



Spatiotemporal transcriptome and metabolome landscapes of cotton fiber during initiation and early development

Received: 20 October 2022

Accepted: 2 January 2025

Published online: 20 January 2025

Check for updates

Xuwu Sun ^{1,2,7}, Aizhi Qin^{2,7}, Xingxing Wang^{1,3,7}, Xiaoyang Ge^{1,3,7}, Zhixin Liu^{2,7}, Chenxi Guo^{2,7}, Xiaole Yu^{2,7}, Xianliang Zhang^{1,3,7}, Yao Lu^{4,7}, Jincheng Yang², Jiuming He ⁵, Yaping Zhou², Yumeng Liu², Mengke Hu², Hao Liu², Zihao Zhao², Guanjing Hu ^{1,3}, Wei Li^{1,3}, Xinshan Zang ^{1,3}, Shuai Dai^{1,3}, Susu Sun², Lenin Yong-Villalobos ⁶, Luis Herrera-Estrella ⁶, Lam-Son Phan Tran⁶ & Xiongfeng Ma ^{1,3}

Cotton fibers are single cells that develop from the epidermal cells in the outer integument of developing seeds. The processes regulating fiber cell development have been extensively studied; however, the spatiotemporal transcriptome and metabolome profiles during the early stages of fiber development remain largely unknown. In this study, we profile the dynamics of transcriptome and metabolome during the early stages of cotton fiber cell development using a combination of spatial transcriptomic, single-cell transcriptomic, and spatial metabolomic analyses. We identify the key genes (e.g., *DOX2*, *KCS19.4*, *BEE3*, and *HOS3.7*) and metabolites (e.g., linoleic acid, spermine, spermidine, and α-linolenic acid) that may regulate the early development of fiber cells. Finally, knockdown and gain-of-function analyses identify the crucial role of *GhBEE3/Gh_A09G062900* in cotton fiber initiation. We also construct a publicly accessible website (<https://cotton.cricaas.com.cn/ovule/>) for visualization of the spatiotemporal gene expression in cotton, providing a reference dataset for further studies on cotton fiber development.

Cotton (*Gossypium* species) is one of the most important cash crops in the world. It is cultivated in about 80 countries, occupies approximately 2.5% of the world's arable land, and generates roughly 500 billion dollars of annual revenue worldwide¹. Upland cotton (*G. hirsutum* L.) is the most widely cultivated allotetraploid cotton variety, accounting for ~ 95% of global cotton fiber production^{2,3}. Cotton fibers are single cells originating on the outer epidermal surface of developing seeds. Approximately 25% of the ovular epidermal cells

differentiate into lint fibers at the early stage of seed development⁴. These fibers develop through four distinct stages, initiation (~2 to 5 days post anthesis; DPA), elongation and primary wall synthesis (3 to 20 DPA), secondary cell wall thickening (16 to 40 DPA), and fiber maturation (40 to 50 DPA)^{5,6}. The first two stages determine the number and length of the fibers produced by each ovule/seed, which determines the final yield of cotton⁷. The latter two stages, associated with cell wall thickening, determine fiber strength and fineness⁷.

¹State Key Laboratory of Cotton Bio-breeding and Integrated Utilization, Institute of Cotton Research of the Chinese Academy of Agricultural Sciences, Anyang, China. ²State Key Laboratory of Crop Stress Adaptation and Improvement, School of Life Sciences, Henan University, Kaifeng, China. ³Zhengzhou Research Base, State Key Laboratory of Cotton Bio-breeding and Integrated Utilization, School of Agricultural Sciences, Zhengzhou University, Zhengzhou, China. ⁴Shanghai OE Biotech Co., Ltd, Shanghai, China. ⁵State Key Laboratory of Bioactive Substance and Function of Natural Medicines, Institute of Materia Medica, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China. ⁶Institute of Genomics for Crop Abiotic Stress Tolerance, Department of Plant and Soil Science, Texas Tech University, Lubbock, TX, USA. ⁷These authors contributed equally: Xuwu Sun, Aizhi Qin, Xingxing Wang, Xiaoyang Ge, Zhixin Liu, Chenxi Guo, Xiaole Yu, Xianliang Zhang, Yao Lu. e-mail: maxiongfeng@caas.cn

However, the regulatory factors that determine these developmental stages remain largely unknown. Understanding the mechanisms determining the cell fate and regulating the initiation and elongation of cotton fiber will provide an opportunity to improve the quality and yield of cotton fiber.

Progress in sequencing and omics technology and functional studies have led to some advances in the identification of transcription factors (TFs) involved in regulating fiber differentiation from ovule epidermal cells. Genes encoding TFs like myeloblastosis-type (MYB-type such as *GhMYB25*⁸, *GhMYB25-like*⁹, and *GhMYB2*¹⁰), homeodomain-leucine zipper (HD-ZIP)-type *homeodomain protein 1* (*GhHD1*)¹¹, defense-related *protodermal factor 1* (*GbPDF1*)¹², and auxin-related *auxin-responsive protein 2* (*ARF2*)¹³ play crucial roles in fiber initiation, while the auxin transporter-encoding gene *pin-formed 3a* (*GhPIN3a*)¹⁴, calcium sensor *calmodulin 7-like* (*GhCaM7-like*)¹⁵, *GhMYB109*¹⁶, glycolysis-related *cytosolic pyruvate kinase 6* (*GhPK6*)¹⁷, HD-ZIP *homeobox 3* (*GhHOX3*)¹⁸, and basic helix-loop-helix (bHLH) TF-encoding *paclobutrazol resistance 1* (*GhPRE1*)¹⁹ regulate the elongation of cotton fibers. However, several studies have demonstrated that the expression of TF-encoding genes responsible for fiber differentiation from ovule epidermal cells is also regulated by phytohormones⁶. Abscisic acid (ABA), auxins, brassinosteroids (BRs), cytokinins, gibberellin (GA), jasmonic acid (JA), and ethylene regulate cotton fiber development²⁰. More specifically, auxins and BRs promote fiber initiation and elongation^{21,22}, GAs and ethylene positively affect fiber elongation²⁰, and JA and GA synergistically promote fiber initiation²³. On the other hand, cytokinins and ABA inhibit cotton fiber elongation²⁴. In addition, several secondary metabolites have also been correlated with cotton fiber development. The flavonoid naringin negatively influences fiber development²⁵, while linolenic acid (C18:3) promotes cotton fiber elongation²⁶. Reactive oxygen species (ROS) and calcium signaling are other crucial factors that regulate cotton fiber development¹⁵.

Although some achievements have been made in the exploration of the regulatory mechanisms of fiber development at the molecular level, these are still unclear to a large extent. Single-cell RNA sequencing (scRNA-seq) is a technology that provides insights into the heterogeneity of gene expression among cells and the dynamics of cell types during development. This technology allows the identification of candidate genes associated with specific traits at the single-cell level^{27–29}. However, applying scRNA-seq in non-model plants, particularly in cotton, is challenging due to the lack of known marker genes for defining each cell type^{30,31}. Qin et al. recently analyzed the single-cell transcriptome during initiation and provided information on the regulation of fiber development, but at the same time, important spatial location information was not available yet³². The emerging spatial transcriptomic (ST) technology is widely used in animals to help overcome this problem by providing markers to associate gene expression with cell location and cell types³³. Similarly, the recently developed spatial metabolomics (SM) based on mass spectrometry imaging (MSI) helps visualize quantification and identification of the distribution of specific metabolites in specific tissues³⁴. SM has demonstrated potential in characterizing the regulatory mechanisms underlying development in various plants^{35–40} but has not yet been applied to cotton. The mechanism of cotton fiber development is complex and cannot be thoroughly studied with a single omics research method. Therefore, the multi-omics analysis of ST, scRNA and SM will help us to further analyze the molecular mechanism of cotton fiber initiation and early development.

In this work, we carry out transcriptome and metabolome profiling of ovules based on ST, scRNA-seq, and SM to systematically analyze the dynamic changes at the single-cell level during fiber initiation and early elongation in cotton (Fig. 1). We then aim to spatially map the cell types and achieve *in situ* visualization of gene and metabolite expression patterns. Further, to integrate the ST and SM datasets, we

consistently set the spatial resolution in the study at 100 μm to identify the marker genes and metabolites that play key roles in regulating fiber development. Finally, we explore the role of a representative marker gene, *GhBEE3*, in regulating the initiation of fiber cells via knockdown and gain-of-function approaches. These findings on the spatio-temporal changes in gene expression and metabolite concentrations during the initiation and elongation of cotton fiber will facilitate further studies on cotton fiber development. In conclusion, we provide a strategic reference for future studies, as well as a methodological model for accelerating the application of single-cell spatial multiomics to diverse crop plants.

Results

Spatio-temporal gene expression patterns in cotton bolls

Cotton fiber initiation occurs from ~2-DPA to 5-DPA. Cotton fiber cells begin to protrude at ~0.5-DPA and display numerous protrusions by 0-DPA, suggesting that the cell fate is determined as early as ~1-DPA or ~1.5-DPA³². In this study, we examined the spatiotemporal anatomical variations at the early stages of fiber development using paraffin-embedded sections of cotton bolls at ~1.5-DPA, ~1-DPA, 0-DPA, 1-DPA, 3-DPA, and 5-DPA (Supplementary Fig. 1). The cross-sections stained with toluidine blue showed that fiber cell initiation starts at 0-DPA. Subsequently, the fiber cells start to elongate from the outer integument at 3-DPA and further elongate at 5-DPA.

We then sequenced the TM-1 cotton boll samples from the above-mentioned stages of development using ST technology to obtain the spatiotemporal gene expression dynamics and elucidate the mechanisms underlying fiber development. A total of eight samples at six stages of development were sequenced, including two cotton bolls collected from two different cotton plants that were maintained as biological replicates at 0-DPA and 5-DPA to ensure accuracy. The RNA integrity number (RIN) of cotton boll sections at all developmental stages was evaluated (Supplementary Table 1), and the qualified samples with RIN > 7 were stained with Hematoxylin-Eosin (H&E) (Fig. 2a), then ST sequencing was performed using 10 × Genomics Visium. The optimal tissue permeabilization conditions were determined through preliminary tissue optimization experiments, including incubating samples for 24 min in a specifically modified permeabilization enzyme cocktail (2% w/v Cellulase R10, 0.4% w/v Macerozyme R10, 1% w/v Pectinase, 1% w/v Hemicellulase, 0.4% Snailase) (Supplementary Fig. 2). Tissue permeabilization was carried out for all the samples subsequently under the optimized conditions, and the cDNA was synthesized from the captured mRNA and used for library preparation. The library had cDNA fragments of the desired length (Supplementary Fig. 3) and was sequenced. Then, the spatial barcodes unique to each spot were used to determine the spatial location of genes. After quality control, our ST experiment using cotton boll samples collected at different DPA identified ~50,000 genes (Supplementary Data 1), which is within the expected range. In detail, we obtained 2651–3998 high-quality spots per section, 5487–11,210 unique molecular identifiers (UMIs) per spot, 3177–5556 expressed genes per spot, and 0.0000115%–0.0000218% mitochondrial genes per spot (Supplementary Fig. 4a–c and Supplementary Data 1). Then, Pearson correlation analysis was conducted among the eight sample sections, and a strong correlation among the samples was obtained (Supplementary Fig. 4d–j). Each spot captured an average of 1–10 cells (Supplementary Fig. 4k).

Subsequent uniform manifold approximation and projection (UMAP) analysis of all spots in the samples from six developmental stages showed 19 cell clusters (Fig. 2b). Next, to assess the spatial organization of the clusters, we mapped the spots to their original coordinates in the tissue sections and annotated them (Fig. 2c). The resulting clusters corresponded to specific regions in the bolls of six DPA. ‘Mesocarp/Diaphragm cells’ (clusters 1, 2, 6, 9, and 14) appeared in the boll slices at most stages, while ‘fiber cells’ (clusters 3, 10, and 19)

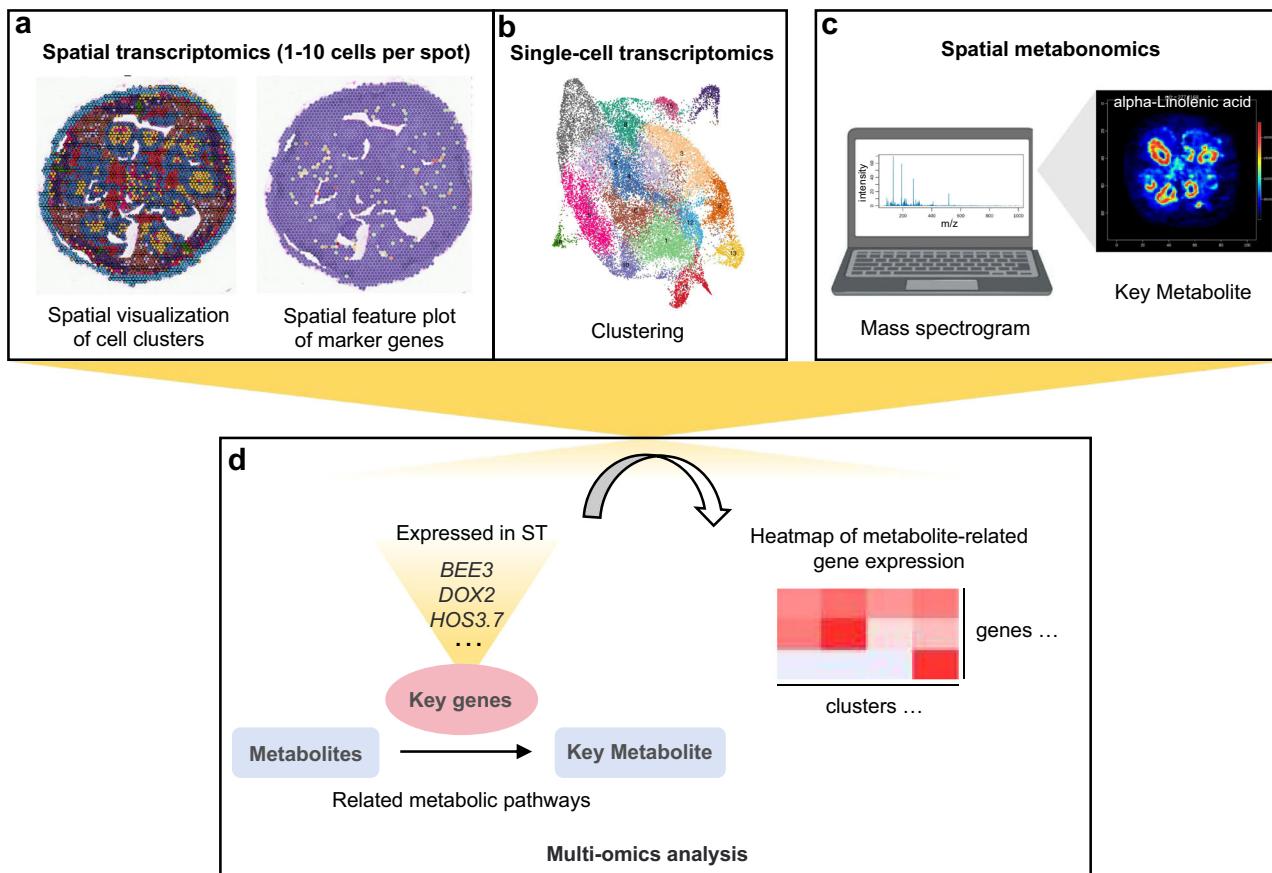


Fig. 1 | A molecular approach towards fiber development atlas in cotton. The study performed (a) spatial transcriptomics to characterize the spatiotemporal transcriptome patterns in cotton bolls from –1.5-DPA to 5-DPA by ST; (b) single-cell transcriptomics (scRNA-seq) to dissect cell heterogeneity in ovules from –1.5-DPA to 1-DPA; (c) spatial metabolomics to identify key metabolites by MSI from

–1.5-DPA to 5-DPA; (d) ST and SM integrated analysis to generate a multi-omic atlas. The genes and clusters (cell types) of interest were selected, and the heatmap of metabolite-related gene expression was generated according to the expression level of each cell cluster (as shown in Fig. 2d).

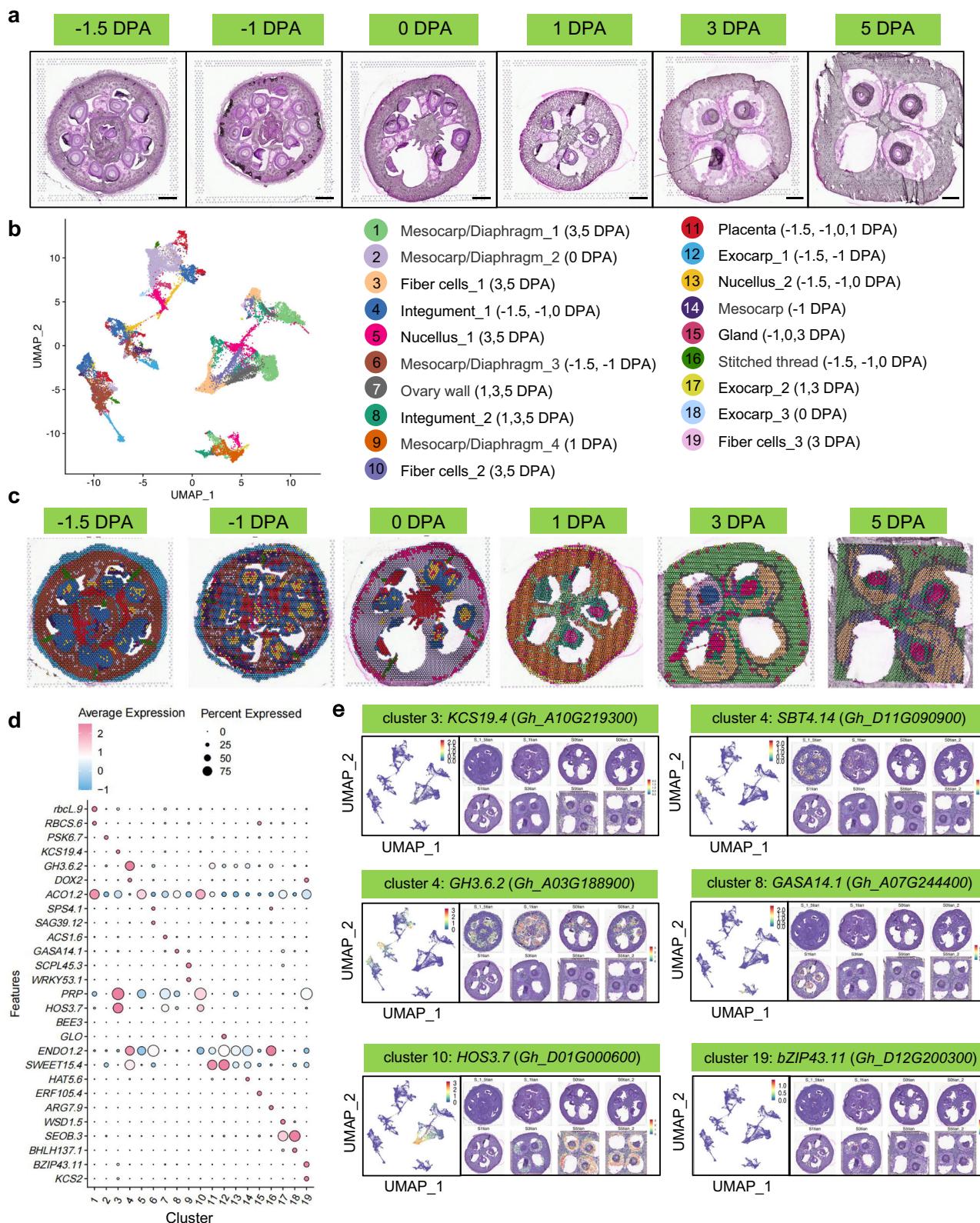
appeared in boll slices at 3-DPA and 5-DPA (Fig. 2b, c and Supplementary Data 2).

To further characterize the clusters, we identified the differentially expressed genes (DEGs) between them (Supplementary Data 3). The DEGs with the highest expression levels in each cell cluster were identified as the key marker genes (Fig. 2d). For example, *gretchen hagen 3* (*GH3.6.2*) (*Gh_A03G188900*), *subtilase 4.14* (*SBT4.14*) (*Gh_D11G090900*), and *alpha dioxygenase 2* (*DOX2*) (*Gh_A09G193100*) were found highly expressed in ‘integument_1’ (cluster 4) (Fig. 2d). The *sugars will eventually be exported transporters 15* (*SWEET15*) (*Gh_D04G160000*) was highly expressed in ‘integument_1’ (cluster 4), and ‘nucellus’ (clusters 5 and 13) (Fig. 2d). The *brassinosteroid enhanced expression 3* (*BEE3*) (*Gh_A09G062900*) was highly expressed in ‘placenta’ (cluster 11), and *GA-stimulated 14.1* (*GASA14.1*) (*Gh_A07G244400*) was highly expressed in ‘integument_2’ (cluster 8) (Fig. 2d). Besides, the *3-ketoacyl-CoA synthase 19.4* (*KCS19.4*) (*Gh_A10G219300*) and *high expression of osmotically responsive 3.7* (*HOS3.7*) (*Gh_D01G000600*) were found to be highly expressed in ‘fiber cells’ (clusters 3 and 10) at two different time points during the early elongation of fiber cells (3-DPA and 5-DPA; Fig. 2e). *Gh_D12G200300*, an ortholog of the *Arabidopsis thaliana* marker gene *basic leucine zipper 43.11* (*bZIP43.11*), showed high expression in ‘fiber cells_3’ (cluster 19) (Fig. 2e).

Further gene ontology (GO) enrichment analysis of the DEGs in the cell clusters indicated that ‘fruit development’ and ‘seed development’ terms were associated with ‘exocarp_1’ (cluster 12), and ‘nucellus_2’ (cluster 13) (Supplementary Table 2). The DEGs in ‘integument_2’ (cluster 8)

(cluster 8) were also associated with the ‘organonitrogen compound biosynthetic process’ (Supplementary Table 2). Meanwhile, the DEGs in ‘mesocarp’ or ‘diaphragm’ (clusters 1, 2, 6, 9, and 14) were associated with ‘macromolecule biosynthetic process’, ‘cellular metabolic process’, ‘fruit development’, ‘organonitrogen compound biosynthetic process’, and ‘response to stimulus’ (Supplementary Table 2), and those of ‘fiber cells’ (clusters 3, 10, and 19) and ‘integument_1’ (cluster 4) were associated with ‘seed trichome differentiation’ (Supplementary Table 2).

To examine the robustness of our ST results, ST data from all cotton boll samples were compared with scRNA-seq data from the outer integument and fiber of Xu142 0-DPA sample by Qin et al.³². Specifically, the AddModuleScore function from Seurat was used to score the expression accumulation of outer integument and fiber enrichment gene datasets from Qin et al.³² in each spot of our ST data. Then, the expression accumulation scores of these gene datasets were plotted as a feature plot for visualization (Supplementary Fig. 5a). It was obvious that these gene datasets were mainly highly expressed in ‘fiber cells’ (clusters 3, and 19) and ‘integument’ (clusters 4, and 8), which was consistent with the results of their research. Further, we performed a Venn diagram to show the correlation between the outer integument and fiber-enriched gene datasets from Qin et al.³² and the gene datasets of ‘fiber cells’ (clusters 3, 10, and 19) and ‘integument’ (clusters 4, and 8) in our ST data, and the results showed that there was a certain correlation between the two gene datasets (Supplementary Fig. 5b). In addition, consistent with previous study by Deng et al.¹², our ST analysis revealed high expression of *PDF1* (*Gh_D04G040400*), a



known key marker gene for fiber initiation, in the 'fiber cells' and 'integument' in the young developing cotton bolls (Supplementary Fig. 5c).

In order to further analyze the mechanisms underlying fiber development, we mapped the gene expression of cotton boll samples at each development stage to explore the genes specifically expressed at each stage (Supplementary Figs. 6–11). The DEGs identified from the

samples for each developmental stage of the cotton boll were shown in Supplementary Data 4–9. The analysis revealed that the cell populations at the six stages of early fiber development shared similar biological processes (Supplementary Tables 3–8); however, each cell population had its own set of uniquely expressed genes (Supplementary Figs. 6–11), from which the marker genes of specific tissue or cell type were identified. For example, *Gh D12G199400*, an ortholog of the

Fig. 2 | Spatiotemporal gene expression during six stages of cotton fiber development. **a** Hematoxylin and Eosin (H&E) stained sections of cotton bolls at -1.5-DPA, -1-DPA, 0-DPA, 1-DPA, 3-DPA, and 5-DPA. The H&E stain experiment was repeated three times. Scale bar, 1 mm. **b** UMAP visualization of cotton boll cells. Each point in the figure represents a spot; different colors indicate different groups of spots. The cell types in each cluster were annotated based on the tissue information in (c). From cluster 1 to cluster 19, the number of spots contained in each cluster is 3540, 3392, 2700, 2469, 1966, 2077, 1572, 1289, 1190, 1114, 1014, 988, 689, 462, 378, 183, 142, 123, and 119, respectively. **c** All clusters obtained by the UMAP analysis are mapped to their specific locations. The mapping shows the localization of 19 clusters. **d** Dotplot shows the expression of the representative marker genes

expressed in each cell cluster. Each row represents a marker gene, and each column represents a cell cluster. The size of the dots represents the percentage of cells in the clusters that express the gene. Gene expression levels are indicated by a color key ranging from blue (low) to pink (high). **e** Annotation of the major clusters of cotton bolls. *KCS19.4*, *HOS3.7*, and *bZIP43.11* are the marker genes of 'fiber cells', while *SBT4.14*, *GH3.6.2*, and *GASA14.1* are the marker genes of 'integument.' The UMAP plot of transcript accumulation for the marker genes and the spatial distribution of these marker genes in cotton bolls are shown. The color intensity indicates the relative transcript level of each marker gene. Source data are provided as a Source Data file.

Arabidopsis teosinte branched 1, cycloidia, and proliferating cell nuclear antigen factor 5.3 (TCP5.3), was found highly expressed in 'mesocarp' (cluster 1) of the -1.5-DPA boll (Supplementary Fig. 6b). *Gh_D09G117000*, an ortholog of the *Arabidopsis fasciclin-like arabinogalactan-protein 11.3 (FLA11.3)*, was highly expressed in 'fiber cells_1' (cluster 2) of the 3-DPA boll (Supplementary Fig. 10b). At 5-DPA, the *Arabidopsis* homologous gene *transmembrane kinase 3 (TMK3)* (*Gh_D09G144600*), was highly expressed in 'fiber cells_1' (cluster 1) (Supplementary Fig. 11b).

ScRNA-seq analysis of gene expression in cotton ovule

We then analyzed the gene expression patterns in cotton ovules at four different early developmental stages from -1.5- to 1-DPA by scRNA-seq to determine the heterogeneity of cells. Four cotton ovules were collected at -1.5-, -1-, 0-, and 1-DPA for library construction and high-throughput sequencing. After filtering, we obtained 7395, 4535, 7528, and 8307 high-quality cells and 52,403, 49,232, 57,608, and 52,379 genes from -1.5-, -1-, 0-, and 1-DPA samples, respectively (Supplementary Table 9). The average number of UMI per cell ranged from 1390 to 4213, the average gene number per cell ranged from 1022 to 2501, and the average proportion of mitochondrial genes per cell ranged from 0.0000068% to 0.0000115% (Supplementary Fig. 12a–c and Supplementary Table 9). We then identified the DEGs for all four early developmental stages from the scRNA-seq data (Supplementary Data 10–14).

Furthermore, to examine the robustness of the sequencing results, we compared the scRNA-seq data for each DPA sample with the corresponding ST data. All the Pearson correlation coefficients (*R*) were > 0.7 with *p*-value < 2.2e-16, which indicated a strong correlation between the two datasets (Supplementary Fig. 12d). We then screened the genes highly expressed in the corresponding cell clusters identified by ST and scRNA-seq (Supplementary Data 15). Subsequent UMAP analysis of data from the four single-cell components (inner integument, outer integument, nucellus, and diaphragm; (Fig. 3a) from ovules at -1.5-, -1-, 0-, and 1-DPA showed 16 cell clusters (Fig. 3b). The percentage of cells in each cluster ranged from 0.8% to 12.86% (Supplementary Table 10). We annotated the 16 cell clusters based on the expression levels of the marker genes and the annotation information of the ST data (Fig. 3b). We further analyzed gene expression and identified the DEGs in each cluster. The expression of the top ten DEGs was presented as a dot plot (Fig. 3c), and the expression patterns of some representative genes were characterized (Fig. 3d). Some of the identified DEGs have been reported to be involved in regulating hormone accumulation, such as the ortholog (*Gh_A05G105700*) of the *Arabidopsis* marker gene *OPDA reductase 2 (OPR2)*, which regulate JA accumulation⁴¹. The cotton homolog (*Gh_D08G075000*) of the *Arabidopsis* gene *indole-3-pyruvate monooxygenase 10.4 (YUC10.4)* has been reported to mediate auxin synthesis⁴². GO analysis of the DEGs in each cluster indicated enrichment of the 'fatty acid biosynthetic process', 'seedling development', and 'post-embryonic development', especially in the 'outer integument' and 'nucellus' (clusters 1, 2, 3, 5, and 15) (Supplementary Table 11). Besides, the 'outer integument' (clusters 3, 7, and 10) had

DEGs related to 'seed trichome differentiation', and 'seed trichome elongation' GO terms (Supplementary Table 11).

In the same way as ST analysis, correlations were compared between the scRNA-seq data from our all cotton boll samples and the scRNA-seq data from outer integument and fiber of Xu142 0-DPA sample by Qin et al.³². The AddModuleScore function was used to score the expression accumulated of outer integument and fiber enrichment gene datasets from Qin et al.³² in each cell of our scRNA-seq data (Supplementary Fig. 13a). The gene datasets of Qin et al.³² were found to be highly expressed in 'outer integument' (clusters 3, 13, and 15) of our scRNA-seq data (Supplementary Fig. 13a), which was consistent with their findings. Further, we performed a Venn diagram to show the correlation between the outer integument and fiber-enriched gene datasets from Qin et al.³² and the gene datasets of 'integument' (clusters 1, 3, 4, 7, 8, 10, 12, 13, and 15) in our ST data, and the results showed that there was a strong correlation between the two gene datasets (Supplementary Fig. 13b). In addition, our scRNA-seq analysis showed that *PDF1*, a known key marker gene for fiber initiation¹², was highly expressed in 0-DPA 'outer integument' (cluster 13) (Supplementary Fig. 13c).

Further analysis using the short time-series expression miner (STEM) program grouped the DEGs into seven based on the expression trend from -1.5-DPA to 1-DPA (Supplementary Fig. 14). We also analyzed the top ten DEGs in each cluster of ovules from these four stages, and identified 33 marker genes specifically expressed at 0-DPA (Supplementary Fig. 15a), such as *xyloglucan endotransglycosylases/hydrolase 22.3 (XTH22.3)* (*Gh_A08G028100*), a member of the *GhXTH1* family⁴³, *expansin 2 (EXPA2)* (*Gh_D10G123800*) and *EXPA4.6 (Gh_A10G149600)*²¹, *auxin-inducible 22 (AUX22)* (*Gh_A05G114100*)⁴⁴, *aminocyclopropane-1-carboxylic acid (ACC) synthase 2.1 (ACS2.1)* (*Gh_A02G179600*)⁴⁵, and *phytosulfokine 6 (PSK6)* (*Gh_A01G235100*)⁴⁶. GO enrichment analysis indicated that compared with those in -1-DPA ovule, the up-regulated genes in 0-DPA ovule were mainly enriched 'translation' and 'ribosomal small subunit assembly' terms (Supplementary Fig. 15b). Besides, the DEGs identified in the different cell clusters included numerous TFs (Supplementary Fig. 15c) probably involved in regulating the early development of fiber cells, including *anthocyaninless 2 (ANL2.3, an HD-ZIP member)* (*Gh_A11G092100*), *ERF008.1 (Gh_A12G080000)*, and *MYB306.11 (Gh_D12G192800)* highly expressed in the 0-DPA 'outer integument cells' (cluster 3) (Supplementary Fig. 15d).

We then analyzed the scRNA-seq data from four stages using Monocle, which orders the single-cell transcriptomes in an unsupervised manner and adopts machine learning to reconstruct the single-cell developmental trajectory along pseudotime. The pseudotime trajectory had three branches, with one main developmental trajectory populated by the cells from the four samples (Fig. 4a). The pseudotime trajectory also grouped the cells into five pseudotime states (Fig. 4a). The proportion of pseudotime states in the four samples indicated that the pseudotime trajectory was consistent with their developmental stage (Fig. 4b). However, cells of some clusters showed a relatively independent distribution along the pseudotime trajectory (Fig. 4c). '-1.5-DPA outer integument cells' (cluster 7) were concentrated in the early stage of the pseudotime

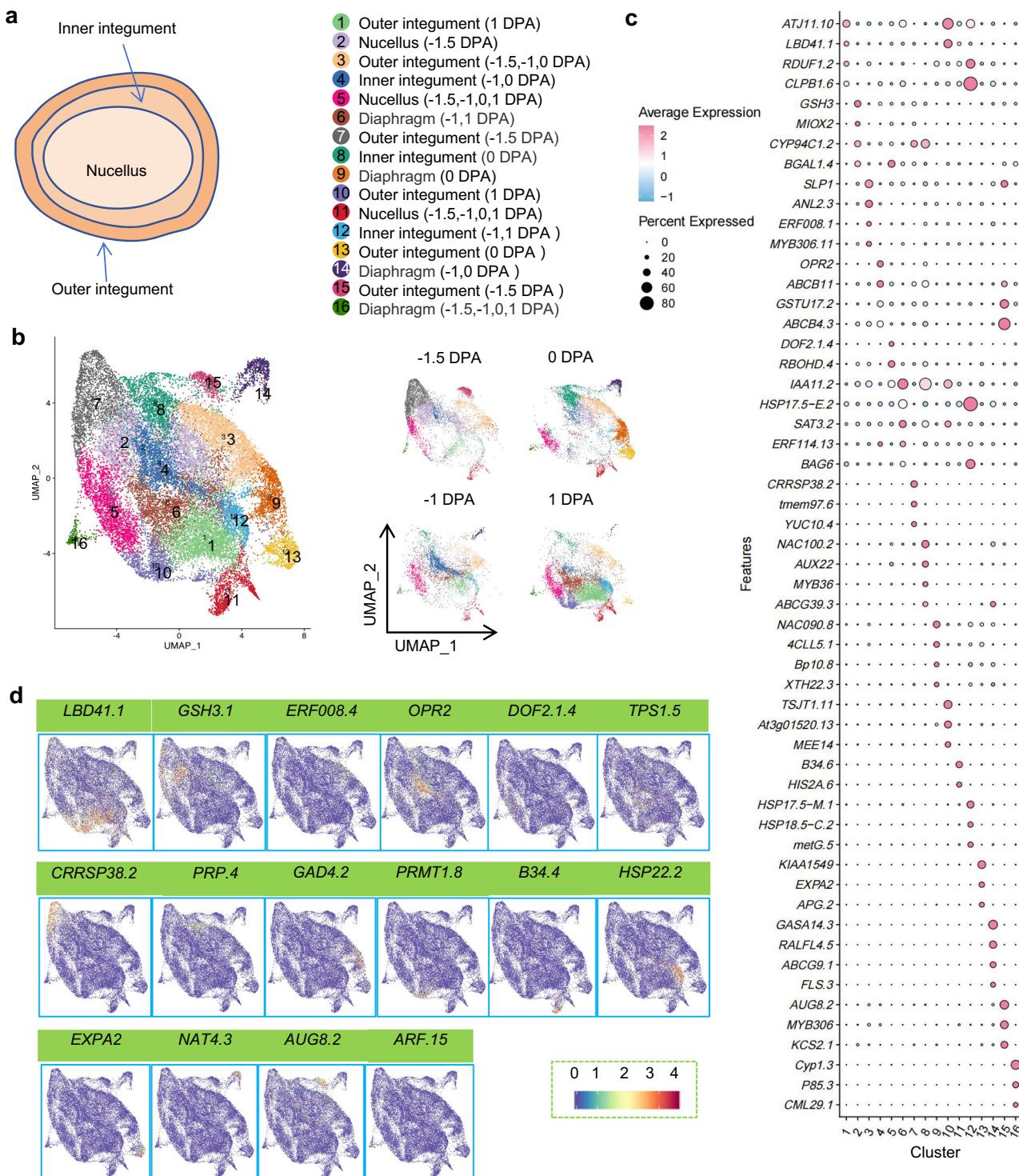


Fig. 3 | Single-cell transcriptome analysis of cotton ovules from -1.5-DPA to 1-DPA. **a** Structure of an ovule. **b** UMAP plots show major clusters based on single-cell transcriptomes of -1.5 DPA, -1 DPA, 0 DPA, and 1 DPA cotton ovules. Each dot represents a cell, and different colors represent different groups of cells. From cluster 1 to cluster 16, the number of cells contained in each cluster is 3571, 3188, 3031, 2616, 2437, 2340, 2189, 1722, 1349, 1268, 1042, 997, 679, 596, 502, and 238, respectively. **c** Dotplot shows the expression of the representative marker genes in

each cell cluster. Each row represents a marker gene, and each column represents a cell cluster. The size of the dots represents the percentage of cells in the clusters that express the gene. Expression levels are indicated by a color key ranging from blue (low) to pink (high). **d** Expression patterns of the representative marker genes from 16 cell clusters. The color intensity indicates the relative transcript level of each marker gene. Source data are provided as a Source Data file.

trajectory, while '1-DPA outer integument cells' (cluster 1) and '0-DPA outer integument cells' (cluster 13) were concentrated in the late stage of the pseudotime trajectory (Fig. 4c), consistent with the actual developmental period of outer integument. Generally, the

distribution characteristics of different cell types along the developmental trajectories suggest the relationships among these cells at different developmental stages. Interestingly, the '-1, 0-DPA inner integument cells' (cluster 4) and the '-1.5, -1, 0, 1-DPA nucellus

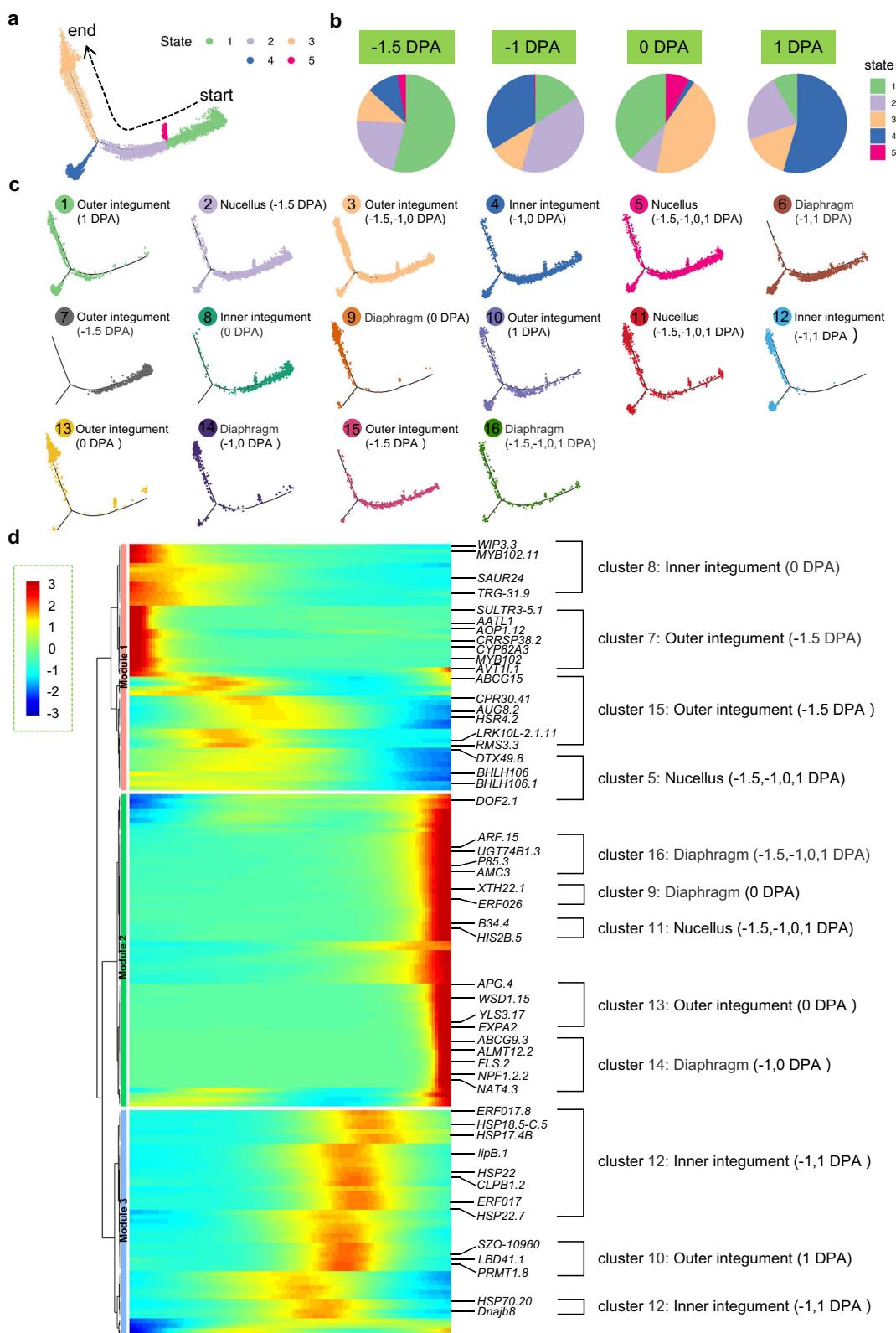


Fig. 4 | Pseudotime analysis of all cotton ovules samples from -1.5-DPA to 1-DPA. **a** A pseudotime trajectory shows the distribution of cell clusters. The dashed arrow indicates the developmental direction of the pseudotime trajectory. **b** Proportion of pseudotime states in the four cotton ovule samples. **c** Pseudotime trajectories colored according to tissue type. **d** Heatmap clustering of key genes for pseudotime.

developmental trajectories in all samples and expression kinetics of genes over pseudotime. The names of genes and cell types that determine the developmental trajectories are shown on the right side of the heatmap. Expression levels are indicated by the color ranging from blue (low) to red (high). Source data are provided as a Source Data file.

cells' (cluster 5) exhibited relatively similar distribution patterns in the developmental trajectory, suggesting similar developmental stages (Fig. 4c).

Next, to analyze the genes that play a decisive role during specific stages of development (early, middle, and late), we screened the key genes in the developmental trajectory based on the scRNA-seq data. The heatmap of these genes showed that the pseudotime patterns could be divided into three modules (Fig. 4d). The first module can be divided into three branches. In the first branch, the expression of genes, mainly from '0-DPA inner integument cells' (cluster 8) and '-1.5-DPA outer integument cells' (cluster 7), was substantially high at the beginning of development but decreased along the pseudotime. In the second branch, the expression of genes, mainly from '-1.5-DPA outer integument cells' (cluster 15), gradually increased to a maximum and then decreased along the pseudotime. In the third branch, the expression of genes, mainly from '-1.5-, -1-, 0-, 1-DPA nucellus cells' (cluster 5), was comparatively low during all developmental periods along the pseudotime and was the weakest at the last stage of pseudotime (Fig. 4d). In the second module, the expression of genes, mainly from '-1.5-, -1-, 0-, 1-DPA diaphragm cells' (cluster 16), '0-DPA diaphragm cells' (cluster 9), '-1.5-, -1-, 0-, 1-DPA nucellus cells' (cluster 11), '0-DPA outer integument cells' (cluster 13), and '-1-, 0-DPA diaphragm cells' (cluster 14), gradually increased along pseudotime, indicating that these cell types were at mature stage of development (Fig. 4d). In the third module, the expression of genes, mainly from '-1-, 1-DPA inner integument cells' (cluster 12), such as the heat shock protein (HSP)-encoding gene *HSP70.20* (*Gh_D13G265800*), gradually increased to the maximum and then decreased quickly along the pseudotime (Fig. 4d). The *HSP70* family has been reported to play a key role in cotton fiber initiation and elongation by maintaining cell homeostasis⁴⁷.

Spatio-temporal metabolites expression patterns in cotton bolls
The ST and scRNA-seq analyses revealed that many genes related to hormone metabolism, such as *GH3.6.2*, *HOS3.7*, and *DOX2* (Fig. 2), play crucial roles in fiber development. However, the internal regulatory mechanisms of fiber development are complex, and transcriptomic techniques alone are not sufficient for their discovery. Therefore, in combination with a multi-stage sampling strategy, we carried out the highly sensitive and wide-coverage imaging, named air flow-assisted desorption electrospray ionization mass spectrometry imaging (AFADESI-MSI), of metabolites in seven developmental stages cotton boll at -1.5-DPA, -1-DPA, 0-DPA, 1-DPA, 2-DPA, 3-DPA, and 5-DPA to explore the metabolic changes during early fiber development and to further understand their underlying mechanisms (Supplementary Fig. 16a). The boll samples at each stage had three biological replicates from different plants, and a total of 21 bolls were used for the study. We detected the spatiotemporal distribution of metabolites by combining the results of spatial shrunken centroids clustering with the safranin O/fast green stained sections of adjacent tissues (Supplementary Fig. 16b, c). Supplementary Fig. 16d shows images of specific metabolites in representative tissues (gland, diaphragm, ovule, and cavity).

The mass spectra obtained from the whole section of cotton bolls in positive and negative ion modes were featured by peaks in the 70–1000 *m/z* range (Supplementary Fig. 17a, b). The main compounds were organic acids, amino acids, polyamines, carbohydrates, choline, nucleotides, nucleosides, sesquiterpenoids, benzenoids, fatty acids (FAs), phosphatidylcholines (PCs), phosphatidylethanolamines (PEs), phosphatidylserines (PSs), phosphatidylglycerols (PGs), phosphatidylinositols (PIs), phosphatidic acids (PAs), etc. Supplementary Fig. 17c–h shows the mass spectra of the specific regions, such as the gland, ovule, and cavity in positive and negative ion modes. Supplementary Fig. 18 shows the spatial distribution of representative metabolites in cotton boll sections based on AFADESI-

MSI imaging. Supplementary Figs. 19–30 shows the imaging of metabolites in some key metabolic pathways, including lipid metabolism (linoleic acid metabolism, fatty acid degradation, glycerolipid metabolism, glycerophospholipid metabolism), carbohydrate metabolism (citrate cycle (TCA cycle), glycolysis /gluconeogenesis, pentose phosphate), amino acid metabolism (arginine and proline metabolism, glycine, serine and threonine metabolism, alanine, aspartate and glutamate metabolism), and nucleotide metabolism (purine metabolism, pyrimidine metabolism). It was worth noting that most metabolites showed highly heterogeneous spatial characteristics, for example, FAs, PAs, and PCs were specifically expressed in the ovule region, and PIs and carbohydrates were specifically expressed in the cavity region. The above results indicated that the AFADESI-MSI technique could indeed perform imaging analysis of metabolites in cotton bolls with high sensitivity, wide coverage, and strong specificity.

Screening of key metabolites for fiber development

To identify metabolites that play a key role in early fiber development, we screened for metabolites that are specifically expressed in ovule from -1.5-DPA to 5-DPA (Supplementary Data 16). These metabolites were mainly fatty acids and conjugates (such as oleic acid and dodecanedioic acid), linoleic acids and derivatives (such as linoleic acid and α-linolenic acid), amines (such as spermine and spermidine), amino acids, peptides, and analogs (such as L-arginine) (Supplementary Fig. 31). In addition, succinic acid, choline, PA-34:1, PA-34:2, PE-44:10, PE-36:4, PC-33:4 were also highly expressed in early developing ovule, while their expression was relatively low in other tissues. The main metabolic pathways involved in these specific metabolites include 'linoleic acid metabolism', 'cutin, suberin and wax biosynthesis', 'α-linolenic acid metabolism', 'glycerophospholipid metabolism', 'β-alanine metabolism', 'biosynthesis of unsaturated fatty acids', 'glutathione metabolism', 'arginine and proline metabolism' (Fig. 5a). We then analyzed the accumulation dynamics of these specific metabolites in ovule at different developmental stages from -1.5-DPA to 5-DPA (Fig. 5b), and imaged the spatial distribution characteristics of some representative metabolites (Fig. 5c). The expression abundance of behenic acid, linoleic acid, PE-44:10, spermine, choline in ovule from -1.5-DPA to 5-DPA was gradually decreased, indicating that these metabolites may be closely related to the initiation of fibers. In contrast, the expression abundance of α-linolenic acid, histidinal, oleic acid, and PA-34:1 were gradually increased in ovule from -1.5-DPA to 5-DPA, suggesting that these metabolites may be closely related to the early elongation of fibers.

We then performed unsupervised principal component analysis (PCA) to compare the metabolite data of cotton ovule in both positive and negative ion modes, and the results showed that there was a significant clustering and grouping trend of endogenous metabolites in different stages of development (Supplementary Fig. 32). The spatiotemporal anatomical variations in the early stages of fiber development showed that fiber cells initiated from cells in the outer integument of the ovule at 0-DPA and then increased significantly in length at 3-DPA, indicating that 3-DPA was the key stage of fiber elongation (Supplementary Fig. 1). Therefore, the differentially produced metabolites (DPMs) that were highly abundant in the ovule of 0-DPA may play a key role in fiber initiation, while the DPMs that were highly abundant in the ovule of 3-DPA may play a key role in fiber elongation.

To further identify key metabolites for fiber initiation, we compared DPMs between 0-DPA ovule and -1.5-DPA ovule (Supplementary Fig. 33 and Supplementary Data 17). Compared with -1.5-DPA ovule, DPMs such as glucaric acid, L-glutamic acid, L-asparagine, D-ribose, L-aspartic acid, D-glucose, fructose 6-phosphate, L-glutamine, oxoglutaric acid, L-threonine, DIMP were up-regulated in 0-DPA ovule. These up-regulated DPMs were mainly involved in 'alanine,

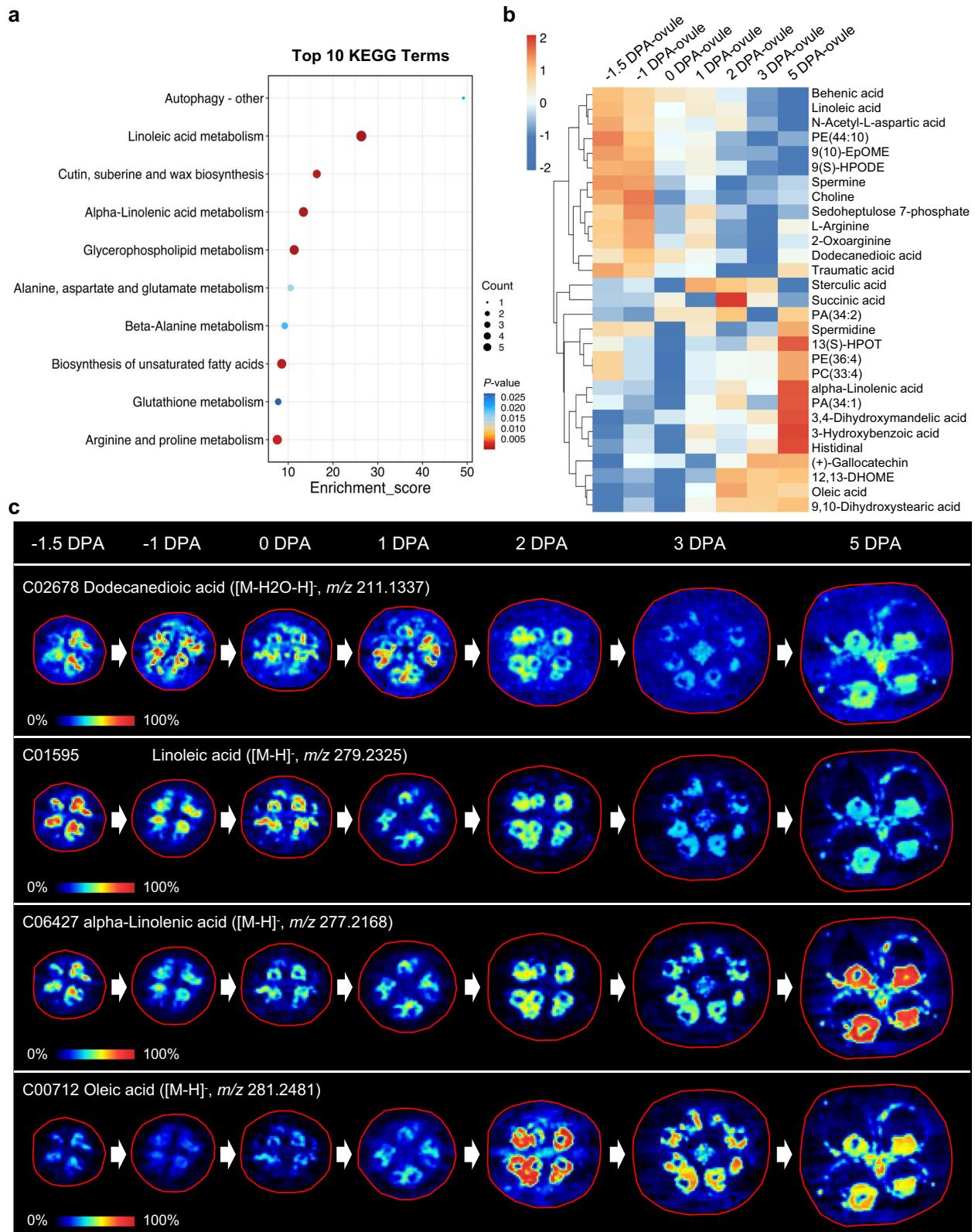


Fig. 5 | Specific expression metabolites identified in cotton ovules from -1.5 DPA to 5-DPA under positive and negative ion mode. **a** Enrichment of Top 10 metabolic pathways of ovule-specific metabolites. KEGG pathway enrichment analysis was performed using a hypergeometric test to evaluate the significance of metabolite enrichment in each pathway. **b** Heatmap shows the pattern of variation in the abundance of specific metabolites identified in ovule at different stages.

Metabolite abundance is indicated by the color ranging from blue (low) to red (high). **c** Mass spectrum imaging of dodecanedioic acid, linoleic acid, alpha-linolenic acid, and oleic acid from Fig. (b). The data shown are representative images of three independent experiments with similar results. Metabolite abundance is indicated by a color key ranging from blue (low) to red (high), and the intensity of the color scale is a relative value. Source data are provided as a Source Data file.

aspartate, and glutamate metabolism,' glyoxylate and dicarboxylate metabolism,' glycine, serine, and threonine metabolism,' and 'histidine metabolism.' However, compared with -1.5-DPA ovule, the DPMs down-regulated in 0-DPA ovule mainly included PA-34:1, L-ornithine, cis-aconitic acid, behenic acid, choline, L-valine, L-arginine, spermine, linoleic acid, α -linolenic acid, L-histidine. These down-regulated DPMs were mainly involved in metabolic pathways including 'linoleic acid metabolism', 'biosynthesis of unsaturated fatty acids', 'arginine and proline metabolism', ' β -alanine metabolism', ' α -linolenic acid metabolism'.

Similarly, to further identify key metabolites for the early elongation of fibers, we compared DPMs between 3-DPA ovule and -1.5-DPA ovule (Supplementary Fig. 34 and Supplementary Data 18). Compared with -1.5-DPA ovule, the DPMs up-regulated in 3-DPA ovule mainly included D-glucose, glucaric acid, oleic acid, L-asparagine, L-glutamic acid, α -linolenic acid, D-ribose, deoxyribose, PI-32:0, glutarate, fructose 6-phosphate, dIMP, 9,10-dihydroxystearic acid, phosphatidate (34:3). These up-regulated DPMs were mainly involved in metabolic pathways including 'pentose phosphate pathway', 'cutin, suberine and wax biosynthesis', 'alanine, aspartate and glutamate metabolism', 'starch and sucrose metabolism', 'biosynthesis of unsaturated fatty acids'. However, compared with the -1.5-DPA ovule, the DPMs down-regulated in the 3-DPA ovule mainly included linoleic acid, L-histidine, L-aspartic acid, behenic acid, L-ornithine, cis-aconitic acid, choline, L-valine, L-arginine, spermine. These down-regulated DPMs were mainly involved in metabolic pathways including 'arginine biosynthesis', 'linoleic acid metabolism', 'pantothenate and CoA biosynthesis', ' β -alanine metabolism', 'arginine and proline metabolism', 'glutathione metabolism'.

In addition, to verify the accuracy of the SM results, we performed liquid chromatography-tandem mass spectrometry (LC-MS)/MS analysis on samples at the same developmental stage as those used in the SM assay and conducted referential target metabolites analysis on some representative metabolites, further proving the metabolites identified in the SM analysis (Supplementary Fig. 35). The annotation of representative target analyte obtained from spatial metabolomics were further identified by MS/MS spectrum (Supplementary Figs. 36–61).

Integrated analysis of transcriptome and metabolome data

Next, we conducted an integrated analysis of the ST and SM data from the early development stage of cotton, and more comprehensively analyzed the internal molecular regulatory mechanism of fiber development from the multi-omics dimension. The SM of this study identified numerous DPMs during early fiber development, of which L-arginine, oxoglutaric acid, L-aspartic acid, α -linolenic acid, linoleic acid, spermine, and spermidine were the significant ones (Fig. 5 and Supplementary Figs. 33–34). Based on the network relationship information of metabolic pathways in the KEGG database, we conducted an integrated analysis of these key DPMs and ST data (Supplementary Data 19). As shown in Fig. 6a, we visualized key metabolites and regulatory genes in the entire α -linolenic acid metabolism pathway. SM analysis revealed that phosphatidylcholine, α -linolenic acid, 13(S)-HPOT, and traumatic acid, key metabolites in the ' α -linolenic acid metabolism' pathway, were heterotopically highly expressed in the ovule. Then, 56 genes encoding metabolic enzymes related to this pathway were screened out from the ST data (Supplementary Data 20). Specifically, *LCAT4* or *PLA2-ALPHA* catalyzes the metabolism of phosphatidylcholine to α -linolenic acid and is highly expressed in ovule. The α -linolenic acid can not only be catalyzed by *DOX2* to 2(R)-HPOT but also by *LOX3.1.3* to 13(S)-HPOT. Then, *HPL* catalyzes the metabolism of 13(S)-HPOT to traumatic acid. In this metabolic pathway, α -linolenic acid can be converted to JA by the catalysis of a series of enzymes. A heatmap of the expression levels of α -linolenic acid-related genes from Supplementary Data 20

was made to study differential transcription at different stages and tissues, and these genes were found to be highly expressed in 3-, 5-DPA nucellus and fiber cells (Fig. 6b). These results were consistent with the high accumulation of α -linolenic acid in 3-DPA and 5-DPA ovule (Fig. 5c), strengthening that α -linolenic acid plays an important role in early fiber elongation as a signature metabolite. In addition, we hypothesized that the genes highly expressed in the fiber cells and integument cells in Fig. 6b (e.g., *DOX2*) play an important role in early fiber development. *DOX2* was highly expressed in 3-DPA fiber cells in ST analysis, and they were also involved in the ' α -linolenic acid metabolism' pathway.

BRs and auxin have been shown to promote fiber initiation and elongation^{21,22}. We examined oxoglutaric acid, L-aspartic acid, linoleic acid, spermine, spermidine, BRs, and auxin, and conducted an integrated analysis with the transcriptome data (Supplementary Figs. 62–63 and Supplementary Data 19). The results showed that oxoglutaric acid-related genes were mainly highly expressed in 0-, 1-DPA integument cells and 3-, 5-DPA fiber cells (Supplementary Fig. 62a), L-aspartic acid-related genes were mainly highly expressed in -1.5-, -1-, 0-, 1-DPA integument and -1-, 1-, 3-, 5-DPA nucellus cells (Supplementary Fig. 62b), linoleic acid-, and spermine and spermidine-related genes were mainly highly expressed in -1.5-, -1-DPA integument and 3-DPA fiber cells (Supplementary Fig. 62c, d), and BR-related genes were mainly highly expressed in placenta and integument (Supplementary Fig. 63a), while auxin synthesis- and transport-related genes were mainly highly expressed in integument and nucellus cells (Supplementary Fig. 63b, c). Thus, these observations suggested that oxoglutaric acid, L-aspartic acid, fatty acid metabolites (such as α -linolenic acid and linoleic acid), amine metabolites (such as spermine and spermidine), auxin, and BR-like substances play an important role in the early development of cotton fibers.

In addition, quantitative strength analysis of correlation was performed on the key KEGG_PATHWAY gene dataset obtained from the integrated analysis of ST and SM in Supplementary Data 19. Specifically, we utilized the AddModuleScore function to score these KEGG_PATHWAY gene datasets (e.g., ' α -linolenic acid metabolism', 'linoleic acid metabolism', and 'brassinosteroid biosynthesis and signal transduction'), and then drew feature maps for visualization to calculate the mean expression values of related genes for integrated analysis in each cell population (Supplementary Fig. 64). Results of the above analysis indicate that there was a strong correlation between the obtained transcriptome dataset and metabolome dataset.

Marker gene expression analysis by ISH

Through ST analysis at the early development stage of cotton bolls, marker genes *DOX2*, *SBT4.14*, *GH3.6.2*, and *SWEET15* were identified in integument cells (Fig. 2d). In order to verify the classification reliability of the clustered cell clusters in ST analysis, ISH was used to analyze the expression patterns of these four integument cells markers in cotton bolls at the early development stage. ISH analysis showed that *DOX2* and *SBT4.14* were highly expressed in -1-DPA outer integument cells, *GH3.6.2* was strongly expressed in 0-DPA outer integument cells, and *SWEET15* was strongly expressed in 3-DPA outer integument cells (Fig. 7b), which was consistent with ST analysis (Fig. 7a). Furthermore, we used ISH to analyze the expression patterns of marker genes *SWEET15* and *DOX2* in cotton bolls at each stage of early development, and found that they were obviously expressed in integument cells at each stage (Supplementary Fig. 65). This result was consistent with the spatiotemporal expression patterns of *DOX2* and *SWEET15* in early development obtained by the ST analysis. The above analysis indicated that the classification of cell clusters in our ST analysis was reliable, and also verified that *DOX2*, *SBT4.14*, *GH3.6.2*, and *SWEET15* were marker genes of integument cells.

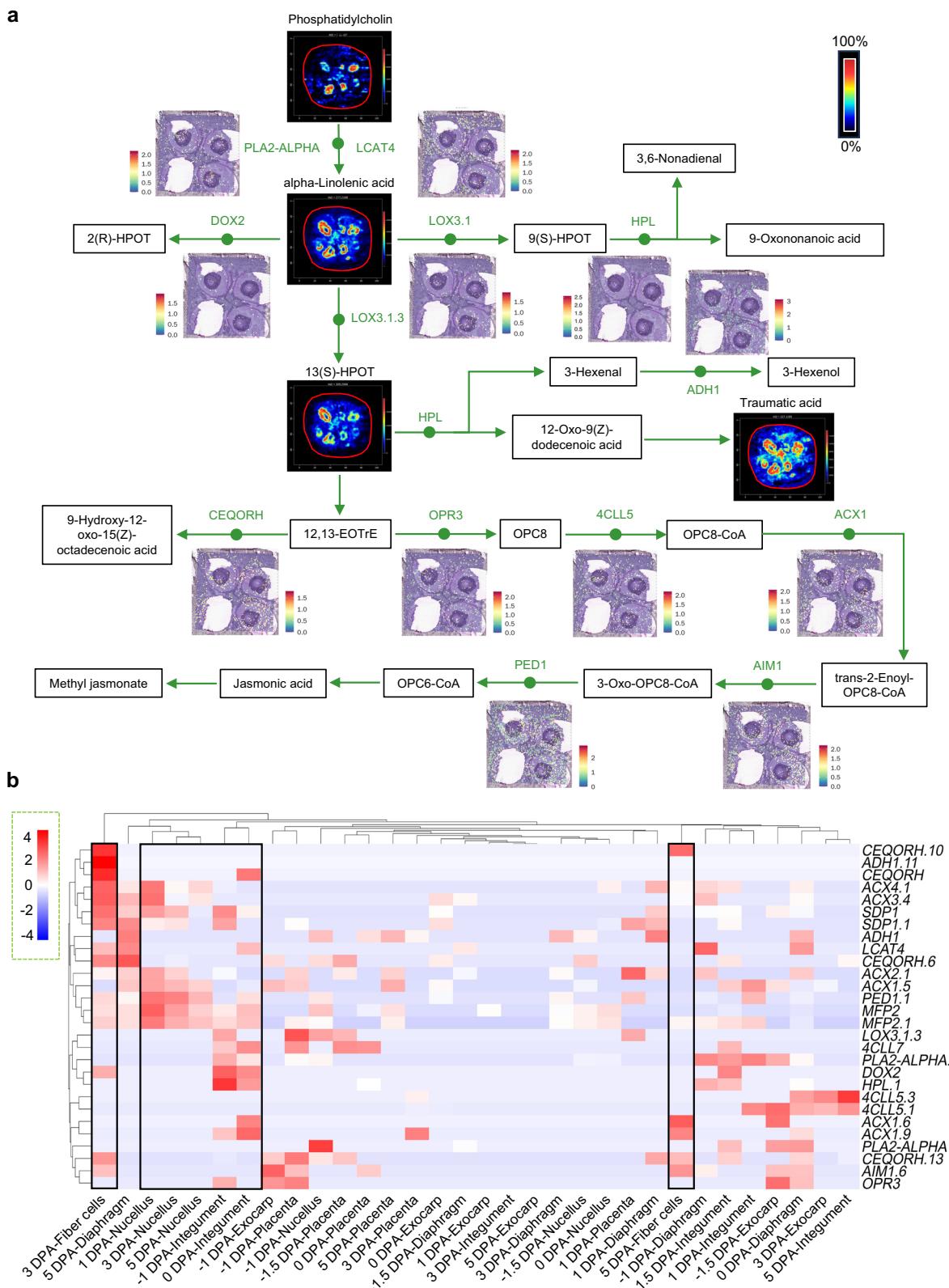


Fig. 6 | Integrated analysis of α -linolenic acid metabolic pathway and genes related to metabolic enzymes in ST. a, MS images of key metabolites and spatial expression images of key genes in α -linolenic acid metabolic pathway (intensity in MS image color scale is relative value, intensity in gene image color scale is log₂

transformed). **b** Heatmap shows the expression level of α -linolenic acid metabolic pathway-related genes identified based on ST of five cotton tissue regions (fiber, integument, placenta, exocarp, nucellus, and diaphragm). Expression levels are indicated by a color key ranging from blue (low) to red (high).

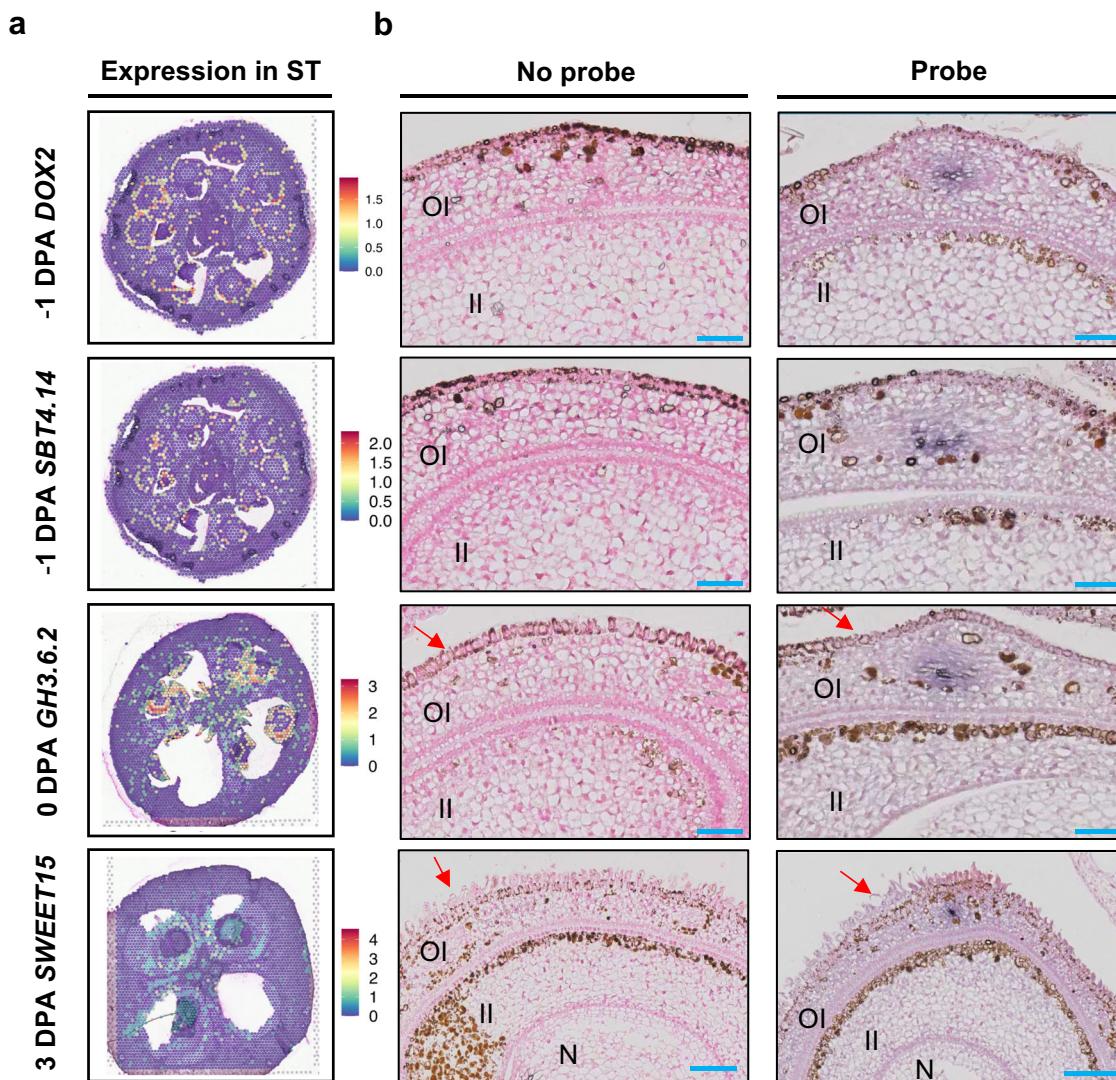


Fig. 7 | Spatiotemporal expression pattern of integument cell marker genes revealed by RNA in situ hybridization. **a** Spatiotemporal expression pattern of *DOX2*, *SBT4.14*, *GH3.6.2*, and *SWEET15*, the marker genes of integument cells based on spatial transcriptomics. The color intensity indicates the relative transcript level of each marker gene. **b** Nitro-blue tetrazolium-stained RNA ISH shows the

spatiotemporal expression patterns of *DOX2*, *SBT4.14*, *GH3.6.2*, and *SWEET15*. No probe was used as a blank control. The positive cells are stained bluish-violet. The data shown are representative images of three independent experiments with similar results. OI, outer integument; II, inner integument; N, nucellus. Red arrows indicate fiber cells. Scale bars, 50 µm in -1 DPA and 0 DPA; 100 µm in 3 DPA.

GhBEE3 promotes cotton fiber cell initiation

BRs have been reported to promote fiber initiation and elongation^{22,48}. In the cotton placenta, we detected higher expression of *GhBEE3*, which encodes an important component of BR signaling⁴⁹ (Supplementary Fig. 9b, c). Therefore, we speculated whether *GhBEE3* might affect the early developmental process of cotton fiber through BR signaling. Therefore, using the Zhongmansuo24 (ZM24) variety as the background, we generated *GhBEE3* (*Gh_A09G062900*) overexpression and RNAi lines and advance them into T3 generation to determine the potential role of *GhBEE3* in the early development of cotton fibers (Supplementary Fig. 66). Bolls at various developmental stages (-1-DPA, 0-DPA, 1-DPA, 2-DPA, and 3-DPA) were collected simultaneously from similar positions of the cotton plant. Then, scanning electron microscopic (SEM) was used to observe the development of fiber cells on ovule surface. It was evident that very few protruding fiber cells on the ovules of ZM24 (control), *GhBEE3*-OE, and *GhBEE3*-RNAi plants at -1-DPA (Fig. 8a). At 0 DPA, the number of protruding fiber cells in *GhBEE3*-OE ovule was significantly higher than ZM24, while those in *GhBEE3*-RNAi were lower (Fig. 8a, b). Similarly, at 1-DPA, the *GhBEE3*-OE ovule surface showed significantly more protruding fiber cells than

ZM24, while the *GhBEE3*-RNAi ovule surface displayed fewer (Fig. 8a, c). Detailed analysis showed quite elongated fiber cells on the ovule surface at 2-DPA and 3-DPA. Besides, the protruding fiber cells in *GhBEE3*-OE were denser than ZM24, while those of *GhBEE3*-RNAi were sparser (Fig. 8a). These results indicated that *GhBEE3* promotes cotton fiber cell initiation.

Discussion

Cotton is one of the most important sources of natural fiber worldwide¹. Therefore, the regulatory mechanism underlying cotton fiber development has been a consistent focus of cotton research. Although researchers have made considerable progress and an array of regulatory genes and metabolites have been identified, our understanding of the regulatory mechanism underlying cotton fiber development remains limited. Cotton fiber develops from a single cell and undergoes a complex series of developmental processes, including the determination of cell fate, the initiation of fiber development, the elongation of individual fibers, and the maturation of fiber cells^{5,6}. Across all of these complex developmental processes, transcriptomic patterns change dynamically at different developmental stages.

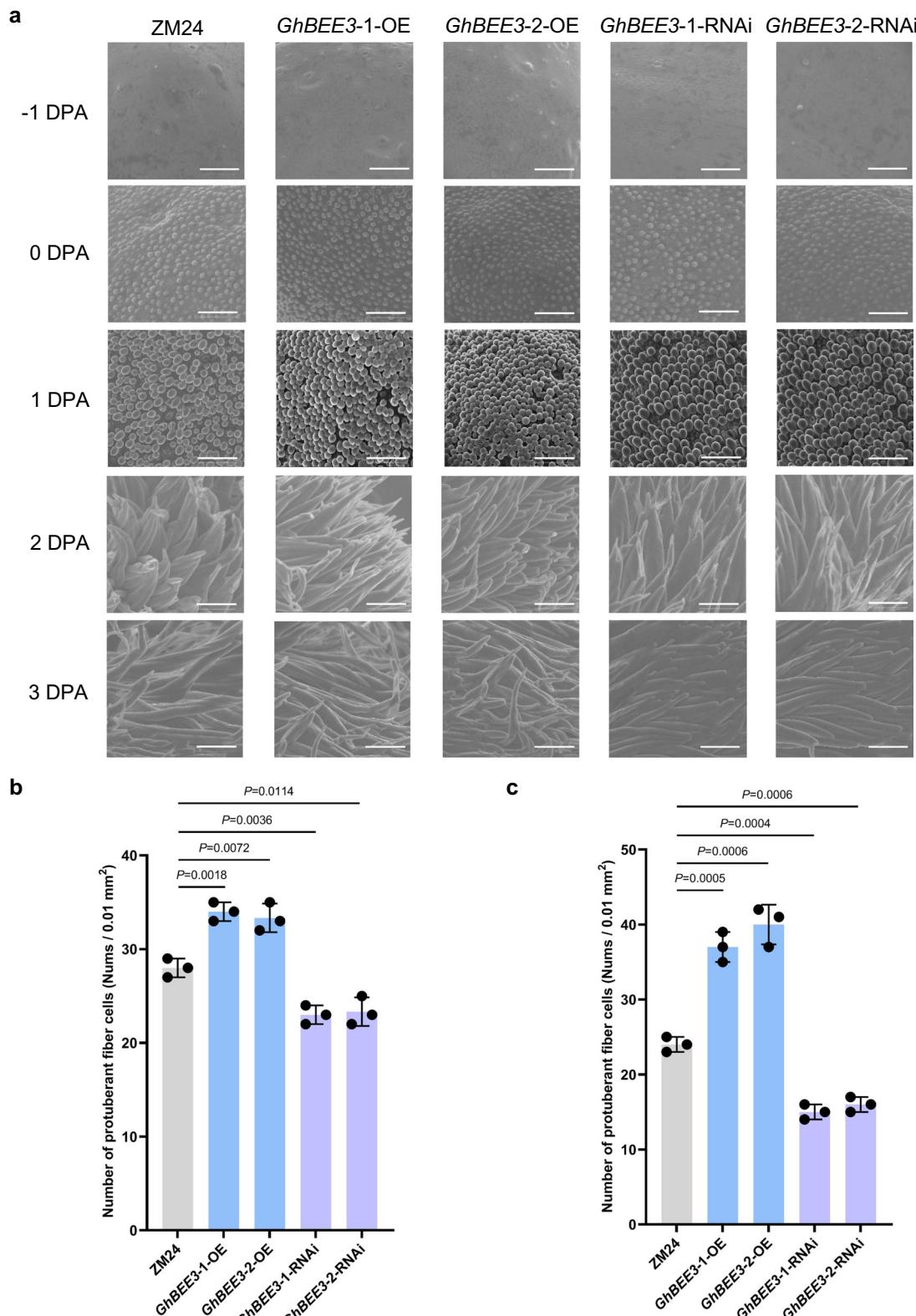


Fig. 8 | GhBEE3 positively regulates cotton fiber initiation. **a** Scanning electron microscopy images show the protuberant fiber cells in the ovules of different cotton lines at different developmental stages. The data shown are representative images of three independent experiments with similar results. Scale bar, 100 μ m. **b** Number of protuberant fiber cells per 0.01 square millimeter in 0-DPA ovules of different lines. Data are means \pm SD from three independent biological replicates ($n = 3$). The error bar is the sample standard deviation relative to the mean. Statistical analysis was assessed by unpaired two-tailed t test, and significant differences were indicated with p -values at confidence intervals of 0.95. **c** Number of protuberant fiber cells per 0.01 square millimeter in 1-DPA ovules of different lines. Data are means \pm SD from three independent biological replicates ($n = 3$). The error bar is the sample standard deviation relative to the mean. Statistical analysis was assessed by unpaired two-tailed t test, and significant differences were indicated with p -values at confidence intervals of 0.95. Source data are provided as a Source Data file.

was assessed by unpaired two-tailed t test, and significant differences were indicated with p -values at confidence intervals of 0.95. **c** Number of protuberant fiber cells per 0.01 square millimeter in 1-DPA ovules of different lines. Data are means \pm SD from three independent biological replicates ($n = 3$). The error bar is the sample standard deviation relative to the mean. Statistical analysis was assessed by unpaired two-tailed t test, and significant differences were indicated with p -values at confidence intervals of 0.95. Source data are provided as a Source Data file.

Specific genes involved in the regulation of fiber development are active at different developmental stages, making the identification and characterization of key regulatory genes challenging. Therefore, it is necessary to conduct in-depth and systematic analyses of the developmental patterns of cotton fiber cells at single-cell resolution. In this study, multi-omics techniques, such as single-cell transcriptomics, spatial transcriptomics, and spatial metabolomics, were used to investigate dynamic transcriptomic and metabolic changes during the early stages of cotton fiber development. In addition, we also preliminarily screened and identified key genes and metabolites expressed at specific developmental stages and in specific tissues. Using these data, we created a graphical and user-friendly database to provide other scientists with information on the expression patterns and dynamics of genes involved in the development of cotton fiber. This information can not only save time during early experimental stages, but can also reduce research costs, accelerate research progress, and promote the genetic improvement of cotton with excellent fiber quality.

Gene expression is spatially specific during the development of organisms; hence, studying gene expression in specific tissues will help us better understand the mechanisms of tissue genesis and development. Recently, the use of ST has allowed successful localization of cells and high-throughput gene detection in plants, such as *Arabidopsis*⁵⁰. Unlike scRNA-seq, ST has the advantage that no specific marker genes are required for cell type identification in non-model crops, such as cotton. Yet, the application of ST in plants is hindered by tissue complexity and permeability challenges. To address these issues, we initially optimized conditions for permeabilizing the cotton boll samples (specifically, modification of permeabilization enzyme cocktail and incubation time) through preliminary tissue optimization experiment, and ensured complete mRNA release from the boll slice, a crucial step in ST (Supplementary Fig. 2). Subsequent library preparation yielded cDNAs with defined lengths (Supplementary Fig. 3), guaranteeing sequencing accuracy and ST precision. A recent study used a combination of ST with scRNA-seq and created a single-cell spatiotemporal transcriptional atlas of intestinal development in humans, mapping morphologic patterns across time, location, and cell compartments⁵¹. The combination of scRNA-seq and ST can enable the study of spatial gene expression to reach the ultra-high resolution single-cell level, which has a promising application prospect in the field of plant tissue cell development.

Recently, Qin et al. conducted scRNA-seq on cotton ovules (-1.5 , -1 , -0.5 , and 0 -DPA)³². They elucidated the regulatory network underlying the initiation of cotton fiber differentiation and provided a genome-wide gene expression during the initiation and early elongation of fiber cells³². Nevertheless, an integrated approach based on scRNA-seq and ST will provide genes' complete temporal and spatial expression data during fiber initiation and elongation. In this study, we analyzed cotton ovules from the same developmental stage (-1.5 , -1 , 0 , 1 , 3 , and 5 -DPA) using both ST and scRNA-seq (Figs. 2, 3). Although we could not obtain high-quality protoplasts from 3 - and 5 -DPA cotton ovules for scRNA-seq, our scRNA-seq still covered samples from the early stages of cotton development. Furthermore, correlation analysis was conducted between ST and scRNA-seq datasets from us and scRNA-seq datasets from Qin et al.³², respectively (Supplementary Figs. 5, and 13). The results both found that there was a strong correlation between the two datasets, which, to some extent confirmed the reliability of our research results.

The spatial information on gene expression obtained via ST combined with multi-stage sampling helped us identify key genes regulating fiber initiation and early elongation, and those highly expressed in specific cell clusters during early boll development. Specifically, we annotated the cell clusters based on spatial barcode information and identified the top ten DEGs for each cell cluster as the

marker genes (Fig. 2). These marker genes are closely related to their cell clusters. The functions of a few marker genes have been studied in cotton and other species, but the potential functions of most of these marker genes have not been fully explored. For example, our ST data identified *KCS19.4*, *FLA11.3*, *TMK3*, and *bZIP43.11* as marker genes of 'fiber cells' (Fig. 2 and Supplementary Figs. 10, 11). It has been reported that *KCS19* is a member of the same gene family as *KCS1* in *Arabidopsis* that plays a role in increasing the level of very-long-chain fatty acids (VLCFAs) and improving the cold tolerance of seedlings⁵². *FLA11.3* acts with *FLA12* to fine-tune the initiation of secondary cell wall development and the balance of lignin and cellulose in the secondary cell wall in *Arabidopsis*⁵³. *TMK3* is a member of the *receptor-like kinase (RLK)* family and a key regulator of cell expansion and proliferation in *Arabidopsis*⁵⁴. The *bZIP43.11* plays a role in regulating *bHLH109* during the induction of somatic embryogenesis (SE) in *Arabidopsis*⁵⁵. *SBT4.14*, *GH3.6.2*, *DOX2*, and *GASA14.1* are marker genes of 'integument cells' (Fig. 2). It has been reported that *Arabidopsis* homologous genes *GH3*, *SBT4.14*, and *Solanum lycopersicum*-homologous *alpha DOX2* were identified to play a role in auxin metabolism, stress responses, and lipid synthesis, respectively^{56–58}. *GASA14.1* has been reported to play a role in regulating the accumulation of ROS in *Arabidopsis*⁵⁹. In addition, we also identified *SWEET15* as the marker gene of 'integument' and 'nucellus cells' (Fig. 2), and *TCP5.3* as the marker gene of 'mesocarp cells' (Supplementary Fig. 6b). Two consecutive field experiments showed that the *GhSWEET15*-upregulated fibers became thinner, longer, and stronger than the wild-type control, suggesting that *SWEET15* promotes fiber elongation⁶⁰. In *Arabidopsis*, *TCP5.3* plays a role in controlling plant thermo-morphogenesis by positively regulating *PIF4* activity⁶¹. Further, GO annotation of the DEGs indicated fatty acid oxidation enrichment in the outer integument cells during early development and the same genes with similar expression levels in fiber cells and the ovary wall (Supplementary Table 2). These results provide insights into the regulatory networks governing cotton fiber differentiation.

In order to verify the classification reliability of the clustered cell clusters in ST analysis, ISH was used to analyze the expression patterns of *DOX2*, *SBT4.14*, *GH3.6.2*, and *SWEET15*, the markers of 'integument cells' in cotton bolls at the early development stage (Fig. 7 and Supplementary Fig. 65). ISH analysis showed that these four genes were expressed in integument cells, which was consistent with our ST analysis (Fig. 7 and Supplementary Fig. 65). In addition, the results of ST analysis revealed that *PDF1*, a key gene known for fiber initiation, was highly expressed in the 'fiber cells' and 'integument cells' regions of young cotton bolls (Supplementary Fig. 5c), consistent with those reported by Deng et al.¹². Previously, *GbPDF1* was preferentially expressed during fiber initiation and early elongation, with the highest transcript accumulation in 5 -DPA fiber cells¹². The results of these reports confirmed that the classification of cell clusters in ST analysis was reliable.

Researchers now adopt MSI instead of gas chromatography-mass spectrometry (GC-MS) to gain insights into the spatial distribution of metabolites in complex and heterogeneous tissues⁶². With various other metabolomics techniques, MSI has been used as an SM approach for *in situ* metabolic profiling⁶³. Several metabolites related to cotton fiber development, such as sugars, hormones, amino acids, and fatty acids, have been reported. For example, sucrose and very long-chain fatty acids (VLCFAs) are known to regulate cotton fiber elongation^{64,65}. However, the influence of small molecule metabolites, such as amino acids, nucleosides, and organic acids, on cotton fiber development remains largely unknown. Therefore, to explore the spatiotemporal changes of metabolites in cotton boll cells and their potential roles in regulating fiber development, we conducted an SM using AFADESI-MSI on samples collected from the same developmental stages of ST (Supplementary Fig. 16). Specifically, we identified several metabolites that play an important role in regulating early development of the

ovule (Fig. 5, Supplementary Fig. 31 and Supplementary Data 16). For example, succinic acid, a four-carbon dihydric acid oxidized by succinate dehydrogenase (SDH) as part of the TCA cycle, mediates ROS metabolism in plants⁶⁶. In order to identify key metabolites regulating fiber initiation and early elongation, we screened the DPMs in 0-DPA and 3-DPA ovules, respectively (Supplementary Figs. 33–34 and Supplementary Data 17–18). α -linolenic acid was detected at high levels at the fiber elongation stage (3-DPA to 5-DPA) (Fig. 5), consistent with the reported roles of linolenic acid in promoting cotton fiber elongation²⁶. JA has been reported to promote fiber development⁶⁷, and α -linolenic acid can be converted to JA in the ‘ α -linolenic acid metabolic’ pathway (Fig. 6a), which seems to indicate that α -linolenic acid is highly likely to promote fiber elongation.

Furthermore, we analyzed the ST and SM data in an integrated manner to explore the gene-metabolite network and elucidate the regulatory pathways at the early stage of fiber development (Fig. 6, Supplementary Figs. 62–63 and Supplementary Data 19–20). In addition, the results obtained by the AddModuleScore function indicate a strong correlation between the transcriptome dataset and metabolome dataset, at least in the case of the pathways related to ‘ α -linolenic acid metabolism,’ ‘linoleic acid metabolism,’ and ‘BR biosynthesis and signal transduction’ (Supplementary Fig. 64).

Analysis of the ST dataset identified a BR-related gene, *GhBEE3*, that was highly expressed in the placenta (Supplementary Fig. 9b, c and Supplementary Fig. 63a). A previous study based on transcriptome analysis revealed that BRs promote cotton fiber elongation by influencing ethylene and cadmium signaling, fatty acid biosynthesis, and cell wall- and cytoskeletal-related gene expression⁶⁸; however, the molecular mechanisms underlying this regulation remain largely unknown. In another independent study, *GhBEE3* has been identified as a positive regulator of BR signaling⁴⁹, which suggests its role in promoting fiber initiation and elongation. Therefore, we produced *GhBEE3*-RNAi and *GhBEE3*-OE lines to assess their function (Supplementary Fig. 66). Detailed analysis showed that at 0-DPA and 1-DPA, the *GhBEE3*-OE ovules had significantly more protruding fiber cells than the ZM24 control, while *GhBEE3*-RNAi ovules had fewer (Fig. 8). These observations indicated that *GhBEE3* regulate fiber initiation in cotton maybe by modulating the BR signaling or other signaling pathways. However, the role of *GhBEE3* and other DEGs in regulating fiber development needs to be investigated.

In summary, based on a rational, multi-DPA sampling strategy and the combined application of ST, scRNA-seq, and SM technologies (Fig. 1), the present study demonstrated the spatiotemporal profiles of transcriptomes and metabolomes during the initiation and elongation of fibers in cotton. We also developed an interactive website to include the transcriptomes and metabolomes generated in this study, providing a reference for further research on cotton fiber development at a single-cell level (Supplementary Fig. 67). At the same time, our study also provides a feasible research strategy for other scientists to conduct similar research, and provides a key methodological model for accelerating the application of single-cell spatial multiomics to diverse crop species.

Methods

Plant materials and sample collection

We used the upland cotton TM-1 (*G. hirsutum* L. acc. Texas Mark-1) to perform ST, scRNA-seq, SM, LC-MS/MS, and ISH experiments and Zhongmiansuo24 (ZM24) to generate the overexpression (OE) and RNA interference (RNAi) lines. All cotton plants were maintained in the experimental field at the Institute of Cotton Research in the Chinese Academy of Agricultural Sciences (CAAS; Anyang, China) under conventional field management measures. Flowers were labeled at the start of flowering, and bolls were sampled –1.5-, –1-, 0-, 1-, 2-, 3-, and 5-DPA to obtain fibers at different stages of development. All samples were collected at 8:00 AM and 8:00 PM. ST experiment and SM

experiment were performed on the same batch of cotton boll tissue samples.

Spatial transcriptomics analysis

All reagents were obtained from Visium Spatial Gene Expression Reagent Kits, and the experiments were operated according to the user guide of “Visium Spatial Protocols-Tissue Preparation Guide (CG000240)”, “Visium Spatial Gene Expression Reagent Kits-Tissue Optimization (CG000238)”, and “Visium Spatial Gene Expression Reagent Kits (CG000239)“.

Sample preparation

The collected cotton fresh bolls (~1.5-DPA, ~1-DPA, 0-DPA, 1-DPA, 3-DPA, and 5-DPA) were embedded directly in optimum cutting temperature (OCT) compound (OE Biotech, Shanghai, Item No. 4583) in a histology plastic cassette, and cut into 10 μ m serial frozen sections at –20 °C on a cryostat microtome (Leica, CM 1950). Subsequently, RNA was extracted and assessed for quality (Supplementary Table 1). The eligible sections with RNA integrity number (RIN) greater than 7 were selected for Hematoxylin-Eosin (H&E) staining (Sigma, USA) and incubated at 37 °C for 5 min prior to imaging. Images were used to confirm that each section included the targeted region. The remaining qualified sample sections were used for subsequent experiments.

Tissue optimization

Adjacent frozen cotton boll sections on 10 \times Genomics Visium array slides were fixed, stained, and permeabilized at different times. The permeabilization enzyme cocktail has been modified to include 2% w/v Cellulase R10, 0.4% w/v Macerozyme R10, 1% w/v Pectinase, 1% w/v Hemicellulase, 0.4% Snailase. Poly adenylated mRNA from the attached tissue section was captured by probes on the slides. Then, add Master Mix containing reverse transcription (RT) reagents and fluorescently labeled nucleotides to the surface of the tissue section to obtain fluorescently labeled cDNA. Next, remove excess tissue, leaving fluorescently labeled cDNA covalently linked to oligonucleotides on the Visium array slides. Then, fluorescently labeled cDNA was visualized. Twenty-four minutes was chosen as the optimal permeabilization time because the fluorescence signal is the strongest and the signal diffusion is the weakest (Supplementary Fig. 2).

Tissue permeabilization, cDNA synthesis, library construction and sequencing

Fixed and stained tissue sections were permeabilized for 24 min with the modified permeabilization enzyme. The primers on the spots capture the poly-adenylated mRNA released from the overlying cells. Then, cDNA synthesis, library construction, and sequencing were performed according to the User Guide of Visium Spatial Gene Expression Reagent Kits. During the library preparation, the cDNA lengths we obtained were distributed within the standard range (Supplementary Fig. 3).

Processing of sequencing data and quantification of gene expression

The raw reads in FASTQ format generated via high-throughput sequencing were processed with the 10 \times Genomics Space Ranger software (version 1.2.0), and the obtained sequences were aligned to the cotton reference genome using the STAR software (version 2.7.1a). The brightfield images of the sections were captured. Then, the spatial barcode information was used to align the reads to specific spots on the tissue section using the H&E images. The total number of spots, the number of reads per spot, the number of detected genes, and the number of unique molecular identifiers (UMIs) were determined to evaluate the quality of the sequencing

reads. Finally, a gene-spot matrix was generated for gene expression analysis.

Quality control and further processing of data

The Seurat4 package was used to perform further quality control and process the data obtained via Space Ranger (version 1.2.0). Since the sequencing data of all samples were of good quality (Supplementary Fig. 4) and more spot expression information was expected, the downstream data processing and analysis were directly based on the matrix results generated by Space Ranger, without further filtering of spots. Then, the SCTransform function was used to normalize and stabilize the variance in the ST data using regularized negative binomial regression, and the data were stored in the SCT assay for further analysis.

Dimensionality reduction and clustering

The batch effects in ST data were corrected with the R package ‘batchelor’⁶⁹ (version 1.3.4), following the mutual nearest neighbors (MNN) approach. Then, the “FindClusters” function of the Seurat4 package was used to scale the gene expression and cluster the cells based on the MNN approach. This strategy removed batch effects in the ST data and enabled the detection of cell subpopulations. Seurat’s “RunUMAP” function, a two-dimensional uniform manifold approximation and projection (UMAP) algorithm, was used to visualize the cells.

Differential gene and enrichment analysis

The “FindMarkers” function (test. use = MAST) in Seurat4 was used to identify the differentially expressed genes (DEGs); the genes with a $|log_2\text{foldchange}| > 0.58$ and a $p\text{-value} < 0.05$ were identified as the DEGs. Then, GO enrichment analyses of DEGs were performed with the annotated gene name of DEGs [<https://pantherdb.org/>]. The reference species is *G. hirsutum*.

Pearson correlation analysis

Correlations were calculated using the summation method to obtain the expression level of each gene across different samples and cell groups. First, the expression counts for each gene was summed for each sample. Second, “1” was added to the summation value to take log10 (i.e., $\log_{10} [\sum + 1]$). Third, the scatter plot was created according to this value, with each point representing one gene. Using the stat.cor() function in the ggpubr package, the correlation value and $p\text{-value}$ were calculated by Pearson analysis.

Preparation of single-cell suspensions from cotton ovules

We optimized reagent and treatment conditions for the dissociation of protoplast from cotton at -1.5-DPA, -1-DPA, 0-DPA, and 1-DPA. The major improvements involved modification in enzyme solution formulation (20 mM MES (pH 5.7), 3% (wt/vol) Cellulase R10, 1% (wt/vol) Macerozyme R10, 1% (wt/vol) Snailase, 1% (wt/vol) Cutinase, 0.4 M Mannitol, and 20 mM KCl) and treatment conditions (dark condition, 28 °C, 20 rpm horizontal shaking). The cotton ovules were cut into 1 mm pieces with a blade and soaked in the enzyme solution for 15 min under a vacuum. The mixture was then strained through a 40 µm cell filter to remove large tissue blocks and obtain a protoplast suspension, which was kept on ice to prevent cell death. The protoplasts were counted with a hemocytometer, and the activity was assayed following trypan blue staining (0.4%). The protoplasts with more than 90% activity rate were used for scRNA-seq library construction.

Construction of scRNA-seq cDNA library

The scRNA-seq library was prepared from the protoplasts using a Chromium Single Cell 3' Gel Beads-in-emulsion (GEM) Library & Gel Bead v3 Kit, following the manufacturer's instructions (10 × Genomics, California, USA).

ScRNA-seq data preprocessing

The raw sequencing data were analyzed using the Cell Ranger software pipeline (version 5.0.0) provided by 10 × Genomics. The sequences were aligned against the *G. hirsutum* TM-1 reference genome⁷⁰ using the STAR aligner (version 2.7.1a), and the transcripts were quantified to obtain the unique molecular identifier (UMI) count matrix. This matrix was further processed using the R package Seurat4. Then, we filtered out cells with UMIs/gene threshold ($>\text{mean value} \pm 2\text{-fold of the standard deviation}$, assuming a Gaussian distribution of each cell's UMI/gene number) and doublets using the DoubletFinder package (version 2.0.2). The data obtained from cells after quality control was normalized using the “NormalizeData” function in Seurat4 to get the normalized count.

Principal component analysis (PCA) was performed with the “RunPCA” function in Seurat4 to reduce the dimensionality, and graph-based clustering was performed using the “FindClusters” function in Seurat4 to cluster the cells based on their gene expression profiles. The final data and the cells were visualized using the “Run UMAP” function in Seurat4. Then, the “FindAllMarkers” function (test. use = bimod) in Seurat4 was used to identify the positive marker genes for each cluster compared with other cells, and the “FindMarkers” function (test. use = MAST) to determine the DEGs ($|log_2\text{foldchange}| > 0.58$ and $p\text{-value} < 0.05$). Further, the GO and KEGG pathway enrichment analyses of DEGs were performed using R based on the hypergeometric distribution.

Pseudotime analysis

The developmental pseudotime was calculated with the Monocle 2 package (version 2.9.0). The scRNA-seq raw count from the Seurat4 object was first converted into the CellDataSet object with the “importCDS” function, and then, the ordering genes ($q\text{-value} < 0.01$) were selected using the “differentialGeneTest” function to order cells along the pseudotime trajectory. The dimensionality reduction and clustering was performed using the “reduceDimension” function, followed by trajectory inference with the “orderCells” function using default parameters. Finally, the expression of each gene was plotted over pseudotime using the “plot_genes_in_pseudotime” function.

Spatial metabolome sample preparation and process

The boll samples at each stage had three biological replicates. The fresh cotton bolls (-1.5-DPA, -1-DPA, 0-DPA, 1-DPA, 2-DPA, 3-DPA, and 5-DPA) were embedded directly in the Leica Cryo-Gel from Leica (Leica Microsystems, Germany, Item No. 14020108926), and subsequently frozen and solidified on dry ice. The continuous sections were sliced using a -20 °C freezing microtome (Leica, CM 1950, Germany) with a thickness of 10 µm. Using a brush, the sections were gently deposited onto Superfrost Plus positive charge anti-slip slides (Thermofisher Scientific, Waltham, MA, USA). The samples were defrosted by placing a finger on the back of the slide, ensuring that the samples were attached to the anti-slip slides. The sections were stored at -80 °C for subsequent MS imaging. Prior to AFADESI-MSI, the frozen tissue sections were dried in a vacuum dryer for 30 mins.

AFADESI-MSI analysis

The MSI was performed on a Q-OT-qIT hybrid mass spectrometer (Orbitrap Fusion Lumos; Thermo Fisher Scientific, United States) coupled with a lab-made AFAL ion source at a spatial resolution of 100 µm. Acetonitrile/water (80:20, V/V; containing 0.1% formic acid; positive ion mode) and acetonitrile/water (80:20; negative ion mode) were used as solvents at a flow rate of 5 µL/min for scanning the tissues. The MSI was carried out in positive and negative ion modes from a mass-to-charge ratio (m/z) of 70–1000, with high mass accuracy (<5 ppm mass error) and a high mass resolution (70,000). Nitrogen (0.6 MPa) was used as spray gas, and the capillary temperature was set to 350 °C. The tissue was scanned continuously along the x-axis at a constant speed (Vx) of 0.2 mm/s, maintaining

a distance of 0.1 mm between two lines (Dy), an X-axis scanning length of 10 mm, and a Y-axis scanning length of 10 mm. The Xcalibur data acquisition and processing system was used to collect data in a sequence based on sample size, step spacing, and scanning speed.

AFADESI-MSI data preprocessing

The acquired raw data (.raw format) were converted into .imzML format using imzMLConverter. Then, the open-source R package Cardinal was used to process the data and visualize the mass spectra. A series of processing steps, such as peak alignment and missing value interpolation, were used to obtain the quantitative data. All spectral images were normalized using the total ion count (TIC) in each pixel. Then, the target regions of interest were selected for subsequent analysis based on the results of spatial shrunken centroids clustering and tissue staining of sections.

LC-MS/MS analysis

To verify the results of the SM analysis, we conducted LC-MS/MS analysis on integument_nucellus and exocarp_mesocarp tissue samples from the same batch used for SM. The metabolomics assay was performed by Shanghai Luming Biotechnology Co., Ltd (Shanghai, China). Six biological replicates were utilized for each group of samples. Briefly, 50 mg samples were added to 1.5 mL Eppendorf tubes containing 600 μ L of methanol-water solution (methanol:water = 7:3, V/V), with L-2-chlorophenylalanine (4 μ g/mL) as an internal standard, and the tubes were subsequently vortexed for 10 s. Whole samples were extracted by ultrasonication for 30 min in an ice-water bath and then centrifuged at 4 °C (13,201 g) for 10 min. The supernatants (150 μ L) from each tube were filtered using 0.22 μ m microfilters and transferred to LC vials. The vials were stored at -80 °C until LC-MS/MS analysis. Quality control (QC) samples were prepared by mixing aliquots of all samples into a pooled sample. An ACQUITY UPLC I-Class plus (Waters Corporation, Milford, MA, USA) fitted with a Q-Exactive mass spectrometer equipped with a heated electrospray ionization (ESI) source (Thermo Fisher Scientific, Waltham, MA, USA) was used to analyze the metabolic profiles in both positive and negative ion modes. An ACQUITY UPLC HSS T3 column (1.8 μ m, 2.1 \times 100 mm) (Waters Corporation, Milford, MA, USA) was employed in both positive and negative modes. The binary gradient elution system consisted of (A) water (containing 0.1% formic acid, V/V) and (B) acetonitrile. Separation was achieved using the following gradient: 0 min, 5% B; 2 min, 5% B; 4 min, 30% B; 8 min, 50% B; 10 min, 80% B; 14-15 min, 100% B; 15.1 min, 5% B; and 16 min, 5% B. The flow rate was 0.35 mL/min, and the column temperature was 45 °C. All samples were kept at 10 °C during the analytical procedure. The injection volume was 4 μ L. For MS, the mass range was from 100 to 1200 m/z . The resolution was set at 70,000 for full MS scans and 17,500 for MS/MS scans. The collision energy was set at 10, 20, and 40 eV. The mass spectrometer was operated as follows: spray voltage, 3800 V (+) and 3000 V (-); sheath gas flow rate, 50 arbitrary units; auxiliary gas flow rate, 10 arbitrary units; capillary temperature, 320 °C; Aux gas heater temperature, 350 °C.

Metabolite annotation and comparative analysis

The ions detected by AFADESI-MSI were annotated following the pySM pipeline and an in-house SmetDB database. SmetDB contains approximately 20,000 endogenous plant metabolites and was constructed by Lumingbio (Shanghai, China). For bulk metabolomic analysis, the original LC-MS data were processed with Progenesis QI V2.3 (Nonlinear Dynamics, Newcastle, UK) for baseline filtering, peak identification, integral, retention time correction, peak alignment, and normalization. The following parameters were applied: 5 ppm precursor tolerance, 10 ppm product tolerance, and 5% production threshold. Compounds were identified using the Metlin and LuMet plant (a mass spectral database containing ~1000 standard substances

constructed by Shanghai Luming Biotechnology Co., Ltd.) databases based on the precise mass-to-charge ratio (m/z), secondary fragments, and isotopic distribution data. The extracted data were then further processed by removing peaks with a missing value (ion intensity = 0) in more than 50% of groups, by replacing the zero value with half of the minimum value, and by screening according to the qualitative results of the compound. Compounds with scores below 36 (out of 60) points were also deemed to be inaccurate and removed. A data matrix was created combining both the positive and negative ion data. PCA was performed for the metabolome data to observe the overall sample distribution and process stability. Orthogonal partial least squares discrimination analysis (OPLS-DA) was used to distinguish the differences in the metabolic profiles and identify the differentially abundant metabolites between groups. Then, the variable importance in projection (VIP) was calculated for each compound to assess their significance in separating the groups. Further, a two-tailed Student's *t* test was applied to compare the samples and determine the differences in metabolites; metabolites with VIP values > 1.0 and *p*-value < 0.05 were identified as the differential ones. The structure-specific pattern ions of the target analyte obtained from spatial metabolomics were further identified by LC-MS/MS after sample homogenizing, and some representative target analytes were demonstrated in Supplementary Figs. 35–61.

Standard solutions and calibration curves

In the pursuit of targeted metabolite analysis, we employed linolenic acid (B21967-20 mg, Shanghai Yuanye Bio-Technology Co., Ltd), lino-leic acid (B21421-0.2 mL, Shanghai Yuanye Bio-Technology Co., Ltd), and succinic acid (B20534-20 mg, Shanghai Yuanye Bio-Technology Co., Ltd) as the subjects of our study. The stock solutions of the standard were prepared by dissolving the compound at a concentration of 1 mg/mL in methanol. To create the primary mixed standard stock solution (MSS), we mixed the individual standard stock solutions and further diluted the resultant mixture with methanol to achieve appropriate concentrations. Calibration curves were generated by serially diluting the MSS with a methanol-2-propanol mixture (1:1, V/V) to achieve the following final concentrations: 2000 ng/mL, 800 ng/mL, 320 ng/mL, 128 ng/mL, 51.2 ng/mL, 20.18 ng/mL, 8.19 ng/mL, 3.28 ng/mL, 1.31 ng/mL, 0.52 ng/mL, 0.21 ng/mL, and 0.08 ng/mL. For quality control purposes, a mixed standard of 50 ng/mL was selected as STD-QC. It is imperative to note that all solutions and samples were stored at -80 °C to ensure their integrity until required for analysis.

Sample pretreatment and metabolite extraction

Metabolite extraction procedures were conducted according to the chemical characteristics of multi-targeted metabolites under investigation. Specifically, 600 μ L of ice-cold methanol-water (7:3, V/V), which included succinic acid-d4 as an internal standard, was added into freeze-dried samples weighing (0.0500 g). Subsequently, the samples were placed within the confined of two steel balls and subjected to grinding using a grinder at a frequency of 60 Hz for 2 min. Following this initial step, the entirety of the samples underwent extraction via ultrasonication for 10 min in an ice-water bath. The resulting mixture was then subjected to centrifugation at 4 °C (13,201 g) for 10 min, after which 300 μ L of the supernatant was decanted into sample vials. This supernatant was subjected to drying under a nitrogen stream and subsequently re-dissolved in 300 μ L of methanol-2-propanol (1:1, V/V), containing 2-chloro-L-phenylalanine as an internal standard. A second round of extraction by ultrasonication for 5 min in an ice-water bath was conducted, followed by centrifugation at 4 °C (15,493 \times g) for 5 min. Finally, 150 μ L of the resulting supernatant was selected for subsequent UPLC-ESI-MS/MS analysis. For quality control purposes, a pooled sample was prepared by mixing aliquots of the individual samples.

UPLC-ESI-MS/MS analysis

The LC procedure was performed on an AB ExionLC system (AB SCIEX, Framingham, MA, United States). An ACQUITY UPLC BEH C18 column (2.1 mm × 100 mm, 1.8 µm) (Waters Corporation, Milford, MA, USA) was used for analysis, with an injection volume of 5 µL. The mobile phase A consisted of water containing 0.1% formic acid, while the mobile phase B comprised methanol with 0.1% formic acid. The elution gradient was as follows, maintaining a flow rate of 0.30 mL/min: 0 min, A/B (95:5, V/V); 1 min, A/B (95:5, V/V); 4 min, A/B (20:80, V/V); 6 min, A/B (0:100, V/V); 7 min, A/B (0:100, V/V); 7.01 min, A/B (95:5, V/V); 8 min, A/B (95:5, V/V). Throughout the analysis, a constant temperature of 4 °C was maintained for all samples, while the column temperature was held at a stable 40 °C. The MS procedure was performed on the AB SCIEX API 6500 + Qtrap System (AB SCIEX, Framingham, MA, United States), incorporating an electrospray ionization (ESI) source that operated in the negative ion mode. Nitrogen served as the collision gas. Key instrumental parameters were configured as follows: Curtain Gas: 30 Psi; IonSpray voltage: -4500 V; Collision Cell Exit Potential: -20 V; Source Temperature: 450 °C; Gas1: 50 Psi; Gas2: 50 Psi.

Targeted metabolites were analyzed in multiple reaction monitoring (MRM) mode. The MRM pairs, declustering potentials (DPs), and collision energies (CEs) were optimized for each analyte. Subsequent data acquisition and further analysis were conducted using the Analyst software, while the quantification of all metabolites was performed using the SCIEX OS-MQ software.

Construction of transcriptome-metabolome association network

Based on the known metabolic pathway information, DEGs with $|\log_2\text{foldchange}| > 0.58$ and $p\text{-value} < 0.05$, and DPMs with $\text{VIP} > 1.0$ and $p\text{-value} < 0.05$ in different tissue regions of cotton boll were selected to construct transcriptome-metabolome association networks. The addmodulescore function in the Seurat4 package was scored to visualize the expression of genes related to metabolite synthesis in spatial organization.

RNA in situ hybridization

In situ hybridization for mRNA was performed using the kit (Boster Biological Technology, Wuhan, China), following the manufacturer's instructions. The cotton bolls were collected, embedded in paraffin, and cut into 5 µm thick sections. The sections were de-waxed using xylene and rehydrated using a series of alcohol gradients. Subsequently, 3% citric acid and concentrated pepsin (two drops) were added to the section and incubated at 37 °C for 10 min to expose the mRNA fragments. The cells were incubated with a DIG-labeled RNA probe (BOSTER) overnight at 60 °C for mRNA hybridization, washed with Saline Sodium Citrate (SSC), blocked with serum, and labeled with alkaline phosphatase-conjugated anti-digoxigenin antibody (anti-Dig-AP, BOSTER, no dilution required) at 37 °C for 1 h. The alkaline phosphatase activity in the cells was detected in the dark using a nitro-blue tetrazolium/5-bromo-4-chloro-3-indolyl-phosphate (NBT/BCIP) chromogenic substrate solution (BOSTER). The sections were analyzed, and the images were captured using an Aperio VERSE 8 (Leica) multi-functional tissue cell analyzer. A blank control, incubated with no probe, was used for comparison. The positive cells are stained bluish-violet. Primers or probes used in this assay are listed in Supplementary Data 21.

Construction of overexpression and RNAi vector and generation of the transgenic materials

The pCAMBIA-2300 plant expression vector with the constitutive CaMV35S promoter was cleaved using *Xba*I and *Pac*I restriction enzymes and used to generate the *GhBEE3*-OE construct by homologous recombination. The pBI121 vector digested with *Sac*I and *Bam*H I restriction enzymes was used to create the *GhBEE3*-RNAi

construct. The constructs were confirmed by sequencing and transferred to *Agrobacterium tumefaciens* LBA4404 cells. Then, *GhBEE3*-OE and *GhBEE3*-RNAi transgenic cotton strains were generated using agrobacterium-mediated genetic transformation. Specifically, cotton seeds were sterilized with 75% (v/v) ethanol 3 times for 1 min each time, then washed with sterile water 3 times for 1 min each time, and finally placed in an incubator at 37 °C for 12–20 h. After the seeds began to germinate, they were transferred to MS medium and cultured in a light incubator (16 h light/8 h dark) at 28 °C for 7 days. Under the aseptic condition, hypocotyls of ZM24 were cut into fragments of about 1 cm, and the segments were infected with *Agrobacterium tumefaciens* at the concentration of $OD_{600} = 0.8$ for 1 min, and then tissue culture in a culture medium. For the identification of positive transgenic plants, the gene-specific primer of *GhBEE3* was used to detect the presence of *GhBEE3*. The putative transgenics were cultured to T3 generation with stable inheritance by adopting the screening method, which was used for further research. The gene-specific primers used in the experiment are listed in Supplementary Data 21.

RNA extraction and quantitative real-time PCR

Total RNA was extracted from the leaves of cotton plants at flowering with FastPure Plant Total RNA Isolation Kit (Vazyme, RC401, USA) and reversely transcribed into cDNA using NovoScript® Plus All-in-one 1st Strand cDNA Synthesis SuperMix (gDNA Purge) (Novoprotein, E047, China). The qRT-PCR was performed using the extracted cDNA and Premix Ex Taq™ II (Takara) on a LightCycler 480 system (Roche Diagnostics, Mannheim, Germany) to detect the *GhBEE3* expression. The thermocycling was carried out as follows: 95 °C for 10 min, followed by 40 cycles of amplification at 95 °C for 10 s and 60 °C for 30 s. Each experiment was repeated at least three times. Primers used in the assay are listed in Supplementary Data 21.

Scanning electron microscopy imaging

Bolls at various developmental stages (~1-DPA, 0-DPA, 1-DPA, 2-DPA, and 3-DPA) were collected simultaneously from similar positions on the cotton plant. The ovules collected from these samples were rapidly imaged on an environmental scanning electron microscope (FEI MLA650 ESEM; Thermo Fisher Scientific, USA) under vacuum mode without any treatment. Each experiment was repeated at least three times.

Statistical analysis

Three biological replicates were maintained for each data set unless stated otherwise, where n indicates the number of biologically independent samples examined. Statistical analysis was assessed by unpaired two-tailed *t* test and significant differences were indicated with *p*-values at confidence intervals 0.95. Data analyses were carried out by GraphPad Prism 9.0.2.

Database website

The Shiny program in R was used to construct an interactive webpage (<https://cotton.cricas.com.cn/ovule/>) to access the cotton scRNA data and ST data generated in this study. The “DimPlot”, “VlnPlot”, “FeaturePlot”, and “SpatialPlot” functions in the Seurat package were used to plot and display the output. The user interface (UI) can employ the Shinythemes package to create custom themes.

The webpage has been divided into two modules (two tabs displayed at the top of the webpage): cotton single-cell transcriptomic data (scRNA-seq data) and cotton spatial transcriptomic data (Spatial Transcriptomics data). Once a tab is opened, the function parameter is displayed on the left and the result on the right. Above the result display area are the cluster display plots/gene violin plots/FeaturePlot. In the function parameter area, the user can set the size of a point, dimensionality reduction method (*t*-SNE/UMAP), grouping information (cluster or sample), and select the gene of interest to adjust and

obtain the desired results. A list of the top 10 marker genes identified appears below the result display area.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Single-cell RNA-sequence and spatial transcriptomics data have been deposited in NCBI under BioProject [PRJNA800600](#). Spatial metabolomics data have been deposited in [METASPACE](#). LC-MS metabolomics and targeted metabolites data are available in the Open Archive for Miscellaneous Data (OMIX) of the National Genomics Data Center (NGDC) under accession [OMIX004901](#) and [OMIX005015](#). Source data are provided in this paper.

Code availability

Customized code used for data analysis can be found at Github [<https://github.com/superly917/landscape-of-cotton-fiber-cells>].

References

- Chen, Z. J. et al. Toward sequencing cotton (*Gossypium*) genomes. *Plant Physiol.* **145**, 1303–1310 (2007).
- Tiwari, S. C. & Wilkins, T. A. Cotton (*Gossypium hirsutum*) seed trichomes expand via diffuse growing mechanism. *Can. J. Bot.* **73**, 746–757 (1995).
- Yang, Z., Qanmber, G., Wang, Z., Yang, Z. & Li, F. *Gossypium* genomics: trends, scope, and utilization for cotton improvement. *Trends Plant Sci.* **25**, 488–500 (2020).
- Stewart, J. M. Fiber initiation on the cotton ovule (*Gossypium Hirsutum*). *Am. J. Bot.* **62**, 723–730 (1975).
- Haigler, C. H., Betancur, L., Stiff, M. R. & Tuttle, J. R. Cotton fiber: a powerful single-cell model for cell wall and cellulose research. *Front. Plant Sci.* **3**, 104 (2012).
- Lee, J. J., Woodward, A. W. & Chen, Z. J. Gene expression changes and early events in cotton fibre development. *Ann. Bot.* **100**, 1391–1401 (2007).
- Kim, H. J. & Triplett, B. A. Cotton fiber growth in planta and in vitro. Models for plant cell elongation and cell wall biogenesis. *Plant Physiol.* **127**, 1361–1366 (2001).
- Wu, Y., Machado, A. C., White, R. G., Llewellyn, D. J. & Dennis, E. S. Expression profiling identifies genes expressed early during lint fibre initiation in cotton. *Plant Cell Physiol.* **47**, 107–127 (2006).
- Walford, S. A., Wu, Y., Llewellyn, D. J. & Dennis, E. S. GhMYB25-like: a key factor in early cotton fibre development. *Plant J.* **65**, 785–797 (2011).
- Wang, S. et al. Control of plant trichome development by a cotton fiber MYB gene. *Plant Cell* **16**, 2323–2334 (2004).
- Walford, S. A., Wu, Y., Llewellyn, D. J. & Dennis, E. S. Epidermal cell differentiation in cotton mediated by the homeodomain leucine zipper gene, GhHD-1. *Plant J.* **71**, 464–478 (2012).
- Deng, F. L. et al. GbPDF1 is involved in cotton fiber initiation via the core cis-element HDZIP2ATATHB2. *Plant Physiol.* **158**, 890–904 (2012).
- Xiao, G. et al. Genome-wide identification of the GhARF gene family reveals that GhARF2 and GhARF18 are involved in cotton fibre cell initiation. *J. Exp. Bot.* **69**, 4323–4337 (2018).
- Zhang, M. et al. Auxin regulates cotton fiber initiation via GhPIN-mediated auxin transport. *Plant Cell Physiol.* **58**, 385–397 (2017).
- Cheng, Y. et al. GhCaM7-like, a calcium sensor gene, influences cotton fiber elongation and biomass production. *Plant Physiol. Biochem.* **109**, 128–136 (2016).
- Pu, L., Li, Q., Fan, X., Yang, W. & Xue, Y. The R2R3 MYB transcription factor GhMYB109 is required for cotton fiber development. *Genetics* **180**, 811–820 (2008).
- Zhang, B. & Liu, J. Y. Cotton cytosolic pyruvate kinase GhPK6 participates in fast fiber elongation regulation in a ROS-mediated manner. *Planta* **244**, 915–926 (2016).
- Shan, C. M. et al. Control of cotton fibre elongation by a homeodomain transcription factor GhHOX3. *Nat. Commun.* **5**, 5519 (2014).
- Zhao, B. et al. Core cis-element variation confers subgenome-biased expression of a transcription factor that functions in cotton fiber elongation. *N. Phytol.* **218**, 1061–1075 (2018).
- Xiao, G. H., Zhao, P. & Zhang, Y. A pivotal role of hormones in regulating cotton fiber development. *Front. Plant Sci.* **10**, 87 (2019).
- Gou, J. Y., Wang, L. J., Chen, S. P., Hu, W. L. & Chen, X. Y. Gene expression and metabolite profiles of cotton fiber during cell elongation and secondary cell wall synthesis. *Cell Res.* **17**, 422–434 (2007).
- Sun, Y. et al. Brassinosteroid signaling affects secondary cell wall deposition in cotton fibers. *Ind. Crops Prod.* **65**, 334–342 (2015).
- Xia, X. C. et al. Cotton (*Gossypium hirsutum*) JAZ3 and SLR1 function in jasmonate and gibberellin mediated epidermal cell differentiation and elongation. *Plant Cell Tiss. Org.* **133**, 249–262 (2018).
- Yamashino, T., Kitayama, M. & Mizuno, T. Transcription of ST2A encoding a sulfotransferase family protein that is involved in jasmonic acid metabolism is controlled according to the circadian clock- and PIF4/PIF5-mediated external coincidence mechanism in *Arabidopsis thaliana*. *Biosci. Biotechnol. Biochem.* **77**, 2454–2460 (2013).
- Tan, J. et al. A genetic and metabolic analysis revealed that cotton fiber cell development was retarded by flavonoid naringenin. *Plant Physiol.* **162**, 86–95 (2013).
- Liu, G. J. et al. Targeted lipidomics studies reveal that linolenic acid promotes cotton fiber elongation by activating phosphatidylinositol and phosphatidylinositol monophosphate biosynthesis. *Mol. Plant* **8**, 911–921 (2015).
- Farmer, A., Thibivilliers, S., Ryu, K. H., Schiebelbein, J. & Libault, M. 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 (2021).
- Satterlee, J. W., Strable, J. & Scanlon, M. J. Plant stem-cell organization and differentiation at single-cell resolution. *Proc. Natl. Acad. Sci. USA* **117**, 33689–33699 (2020).
- Liu, Z. et al. Global dynamic molecular profiling of stomatal lineage cell development by single-cell RNA sequencing. *Mol. Plant* **13**, 1178–1193 (2020).
- Hou, X. L., Wang, L., Ding, Y. L., Xie, Q. & Diao, H. Y. Current status and recent advances of next generation sequencing techniques in immunological repertoire. *Genes Immun.* **17**, 153–164 (2016).
- Bawa, G., Liu, Z., Yu, X., Qin, A. & Sun, X. Single-cell RNA sequencing for plant research: insights and possible benefits. *Int. J. Mol. Sci.* **23**, 4497 (2022).
- Qin, Y. et al. Single-cell RNA-seq reveals fate determination control of an individual fibre cell initiation in cotton (*Gossypium hirsutum*). *Plant Biotechnol. J.* **20**, 2372–2388 (2022).
- Wong, K., Navarro, J. F., Bergenstråhlé, L., Ståhl, P. L. & Lundeberg, J. ST Spot Detector: a web-based application for automatic spot and tissue detection for spatial Transcriptomics image datasets. *Bioinformatics* **34**, 1966–1968 (2018).
- Luo, Z. et al. Air flow-assisted ionization imaging mass spectrometry method for easy whole-body molecular imaging under ambient conditions. *Anal. Chem.* **85**, 2977–2982 (2013).
- Nakabayashi, R. et al. Spatial metabolomics using imaging mass spectrometry to identify the localization of asparaginase A in *Asparagus officinalis*. *Plant Biotechnol.* **38**, 311–315 (2021).
- Tong, Q. et al. Biosynthesis-based spatial metabolome of *Salvia miltiorrhiza* Bunge by combining metabolomics approaches with mass spectrometry-imaging. *Talanta* **238**, 123045 (2022).

37. Zhang, T. et al. Chemical imaging reveals diverse functions of tricarboxylic acid metabolites in root growth and development. *Nat. Commun.* **14**, 2567 (2023).
38. Zhang, C. et al. In situ characterisation of phytohormones from wounded *Arabidopsis* leaves using desorption electrospray ionisation mass spectrometry imaging. *Analyst* **146**, 2653–2663 (2021).
39. Li, B. et al. Interrogation of spatial metabolome of *Ginkgo biloba* with high-resolution matrix-assisted laser desorption/ionization and laser desorption/ionization mass spectrometry imaging. *Plant Cell Environ.* **41**, 2693–2703 (2018).
40. Dong, Y. et al. High mass resolution, spatial metabolite mapping enhances the current plant gene and pathway discovery toolbox. *New Phytol.* **228**, 1986–2002 (2020).
41. Chini, A. et al. An OPR3-independent pathway uses 4,5-didehydrojasmonate for jasmonate synthesis. *Nat. Chem. Biol.* **14**, 171–178 (2018).
42. Wójcikowska, B. et al. LEAFY COTYLEDON2 (LEC2) promotes embryogenic induction in somatic tissues of *Arabidopsis*, via YUCCA-mediated auxin biosynthesis. *Planta* **238**, 425–440 (2013).
43. Lee, J. et al. Xyloglucan endotransglycosylase/hydrolase genes in cotton and their role in fiber elongation. *Planta* **232**, 1191–1205 (2010).
44. Shani, E. et al. Plant stress tolerance requires auxin-sensitive Aux/IAA transcriptional repressors. *Curr. Biol.* **27**, 437–444 (2017).
45. Schellingen, K. et al. Cadmium-induced ethylene production and responses in *Arabidopsis thaliana* rely on ACS2 and ACS6 gene expression. *BMC Plant Biol.* **14**, 214 (2014).
46. Han, J., Tan, J., Tu, L. & Zhang, X. A peptide hormone gene, GhPSK promotes fibre elongation and contributes to longer and finer cotton fibre. *Plant Biotechnol. J.* **12**, 861–871 (2014).
47. Sable, A. et al. Inhibition of Heat Shock proteins HSP90 and HSP70 induce oxidative stress, suppressing cotton fiber development. *Sci. Rep.* **8**, 3620 (2018).
48. Sun, Y. et al. Brassinosteroid regulates fiber development on cultured cotton ovules. *Plant Cell Physiol.* **46**, 1384–1391 (2005).
49. Friedrichsen, D. M. et al. Three redundant brassinosteroid early response genes encode putative bHLH transcription factors required for normal growth. *Genetics* **162**, 1445–1456 (2002).
50. Xia, K. et al. The single-cell stereo-seq reveals region-specific cell subtypes and transcriptome profiling in *Arabidopsis* leaves. *Dev. Cell* **57**, 1299–1310 (2022).
51. Fawkner-Corbett, D. et al. Spatiotemporal analysis of human intestinal development at single-cell resolution. *Cell* **184**, 810–826 (2021).
52. Chen, L. et al. AKR2A interacts with KCS1 to improve VLCFAs contents and chilling tolerance of *Arabidopsis thaliana*. *Plant J.* **103**, 1575–1589 (2020).
53. Ma, Y. et al. FLA11 and FLA12 glycoproteins fine-tune stem secondary wall properties in response to mechanical stresses. *N. Phytol.* **233**, 1750–1767 (2022).
54. Chen, H. et al. Crystal structure of the extracellular domain of the receptor-like kinase TMK3 from *Arabidopsis thaliana*. *Acta Crystallogr. F. Struct. Biol. Commun.* **76**, 384–390 (2020).
55. Nowak, K. & Gaj, M. D. Stress-related function of bHLH109 in somatic embryo induction in *Arabidopsis*. *J. Plant Physiol.* **193**, 119–126 (2016).
56. Pierdonati, E. et al. Cytokinin-dependent control of GH3 group II family genes in the *Arabidopsis* root. *Plants* **8**, 94 (2019).
57. Kuroha, T. et al. Identification of *Arabidopsis* subtilisin-like serine protease specifically expressed in root stele by gene trapping. *Physiol. Plant* **137**, 281–288 (2009).
58. Bannenberg, G. et al. Functional analysis of alpha-DOX2, an active alpha-dioxygenase critical for normal development in tomato plants. *Plant Physiol.* **151**, 1421–1432 (2009).
59. Sun, S. et al. GASA14 regulates leaf expansion and abiotic stress resistance by modulating reactive oxygen species accumulation. *J. Exp. Bot.* **64**, 1637–1647 (2013).
60. Ding, X. et al. Sucrose enhanced reactive oxygen species generation promotes cotton fibre initiation and secondary cell wall deposition. *Plant Biotechnol. J.* **19**, 1092–1094 (2021).
61. Han, X. et al. *Arabidopsis* transcription factor TCP5 controls plant thermomorphogenesis by positively regulating PIF4 activity. *iScience* **15**, 611–622 (2019).
62. Spanik, I. & Machynakova, A. Recent applications of gas chromatography with high-resolution mass spectrometry. *J. Sep. Sci.* **41**, 163–179 (2018).
63. Sun, C. L. et al. Spatially resolved metabolomics to discover tumor-associated metabolic alterations. *Proc. Natl. Acad. Sci. USA* **116**, 52–57 (2019).
64. He, P. et al. Comprehensive analyses of ZFP gene family and characterization of expression profiles during plant hormone response in cotton. *BMC Plant Biol.* **19**, 329 (2019).
65. Ahmed, M. et al. Overexpression of a sucrose synthase gene indirectly improves cotton fiber quality through sucrose cleavage. *Front. Plant Sci.* **11**, 476251 (2020).
66. Huang, S. & Millar, A. H. Succinate dehydrogenase: the complex roles of a simple enzyme. *Curr. Opin. Plant Biol.* **16**, 344–349 (2013).
67. Wang, L. M. et al. Comparative transcriptomics reveals jasmonic acid-associated metabolism related to cotton fiber initiation. *Plos ONE* **10**, e0129854 (2015).
68. Yang, Z. R. et al. PAG1, a cotton brassinosteroid catabolism gene, modulates fiber elongation. *N. Phytol.* **203**, 437–448 (2014).
69. Haghverdi, L., Lun, A. T. L., Morgan, M. D. & Marioni, J. C. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* **36**, 421–427 (2018).
70. Wang, M. J. et al. Reference genome sequences of two cultivated allotetraploid cottons, *Gossypium hirsutum* and *Gossypium barbadense*. *Nat. Genet.* **51**, 224–229 (2019).

Acknowledgements

This study was funded by support from the National Key R & D Program (Grant No. 2021YFF1000100 and 2022YFD1200300), the Fundamental Research Funds of State Key Laboratory of Cotton Bio-breeding and Integrated Utilization (CB2021E02), the National Natural Science Foundation of China (Grant No. 32072114 and 31670233), the CAAS Elite Youth Program, the Top Talent Project of Henan Province (ZYY-CYU202012146), the Innovation Project of CAAS (Grant No. CAAS-ASTIP-ICR-KP-2021-01), the Major Science and Technology Program of Changji Hui Autonomous Prefecture (Grant No. 2021Z01), the China Agriculture Research System (CARS-15-07), the Natural Science Foundation of Henan Province (Grant No. 202300410550). We thank Dr. Jiamin Gao for their help in data processing.

Author contributions

Conceptualization: X.S. and X.M.; Methodology: Z.L., A.Q., X.Z., Y.L. (Yao Lu), and J.H.; Formal Analysis: X.S., C.G., and A.Q.; Investigation: X.Y., Y.Z., Y.L. (Yumeng Liu), J.Y., M.H., H.L., Z.Z., G.H., W.L., X.Z., S.D., and S.S.; Resources: X.S., X.W., X.G., and X.M.; Writing: X.S., Z.L., and A.Q.; Writing – Review & Editing: X.S., A.Q., L.H.E., X.W., L.Y.V., and L.S.P.T.; Supervision: X.S. and X.M.; Project Administration: X.S. and X.M.; Funding Acquisition: X.S. and X.M.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at
<https://doi.org/10.1038/s41467-025-55869-z>.

Correspondence and requests for materials should be addressed to Xiongfeng Ma.

Peer review information *Nature Communications* thanks Yuxian Zhu, and the other anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at
<http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025, corrected publication 2025