CellPress
Partner Journal

Molecular Plant
Research report

Single-cell RNA sequencing reveals a hierarchical transcriptional regulatory network of terpenoid biosynthesis in cotton secretory glandular cells

Jia-Ling Lin[1,2,9], Longxian Chen[1,9], Wen-Kai Wu[1], Xiao-Xiang Guo[1], Cheng-Hui Yu[3], Min Xu[1,2], Gui-Bin Nie[1], Jun-ling Dun[4], Yan Li[5,6], Baofu Xu[5,6], Ling-Jian Wang[1], Xiao-Ya Chen[1,2,7], Wei Gao[8,*] and Jin-Quan Huang[1,*]

[1]State Key Laboratory of Plant Molecular Genetics, CAS Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology and Ecology, University of CAS, Chinese Academy of Sciences, Shanghai 200032, China

[2]School of Life Science and Technology, ShanghaiTech University, Shanghai 200031, China

[3]Chongqing Key Laboratory of Micro-Nano Systems and Intelligent Transduction, Collaborative Innovation, National Research Base of Intelligent Manufacturing Service, Chongqing Technology and Business University, Chongqing 400067, China

[4]Analytical Applications Center, Shimadzu (China) Co., Ltd., Shanghai 200233, China

[5]Shandong Laboratory of Yantai Drug Discovery, Bohai Rim Advanced Research Institute for Drug Discovery, Yantai 264117, Shandong, China

[6]State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China

[7]Shanghai Key Laboratory of Plant Functional Genomics and Resources, Shanghai Chenshan Botanical Garden, Shanghai 201602, China

[8]National Key Laboratory of Cotton Bio-breeding and Integrated Utilization (Henan University), Henan 475004, China

[9]These authors contributed equally to this article.

***Correspondence: Wei Gao (gaowei021@163.com), Jin-Quan Huang (huangjinquan@cemps.ac.cn)**

https://doi.org/10.1016/j.molp.2023.10.008

ABSTRACT

Plants can synthesize a wide range of terpenoids in response to various environmental cues. However, the specific regulatory mechanisms governing terpenoid biosynthesis at the cellular level remain largely elusive. In this study, we employed single-cell RNA sequencing to comprehensively characterize the transcriptome profile of cotton leaves and established a hierarchical transcriptional network regulating cell-specific terpenoid production. We observed substantial expression levels of genes associated with the biosynthesis of both volatile terpenes (such as β-caryophyllene and β-myrcene) and non-volatile gossypol-type terpenoids in secretory glandular cells. Moreover, two novel transcription factors, namely GoHSFA4a and GoNAC42, are identified to function downstream of the *Gossypium PIGMENT GLAND FORMATION* genes. Both transcription factors could directly regulate the expression of terpenoid biosynthetic genes in secretory glandular cells in response to developmental and environmental stimuli. For convenient retrieval of the single-cell RNA sequencing data generated in this study, we developed a user-friendly web server . Our findings not only offer valuable insights into the precise regulation of terpenoid biosynthesis genes in cotton leaves but also provide potential targets for cotton breeding endeavors.

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

INTRODUCTION

Terpenoids, a diverse group of metabolites, comprise >50 000 naturally occurring molecules in living organisms (Thulasiram et al., 2007). Some terpenoids play essential roles as phytohormones, membrane constituents, and photosynthesis-associated components within plants. However, most terpenoids have independently evolved within different plant lineages to adapt to specific ecological niches and thus serve as specialized metabolites. These specialized terpenoids serve various functions, including attracting pollinators and seed dispersers, defending against pathogens and herbivores (Pichersky and Raguso, 2018; Zozina et al., 2018; Borisova-Mubarakshina et al., 2019; Fan et al., 2021). They are synthesized in specific

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

organs within distinct cell types, which prevents self-toxicity in plants and optimizes resource utilization (Kliebenstein, 2004; Wang et al., 2019; Sabbadin et al., 2021).

Cotton plants possess lysigenous glands in their leaves, stems, and seeds, which accumulate various non-volatile terpenoids such as gossypol, heliocides, and hemigossypolone. These terpenoids are derived from the sesquiterpene (+)-δ-cadinene via the mevalonate (MVA) pathway (Altman et al., 1990; Hagenbucher et al., 2013; Mellon et al., 2014; Tian et al., 2018; Huang et al., 2021). Besides, in response to insect herbivory, cotton plants release a complex blend of volatile terpenes, including *trans*-β-ocimene, β-myrcene, α-pinene, and β-caryophyllene. This defensive mechanism acts against chewing herbivores like *Spodoptera exigua* or *Helicoverpa zea*, as well as piercing-sucking herbivores such as *Lygus hesperus* (Loughrin et al., 1994, 1995; McCall et al., 1994; Pare and Tumlinson, 1997; Rodriguez-Saona et al., 2002; Holopainen and Gershenzon, 2010; Rosenkranz et al., 2021). While several studies have reported on the identification of terpenoid biosynthetic enzymes (Tian et al., 2018; Huang et al., 2020; Lin et al., 2023) and transcription factors governing pigment gland parenchyma differentiation (Ma et al., 2016; Janga et al., 2019; Gao et al., 2020; Wu et al., 2021; Zang et al., 2021), the coordination and spatiotemporal regulation of different terpenoids in cotton plants, including non-volatile gossypol-type terpenoids, volatile monoterpenes, and sesquiterpenes, remain poorly understood.

The integration of multi-tissue RNA sequencing with metabolite analysis has significantly contributed to the identification of key enzymes involved in natural product synthesis and the elucidation of metabolic pathways (Nützmann et al., 2016; Tian et al., 2018; Huang et al., 2020; Lin et al., 2023). However, it is important to note that conventional multi-tissue RNA sequencing (RNA-seq) approaches, by averaging gene expression profiles across all cells within a sample, inevitably sacrifice the ability to capture cellular heterogeneity and perform transcriptomic profiling at the individual cell level. In contrast, the emergence of single-cell RNA-seq (scRNA-seq) has revolutionized the field by enabling the profiling of individual cells, thereby offering unprecedented opportunities for various applications. These applications encompass the identification of novel cell types (Liu et al., 2020; Kim et al., 2021; Tenorio Berrío et al., 2022), the reconstruction of developmental trajectories (Denyer et al., 2019; Zhang et al., 2019, 2021a), and the elucidation of gene regulatory networks (Dorrity et al., 2021; Roszak et al., 2021). Notably, scRNA-seq has been effectively employed to unravel the intricate transcriptomic landscape of diverse crop species (Satterlee et al., 2020; Wang et al., 2020; Bezrutczyk et al., 2021; Sun et al., 2022) and to provide valuable insights into the specialized metabolism of medicinal plants (Kang et al., 2022; Li et al., 2023; Sun et al., 2023). The inherent single-cell resolution of scRNA-seq facilitates its integration with other omics datasets, such as multi-tissue RNA-seq data, thereby enabling the identification of pivotal enzymes or transcription factors involved in complex metabolic pathways. This integration not only enhances our understanding of cellular heterogeneity but also provides crucial insights into the regulatory mechanisms governing specialized metabolic processes.

In the present study, we generated a comprehensive, high-resolution transcriptomic landscape of cotton leaves using scRNA-seq. We identified 14 distinct cell clusters representing six major cell populations. By integrating the scRNA-seq results with multi-tissue omic data and reverse genetics, we unveiled a hierarchical transcriptional cascade that regulates the synthesis of both volatile terpenes and non-volatile terpenoids at the single-cell level. Our findings provide insights into the metabolic activities and regulatory mechanisms in secretory glandular cells (SGCs) of cotton plants. Additionally, we identified cell-type-specific marker genes, which can serve as valuable resources for future investigations on cotton leaf development. To facilitate future studies and explorations, we developed a web server, accessible at https://www.cottonleafatlas.com, for convenient data access.

RESULTS AND DISCUSSION

Single-cell expression atlas of cotton leaves

To establish a comprehensive single-cell transcriptomic atlas of *Gossypium hirsutum* leaves, we isolated protoplasts from two 15-day-old developing true leaves and subjected them to droplet-based scRNA-seq on the 10× Genomics platform. The raw data were preprocessed using the Cell Ranger pipeline (Supplemental Table 1); subsequently, the transcriptome data of 28 655 cells were successfully obtained from three biological replicates.

Before performing downstream analyses, we implemented rigorous filtering criteria for quality control. Cells with a low count of detected genes and those with extensive contamination from mitochondrial or chloroplast genes were excluded from further analyses (Supplemental Figure 1). Doublets or multiplets with abnormally elevated gene expression were also excluded. Potential confounding factors such as cell-cycle heterogeneity and the effect of enzymatic treatment were also considered to ensure robustness in subsequent clustering analyses (Supplemental Figures 2A and 2B). The three independent samples were integrated for comprehensive analyses after confirming the absence of batch effects among them (Supplemental Figure 2C).

After quality control, 22 163 high-quality cells were profiled, with a median number of 1319 genes expressed. Unsupervised clustering using Seurat revealed 14 distinct cell clusters (Figure 1A; Supplemental Tables 2 and 3). The expression patterns of the cotton genes homologous to well-characterized marker genes in other plant species were scrutinized to assign cell identities to the clusters (Supplemental Table 4; Supplemental Figure 3A). Consequently, the following six distinct cell populations were identified: mesophyll cells (MCs), epidermal cells (ECs), guard cells (GCs), vascular cells (VCs), proliferating cells (PCs), and SGCs (Figure 1A and Supplemental Figure 3B). These findings are consistent with the heterogeneity observed in cellular composition under light microscopy (Figure 1B).

The MC population consisted of five clusters and showed high expression of photosynthesis-associated genes, such as *CARBONIC ANHYDRASE 1* (*CA1*) and *LIGHT HARVESTING COMPLEX OF PHOTOSYSTEM II* (*LHCB*) (Figure 1C and Supplemental Figure 3A) (Sawchuk et al., 2008; Endo et al., 2014). Cluster 2 was assigned as ECs because it exhibited

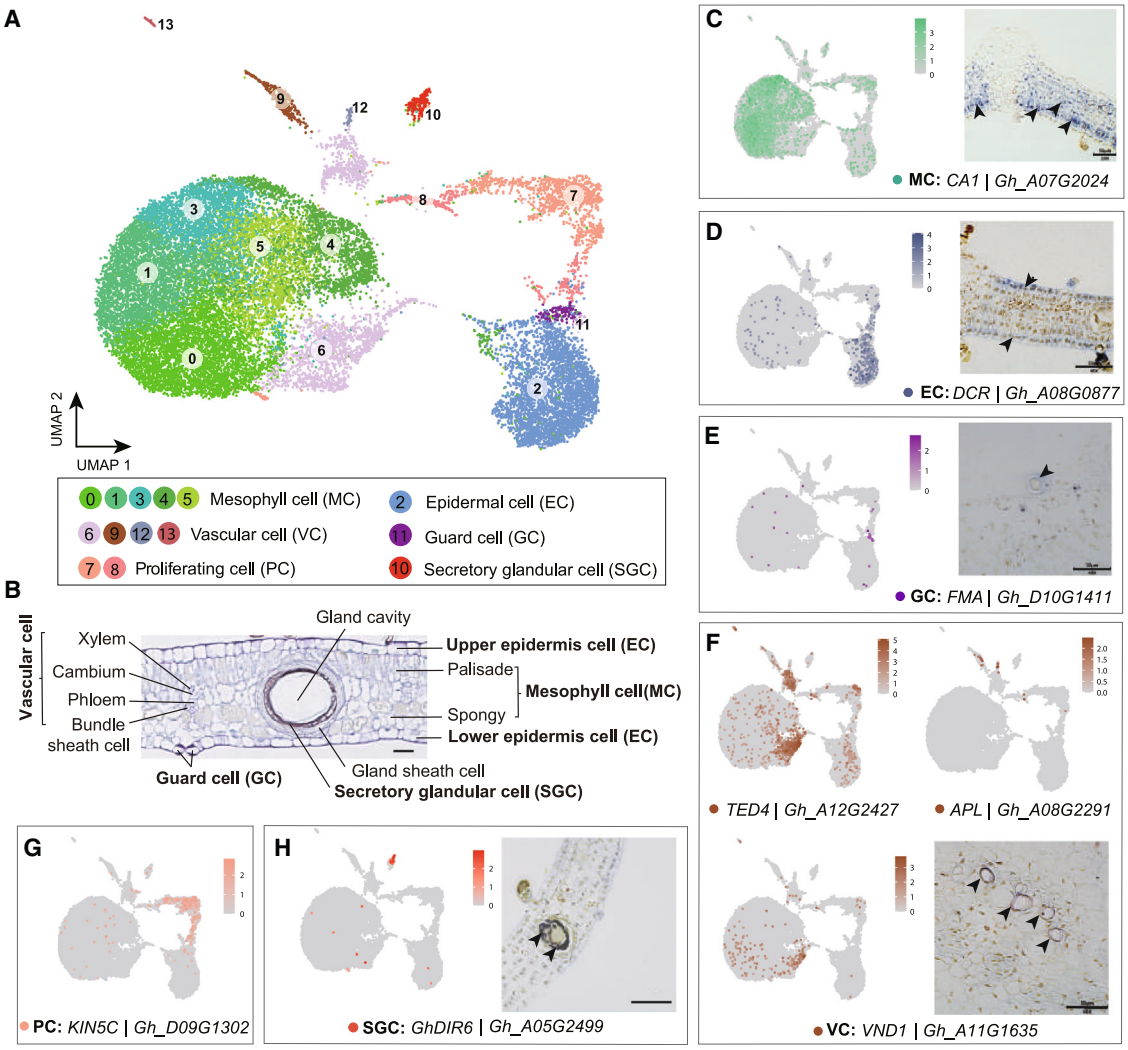

Figure 1. Cell heterogeneity in true leaves of *Gossypium hirsutum*.

(A) Uniform manifold approximation and projection (UMAP) visualization of 14 cell clusters. Each dot represents an individual cell. Colors indicate corresponding cell clusters. n = 22 163.

(B) Cross-section of a cotton leaf with cell types indicated. Scale bar, 20 μm.

(C–H) UMAP visualization and RNA *in situ* hybridization of representative cell-type-specific marker genes in cotton leaves. Colors in UMAP plot represent scaled expression levels of given genes in individual cells. In the hybridized sections, the identified cell types are denoted by black arrows. Scale bar, 50 μm.

the expression of epidermal-specific genes, such as *DEFECTIVE IN CUTICULAR RIDGES* (*DCR*), *3-KETOACYL-COA SYNTHASE 17* (*KCS*), and wax-associated gene *ECERIFERUM3* (*CER3*) (Figure 1D and Supplemental Figure 3A) (Zhang et al., 2021a; Kim et al., 2021; Sun et al., 2023). All GC markers, including cell-specific transcription factors (TFs) *FAMA* and *WRKY6*, were enriched in cluster 11 (Figure 1E) (Rusconi et al., 2013; Kim et al., 2021; Tenorio Berrío et al., 2022; Sun et al., 2023). The VC population comprised four clusters and displayed companion cell-, phloem-, and xylem-specific marker genes (Figure 1F) (Endo et al., 2018; Smet et al., 2019; Zhang et al., 2021a; Lopez-Anido et al., 2021). Clusters 7 and 8 were designated as the PC population owing to their enriched expression of cell-cycle-related genes, such as syntaxin-related protein gene *KN*, *CYCLIN-DEPENDENT KINASE B2* (*CDKB2*), and kinesin-like protein gene *KIN5C* (Figure 1G) (Sun et al., 2023). Notably, within the PC population, distinct regions exhibited specific expression of marker genes associated with MCs, ECs, and VCs (Figures 1C, 1D, and 1F), suggesting cellular heterogeneity within the PC population. Therefore, after mitigating the effects of cell-cycle heterogeneity, we performed a clustering analysis specifically on the PC population, which yielded five subclusters (Supplemental Figure 3C). In subclusters 0, 1, and 4, marker genes such as *CARBONIC ANHYDRASE 1* (*CA1*) and *LIGHT HARVESTING COMPLEX OF PHOTOSYSTEM II (LHCB*), which are indicative of MCs, displayed upregulated expression (Supplemental Figure 3D). Additionally, marker genes for ECs and VCs were highly expressed in subclusters 2 and 3, respectively. This observation indicated that the PC population encompassed three distinct subpopulations, representing the differentiation state of MCs, ECs, and VCs.

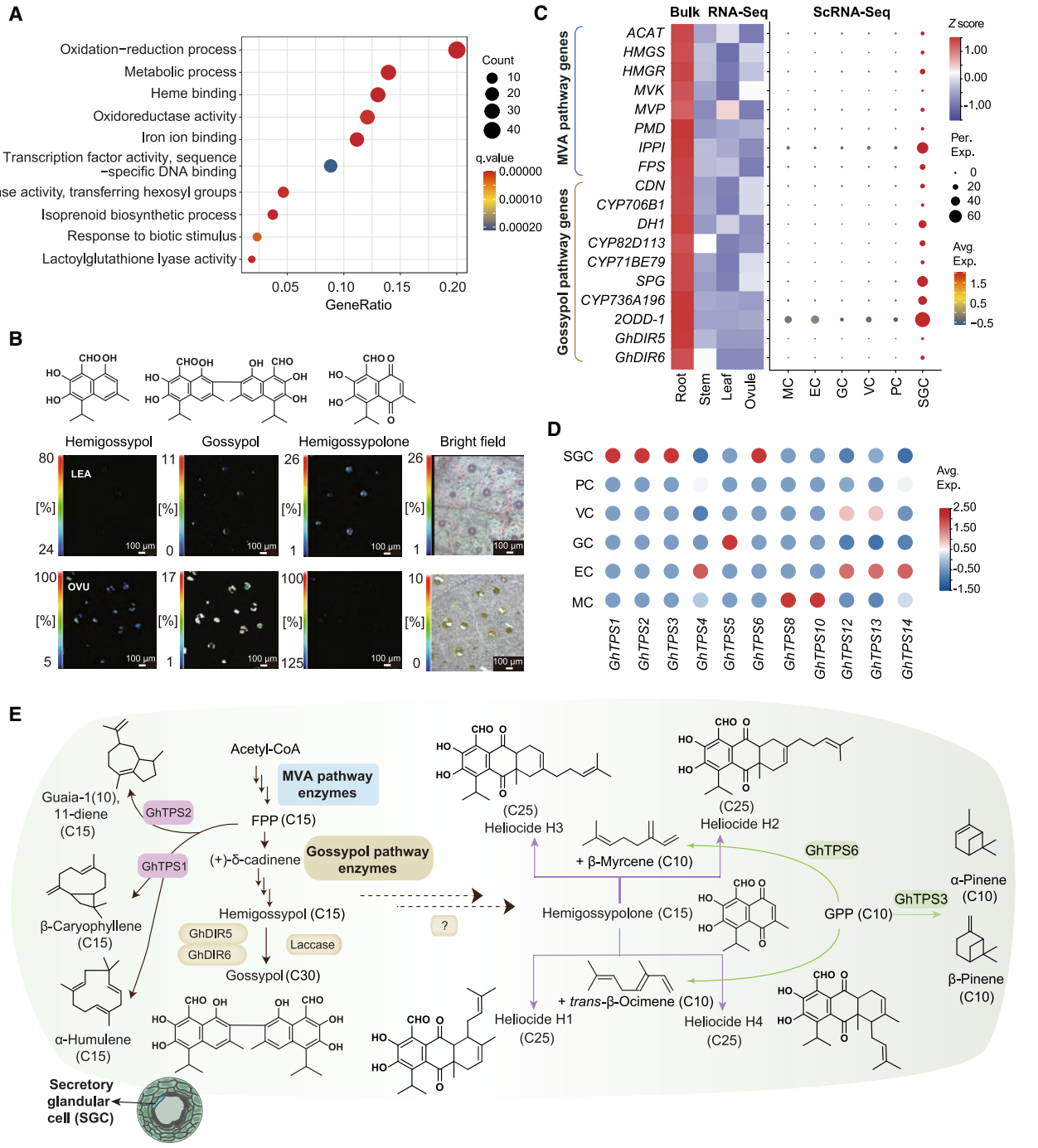

Figure 2. Non-volatile terpenoids and volatile terpenes are exclusively biosynthesized within SGCs.

(A) Enriched GO terms attributed to marker genes of SGCs. The color of each dot represents the statistical significance (q value) of the enriched terms, while the diameter of the dot indicates the number of genes associated with each term.

(B) MALDI–MS images depicting the spatial distribution of gossypol-type sesquiterpene aldehydes in both cotton leaves and ovules. LEA, leaf; OVU, 30-day-old ovule. Scale bar, 100 μm.

(C) Expression patterns of the MVA and gossypol pathway genes for bulk and single-cell transcriptomes. For the single-cell gene expression heatmap, the color of each dot represents the scaled expression levels of the respective genes, while the diameter of the dot indicates the percentage of cells in each cluster expressing the corresponding gene. The gene order is presented sequentially from upstream to downstream of the metabolic pathways. *ACAT*, acyl CoA-cholesterol acyltransferase; *HMGS*, 3-hydroxy-3-methylglutaryl-coenzyme-A (HMG-CoA) synthase; *HMGR*, HMG-CoA reductase; *MVK*, mevalonate kinase; *MVP*, phosphomevalonate kinase; *PMD*, diphosphomevalonate decarboxylase; *IPPI*, isopentenyl diphosphate isomerase; *FPS*, farnesyl diphosphate synthase; *CDN*, (+)-δ-cadinene synthase; *CYP106B1/CYP82D113/CYP71BE79/CYP736A196*, cytochrome P450

(legend continued on next page)

The SGC population was distinguished by robust expression of the *Gossypium PIGMENT GLAND FORMATION GENE* (*GoPGF*), along with secretory dirigent protein genes, namely *GhDIR5* and *GhDIR6*, which play pivotal roles in the gossypol biosynthesis pathway (Figure 1H and Supplemental Figure 3A) (Ma et al., 2016; Gao et al., 2020; Lin et al., 2023). The SGC cluster exhibited a conspicuous demarcation from the other cell clusters on the uniform manifold approximation and projection plot, indicating its distinctive transcriptome profile.

To validate the accuracy of our cell-type definition, we performed RNA *in situ* hybridization on the tissue sections of the cotton leaves and examined the expression patterns of the aforementioned marker genes (Figures 1C–1H). The results supported the fidelity of our cell-type classification. Additionally, specifically expressed genes in each cell type were identified and used to perform Gene Ontology (GO) enrichment analysis (Supplemental Tables 5 and 6). Notably, the specifically expressed genes in each cell type exhibited significant enrichments of the GO terms that were highly relevant to their corresponding biological functions (Figure 2A and Supplemental Figure 4). For instance, MC-specific genes were markedly enriched for photosynthesis-related terms (Supplemental Figure 4A; Supplemental Table 6), whereas SGC-specific genes exhibited enrichments in the isoprenoid biosynthetic process (Figure 2A; Supplemental Table 6) (Loudya et al., 2021), indicating their active state in secondary metabolism.

SGC-specific biosynthesis of non-volatile terpenoids and volatile terpenes

To determine the spatial distribution of the gossypol pathway in cotton leaves, we examined the expression patterns of the characterized genes involved in this pathway at the single-cell level. Consistent with the gland-specific distribution of gossypol-type terpenoids (Figure 2B), we discovered that all tailoring enzymes in the gossypol pathway, as previously identified through coexpression analyses using bulk RNA-seq data (Tian et al., 2018; Huang et al., 2020; Lin et al., 2023), were exclusively expressed in SGCs (Figure 2C). While all leaf cells expressed enzymes involved in the MVA and methylerythritol phosphate (MEP) pathways (Supplemental Figure 5), which are responsible for synthesizing essential terpenoids such as photosynthetic pigments and phytosterols (Vranová et al., 2013), a set of MVA pathway genes showed specific expression in SGCs and displayed a similar expression pattern across multiple tissues to the gossypol pathway genes (Figure 2C), suggesting a potential coordinated transcriptional regulation of MVA and gossypol pathways (Tian et al., 2018; Huang et al., 2020).

In addition to the gossypol biosynthesis-related enzymes, we observed distinct and remarkable expression patterns of four terpene synthases (TPSs) in SGCs (Figure 2D). Among these TPSs, GhTPS1 and GhTPS2 are classified as sesquiterpene synthases utilizing farnesyl diphosphate derived from the MVA pathway, while GhTPS3 and GhTPS6 are categorized as monoterpene synthases using geranyl diphosphate originating from the MEP pathway (Supplemental Figure 6A). GhTPS1 is responsible for the synthesis of volatile sesquiterpenes, namely β-caryophyllene and α-humulene (Yang et al., 2013), whereas GhTPS2 primarily catalyzes the formation of guaia-1(10),11-diene (Yang et al., 2013). GhTPS3 is implicated in the biosynthesis of volatile monoterpenes, including α-pinene and β-pinene (Yang et al., 2013). In contrast, GhTPS6 demonstrates specificity in catalyzing the synthesis of β-myrcene (Huang et al., 2018).

Previous studies have demonstrated that the linear monoterpene β-myrcene, a conjugated diene synthesized by monoterpene synthases, undergoes a Diels–Alder reaction with hemigossypolone, a dieneophile, resulting in the formation of the [4+2] cycloaddition products heliocides 2 and 3 (Park et al., 2019). The Diels–Alder reaction, a widely used and powerful tool in organic synthesis, is extensively utilized for synthesizing intricate pharmaceutical and biologically active compounds (Desimoni, 1983; Briou et al., 2021). However, the specific monoterpene synthases associated with the synthesis of heliocides in cotton are yet to be elucidated. Consequently, we employed virus-induced gene silencing (VIGS) techniques to investigate the functional roles of SGC-localized monoterpene synthase GhTPS6, as it possesses the ability to catalyze the formation of β-myrcene from geranyl diphosphate. Silencing of *GhTPS6* significantly reduced β-myrcene accumulation and the subsequent Diels–Alder reaction products heliocides 2 and 3, while leading to an increase in the levels of *trans*-β-ocimene, heliocide 1, and heliocide 4 (Supplemental Figures 6B–6D). These findings highlight the important roles of GhTPS6 in providing the required diene substrates for the Diels–Alder reaction in cotton plants, facilitating complex metabolic interactions between volatile and non-volatile metabolites. Altogether, our results provide insights into the spatial distribution of biosynthetic enzymes involved in the biosynthesis of both volatile terpenes and non-volatile terpenoids, confirming their SGC-specific production in cotton leaves (Figure 2E).

Biosynthesis of volatile terpenes and non-volatile terpenoids is concurrently regulated by GoHSFA4a and GoNAC42

Based on the scRNA-seq analysis, we identified a set of genes that are specifically expressed in SGCs, providing an opportunity to explore novel regulators in terpenoid biosynthesis via reverse genetics. We used stringent selection criteria ($\log_{2\ \text{fold change}} > 0.5$, pct1/pct2 > 5, and adjusted *P value* < 0.01) and identified 20 highly expressed TFs in SGCs (Figure 3A and Supplemental Figure 7A). Notably, 10 of these TFs showed minimal expression in glandless cotton leaves (Figure 3A and

monooxygenase; *DH1*, alcohol dehydrogenase-1; *SPG*, specialized glyoxalase I; *2ODD-1*, 2-oxoglutarate/Fe(II)-dependent dioxygenase-1; *GhDIR5*/*GhDIR6*, dirigent proteins.

(D) Heatmap showing the expression of TPS genes in each cell type. The color of each dot represents the scaled expression levels of the respective genes.

(E) Biosynthesis of non-volatile terpenoids (gossypol, hemigossypolone, and heliocides 1–4) and volatile terpenes (α-humulene, β-caryophyllene, β-myrcene, *trans*-β-ocimene, α-pinene, and β-pinene) in SGCs of cotton leaves. Enzymes responsible for specific biosynthesis steps leading to the formation of non-volatile terpenoids and volatile terpenes are depicted. Unidentified reactions are denoted by dashed arrows.

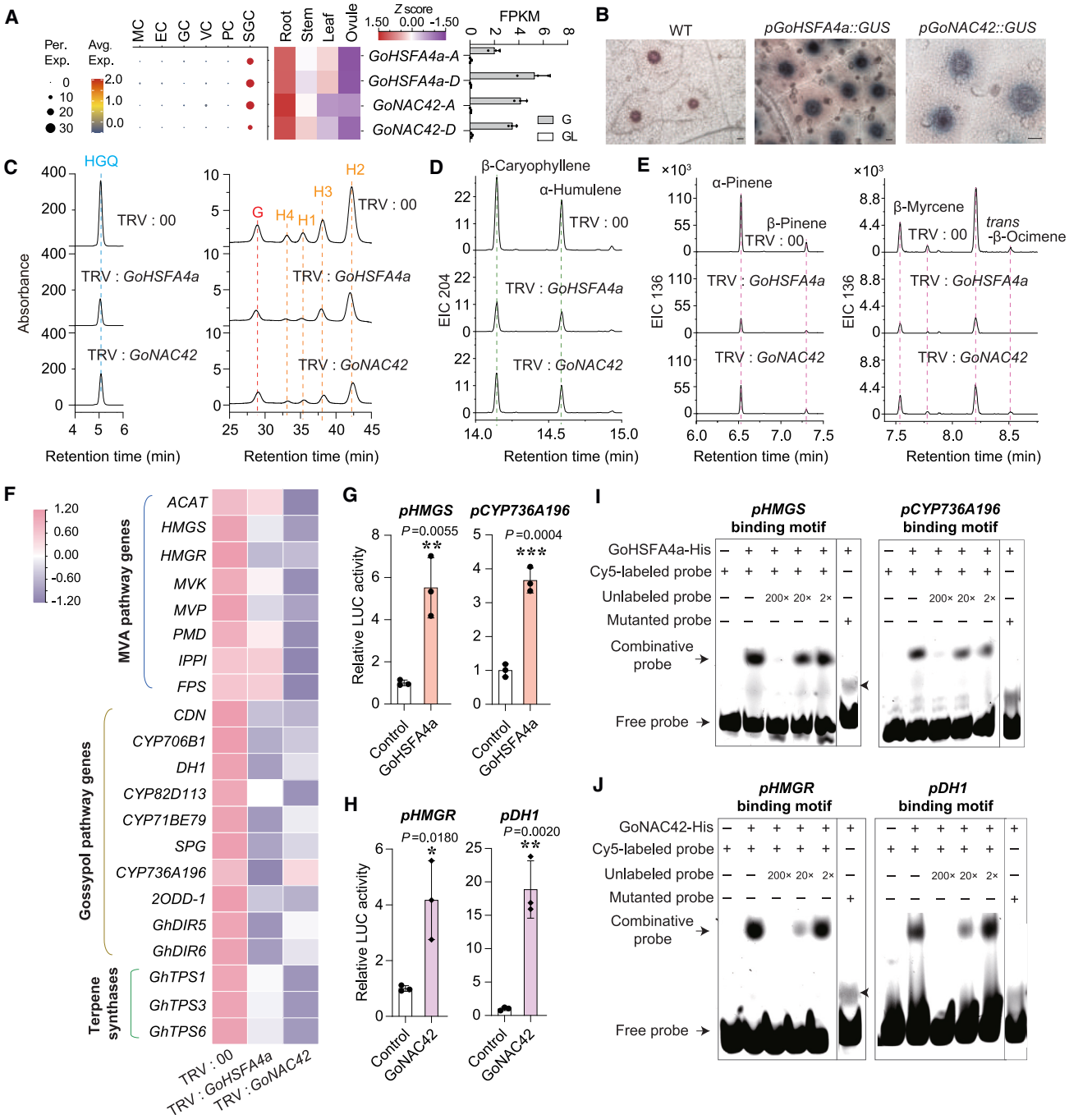

Figure 3. Terpenoid biosynthesis regulated by SGC-enriched *GoHSFA4a* and *GoNAC42*.

(A) Expression patterns of *GoHSFA4a* (*Gh_A05G3255/Gh_D04G0354*) and *GoNAC42* (*Gh_A06G1947/Gh_D06G2096*) in both bulk and single-cell transcriptomes, along with their relative expression levels in leaves of the glandless cotton cultivar CCRI12gl (GL) compared with the gland cultivar CCRI12 (G). Data are presented as mean ± standard deviation with three biological replicates (n = 3). FPKM, fragments per kilobase of transcript per million mapped reads.

(B) GUS staining analysis of transgenic cotton leaves harboring *pGoHSFA4a::GUS* and *pGoNAC42::GUS* constructs compared to the control. Scale bars, 50 μm.

(C) Decreased gossypol-type non-volatile terpenoids content in leaves after VIGS of *GoHSFA4a* and *GoNAC42*. Peaks corresponding to different terpenoids are indicated by dotted lines. HGQ, hemigossypolone; G, gossypol. H1–4 are heliocides. The statistical analysis, which illustrates the relative amounts of non-volatile terpenoids, is displayed in Supplemental Figures 9A and 9B.

(D and E) Gas chromatography–mass spectrometry profiles of the extracts derived from cotton leaves subjected to VIGS-mediated silencing of *GoHSFA4a* and *GoNAC42* in comparison to the control leaves. The extracted-ion chromatogram at *m/z* 204 illustrates the presence of the volatile sesquiterpenes **(D)** β-caryophyllene and α-humulene. The extracted-ion chromatogram at *m/z* 136 corresponds to the volatile monoterpenes

(legend continued on next page)

Supplemental Figure 7B), which lack the expression of secondary metabolism pathway genes. These TFs included well-known regulators, such as GoPGF, responsible for gland development, and the MYB TF CGP1, involved in gland pigmentation (Supplemental Figure 7B) (Ma et al., 2016; Gao et al., 2020). Furthermore, by conducting correlation analyses using multi-tissue RNA-seq data, we identified two candidates, HSF TFs (Gh_A05G3255/Gh_D04G0354) and NAC TFs (Gh_A06G1947/Gh_D06G2096), that showed a strong correlation with gossypol pathway genes (Figure 3A and Supplemental Figure 7C). Gh_A05G3255/Gh_D04G0354 are homologous to AtHSFA4a (AT4G18880), which belongs to the HSFA4 subfamily, whereas Gh_A06G1947/Gh_D06G2096 are homologous to AtNAC042 (AT2G43000) (Supplemental Figure 8). To simplify, we designated Gh_A05G3255/Gh_D04G0354 as GoHSFA4a and Gh_A06G1947/Gh_D06G2096 as GoNAC42

Transgenic cotton lines carrying GUS reporter genes driven by the promoters of *GoHSFA4a* and *GoNAC42* exhibited strong GUS activity specifically in the pigment glands (Figure 3B), supporting their specific expression in SGCs. Further investigation using VIGS technology showed a significant reduction in the levels of MVA-derived terpenoids, including gossypol-type non-volatile terpenoids (Figure 3C and Supplemental Figures 9A and 9B), volatile sesquiterpenes β-caryophyllene and α-humulene (Figure 3D and Supplemental Figures 9C and 9D), and MEP-derived monoterpenes α-pinene, β-pinene, β-myrcene, and *trans*-β-ocimene (Figure 3E and Supplemental Figures 9C and 9D). This was accompanied by the downregulation of related biosynthesis genes (Figure 3F and Supplemental Figures 9E and 9F). These findings reveal a tightly coordinated transcriptional regulation of genes encoding enzymes responsible for the biosynthesis of non-volatile terpenoids and volatile terpenes in SGCs. Despite the reduced terpenoid levels, no visible difference was found in the glandular developments between the control and VIGS plants (Supplemental Figure 10).

Although members of the HSFA4 subfamily are considered activators of heat stress gene expression (Lin et al., 2018; von Koskull-Döring et al., 2007), we found that heat stress does not affect *GoHSFA4a* expression (Supplemental Figure 11). To investigate the specific biological role of GoHSFA4a, we generated mutant lines using CRISPR–Cas9-mediated genome editing. Nucleotide deletions were detected in target-containing amplicons of CR-*Gohsfa4a*-positive lines, and no mutation was found in the potential off-target sites (Supplemental Figures 12A–12C; Supplemental Table 7). The genetically edited plants exhibited a significant reduction in terpenoid production compared to wild-type plants (Supplemental Figures 12D and 12E), confirming the role of GoHSFA4a in terpenoid metabolism. Remarkably, the development of glandular structures remained unaffected (Supplemental Figures 12F and 12G).

Subcellular localization studies showed that both GoHSFA4a and GoNAC42 are localized in the nucleus (Supplemental Figure 13A), supporting their role as TFs. Then, sequence analysis was performed using PlantPAN 4.0 (http://plantpan.itps.ncku.edu.tw/plantpan4/TF_info.php) to predict the *cis*-elements of GoHSFA4a and GoNAC42, which indicated their binding with the promoter regions of MVA and gossypol pathway genes. Promoters activated by GoNAC42 and GoHSFA4a were predicted to have one or multiple binding sites. Dual-luciferase (LUC) assay showed that these TFs activate the expression of MVA and gossypol pathway genes (Figures 3G and 3H and Supplemental Figure 13B). To investigate the direct interaction between TFs and the promoter regions of their target genes, we performed electrophoretic mobility shift assays and yeast one-hybrid assays. Our results showed that GoHSFA4a directly binds to the detected HSF binding motifs in the promoter regions of the MVA pathway gene *HMGS* and the gossypol pathway gene *CYP736A196* (Figure 3I and Supplemental Figure 13C), whereas GoNAC42 binds to NAC binding motifs in the promoter regions of the MVA pathway gene *HMGR* and the gossypol pathway gene *DH1* (Figure 3J and Supplemental Figure 13D). Altogether, these findings indicated that GoHSFA4a and GoNAC42 activate the specific MVA and gossypol pathway genes by binding to their promoter regions, thus regulating the biosynthesis of gossypol-type metabolites.

GoHSFA4a and *GoNAC42* are directly activated by GoPGF in SGCs

Despite the crucial role of GoPGF in gland formation and gossypol biosynthesis (Ma et al., 2016), the signal transduction mechanisms involved in this process remain incompletely understood. To investigate the relationship of GoPGF with GoHSFA4a and GoNAC42, we evaluated the expression of *GoHSFA4a* and *GoNAC42* in *GoPGF*-silenced leaves. Silencing of *GoPGF* resulted in a significant decrease in the mRNA levels of both genes compared to those in the control (Figure 4A). Through chromatin immunoprecipitation followed by sequencing (ChIP-seq) analysis using transgenic cotton lines expressing *35S::GoPGF-GFP*, we found that GoPGF directly binds to the promoters of *GoHSFA4a* and *GoNAC42* (Figure 4B). Moreover, through electrophoretic mobility shift assays, yeast one-hybrid assays, and microscale thermophoresis assays, we established the physical interaction between GoPGF and the promoter regions of *GoHSFA4a* and *GoNAC42*, which harbor the G-box and

(E) α-pinene, β-pinene, β-myrcene, and *trans*-β-ocimene. Peaks corresponding to different volatile terpenes are indicated by dotted lines. The statistical analysis, which illustrates the relative amounts of volatile terpenes, is displayed in Supplemental Figures 9C and 9D.
(F) A comprehensive global heatmap displaying the expression levels of MVA and gossypol pathway genes, along with SGC-enriched terpene synthases, in cotton leaves subjected to VIGS-mediated silencing of *GoHSFA4a* and *GoNAC42* compared to control (TRV:00) plants. Gene expression levels were quantified using qRT–PCR analysis.
(G and H) Dual-LUC assays indicating the activation effect of GoHSFA4a on promoters of *HMGS* and *CYP736A196* **(G)** and the activation effect of GoNAC42 on promoters of *HMGR* and *DH1* **(H)**. The LUC activities were normalized to REN, and values relative to the empty vector control (normalized to 1) are shown. Data are presented as mean ± standard deviation with three biological replicates ($n = 3$).
(I and J) Electrophoretic mobility shift assay experiments showing the specific binding of GoHSFA4a to the predicted binding motif in *HMGS* and *CYP736A196* promoters **(I)** and the specific binding of GoNAC42 to the predicted binding motif in *HMGR* and *DH1* promoters **(J)**. Sequences of predicted binding motifs are listed in Supplemental Table 10.

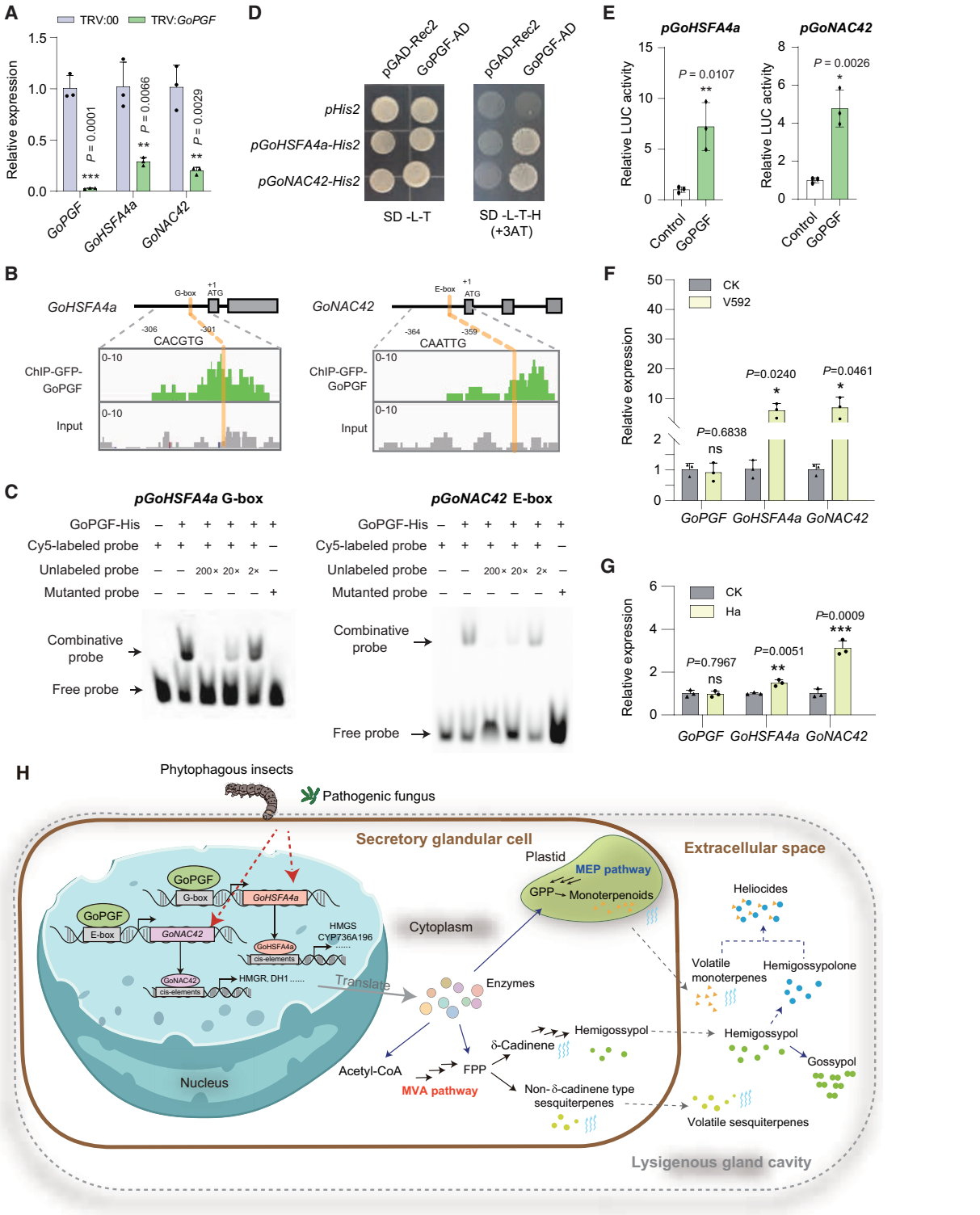

Figure 4. Regulation of *GoHSFA4a* and *GoNAC42* by GoPGF and biotic stress signals.

(A) The transcript levels of *GoHSFA4a* and *GoNAC42* reduced substantially in the leaves of *GoPGF*-silenced (TRV:*GoPGF*) plants compared with the control (TRV:00) leaves (mean ± standard deviation, $n = 3$ independent experiments).

(B) ChIP-seq analysis reveals significant enrichment of GoPGF in the promoter regions of *GoHSFA4a* (left) and *GoNAC42* (right).

(C and D) GoPGF exhibited binding affinity to the G-box sequence within the *GoHSFA4a* promoter and the E-box sequence within the *GoNAC42* promoter, as shown by electrophoretic mobility shift assays **(C)** and Y1H assays **(D)**.

(legend continued on next page)

E-box regulatory elements, respectively (Figure 4C and 4D and Supplemental Figure 14).

To investigate the activation effect of GoPGF on *GoHSFA4a* and *GoNAC42*, we performed a dual-LUC assay in *Nicotiana benthamiana* using a transient expression system. Coexpression of GoPGF resulted in a significant induction of LUC reporters driven by the promoters of *GoHSFA4a* and *GoNAC42*, exhibiting approximately seven- and five-fold increases, respectively, compared with those of the negative control (Figure 4E). Notably, our investigation into the hierarchy between GoHSFA4a and GoNAC42 showed no reciprocal effect on the expression of the other (Supplemental Figure 15).

Altogether, these findings unequivocally reveal that GoPGF directly activates the transcription of *GoHSFA4a* and *GoNAC42*, which in turn regulate terpenoid biosynthesis by selectively activating genes involved in non-volatile terpenoids and volatile terpenes biosynthesis.

GoHSFA4a and GoNAC42 orchestrate the stress-induced production of both volatile terpenes and non-volatile terpenoids

Transcriptional gene regulatory networks establish connections between regulatory proteins and target genes, governing precise spatiotemporal patterns of gene expression (Smith et al., 2016; Chronis et al., 2017). These networks undergo reconfiguration during dynamic processes, such as development or stress induction, enabling cell-type-specific expression. Terpenoid phytoalexins are accumulated in response to pathogenic fungi and phytophagous insect infestation (Chen et al., 1995; Liu et al., 1999; Huang et al., 2018; Tian et al., 2018). In response to pathogen invasion and insect infestation, the levels of terpenoids such as gossypol, β-myrcene, and *trans*-β-ocimene were significantly increased, accompanied by concurrent upregulation of terpenoid biosynthesis genes (Supplemental Figures 16 and 17).

Although the expression of GoNAC42 and GoHSFA4a showed a significant increase in response to pathogen and insect invasion, the expression level of their upstream regulator, GoPGF, remained relatively stable or slightly decreased (Figure 4F and 4G). These findings imply that GoHSFA4a and GoNAC42 may have a more prominent role as regulators in the defensive network, specifically in response to pathogen or insect herbivore stimuli, in comparison to GoPGF. The regulation of terpenoid biosynthetic genes by GoHSFA4a and GoNAC42 may establish a connection between plant stress responses and metabolic alterations. It is plausible that the invading pathogen or insect hijacks the activity of GoPGF, leading to its inhibition (Chen et al., 2023). Nonetheless, further comprehensive investigations are warranted to fully delineate the intricate regulatory mechanisms underlying plant defense responses.

In summary, we used scRNA-seq technology to establish the transcriptome landscape of cotton leaves and examined the biosynthesis of specialized terpenoids within the SGCs of cotton plants. Based on the scRNA-seq data, assisted by multi-tissue RNA-seq analysis, we discovered two previously unidentified TFs, GoHSFA4a and GoNAC42. They directly regulate the expression of terpenoid biosynthesis genes (such as *HMGS*, *HMGR*, *CYP736A196,* and *DH1*) and function downstream of *GoPGF* within SGCs in response to developmental and environmental signals (Figure 4H). These findings provide insights into the precise localization of volatile terpene and non-volatile terpenoid enzymes and reveal a hierarchical transcriptional network that regulates terpenoid biosynthesis at the single-cell level.

METHODS

Plant growth conditions

The upland cotton (*G. hirsutum*) variety R15 and the genetically modified cotton lines were cultivated in a greenhouse under a photoperiod of 16-h light and 8-h darkness, with a temperature of 28°C. The *N. benthamiana* plants were grown under a 16-h light and 8-h dark photoperiod, with a temperature of 22°C.

Protoplast isolation and scRNA-seq library construction

To isolate protoplasts, the first and secondary true leaves of *G. hirsutum* at 15 days old were subjected to sterilization followed by chopping. The chopped leaves were then digested for a duration of 3 h in an incubator shaker using an enzyme solution comprised of 1% cellulase R-10, 0.15% macerozyme R-10, 0.45 M mannitol, and 20 mM MES adjusted to a pH range of 5.6–5.8. The resulting protoplasts were subsequently filtered through 40-μm strainers to eliminate impurities and clumps. These protoplasts were then centrifuged at a force of 300 *g* for 5 min at a temperature of 4°C and subsequently washed twice employing a protoplasting solution. Subsequently, the viability of the protoplasts was assessed through staining with 0.04% trypan blue, followed by cell counting under a microscope. Finally, the suspension volume was adjusted to achieve a density of 1800 cells per μL. For the construction of the scRNA-seq library, Chromium Next GEM Single Cell 3' Reagent Kits v.3.1 were utilized, and the generated libraries were sequenced using an Illumina NovaSeq 6000 sequencer.

(E) Activation effects of GoPGF on the promoters of *GoHSFA4a* (left) and *GoNAC42* (right) were evaluated using dual-LUC assays (mean ± standard deviation, $n = 3$ independent experiments).
(F and G) Relative expressions of *GoPGF*, *GoHSFA4a*, and *GoNAC42* in cotton leaves infected with *Verticillium dahliae* isolate V592 **(F)** or damaged by *Helicoverpa armigera* (Ha) **(G)** compared with those in the control plants. The expression in undamaged plants (CK) was normalized to 1. Statistical analysis was performed using Student's *t*-test, and significance is denoted as $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, and $^{****}P < 0.0001$. Data are presented as mean ± standard deviation ($n = 3$ biological replicates).
(H) Hierarchical transcriptional cascade model illustrating the regulation of gossypol-type non-volatile terpenoid synthesis and volatile terpene production in SGCs of cotton leaves. GoPGF binds directly to the promoters of *GoHSFA4a* and *GoNAC42*, thus regulating their constitutive expression. Pathogens and insect herbivores induce the expression of *GoHSFA4a* and *GoNAC42*, whereas *GoPGF* remains unresponsive to these stimuli. Subsequently, GoHSFA4a and GoNAC42 activate the MVA pathway genes (such as *HMGS* and *HMGR*) and gossypol pathway genes (such as *CYP736A196* and *DH1*), initiating the synthesis of hemigossypol, which undergoes homocoupling in the extracellular space of gland cavity to form gossypol. Additionally, GoHSFA4a and GoNAC42 regulate the biosynthesis of volatile terpenes through the modulation of terpene synthases. Volatile monoterpenes (such as β-myrcene and *trans*-β-ocimene) and hemigossypolone undergo Diels–Alder reactions to form heliocides, thereby providing evidence of the interplay between volatile terpenes and non-volatile terpenoids. The volatilization of terpenes is denoted by curved arrowheads.

Processing of scRNA-seq data

The reference genome (NBI_Gossypium_hirsutum_v1.1) and annotation files for *G. hirsutum* were obtained from https://www.cottongen.org. Raw sequencing data were aligned to the reference genome using Cell Ranger (v.7.1.0), and the resulting gene–cell matrices were used for subsequent analyses. Doublets were identified and removed from the scRNA-seq datasets using DoubletFinder (v.2.0.3) (McGinnis et al., 2019). We utilized Seurat (v.4.0.4) to integrate the three samples and perform initial quality control. Cells were filtered based on thresholds for gene counts (400–10 000), unique molecular identifier (UMI) counts (<50 000), and percentages of mitochondrial (<5%) and chloroplast genes (<20%). After filtering, the remaining cells underwent initial dimensionality reduction and clustering analysis. Notably, cell clusters with relatively low UMI counts but high percentages of mitochondrial or chloroplast genes, indicative of low-quality cells, were excluded from subsequent analyses (Supplemental Figure 1).

To account for the impact of cell-cycle heterogeneity, cell-cycle scores were calculated using the "CellCycleScoring" function. The cell-cycle genes were adapted from a previously published paper (Supplemental Table 8) (Zhang et al., 2021b). The influence of cell-cycle effects was regressed out using the "ScaleData" function.

For cell clustering, normalized and scaled data were utilized to identify highly variable features using the "FindVariableFeatures" function (nfeatures = 3000). Dimensionality reduction was performed using these features, followed by cell clustering. The resulting cell clusters were visualized using the uniform manifold approximation and projection method and annotated by examining the expression of known marker genes. The "FindMarkers" function was applied to identify genes specifically expressed in each cell cluster or population. Candidate genes were further refined based on criteria including avg_$\log_{2\ \text{fold change}}$ (>0.5), adjusted *P value* (<0.01), and pct.1/pct.2 (>2) (Supplemental Table 5). GO enrichment analysis was conducted for cell-type-enriched genes using the website https://cottonfgd.net/ . For cell types with over 300 specifically expressed genes, only the top 300 genes were used for GO analysis.

Evaluation of enzymatic treatment on cell clustering

Bulk RNA-seq libraries were generated from pooled protoplasts and intact cotton leaves, followed by sequencing on an Illumina Novaseq 6000 sequencer. Raw reads were preprocessed using fastp (v.0.23.2) (Chen et al., 2018) and subsequently aligned to the reference genome using hisat2 (v.2.1.0) (Kim et al., 2019). FeatureCounts (v.2.0.1) (Liao et al., 2014) was utilized to quantify gene expression, and DESeq2 (v.1.32.0) (Love et al., 2014) was employed to identify differentially expressed genes (DEGs) between protoplast and leaf samples. To assess the impact of enzymatic treatment on cell clustering and cell-type annotation, we reanalyzed the scRNA-seq data by excluding the DEGs from the highly variable features in the "FindVariableFeatures" step. The DEG-excluded subset of highly variable genes was then used for dimensional reduction and cell clustering. We utilized the same marker genes to annotate the cell clusters. The newly defined cell identities were compared with the previous results to evaluate the influence of transcriptional perturbations arising from enzymatic treatment on single-cell clustering and cell-type definition (Supplemental Table 9).

RNA *in situ* hybridization

The experimental procedure for scRNA-seq sample collection involved the collection of leaves from 15-day-old cotton plants. The probe sequences for all genes used in the study can be found in Supplemental Table 10. Tissue sections were subjected to an overnight incubation at 42°C with RNA digoxigenin-labeled mature probes (Servicebio). Subsequently, the slides were incubated with an anti-digoxigenin alkaline-phosphatase-conjugated antibody (Roche) at a dilution rate of 1:2000. Alkaline phosphatase substrate deposition was performed using the NBT/BCIP reaction, and the sections were imaged using an Olympus BX50 microscope equipped with a Spot Insight 6 camera.

Sample preparation for matrix-assisted laser desorption/ionization mass spectrometry imaging

The cotton ovules were embedded in 10% gelatin (wt/vol) solutions and rapidly frozen using liquid nitrogen for 2 min to obtain frozen gelatin blocks suitable for sectioning. Subsequently, sections with a thickness of 10 μm were obtained using a freezing microtome equipped with a retraction system (Leica Biosystems CM1950, Wetzlar, Germany) at a temperature of 20°C. The resultant sections were directly placed on indium-tin-oxide-coated glass slides, which were predried in a 50-mL centrifuge tube containing desiccant for 20 min prior to matrix application. As for cotton leaves, they were directly affixed onto indium tin oxide conductive glass using conductive double-sided adhesive. For the uniform application of the MALDI (matrix-assisted laser desorption/ionization) matrix, an iMLayer matrix sublimation apparatus (Shimadzu, Kyoto, Japan) was utilized. In positive-mode MALDI experiments, an α-cyano-4-hydroxycinnamic acid matrix was applied with a thickness of 0.9 μm. Following matrix deposition, the slide was placed inside a sealed container containing methanol vapor to facilitate matrix recrystallization, with a recrystallization time of 5 min.

MALDI–MSI analysis

All MALDI–MSI (matrix-assisted laser desorption/ionization mass spectrometry imaging) data were acquired utilizing an iMScope QT instrument (Shimadzu, Kyoto, Japan) equipped with an integrated optical microscope (magnification: ×5, ×10, and ×40), an atmospheric-pressure MALDI source, and a quadrupole-time-of-flight analyzer. The scanning area was determined using the optical microscope, and the sample sections were irradiated with a 355-nm Nd:YAG laser, employing the following parameters: laser shots, 700; repetition rate, 2000 Hz; laser diameter, 0 (5 μm); laser intensity, 45; detector voltage, 2.4 kV; and pitch (spatial resolution), 5 × 5 μm. The instrument was operated in negative mode, and spectra were acquired within the *m/z* range 150–650. The laser was calibrated using ink, ensuring accurate mass calibration of the MALDI–MSI instrument with the 2,5-ihydroxybenzoic acid matrix prior to each experiment. All acquired data were subjected to analysis using IMAGEREVEAL MS software (Shimadzu, Kyoto, Japan). The same software was employed for data visualization and relative quantitative analysis. A tolerance of 10 ppm was set for MALDI–MSI, allowing exclusion of ions adjacent to the target ion to minimize interference.

Histochemical GUS staining

The staining assay was conducted using a previously established protocol (Jefferson et al., 1987). Leaves from cotton plants containing the *pGoHSFA4a::GUS* and *pGoNAC42::GUS* constructs were subjected to vacuum infiltration using GUS staining buffer (sodium phosphate 50 mM [pH 7.0], EDTA 10 mM [pH 8.0], methanol 20%, TritonX-100 1%, potassium ferrocyanide 2 mM, potassium ferricyanide 2 mM, X-Gluc 1 mg/mL) for a duration of 30 min, followed by incubation at 37°C for 12 h. The stained materials were subsequently cleared using 70% ethanol at 25°C until the negative control was rendered colorless.

VIGS analysis

The VIGS assays were conducted following the previously published protocol (Gao et al., 2011). DNA fragments of the target genes, ranging from 300 to 500 bp in length, were amplified through polymerase chain reaction (PCR) using gene-specific primers as listed in Supplemental Table 10. Quantification of (+)-δ-cadinene-derived terpenoids was carried out using high-performance liquid chromatography, while volatile organic terpenoids were quantified using gas chromatography–mass spectrometry (GC–MS), utilizing the methods described in previous studies (Tian et al., 2018; Huang et al., 2020).

Gene expression analysis

Total RNA from cotton leaves was extracted and purified using the RNAprep Pure Plant Plus Kit (DP441, TIANGEN). A total of 2 μg purified RNA was subjected to reverse transcription using the EasyScript One-Step gDNA removal and cDNA Synthesis SuperMix (AE311-03, TransGen Biotech). Subsequently, quantitative real-time PCR was performed utilizing the SYBR Green Premix Pro Taq HS qPCR Kit (AG11701, Accurate Biotechnology). The cotton histone gene GhHIS3 (*Gh_D03G0370*) was employed as the internal control for normalization purposes (Shan et al., 2014).

Off-target analysis

Potential off-target sites in the cotton genome were identified using the CRISPR-P 2.0 Web tool (https://crispr.hzau.edu.cn/CRISPR2/) as described previously (Liu et al., 2017). Blast hits containing potential off-target sequences (with no more than five mismatched bases and located within the CDS region) were selected and amplified (Supplemental Table 7). PCR products were then subjected to Sanger sequencing for further analysis.

Dual-LUC assay

The upstream 2000-bp promoters of *GoHSFA4a*, *GoNAC42*, MVA pathway genes, and gossypol biosynthesis genes were amplified from genomic DNA extracted from *G. hirsutum* cv R15. These promoters were then ligated to the firefly LUC reporter gene. Subsequently, analysis of the LUC and fluorescence detection were conducted following the previously described protocol (Cao et al., 2020). The LUC activity was normalized to that of the internal control gene REN.

Electrophoretic mobility shift assay

The coding sequences of target proteins were cloned into the pET32a vector and subsequently expressed in *E. coli* strain BL21 (DE3). The recombinant protein containing a His tag was induced by supplementing 500 μM isopropyl-D-1-thiogalactopyranoside at 16°C for 20 h. Ni-NTA agarose beads (Invitrogen Life Technologies) were utilized to purify the protein. Cy5-labeled probes and mutant probes were synthesized by Sangon (Shanghai, China) as per the sequences listed in Supplemental Table 10. The purified protein was incubated with probes at 30°C for 30 min and subsequently separated by 5% native polyacrylamide gel electrophoresis in Tris/borate/EDTA buffer under 10 V/cm at 4°C. Fluorescence signals were detected using an image scanner (FLA-9000; Fujifilm).

Microscale thermophoresis

The coding sequences of *GoPGF* were cloned into the pET32a vector and subsequently expressed in *E. coli* strain BL21 (DE3). The expressed proteins were purified and labeled following the protocol of the Protein Labeling Kit RED-NHS 2nd Generation (MO-LO11, NanoTemper Tech). The interaction analysis of biomolecules was performed using the microscale thermophoresis Monolith NT.115 instrument.

Yeast one-hybrid

Three consecutive copies of the binding motifs in the promoters were cloned into pHIS2.1 vectors. The complete coding sequences of target proteins were recombined into pGADT7-Rec2 plasmids. The empty pGADT7-Rec2 vector and pHIS2.1 vectors served as negative controls. These recombination plasmids were assembled and subsequently introduced into the yeast strain AH109. Clones were cultured on selective medium SD/-Trp/-Leu at 30°C. After 2 days, the clones were transferred to selection medium (SD/-Trp/-Leu/-His medium) and grown for 3 days.

ChIP-seq assay

ChIP-seq was performed on 2-week-old *G. hirsutum* plantlets expressing the GoPGF-GFP fusion protein. The plantlets were treated with 1% (v/v) formaldehyde for 15 min to cross-link the proteins and DNA. Quenching of the cross-linking reaction was achieved by adding glycine to a final concentration of 125 mM and incubating for 5 min. The cross-linked plantlets were ground to a fine powder in liquid nitrogen, and the nuclei were isolated using a nucleus lysis buffer containing 0.1% SDS, 50 mM Tris–HCl (pH 8), and 10 mM EDTA (pH 8). The chromatin was then sonicated for 7 min using a Covaris S220 instrument with the following settings: peak power, 175; cycles/burst, 200; and duty factor, 20. This resulted in shearing of the chromatin into fragments of approximately 200–500 bp in length. The sonicated chromatin was subjected to immunoprecipitation using anti-GFP antibodies (Abcam, ab290). The immunocomplexes were then captured using Dynabead protein A and incubated for 2 h at 4°C with rotation. To remove non-specific binding, the immunoprecipitated material was washed six times for 5 min each with a ChIP dilution buffer containing 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris–HCl (pH 8), 167 mM NaCl, and protease inhibitors. Subsequently, two washes were performed using TE buffer (1 mM Tris–HCl [pH 8], 1 mM EDTA [pH 8]). The DNA fragments bound to the immunocomplexes were eluted by incubating with freshly prepared elution buffer (1% SDS, 0.1 M $NaHCO_3$) at 65°C for two 15-min incubations. The eluted chromatin was then subjected to reverse cross-linking by adding 16 μL 5 M NaCl and incubating overnight at 65°C. To remove RNA and proteins, the reverse cross-linked chromatin was treated with RNase and proteinase K and incubated for 3 h at 50°C. The DNA was then extracted using phenol-chloroform, and ethanol precipitation in the presence of GlycoBlue was performed to concentrate and purify the DNA. The resulting DNA pellet was resuspended in 10 μL nuclease-free water. Libraries for high-throughput sequencing were generated using 10 ng purified DNA and the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs). The quality of the libraries was assessed using the Agilent 2100 Bioanalyzer, and the libraries were sequenced using a 1 × 75 bp or similar length protocol on the NextSeq 500 platform (Illumina).

Data analysis and statistics

Compound structures were generated using ChemDraw 19.0 software. Chromatograms were plotted using Origin 2022. The quantification and statistical analysis of the compounds were performed using GraphPad Prism 8.2.1. All experimental data were expressed as the mean ± standard deviation, and statistical significance was determined using a two-tailed unpaired Student's *t*-test. The significance levels were denoted as $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, and $^{****}P < 0.0001$.

DATA AVAILABILITY

The authors hereby declare that all relevant data supporting the findings of this study are accessible within the research article and its supplemental information. The single-cell sequencing data produced in this study have been deposited in the National Center for Biotechnology Information under the accession number GEO: GSE243419. Additionally, the data have also been deposited in the Genome Sequence Archive at the National Genomics Data Center and the China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences (GSA: CRA012990), and are publicly accessible at https://ngdc.cncb.ac.cn/gsa. Publicly available RNA-seq data can be obtained from the Sequence Read Archive repository (https://www.ncbi.nlm.nih.gov/sra) under the accession numbers SRA: PRJNA248163 and PRJNA600707. Genome sequence data and gene identification can be found in the Cottongen database (https://www.cottongen.org/). Additionally, datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Molecular Plant Online*.

FUNDING

The research was supported by grants from the National Key R&D Program of China (2022YFF1001400) to J.-Q.H.; the National Natural Science Foundation of China (32388201) to X.-Y.C.; the Chinese Academy of

Sciences (XDB27020207) to X.-Y.C.; the Foundation of Youth Innovation Promotion Association of the Chinese Academy of Sciences to J.-Q.H.; the Young Elite Scientists Sponsorship Program by CAST (2019QNRC001) to J.-Q.H.; the Yunnan Revitalization Talent Support Program "Top Team" Project (202305AT350001) to X.-Y.C.; Winall Hi-tech Seed Co., Ltd. (GMLM2023) to X.-Y.C.; and the National Natural Science Foundation of China (32172041) to W.G.

AUTHOR CONTRIBUTIONS

J.-Q.H., J.-L.L., X.-Y.C., and W.G. conceived and designed this study; B.X., Y.L., and L.-J.W. provided advice; J.-L.L., W.-K.W., C.-H.Y., M.X., X.-X.G., and G.-B.N. performed the experiments; L.-X.C., J.-Q.H., J.-L.L., and W.-K.W. analyzed the scRNA-seq data; J.-l.D. and J.-L.L. performed the MALDI–MS experiments; W.G. performed the RNA *in situ* hybridization and ChIP-seq; and J.-L.L., L.-X.C., W.G., and J.-Q.H wrote the manuscript draft. X.-Y.C. revised the manuscript.

ACKNOWLEDGMENTS

We would like to express our gratitude to Personalbio Co., Ltd. (Shanghai, China) for providing scRNA-seq data and to Dr. Qi-Qi Jin from the Center for Excellence in Molecular Cell Science, CAS, for her invaluable assistance with bioinformatics analysis. Furthermore, we sincerely thank Wen-Juan Cai and Shui-Ning Yin for their valuable support in conducting microscopy analyses. Additionally, we extend our sincere thanks to Li Liu for her support throughout the microscale thermophoresis assays. No conflict of interest is declared.

Received: April 6, 2023
Revised: September 24, 2023
Accepted: October 12, 2023
Published: October 17, 2023

REFERENCES

Altman, D.W., Stipanovic, R.D., and Bell, A.A. (1990). Terpenoids in foliar pigment glands of A, D, and AD genome cottons: introgression potential for pest resistance. J. Hered. **81**:447–454.

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

Borisova-Mubarakshina, M.M., Vetoshkina, D.V., and Ivanov, B.N. (2019). Antioxidant and signaling functions of the plastoquinone pool in higher plants. Physiol. Plantarum **166**:181–198.

Briou, B., Améduri, B., and Boutevin, B. (2021). Trends in the Diels-Alder reaction in polymer chemistry. Chem. Soc. Rev. **50**:11055–11097.

Cao, J.F., Zhao, B., Huang, C.C., Chen, Z.W., Zhao, T., Liu, H.R., Hu, G.J., Shangguan, X.X., Shan, C.M., Wang, L.J., et al. (2020). The miR319-targeted GhTCP4 promotes the transition from cell elongation to wall thickening in cotton fiber. Mol. Plant **13**:1063–1077.

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

Chen, X.-Y., Chen, Y., Heinstein, P., and Davisson, V.J. (1995). Cloning, expression, and characterization of (+)-δ-cadinene synthase: a catalyst for cotton phytoalexin biosynthesis. Arch. Biochem. Biophys. **324**:255–266.

Chen, X., Liu, Y.Q., Wu, M.N., Yan, L., Chen, C.Y., Mu, Y.P., Liu, Y.J., Wang, M.Y., Chen, X.Y., and Mao, Y.B. (2023). A highly accumulated secretory protein from cotton bollworm interacts with basic helix-loop-helix transcription factors to dampen plant defense. New Phytol. **237**:265–278.

Chronis, C., Fiziev, P., Papp, B., Butz, S., Bonora, G., Sabri, S., Ernst, J., and Plath, K. (2017). Cooperative binding of transcription factors orchestrates reprogramming. Cell **168**:442–459.e20.

Denyer, T., Ma, X., Klesen, S., Scacchi, E., Nieselt, K., and Timmermans, M.C.P. (2019). Spatiotemporal developmental trajectories in the *Arabidopsis* root revealed using high-throughput single-cell RNA sequencing. Dev. Cell **48**:840–852.e5.

Desimoni, G. (1983). Natural products synthesis: through pericyclic reactions. Acs Monograph 180 (American Chemical Society).

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

Endo, M., Shimizu, H., Nohales, M.A., Araki, T., and Kay, S.A. (2014). Tissue-specific clocks in *Arabidopsis* show asymmetric coupling. Nature **515**:419–422.

Endo, S., Iwamoto, K., and Fukuda, H. (2018). Overexpression and cosuppression of xylem-related genes in an early xylem differentiation stage-specific manner by the AtTED4 promoter. Plant Biotechnol. J. **16**:451–458.

Fan, H., Liu, Y., Li, C.-Y., Jiang, Y., Song, J.-J., Yang, L., Zhao, Q., Hu, Y.-H., Chen, X.-Y., and Xu, J.-J. (2021). Engineering high coenzyme Q10 tomato. Metab. Eng. **68**:86–93.

Gao, X., Britt, R.C., Jr., Shan, L., and He, P. (2011). Agrobacterium-mediated virus-induced gene silencing assay in cotton. J. Vis. Exp. **54**:2938.

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.

Hagenbucher, S., Olson, D.M., Ruberson, J.R., Wäckers, F.L., and Romeis, J. (2013). Resistance mechanisms against arthropod herbivores in cotton and their interactions with natural enemies. Crit. Rev. Plant Sci. **32**:458–482.

Holopainen, J.K., and Gershenzon, J. (2010). Multiple stress factors and the emission of plant VOCs. Trends Plant Sci. **15**:176–184.

Huang, G., Huang, J.Q., Chen, X.Y., and Zhu, Y.X. (2021). Recent advances and future perspectives in cotton research. Annu. Rev. Plant Biol. **72**:437–462.

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

Huang, X.-Z., Xiao, Y.-T., Köllner, T.G., Jing, W.-X., Kou, J.-F., Chen, J.-Y., Liu, D.-F., Gu, S.-H., Wu, J.-X., Zhang, Y.-J., and Guo, Y.Y. (2018). The terpene synthase gene family in *Gossypium hirsutum* harbors a linalool synthase GhTPS12 implicated in direct defence responses against herbivores. Plant Cell Environ. **41**:261–274.

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.

Jefferson, R.A., Kavanagh, T.A., and Bevan, M.W. (1987). GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J. **6**:3901–3907.

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

Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. **37**:907–915.

Kim, J.-Y., Symeonidi, E., Pang, T.Y., Denyer, T., Weidauer, D., Bezrutczyk, M., Miras, M., Zöllner, N., Hartwig, T., Wudick, M.M.,

et al. (2021). Distinct identities of leaf phloem cells revealed by single cell transcriptomics. Plant Cell **33**:511–530.

Kliebenstein, D.J. (2004). Secondary metabolites and plant/environment interactions: a view through *Arabidopsis thaliana* tinged glasses. Plant Cell Environ. **27**:675–684.

Li, C., Wood, J.C., Vu, A.H., Hamilton, J.P., Rodriguez Lopez, C.E., Payne, R.M.E., Serna Guerrero, D.A., Gase, K., Yamamoto, K., Vaillancourt, B., et al. (2023). Single-cell multi-omics in the medicinal plant *Catharanthus roseus*. Nat. Chem. Biol. **19**:1031–1041.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics **30**:923–930.

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., Wang, L.-J., et al. (2023). Dirigent gene editing of gossypol enantiomers for toxicity-depleted cotton seeds. Nat. Plants **9**:605–615.

Lin, K.F., Tsai, M.Y., Lu, C.A., Wu, S.J., and Yeh, C.H. (2018). The roles of *Arabidopsis HSFA2, HSFA4a*, and *HSFA7a* in the heat shock response and cytosolic protein response. Bot. Stud. **59**:15.

Liu, H., Ding, Y., Zhou, Y., Jin, W., Xie, K., and Chen, L.-L. (2017). CRISPR-P 2.0: An improved CRISPR-Cas9 tool for genome editing in plants. Mol. Plant **10**:530–532.

Liu, J., Benedict, C.R., Stipanovic, R.D., and Bell, A.A. (1999). Purification and characterization of *S*-adenosyl-l-methionine: desoxyhemigossypol-6-*O*-methyltransferase from cotton plants. an enzyme capable of methylating the defense terpenoids of cotton. Plant Physiol. **121**:1017–1024.

Liu, Z., Zhou, Y., Guo, J., Li, J., Tian, Z., Zhu, Z., Wang, J., Wu, R., Zhang, B., Hu, Y., et al. (2020). Global dynamic molecular profiling of stomatal lineage cell development by single-cell RNA sequencing. Mol. Plant **13**:1178–1193.

Lopez-Anido, C.B., Vatén, A., Smoot, N.K., Sharma, N., Guo, V., Gong, Y., Anleu Gil, M.X., Weimer, A.K., and Bergmann, D.C. (2021). Single-cell resolution of lineage trajectories in the *Arabidopsis* stomatal lineage and developing leaf. Dev. Cell **56**:1043–1055.e4.

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

Loughrin, J.H., Manukian, A., Heath, R.R., and Tumlinson, J.H. (1995). Volatiles emitted by different cotton varieties damaged by feeding beet armyworm larvae. J. Chem. Ecol. **21**:1217–1227.

Loughrin, J.H., Manukian, A., Heath, R.R., Turlings, T.C., and Tumlinson, J.H. (1994). Diurnal cycle of emission of induced volatile terpenoids by herbivore-injured cotton plant. Proc. Natl. Acad. Sci. USA **91**:11836–11840.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. **15**:550.

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

McCall, P.J., Turlings, T.C., Loughrin, J., Proveaux, A.T., and Tumlinson, J.H. (1994). Herbivore-induced volatile emissions from cotton (*Gossypium hirsutum* L.) seedlings. J. Chem. Ecol. **20**:3039–3050.

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.e4.

Mellon, J.E., Dowd, M.K., Beltz, S.B., and Moore, G.G. (2014). Growth inhibitory effects of gossypol and related compounds on fungal cotton root pathogens. Lett. Appl. Microbiol. **59**:161–168.

Nützmann, H.W., Huang, A., and Osbourn, A. (2016). Plant metabolic clusters - from genetics to genomics. New Phytol. **211**:771–789.

Pare, P.W., and Tumlinson, J.H. (1997). De Novo biosynthesis of volatiles induced by insect herbivory in cotton plants. Plant Physiol. **114**:1161–1167.

Park, S.H., Scheffler, J., Scheffler, B., Cantrell, C.L., and Pauli, C.S. (2019). Chemical defense responses of upland cotton, *Gossypium hirsutum* L. to physical wounding. Plant Direct **3**, e00141.

Pichersky, E., and Raguso, R.A. (2018). Why do plants produce so many terpenoid compounds? New Phytol. **220**:692–702.

Rodriguez-Saona, C., Crafts-Brandner, S.J., Williams, L., 3rd, and Paré, P.W. (2002). Lygus hesperus feeding and salivary gland extracts induce volatile emissions in plants. J. Chem. Ecol. **28**:1733–1747.

Rosenkranz, M., Chen, Y., Zhu, P., and Vlot, A.C. (2021). Volatile terpenes–mediators of plant-to-plant communication. Plant J. **108**:617–631.

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

Rusconi, F., Simeoni, F., Francia, P., Cominelli, E., Conti, L., Riboni, M., Simoni, L., Martin, C.R., Tonelli, C., and Galbiati, M. (2013). The *Arabidopsis thaliana MYB60* promoter provides a tool for the spatio-temporal control of gene expression in stomatal guard cells. J. Exp. Bot. **64**:3361–3371.

Sabbadin, F., Urresti, S., Henrissat, B., Avrova, A.O., Welsh, L.R.J., Lindley, P.J., Csukai, M., Squires, J.N., Walton, P.H., Davies, G.J., et al. (2021). Secreted pectin monooxygenases drive plant infection by pathogenic oomycetes. Science **373**:774–779.

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

Sawchuk, M.G., Donner, T.J., Head, P., and Scarpella, E. (2008). Unique and overlapping expression patterns among members of photosynthesis-associated nuclear gene families in Arabidopsis. Plant Physiol. **148**:1908–1924.

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

Smet, W., Sevilem, I., de Luis Balaguer, M.A., Wybouw, B., Mor, E., Miyashima, S., Blob, B., Roszak, P., Jacobs, T.B., Boekschoten, M., et al. (2019). DOF2.1 controls cytokinin-dependent vascular cell proliferation downstream of TMO5/LHW. Curr. Biol. **29**:520–529.e6.

Smith, Z.D., Sindhu, C., and Meissner, A. (2016). Molecular features of cellular reprogramming and development. Nat. Rev. Mol. Cell Biol. **17**:139–154.

Sun, S., Shen, X., Li, Y., Li, Y., Wang, S., Li, R., Zhang, H., Shen, G., Guo, B., Wei, J., et al. (2023). Single-cell RNA sequencing provides a high-resolution roadmap for understanding the multicellular compartmentation of specialized metabolism. Nat. Plants **9**:179–190.

Sun, X., Feng, D., Liu, M., Qin, R., Li, Y., Lu, Y., Zhang, X., Wang, Y., Shen, S., Ma, W., and Zhao, J. (2022). Single-cell transcriptome reveals dominant subgenome expression and transcriptional response to heat stress in Chinese cabbage. Genome Biol. **23**:262.

Tenorio Berrío, R., Verstaen, K., Vandamme, N., Pevernagie, J., Achon, I., Van Duyse, J., Van Isterdael, G., Saeys, Y., De Veylder,

L., Inzé, D., and Dubois, M. (2022). Single-cell transcriptomics sheds light on the identity and metabolism of developing leaf cells. Plant Physiol. **188**:898–918.

Thulasiram, H.V., Erickson, H.K., and Poulter, C.D. (2007). Chimeras of two isoprenoid synthases catalyze all four coupling reactions in isoprenoid biosynthesis. Science **316**:73–76.

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

von Koskull-Döring, P., Scharf, K.-D., and Nover, L. (2007). The diversity of plant heat stress transcription factors. Trends Plant Sci. **12**:452–457.

Vranová, E., Coman, D., and Gruissem, W. (2013). Network analysis of the MVA and MEP pathways for isoprenoid synthesis. Annu. Rev. Plant Biol. **64**:665–700.

Wang, S., Alseekh, S., Fernie, A.R., and Luo, J. (2019). The structure and function of major plant metabolite modifications. Mol. Plant **12**:899–919.

Wang, Y., Huan, Q., Chu, X., Li, K., and Qian, W. (2020). Single-cell transcriptome analyses recapitulate the cellular and developmental responses to abiotic stresses in rice. Preprint at bioRxiv.

Wu, C., Cheng, H., Li, S., Zuo, D., Lin, Z., Zhang, Y., Lv, L., Wang, Q., and Song, G. (2021). Molecular cloning and characterization of *GhERF105*, a gene contributing to the regulation of gland formation in upland cotton (*Gossypium hirsutum* L.). BMC Plant Biol. **21**:102.

Yang, C.Q., Wu, X.M., Ruan, J.X., Hu, W.L., Mao, Y.B., Chen, X.Y., and Wang, L.J. (2013). Isolation and characterization of terpene synthases in cotton (*Gossypium hirsutum*). Phytochemistry **96**:46–56.

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, T.-Q., Xu, Z.-G., Shang, G.-D., and Wang, J.-W. (2019). A single-cell RNA sequencing profiles the developmental landscape of *Arabidopsis* root. Mol. Plant **12**:648–660.

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

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

Zozina, V.I., Covantev, S., Goroshko, O.A., Krasnykh, L.M., and Kukes, V.G. (2018). Coenzyme Q10 in cardiovascular and metabolic diseases: current state of the problem. Curr. Cardiol. Rev. **14**:164–174.