesis of gossypol in upland cotton (*Gossypol hirsutum* L).

(A) Morphologies of the seedlings of cultivars CKG and CKN. The speckles, distributed onto the cotyledon leaf and stem surfaces, are pigment glands. The yellow bars denote a scale of 1 mm in length. CKG and CKN represent the glandular and glandless cultivars Coker 312, respectively.

(B) The density of glands and the contents of total/ (+)-gossypol/ (-)-gossypol involved in the tissues of cultivars CKG and CKN. The values showed as mean ± SD (n=6). The different letters above the error bars denote the statistical significance of differences ($P<0.05$). n. a. represents that the value is extremely low to zero or cannot be detected.

(C) The abundance of mRNAs involved in gossypol synthesis. The MVA and gossypol pathways are indicated by green and orange arrows, respectively. Number (1), (2), (3), (4), (5), (6), (7), (8), (9) and (10) denote, in order, Gossypol, Hemigossypol, Desoxyhemigossypol, 3-hydroxy-furocalamen-2-one, Furocalamen-2-one, 8,11-dihydroxy-7-keto-δ-cadinene, 8-hydroxy-7-keto-δ-cadinene, 7-keto-δ-cadinene, 7-hydroxy-(+)-δ-cadinene, (+)-δ-cadinene. The substrates are abbreviated as follows, Acetyl-CoA, Acetyl-coenzyme-A; Acetoacetyl-CoA, Acetoacetyl-coenzyme-A; HGM-CoA, 3-hydroxy-3-methylglutaryl-coenzyme-A; MVA, Mevalonate; MVA-5-P, Mevalonate-5-phospate; MVA-5-PP, Mevalonate-5-diphospate; IPP, Isopentenyl diphosphate; DMAPP, Dimethylallyl diphosphate; FPP, Farnesyl diphosphate. Additionally, the enzymes are abbreviated as follows, ACAT, acyl CoA-cholesterol acyltransferase; HMGS, 3-hydroxy-3-methylglutaryl-coenzyme-A (HMG-CoA) synthase; HMGR, 3-hydroxy-3-methylglutaryl-coenzyme-A (HMG-CoA) reductase; MVK, mevalonate kinase; MVP, phosphomevalonate kinase; PMD, diphosphomevalonate decarboxylase; IPPI, IPP isomerase; FPS, FPP synthase; CDNC, (+)-δ-cadinene synthase; DH1, alcohol dehydrogenase 1; CYP706B1, CYP82D113, CYP71BE79, three cytochrome P450s in the gossypol biosynthesis pathway; 2-ODD-1, 2-oxoglutarate/Fe(II)-dependent dioxygenase-1. The letters L, S, and R after the cultivar name denote abbreviations for cotyledon leaves, stems, and roots, respectively. The colors from blue to red depended on the abundance of mRNAs, which ranged from less to more. Each replicate had mixed for three seedling individuals. Three replicates were counted in the biological samples.

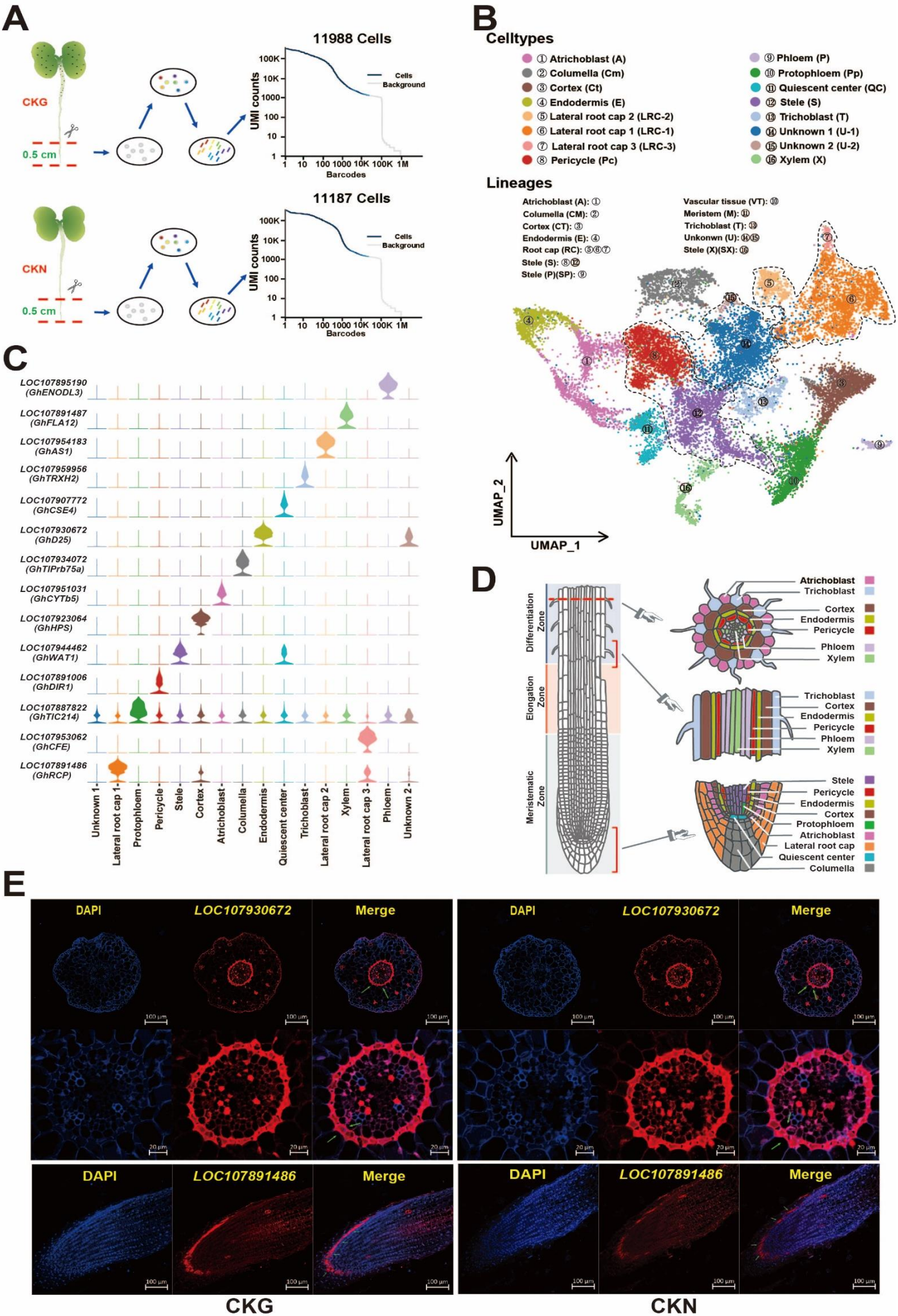

A
CKG
0.5 cm
11988 Cells
UMI counts
Cells
Background
Barcodes
CKN
0.5 cm
11187 Cells
UMI counts
Cells
Background
Barcodes
B
Celltypes
① Atrichoblast (A)
② Columella (Cm)
③ Cortex (Ct)
④ Endodermis (E)
⑤ Lateral root cap 2 (LRC-2)
⑥ Lateral root cap 1 (LRC-1)
⑦ Lateral root cap 3 (LRC-3)
⑧ Pericycle (Pc)
⑨ Phloem (P)
⑩ Protophloem (Pp)
⑪ Quiescent center (QC)
⑫ Stele (S)
⑬ Trichoblast (T)
⑭ Unknown 1 (U-1)
⑮ Unknown 2 (U-2)
⑯ Xylem (X)
Lineages
Atrichoblast (A): ①
Columella (CM): ②
Cortex (CT): ③
Endodermis (E): ④
Root cap (RC): ⑤⑥⑦
Stele (S): ⑧⑫
Stele (P)(SP): ⑨
Vascular tissue (VT): ⑩
Meristem (M): ⑪
Trichoblast (T): ⑬
Unkonwn (U): ⑭⑮
Stele (X)(SX): ⑯
UMAP_2
UMAP_1
C
LOC107895190 (GhENODL3)
LOC107891487 (GhFLA12)
LOC107954183 (GhAS1)
LOC107959956 (GhTRXH2)
LOC107907772 (GhCSE4)
LOC107930672 (GhD25)
LOC107934072 (GhTIPrb75a)
LOC107951031 (GhCYTb5)
LOC107923064 (GhHPS)
LOC107944462 (GhWAT1)
LOC107891006 (GhDIR1)
LOC107887822 (GhTIC214)
LOC107953062 (GhCFE)
LOC107891486 (GhRCP)
Unknown 1
Lateral root cap 1
Protophloem
Pericycle
Stele
Cortex
Atrichoblast
Columella
Endodermis
Quiescent center
Trichoblast
Lateral root cap 2
Xylem
Lateral root cap 3
Phloem
Unknown 2
D
Differentiation Zone
Elongation Zone
Meristematic Zone
Atrichoblast
Trichoblast
Cortex
Endodermis
Pericycle
Phloem
Xylem
Trichoblast
Cortex
Endodermis
Pericycle
Phloem
Xylem
Stele
Pericycle
Endodermis
Cortex
Protophloem
Atrichoblast
Lateral root cap
Quiescent center
Columella
E
DAPI
LOC107930672
Merge
DAPI
LOC107930672
Merge
DAPI
LOC107891486
Merge
DAPI
LOC107891486
Merge
CKG
CKN

Figure 2 Analysis of heterogeneous cell types in cotton root.

(A) Overview of the single-cell library preparation process. The distal roots in length of 0.5 cm were sampled.

(B) UMAP visualization of 16 cell clusters derived from 12 cell lineages. Different colors represent the various cell clusters and a black, dotted line surrounds cells of the same lineage. Each colored dot denotes an individual cell. The letters in parentheses indicate the cell types detected: LRC, Lateral root cap; Pp, Protophloem; Pc, Pericycle; S: Stele; Ct, Cortex; A, Atrichoblast; Cm, Columella; E, Endodermis; QC, Quiescent center; T, Trichoblast; X, Xylem; P, Phloem; U, Unknown.

(C) The violin plot exhibits the expression patterns of representative cell-type marker genes in each of 16 cell clusters.

(D) A simulation of the spatial structuring of cell types constituting the cotton root. The cotton root structure was modeled after the Arabidopsis tap root using the online tool *figshare* (Bouché, 2017). The colors that denote the corresponding cell clusters are in accordance with panel **B**.

(E) RNA fluorescent in situ hybridization (FISH) of *LOC107930672 (GhD25)* and *LOC107891486 (GhRCP)* in cotton seedling roots. DAPI, denotes 4', 6-diamidino-2-phenylindole. The green arrows highlight the targeted cells and scale marks can be found in the lower right corner of each panel.

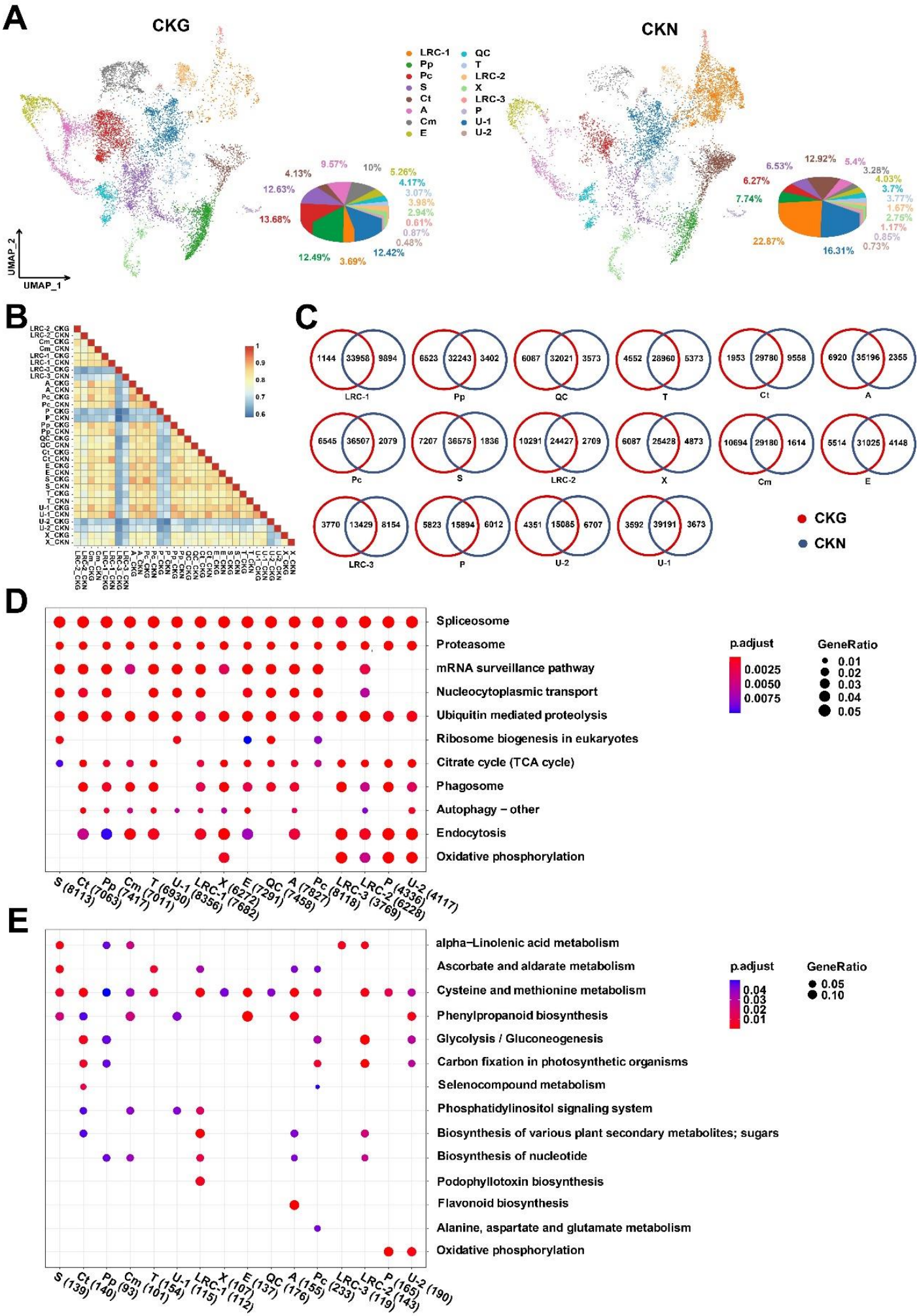

A
CKG
CKN
LRC-1
Pp
Pc
S
Ct
A
Cm
E
QC
T
LRC-2
X
LRC-3
P
U-1
U-2
UMAP_2
UMAP_1
B
C
CKG
CKN
D
Spliceosome
Proteasome
mRNA surveillance pathway
Nucleocytoplasmic transport
Ubiquitin mediated proteolysis
Ribosome biogenesis in eukaryotes
Citrate cycle (TCA cycle)
Phagosome
Autophagy – other
Endocytosis
Oxidative phosphorylation
p.adjust
GeneRatio
E
alpha−Linolenic acid metabolism
Ascorbate and aldarate metabolism
Cysteine and methionine metabolism
Phenylpropanoid biosynthesis
Glycolysis / Gluconeogenesis
Carbon fixation in photosynthetic organisms
Selenocompound metabolism
Phosphatidylinositol signaling system
Biosynthesis of various plant secondary metabolites; sugars
Biosynthesis of nucleotide
Podophyllotoxin biosynthesis
Flavonoid biosynthesis
Alanine, aspartate and glutamate metabolism
Oxidative phosphorylation

Figure 3 Analysis of the diversity and conservation of gene expression within cell clusters of different types between cultivars CKG and CKN.

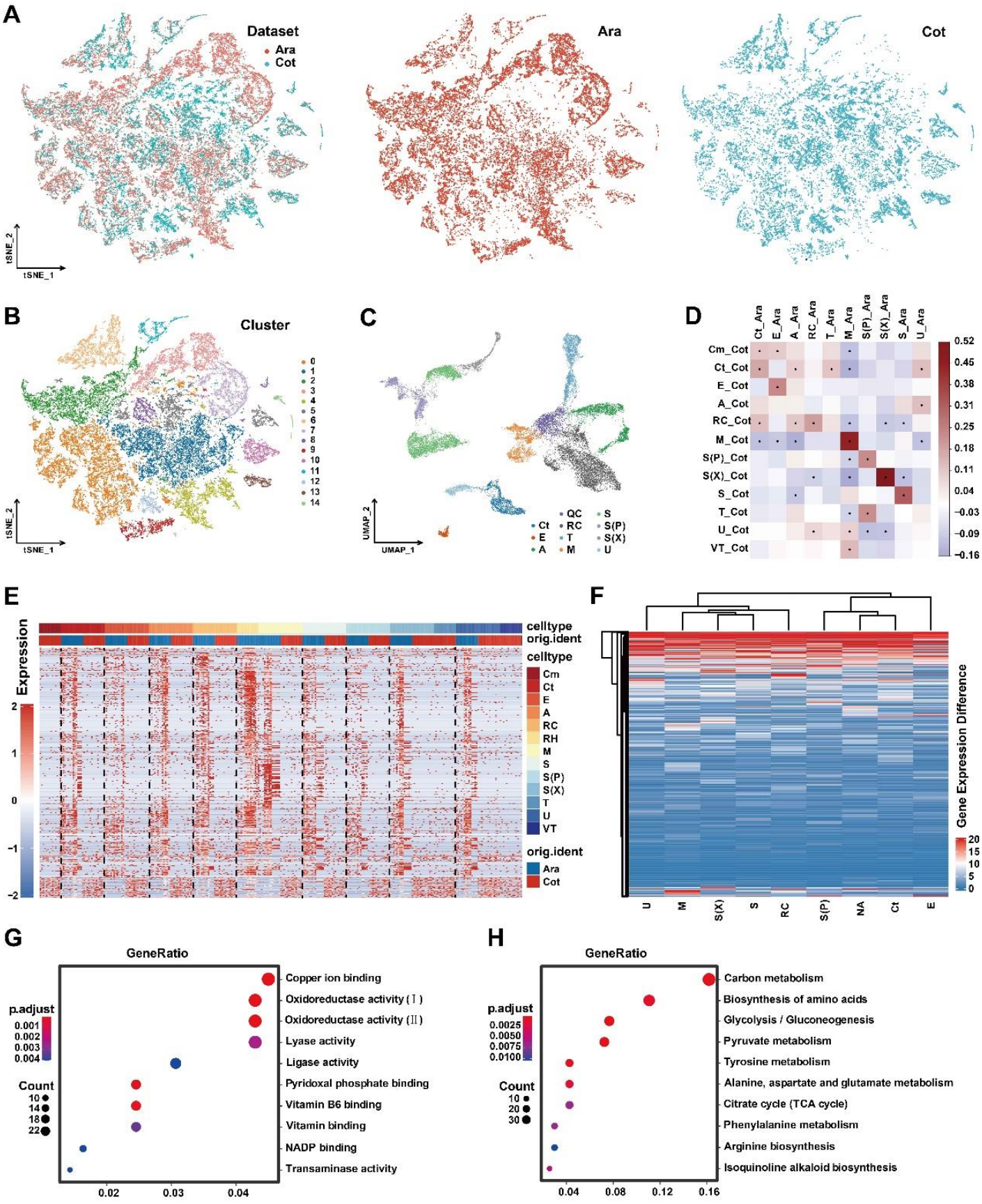
(A) A UMAP visualization and pie chart of cell types between CKG and CKN. Each colored dot denotes an individual cell. The letters in parentheses denote different cell types: LRC, Lateral root cap; Pp, Protophloem; Pc, Pericycle; S: Stele; Ct, Cortex; A, Atrichoblast; Cm, Columella; E, Endodermis; QC, Quiescent center; T, Trichoblast; X, Xylem; P, Phloem; U, Unknown.

(B) A heat map exhibiting genes that show highly correlated expression between CKG and CKN among cell types.

(C) A Venn diagram showing the numbers of genes expressed in common between CKG and CKN for 16 distinct cell types.

(D) Scatter plots showing the top KEGG enrichments of shared genes for each cell type between CKG and CKN.

(E) Scatter plots showing the top KEGG enrichment of differentially expressed genes between CKG and CKN for each cell type. A hypergeometric distribution test used to determine significant enrichment.


A
Dataset
Ara
Cot
tSNE_2
tSNE_1
Ara
Cot
B
Cluster
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
tSNE_2
tSNE_1
C
QC
S
Ct
RC
S(P)
E
T
S(X)
A
M
U
UMAP_2
UMAP_1
D
Ct_Ara
E_Ara
A_Ara
RC_Ara
T_Ara
M_Ara
S(P)_Ara
S(X)_Ara
S_Ara
U_Ara
Cm_Cot
Ct_Cot
E_Cot
A_Cot
RC_Cot
M_Cot
S(P)_Cot
S(X)_Cot
S_Cot
T_Cot
U_Cot
VT_Cot
0.52
0.45
0.38
0.31
0.25
0.18
0.11
0.04
−0.03
−0.09
−0.16
E
Expression
celltype
orig.ident
celltype
Cm
Ct
E
A
RC
RH
M
S
S(P)
S(X)
T
U
VT
orig.ident
Ara
Cot
F
Gene Expression Difference
U
M
S(X)
S
RC
S(P)
NA
Ct
E
G
GeneRatio
p.adjust
Count
Copper ion binding
Oxidoreductase activity (I)
Oxidoreductase activity (II)
Lyase activity
Ligase activity
Pyridoxal phosphate binding
Vitamin B6 binding
Vitamin binding
NADP binding
Transaminase activity
H
GeneRatio
p.adjust
Count
Carbon metabolism
Biosynthesis of amino acids
Glycolysis / Gluconeogenesis
Pyruvate metabolism
Tyrosine metabolism
Alanine, aspartate and glutamate metabolism
Citrate cycle (TCA cycle)
Phenylalanine metabolism
Arginine biosynthesis
Isoquinoline alkaloid biosynthesis

214

Figure 4 Evolutionary conservation and divergence of cell types between cotton and Arabidopsis.

(A) tSNE visualization of integrated cotton (CKG and CKN) and Arabidopsis (Denyer et al., 2019; Jean-Baptiste et al., 2019; Wendrich et al., 2020; Zhang et al., 2019) datasets. Cotton and Arabidopsis are abbreviated as Cot and Ara, respectively.

(B) tSNE visualization of cell clusters after alignment between integrated cotton and Arabidopsis datasets.

(C) UMAP visualization of cell lineages that were delineated from the merged cotton and Arabidopsis datasets. Ct, Cortex; E, Endodermis; A, Atrichoblast; QC, Quiescent center; RC, Root cap; T, Trichoblast; M, Meristem; S, Stele; S(P), Phloem; S(X), Xylem; U, Unknown; Pp, Protophloem; Pc, Pericycle; Cm, Columella; VT, vascular tissue;

(D) Heat map showing Pearson's correlations between cotton and Arabidopsis cell lineages.

(E) Orthologous marker gene expression patterns among cell lineages between cotton and Arabidopsis.

(F) Heat map of differentially expressed genes between cotton and Arabidopsis for the top 550 divergently expressed genes. Divergent expression is defined as an expression difference >5-fold change.

(G) GO annotation of divergent expression patterns.

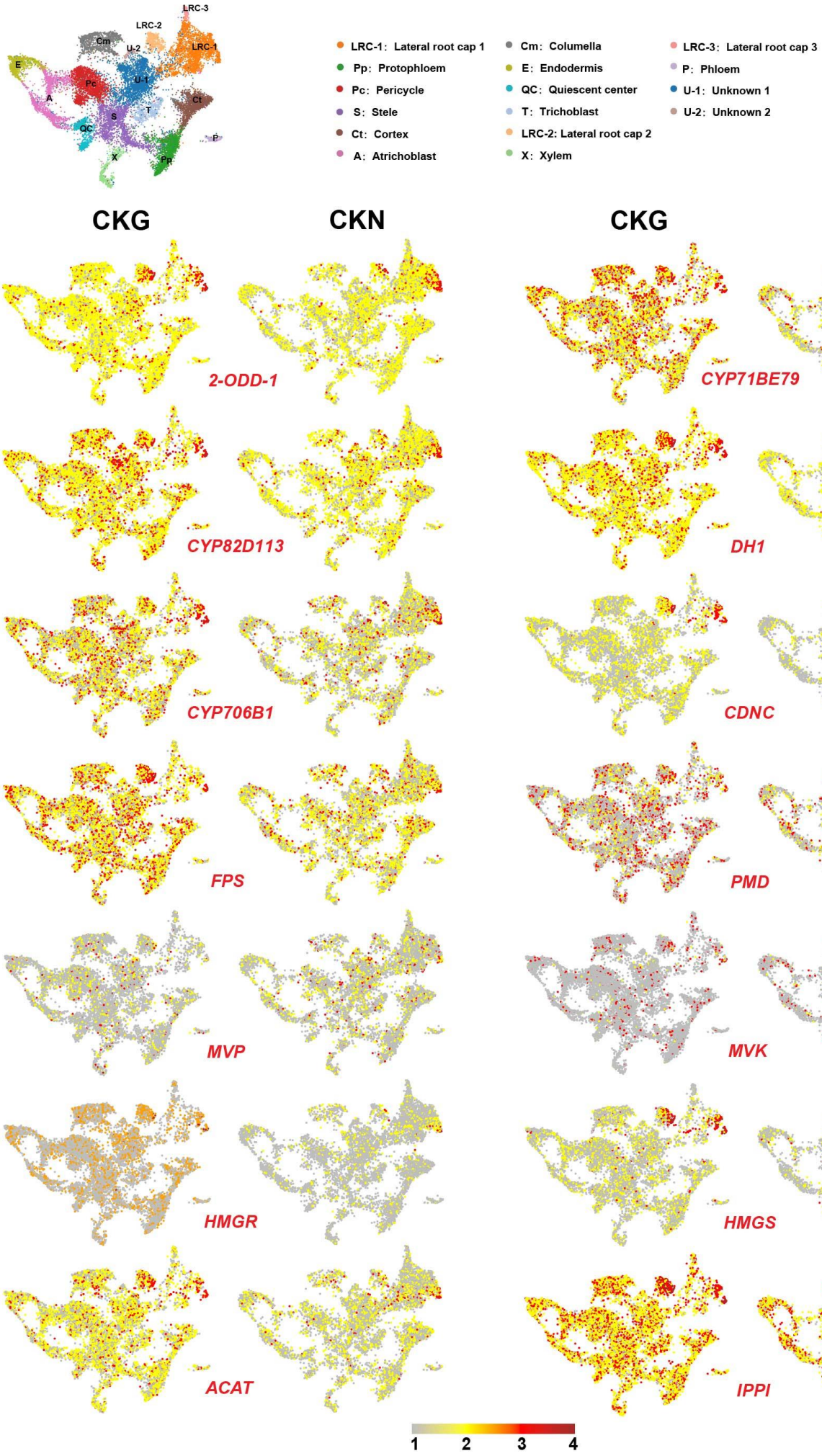
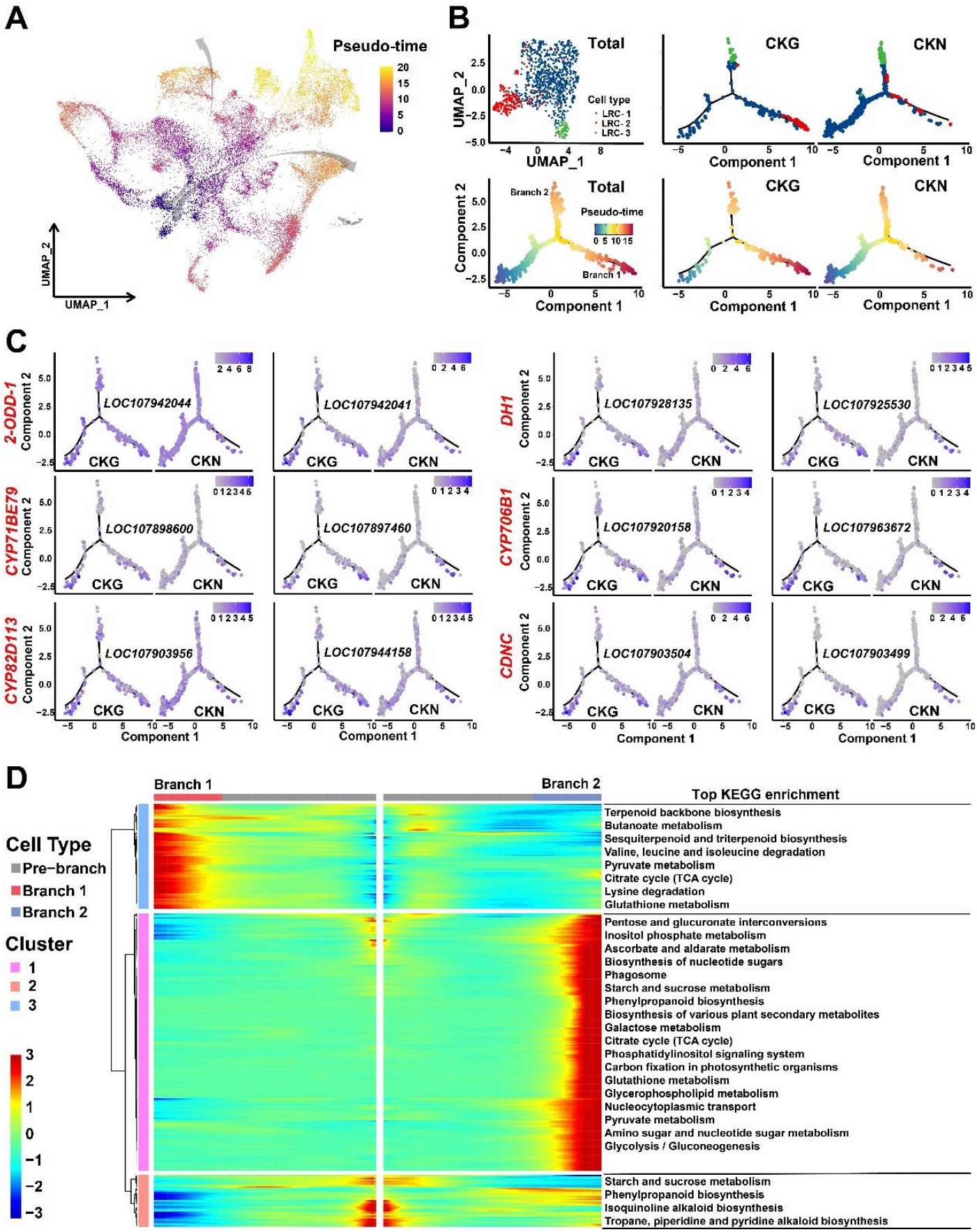

LRC-3
LRC-2
Cm
U-2
LRC-1
E
Pc
U-1
A
Ct
S
T
QC
P
X
Pp
LRC-1: Lateral root cap 1
Pp: Protophloem
Pc: Pericycle
S: Stele
Ct: Cortex
A: Atrichoblast
Cm: Columella
E: Endodermis
QC: Quiescent center
T: Trichoblast
LRC-2: Lateral root cap 2
X: Xylem
LRC-3: Lateral root cap 3
P: Phloem
U-1: Unknown 1
U-2: Unknown 2
CKG
CKN
CKG
CKN
2-ODD-1
CYP71BE79
CYP82D113
DH1
CYP706B1
CDNC
FPS
PMD
MVP
MVK
HMGR
HMGS
ACAT
IPPI
1
2
3
4

Figure 5 ScRNA-seq analysis of the gossypol synthesis pathway.

The cell clusters were checked from the left upper model. The abundance of gene expression in individual cell was showed as the color. The color bar ranging from grey to read exhibited increased gene expression, as located at the bottom of the figure. ACAT, acyl CoA-cholesterol acyltransferase; HMGS, 3-hydroxy-3-methylglutaryl-coenzyme-A (HMG-CoA) synthase; HMGR, 3-hydroxy-3-methylglutaryl-coenzyme-A (HMG-CoA) reductase; MVK, mevalonate kinase; MVP, phosphomevalonate kinase; PMD, diphosphomevalonate decarboxylase; IPPI, IPP isomerase; FPS, FPP synthase; CDNC, (+)-δ-cadinene synthase; DH1, one alcohol dehydrogenase; CYP706B1, CYP82D113, CYP71BE79 are three cytochrome P450 enzymes in the gossypol biosynthesis pathway; 2-ODD-1, one 2-oxoglutarate/Fe(II)-dependent dioxygenase.


A
Pseudo-time
20
15
10
5
0
UMAP_2
UMAP_1
B
Total
Cell type
LRC- 1
LRC- 2
LRC- 3
UMAP_2
UMAP_1
CKG
CKN
Component 1
Component 2
Branch 2
Branch 1
Pseudo-time
C
2-ODD-1
LOC107942044
LOC107942041
CYP71BE79
LOC107898600
LOC107897460
CYP82D113
LOC107903956
LOC107944158
DH1
LOC107928135
LOC107925530
CYP706B1
LOC107920158
LOC107963672
CDNC
LOC107903504
LOC107903499
CKG
CKN
Component 1
Component 2
D
Branch 1
Branch 2
Top KEGG enrichment
Cell Type
Pre−branch
Branch 1
Branch 2
Cluster
1
2
3
Terpenoid backbone biosynthesis
Butanoate metabolism
Sesquiterpenoid and triterpenoid biosynthesis
Valine, leucine and isoleucine degradation
Pyruvate metabolism
Citrate cycle (TCA cycle)
Lysine degradation
Glutathione metabolism
Pentose and glucuronate interconversions
Inositol phosphate metabolism
Ascorbate and aldarate metabolism
Biosynthesis of nucleotide sugars
Phagosome
Starch and sucrose metabolism
Phenylpropanoid biosynthesis
Biosynthesis of various plant secondary metabolites
Galactose metabolism
Citrate cycle (TCA cycle)
Phosphatidylinositol signaling system
Carbon fixation in photosynthetic organisms
Glutathione metabolism
Glycerophospholipid metabolism
Nucleocytoplasmic transport
Pyruvate metabolism
Amino sugar and nucleotide sugar metabolism
Glycolysis / Gluconeogenesis
Starch and sucrose metabolism
Phenylpropanoid biosynthesis
Isoquinoline alkaloid biosynthesis
Tropane, piperidine and pyridine alkaloid biosynthesis

241

Figure 6 Differing developmental trajectories of lateral root cap cell-type populations.

(**A**) Trajectory of cells derived from the combined scRNA data after merging the data for CKG and CKN. Each colored dot indicates a single cell.

(**B**) Trajectory analysis on three cell types including LRC-1, -2 and -3. The analysis, performed using *Monocle 2*, detected a crucial branch point in cell differentiation trajectories. LRC, later root cap.

(**C**) Expression of representative genes involved in gossypol synthesis projected on the pseudo-time trajectory. The level of a gene's expression across cells is denoted by varying degrees of color intensity in each single cell.

(**D**) Heat map exhibiting the expression of the 296 most significant DEGs ($q<1e^{-4}$) along the pseudo-time trajectories. The DEGs were gathered into three clusters according to their expression patterns. Each row denotes one gene and the color bar defines its expression level. Genes belonging to clusters 1, 2, and 3 were enriched in cells on branch 1, branch 2, and the pre-branching part of the trajectories, respectively. The table adhering to the right of the heat map demonstrates the most strongly enriched KEGG pathways for each cluster.

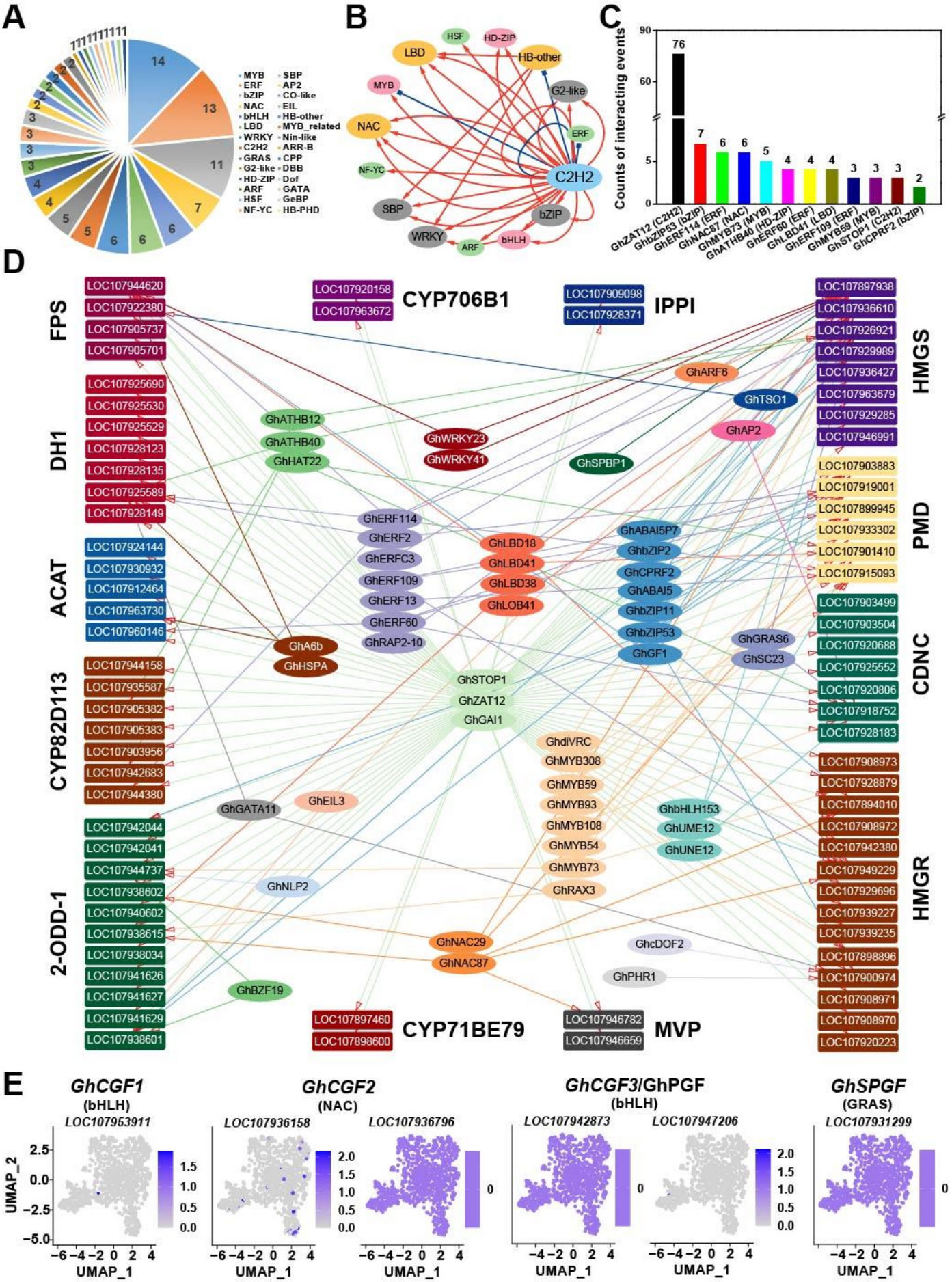

A
B
C
D
E
MYB
ERF
bZIP
NAC
bHLH
LBD
WRKY
C2H2
G2-like
HD-ZIP
ARF
HSF
NF-YC
SBP
AP2
CO-like
EIL
HB-other
MYB_related
Nin-like
ARR-B
CPP
DBB
Dof
GATA
GeBP
HB-PHD
Counts of interacting events
GhZAT12 (C2H2)
GhbZIP53 (bZIP)
GhERF114 (ERF)
GhNAC87 (NAC)
GhMYB73 (MYB)
GhATHB40 (HD-ZIP)
GhERF60 (ERF)
GhLBD41 (LBD)
GhERF109 (ERF)
GhMYB59 (MYB)
GhSTOP1 (C2H2)
GhCPRF2 (bZIP)
FPS
DH1
ACAT
CYP82D113
2-ODD-1
CYP706B1
IPPI
CYP71BE79
MVP
HMGS
PMD
CDNC
HMGR
GhATHB12
GhATHB40
GhHAT22
GhWRKY23
GhWRKY41
GhARF6
GhTSO1
GhAP2
GhSPBP1
GhERF114
GhERF2
GhERFC3
GhERF109
GhERF13
GhERF60
GhRAP2-10
GhLBD18
GhLBD41
GhLBD38
GhLOB41
GhABAI5P7
GhbZIP2
GhCPRF2
GhABAI5
GhbZIP11
GhbZIP53
GhGF1
GhGRAS6
GhSC23
GhA6b
GhHSPA
GhSTOP1
GhZAT12
GhGAI1
GhdiVRC
GhMYB308
GhMYB59
GhMYB93
GhMYB108
GhMYB54
GhMYB73
GhRAX3
GhbHLH153
GhUME12
GhUNE12
GhGATA11
GhEIL3
GhNLP2
GhNAC29
GhNAC87
GhcDOF2
GhPHR1
GhBZF19
E
GhCGF1
(bHLH)
LOC107953911
GhCGF2
(NAC)
LOC107936158
LOC107936796
GhCGF3/GhPGF
(bHLH)
LOC107942873
LOC107947206
GhSPGF
(GRAS)
LOC107931299
UMAP_1
UMAP_2

Figure 7 Transcription factors (TFs) regulating gossypol synthesis.

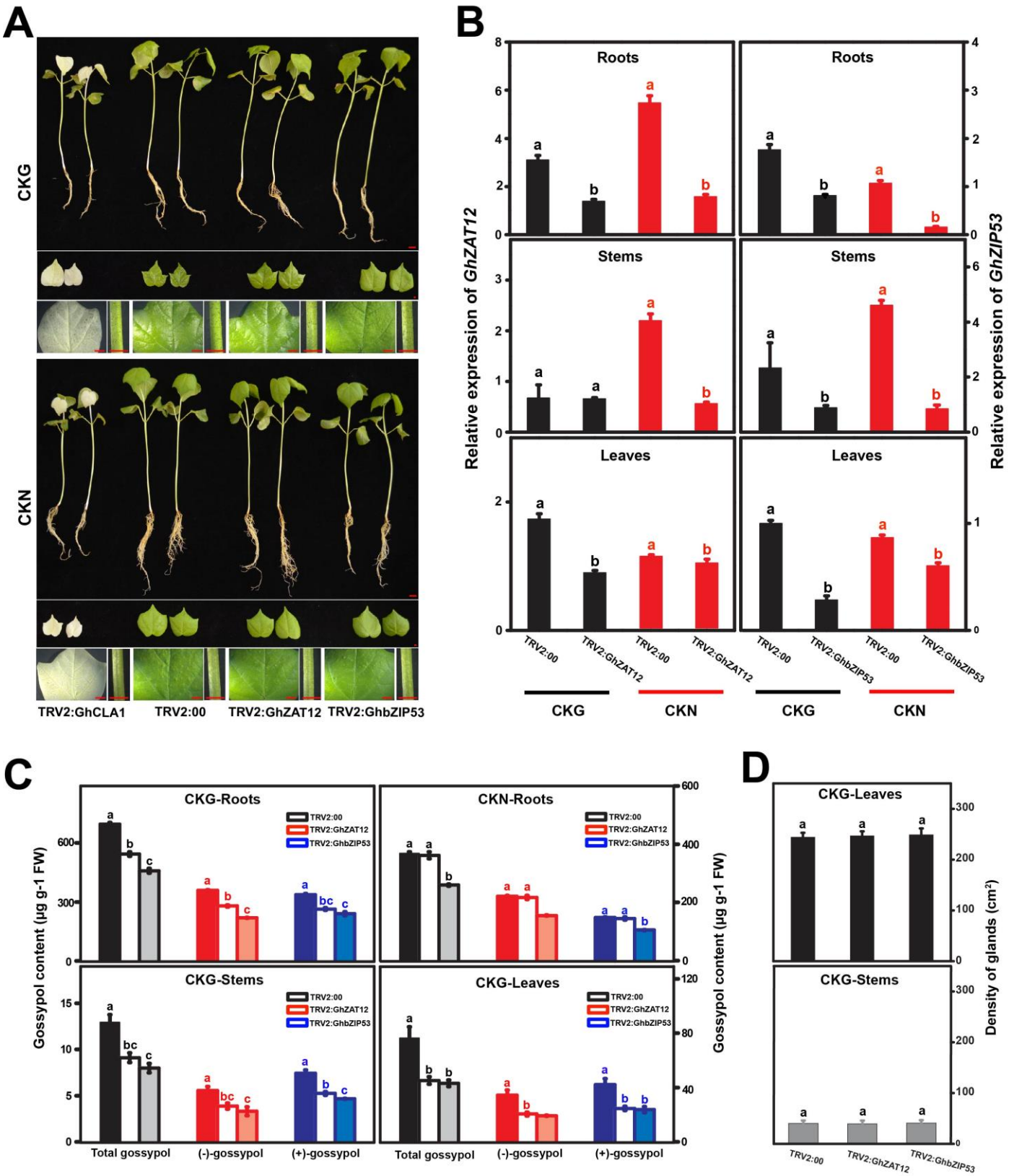
(A) The pie chart shows the categories of 115 TFs involving regulating the gene expression of gossypol synthesis, which was screened from scRNA-seq data. The numbers around the pie denote the counts of each TF family member.

(B) Core TF-TF interaction analysis in TF type. The red arrow denotes promotion and the blue 'T' shape denotes inhibition. Ellipse shows TF type, and the bigger area indicates more important interacting relationship with other TF types. Same color exhibits the same degree of importance in the interaction network.

(C) Top twelve TFs in the TF-gene interacting network of gossypol synthesis with the highest counts of interaction events.

(D) Profiles of core TF-gene interacting network of gossypol synthesis. The ellipse and round rectangle denotes TF and gene, respectively. The color of the ellipse presents TF type as followed, AP2: red purple, ARF: orange, bHLH: green blue, bZIP: blue, C2H2: shallow green, CPP: blue, DBB: green, Dof: shallow purple, EIL: shallow red, ERF: purple, G2-like: shallow grey, GATA: grey, GRAS deep purple, HD-ZIP: yellow-green, HSF: Deep orange, LBD: red, MYB: shallow orange, NAC: orange, Nin-like: shallow blue, SBP: deep green, WRKY: deep red.

(E) Single-cell gene expression feature of reported core TF involved pigment development among cell type LRC-1, LRC-2 and LRC-3.The expression of genes reflected on color according to each bar.


A
CKG
CKN
TRV2:GhCLA1
TRV2:00
TRV2:GhZAT12
TRV2:GhbZIP53
B
Roots
Stems
Leaves
Relative expression of GhZAT12
Relative expression of GhZIP53
TRV2:00
TRV2:GhZAT12
TRV2:GhbZIP53
CKG
CKN
C
CKG-Roots
CKN-Roots
CKG-Stems
CKG-Leaves
Gossypol content (μg g-1 FW)
Total gossypol
(-)-gossypol
(+)-gossypol
D
CKG-Leaves
CKG-Stems
Density of glands (cm²)

274

Figure 8 Functional verification of transcription factors *GhZAT12* and *GhbZIP53* and their phenotypes on gossypol synthesis

(A) Seeding phenotypes on TRV2:CLA1, TRV2:00, TRV2:GhZAT12 and, TRV2:GhbZIP53, involving cultivar CKG and CKN. The lengths of the red bars denote 1cm.

(B) Relative gene expression of TRV2:00, TRV2:GhZAT12 and, TRV2:GhbZIP53, in the organs of the seedlings. The values are denoted as the means± SD (n=3). The significances are marked with the different lowercase letters.

(C) Gossypol contents, Total gossypol, (-)-gossypol and (+)-gossypol contents in the organs of seedling involving TRV2:00, TRV2:GhZAT12 and, TRV2:GhbZIP53. The values are denoted as the means± SD (n=3). The significances are marked with the different lowercase letters.

(D) Density of gland onto leaves and stems in CKG seedlings TRV2:00, TRV2:GhZAT12 and, TRV2:GhbZIP53. The values are shown as the means± SD (n=6). The significances are marked with the different lowercase letters.