



Spatiotemporal transcriptome and metabolome landscapes of cotton somatic embryos

Received: 20 October 2022

Accepted: 2 January 2025

Published online: 20 January 2025

Check for updates

Xiaoyang Ge^{1,2,5}, Xiaole Yu^{3,5}, Zhixin Liu^{3,5}, Jiachen Yuan^{1,5}, Aizhi Qin^{3,5}, Ye Wang^{2,5}, Yanli Chen^{2,5}, Wenqiang Qin^{2,5}, Yumeng Liu³, Xingxing Liu¹, Yaping Zhou³, Peng Wang², Jincheng Yang³, Hao Liu³, Zihao Zhao³, Mengke Hu³, Yixin Zhang³, Susu Sun³, Luis Herrera-Estrella^④, Lam-Son Phan Tran⁴, Xuwu Sun^{③✉} & Fuguang Li^{①,2✉}

Somatic embryogenesis (SE) is a developmental process related to the regeneration of tissue-cultured plants, which serves as a useful technique for crop breeding and improvement. However, SE in cotton is difficult and elusive due to the lack of precise cellular level information on the reprogramming of gene expression patterns involved in somatic embryogenesis. Here, we investigate the spatial and single-cell expression profiles of key genes and the metabolic patterns of key metabolites by integrated single-cell RNA-sequencing (scRNA-seq), spatial transcriptomics (ST), and spatial metabolomics (SM). To evaluate the results of these analyses, we functionally characterized the potential roles of two representative marker genes, *AATP1* and *DOX2*, in the regulation of cotton somatic embryo development. A publicly available web-based resource database (<https://cotton.cricaas.com.cn/somaticembryo/>) in this study provides convenience for future studies of the expression patterns of marker genes at specific developmental stages during the process of SE in cotton.

Cotton (*Gossypium* spp.) is one of the most important commodity crops in the world¹. However, genetic transformation of cotton relies on a complex, time-consuming, and genotype-dependent process of somatic embryo regeneration². Somatic embryogenesis (SE) is a developmental process in which somatic cells dedifferentiate into totipotent embryonic stem cells that can give rise to somatic embryos under appropriate conditions, often via the formation of undifferentiated callus tissues³. Although SE is one of the most well-studied regeneration processes in plants, there is still lack of fundamental, cellular-level information about the reprogramming of gene

expression patterns involved in somatic embryo formation⁴. In cotton, SE involves indirect embryogenesis⁵, in which hypocotyl or other explant cells are cultivated under *in vitro* conditions to generate undifferentiated callus tissues. Next, callus cells are induced to form embryonic cells that have the capacity to develop into an initial somatic embryo to form globular embryo⁶. Subsequently, the globular embryos differentiate into torpedo embryos in an orderly process that includes the formation of well-developed vascular bundles, which are essential to maintain the shape of torpedo embryos⁷. Cotyledons develop at the top of the torpedo embryo to form a cotyledonary

¹Zhengzhou Research Base, State Key Laboratory of Cotton Bio-breeding and Integrated Utilization, School of Agricultural Sciences, Zhengzhou University, Zhengzhou 450000, China. ²State Key Laboratory of Cotton Bio-breeding and Integrated Utilization, Institute of Cotton Research of the Chinese Academy of Agricultural Sciences, Anyang 455000, China. ³State Key Laboratory of Crop Stress Adaptation and Improvement, School of Life Sciences, Henan University, 85 Minglun Street, Kaifeng 475001, China. ⁴Institute of Genomics for Crop Abiotic Stress Tolerance, Department of Plant and Soil Science, Texas Tech University, Lubbock, TX 79409, USA. ⁵These authors contributed equally: Xiaoyang Ge, Xiaole Yu, Zhixin Liu, Jiachen Yuan, Aizhi Qin, Ye Wang, Yanli Chen, Wenqiang Qin. ✉e-mail: sunxuwu@henu.edu.cn; aylifug@caas.cn

embryo⁸. On suitable medium, cotyledonary embryos can develop into a mature stage capable of forming roots and buds⁹. These materials are suitable for transplanting to develop into regenerated plants, thus completing the process of in vitro regeneration¹⁰.

Unlike many other plants, in vitro regeneration in cotton is still a difficult and complex process². In cotton, a large number of genes are differentially expressed and interact with each other during different developmental stages and in different tissues to precisely regulate the development of somatic embryos¹¹. Research suggests that the *leafy cotyledon* (*LEC*) gene regulates many aspects of embryonic development, including cotyledon development and seed maturation^{12,13}, while *wuschel* (*WUS*) is involved in regulating callus differentiation¹⁴. In addition, genomic research suggests that microRNAs (miRNAs) participate in and regulate the process of plant SE¹⁵. Analyses of global transcriptional changes during SE have been conducted in several plant species, including maize (*Zea mays*), *Arabidopsis thaliana*, soybean (*Glycine max*), and several woody species, among others^{16–19}. Recently, RNA sequencing (RNA-seq) has been utilized to study cell wall characteristics during the differentiation of *A. thaliana* endosperm cells and embryos²⁰, and to obtain transcriptome maps of maize seed development²¹. However, few studies have sought to dissect this fundamental developmental process at the single-cell level.

Single-cell RNA-seq (scRNA-seq) is an innovative technique for studying the complex process of somatic embryo development at a single-cell resolution. To date, scRNA-seq has been applied to *A. thaliana*, rice (*Oryza sativa*), and maize, among many other plant species^{22–26}. Spatial transcriptomics (ST) methods are well-suited for the visualization of gene expression patterns obtained from scRNA-seq data²⁷. Emerging spatial metabolomics (SM) techniques utilize mass spectrometry imaging (MSI) to directly study the spatial distribution of both known and unknown metabolites in biological tissues, and to create metabolic imaging maps^{28–31}. SM has been used to track metabolite movement resulting from nematode infection in tomato (*Solanum lycopersicum*) roots³², and to document the distribution of specific metabolites in ginseng (*Panax ginseng*)³³ and maize roots³⁴.

Previous studies reported that both polyamines and hydrogen peroxide (H_2O_2) play key roles in the transformation of cotton embryogenic callus into somatic embryos³⁵. Phytohormones, such as zeatin (a type of cytokinins³⁶), are also important for inducing direct SE in cotton³⁷. However, to date, there are no published reports on the application of ST to study the development of plant somatic embryos. Early embryonic development primarily involves differentiation and the development of specific cell types and tissues. Precise detection of the dynamics of transcriptomic and metabolic changes undergone by specific cell types will aid in studies of the metabolic and genetic regulation of cellular development.

Here, we use a multi-omics strategy to perform detailed transcriptomic and metabolomic characterization of the four key stages of somatic embryo development in cotton (e.g., callus, globular embryo, torpedo embryo, and cotyledonary embryo) based on scRNA-seq, ST, and SM. To integrate the ST and SM datasets, we consistently set the spatial resolution in the study at 100 μ m. We screen and identify key genes and metabolites that regulate somatic embryo development (Fig. 1). Furthermore, we integrated metabolite (e.g., polyamines and phytohormones) screening data with transcriptional data from related genes to achieve a multiplex analysis of gene-metabolite correlations.

Results

Spatial dynamics of gene expression during somatic embryo development

We first utilized ST to comprehensively assess the spatial expression patterns of genes during somatic embryo development in cotton. Specifically, ST was used to conduct a comprehensive transcriptome analysis of different cotton tissues across all stages of SE. First, globular, torpedo, and cotyledonary embryonic tissues were sectioned and

frozen (Supplementary Fig. 1). Three sampled embryonic stages consisted of three sections, each from three different embryos at the same developmental stage, and each spot on each capture area corresponding to 1–10 cells (Fig. 2a–c). After quality control, we ended up with 801–1217 high-quality spots per section, with an average of 11,857–16,401 unique molecular identifiers (nUMIs) per spot, and an average of 6287–8001 expressed genes per spot (Supplementary Fig. 2a–c, Supplementary Data 1).

We used uniform manifold approximation and projection (UMAP) to divide the data into 13 cell clusters, which were then mapped back to its original position in the tissue section (Figs. 2d and e). Each of the 13 clusters was named based on its spatial location. We also constructed a dotplot of the most highly expressed genes in each cluster (Fig. 2f). These highly expressed, cluster-specific genes may contribute to the development of specific cell clusters. For example, *Gh_D05G159100*, which is highly expressed in ‘pro-embryogenic cell’ (cluster 6), is a homolog of *wuschel-related homeobox 9* (*WOX9*) that plays a role in embryogenesis and meristems in plants³⁸. In ‘cotyledon cell’ (cluster 9), *Gh_A07G040500* exhibits high homology with *axial regulator YABBY 5* (*YABS*)³⁹. In ‘cortex cell_12’ (cluster 12), *Gh_D08G133100* is homologous to *YUCCA10* (*YUC10*), which encodes a flavin monooxygenase and is associated with both embryogenesis and auxin synthesis in *A. thaliana*⁴⁰.

The differentially expressed genes (DEGs) were identified across the 13 clusters (Supplementary Data 2), and gene ontology (GO) enrichment analysis was conducted on the DEGs (Supplementary Table 1). Genes associated with the ‘regulation of post-embryonic development’ were highly expressed in ‘cotyledon cell’ (cluster 9), while genes associated with ‘defense response to other organism’ were highly expressed in ‘cortex cell_11’ (cluster 11) (Supplementary Table 1). Both *DOX2* (*Gh_A09G193100*) and *WOX9* (*Gh_D05G159100*) were highly expressed in ‘pro-embryogenic cell’ (cluster 6), and this cluster of cells is involved in ‘shoot system development’ (Supplementary Table 1). These results suggest that DEGs may play specific physiological functions during the different stages of SE.

Spatial expression dynamics of genes regulating the initial development of somatic embryos

The initial induction of callus cells and the formation of embryogenic cells are the most critical stages for the development of somatic embryos in cotton⁴¹. Because scRNA-seq has proven useful in the study of different cell types in callus tissues of *A. thaliana*⁴², we conducted an ST analysis of the callus induced from cotton hypocotyl (Fig. 3, Supplementary Data 3). As a result, six cell clusters were obtained (Fig. 3a).

Both the homolog of the vascular tissue marker gene *aintegumenta* (*ANT*) (*Gh_A11G229400*) and the homolog of the stem cell formation-related gene *wuschel-related homeobox 13* (*WOX13*) (*Gh_A02G209400*) were expressed in cluster 1^{43,44}, suggesting that cluster 1 contained vascular tissue cells from explants and callus founder cells. Based on a previous identification of callus cell types⁴², cluster 1 was identified as ‘explant vasculature and callus founder cells_1’. Similarly, a homolog of the procambium developmental marker gene *homeobox-leucine zipper protein ATHB-15* (*ATHB-15*) (*Gh_D05G130700*) was expressed in ‘explant vasculature and callus founder cells_3’ (cluster 3)⁴⁵, and homologs of the vascular cambium developmental marker genes *wuschel-related homeobox 4* (*WOX4*) (*Gh_A05G188600*) and *WOX13* were expressed in ‘explant vasculature and callus founder cells_6’ (cluster 6)^{44,46}. The epidermal cell marker gene *ethylene responsive element binding factor 4* (*ERF4*) (*Gh_A10G019700*) was expressed in ‘outer cell layer of the explant’ (cluster 2)⁴⁷. A homolog of the embryonic root apical meristem marker gene *like auxin resistant 2* (*LAX2*) (*Gh_A01G254000*) was expressed in ‘middle cell layer of callus cells’ (cluster 4). A homolog of *WOX4* was expressed in ‘inner cell layer of callus cells’ (cluster 5)⁴⁶.

Interestingly, ‘explant vasculature and callus founder cells_3’ (cluster 3) was not clustered together with ‘explant vasculature and

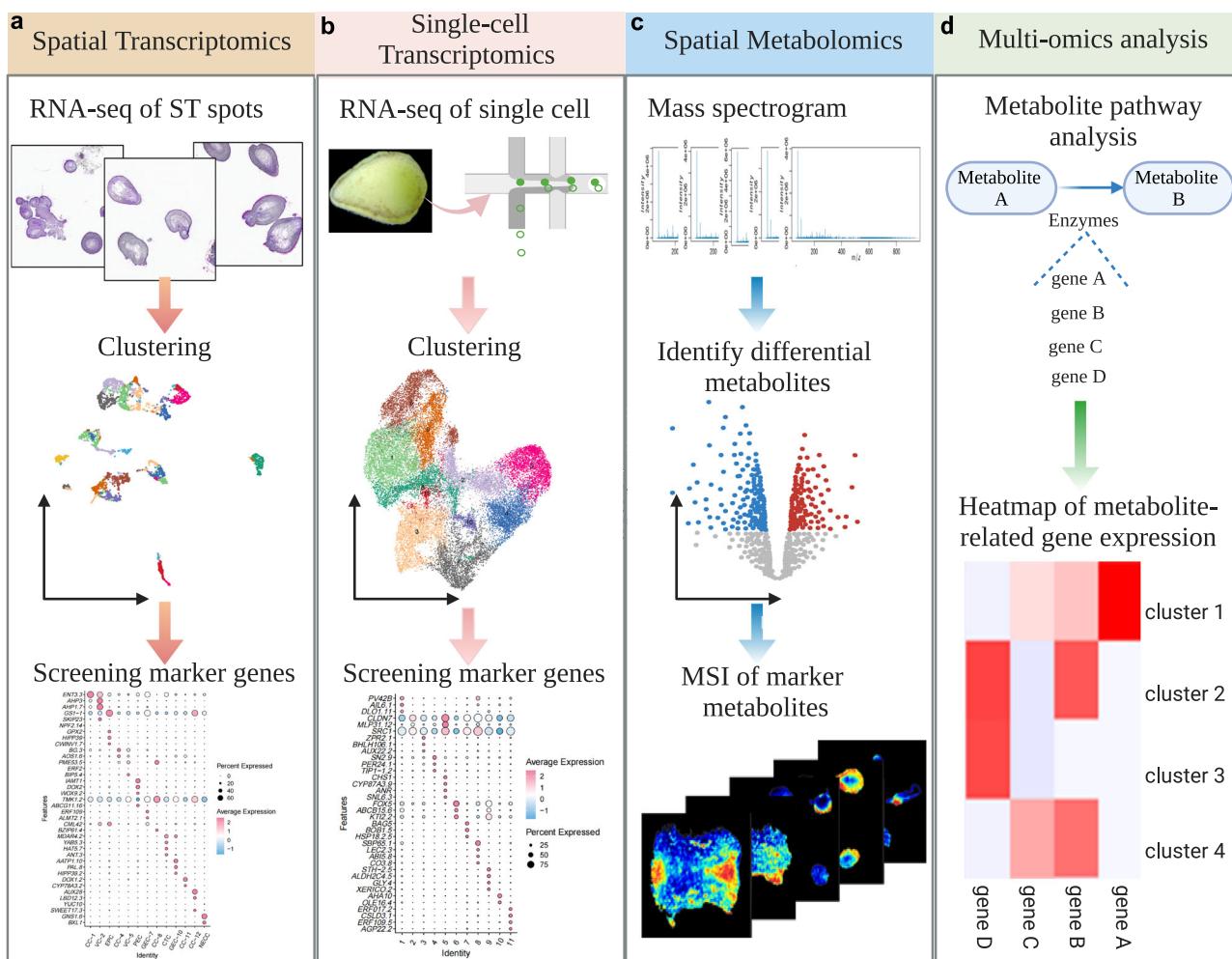


Fig. 1 | A molecular approach for the development of SE in cotton. **a** ST analysis of cotton non-embryogenic callus tissues, globular embryos, torpedo embryos, and cotyledonary embryos. **b** Analysis of the heterogeneity of somatic embryo cell types in cotton by scRNA-seq. **c** SM analysis of metabolic pathways related to SE.

d Creation of a multi-omic cotton SE atlas. The genes and clusters (cell types) of interest were selected, and a heatmap was generated based on the expression levels of the metabolite-related genes in each cell cluster.

callus founder cells_1' (cluster 1) or 'explant vasculature and callus founder cells_6' (cluster 6) (Supplementary Table 2), which suggests that similar cell types may be found in multiple cell clusters and multiple developmental stages. The expression of marker genes associated with each cell cluster is presented in Fig. 3b, c. Further GO analysis indicated that the majority of DEGs in 'explant vasculature and callus founder cells_3' and 'inner cell layer of callus cells' (clusters 3 and 5) are involved in the 'response to hypoxia' and 'Defense response to other organism' (Supplementary Table 2).

Gene expression patterns across multiple developmental stages of cotton somatic embryos

To further study the relationship between gene expression and tissue development in cotton somatic embryos, we divided all the samples from the three developmental stages into 9 sections and performed a Pearson correlation analysis (Supplementary Fig. 2d). We then analyzed the GO enrichment terms and spatial expression of DEGs in different cell clusters within the 9 sections (Supplementary Figs. 3–11, Supplementary Tables 3–11, and Supplementary Data 4–12).

In section 1, a homolog of the vascular tissue marker gene *WOX13* was highly expressed in 'non-embryogenic callus cell' (cluster 1)⁴⁴. Both the homolog of the early embryonic vascular marker gene *ATHB-8* (*Gh_D05G049400*) and the homolog of the SE-inducing gene *baby*

(*BBM*) (*Gh_D08G251600*) were highly expressed in 'globular embryo cell' (cluster 2)^{48,49} (Supplementary Data 4).

In section 2, both the homolog of the embryonic epidermal cell marker gene *protodermal factor 1* (*PDF1*) (*Gh_D04G040400*) and the homolog of the gene responsible for embryonic pattern formation (*WOX9*)⁵⁰ were highly expressed in 'pro-embryogenic cells_3' (cluster 3). A homolog of the SE marker gene *AP2-like ethylene-responsive transcription factor 6* (*AIL6*) (*Gh_D10G104500*)⁵¹ was highly expressed in 'pro-embryogenic cells_4' (cluster 4). A homolog of the SE-associated marker gene *BBM*⁴⁸ was highly expressed in 'pro-embryogenic cells_5' (cluster 5). Both the homolog of *auxin response factor 5* (*ARF5*) (*Gh_A01G116800*) and the homolog of *ANT* associated with embryonic vascular tissue formation were highly expressed in 'globular embryo cell_1' (cluster 1)^{43,52}. A homolog of the SE marker gene *floral homeotic protein apetala 2* (*AP2*) (*Gh_A11G207700*)⁵³ was highly expressed in 'globular embryo cell_2' (cluster 2). Finally, the homologs of both *WOX9* and *WOX13*^{38,44} were highly expressed in 'globular embryo cell_6' (cluster 6) (Supplementary Data 5).

Genes regulating cambium development in cotton somatic embryos

The spatial patterns of division and differentiation of stem cells are the basis of plant growth and development⁵⁴. Furthermore, the differentiation of embryogenic cells marks the beginning of embryonic

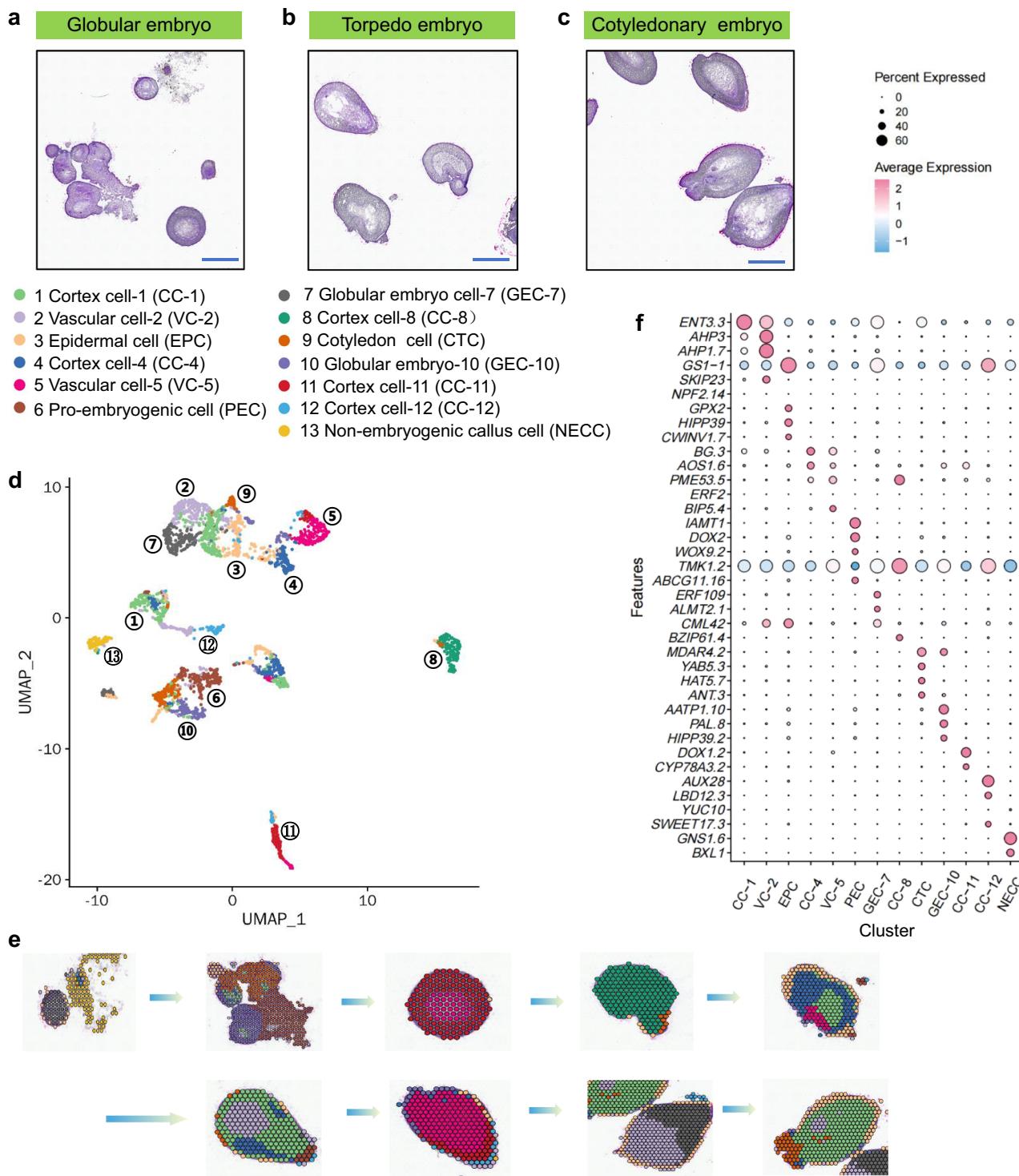


Fig. 2 | Spatial analysis of the three stages of somatic embryogenesis in cotton. a–c H&E (hematoxylin-eosin) stained sections of cotton globular, torpedo, and cotyledonary embryos. The H&E stain experiment was repeated three times. Scale bars (blue line) in (a), (b), and (c) = 400 µm. d UMAP visualization of somatic embryonic cells. Each number represents a different cell type, with each cell type indicated by a different color. From cluster 1 to cluster 13, the number of spots contained in each cluster is 443, 313, 286, 267, 229, 210, 200, 190, 183, 181, 143, 115,

and 101, respectively. e All clusters in (d) were mapped to their spatial locations, with each of the 13 clusters located in a different organizational region. f Dotplot shows the expression of the representative marker genes expressed in each 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). Source data are provided as a Source Data file.

development⁵⁵. During embryonic development and maturation, differentiation begins at the shoot and root meristems at the base of the embryo, as well as in the cambium⁵⁶. The differentiation of cambial cells is essential for the formation of the cortex and the development of vascular tissue⁴³.

We found that many well-known transcription factor (TF)-encoding genes were highly expressed in cambial clusters. Many of these, including *revoluta* (*REV*) (*Gh_D13G262600*), *short root* (*SHR*) (*Gh_D03G030200*), *MP/ARFs*, and *ATHB-15*, are associated with cambium differentiation and vascular development in *A. thaliana* and

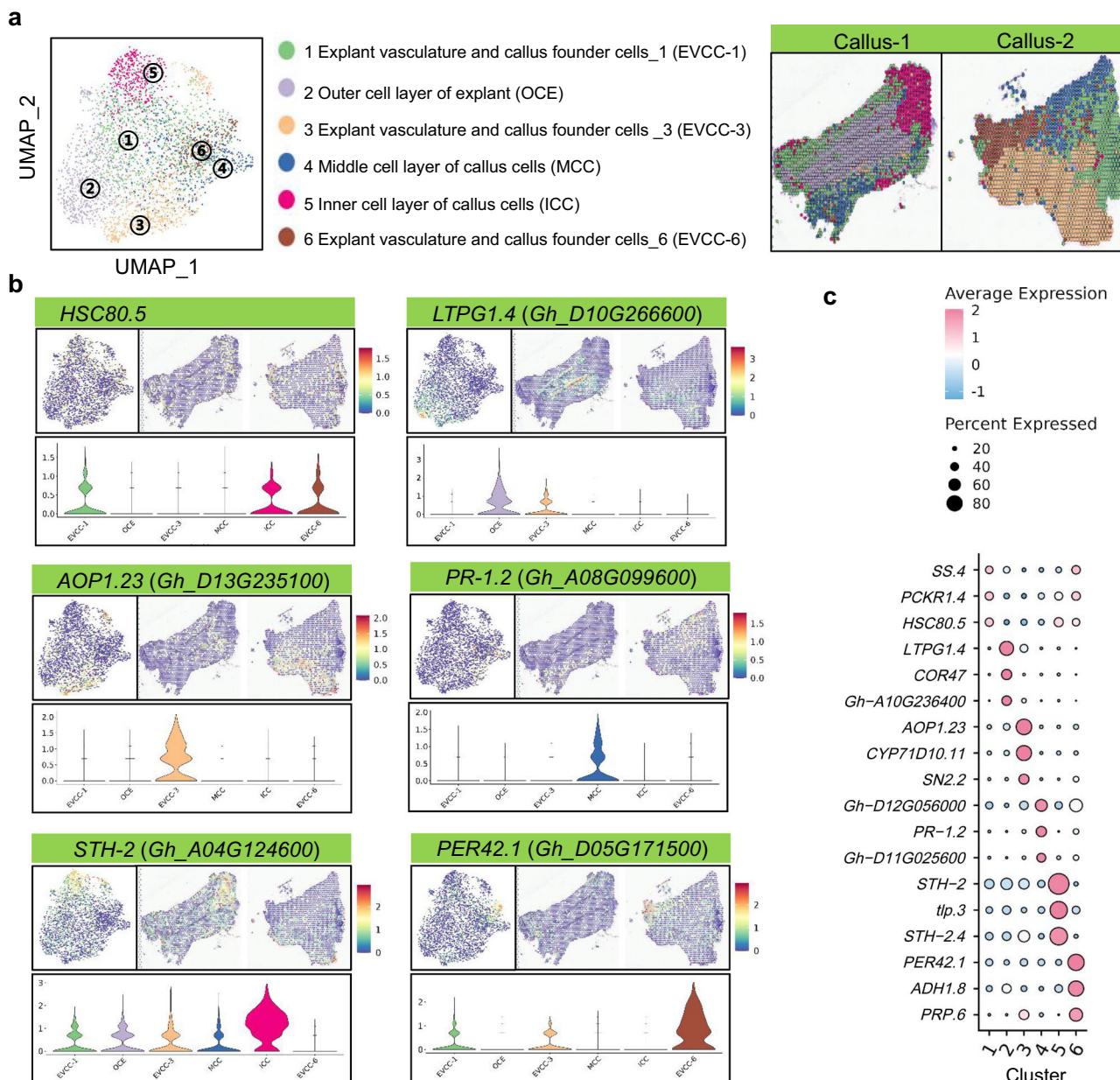


Fig. 3 | Spatial analysis of non-embryogenic callus tissue. **a** UMAP visualization of callus cells. The left figure is the UMAP plot of dimensionality reduction and clustering of each point in callus 1 and callus 2, while the right figure is the spatial position corresponding to different clusters in the sample. The cell type corresponding to each cluster is labeled on the right side of the UMAP plot. From cluster 1 to cluster 6, the number of spots contained in each cluster is 947, 694, 590, 466, 351, and 229, respectively. **b** The spatial expression position map for the marker genes HSC80.5, LTPG1.4, AOP1.23, PR-1.2, STH-2, and PER42.1 in callus_1 and callus_2.

corresponding to the UMAP and genes expressed in the violin map. The UMAP plot of transcript accumulation for the marker genes and the spatial distribution of these marker genes in callus are shown. The color intensity indicates the relative transcript level of each marker gene. **c** Dotplot of the expression of marker genes in each 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). Source data are provided as a Source Data file.

rice^{57–60}. Several TFs of the HB-HD-ZIP gene family, such as *homeobox-leucine zipper protein HAT2* (*HAT2*) (*Gh_A08G191700*), *ATHB-13* (*Gh_D02G129400*), *ATHB-20* (*Gh_D09G022200*), and *ATHB-6* (*Gh_A02G158400*), were also identified in the cambial clusters. The newly discovered TFs associated with the cambium, cotyledon primordium, and hypophysis are listed in Supplementary Data 13 and 14.

ScRNA-seq analysis of gene expression during somatic embryo development

To evaluate the transcriptional heterogeneity of individual cells, we conducted scRNA-seq on protoplasts isolated from the samples used

for ST. Raw data were filtered using a linear model fitting curve and a violin map of the number of genes (nGene) and the nUMIs per cell (Supplementary Fig. 12a–c). After quality control and screening 8842, 9031, and 9037 single-cell transcripts were identified in globular, torpedo, and cotyledonary embryos, with an average of 2339, 1992, and 2146 genes per cell, respectively (Supplementary Data 15). A total of 11 cell clusters were identified after the UMAP analysis (Fig. 4a), and a dotplot of the expression of several highly expressed marker genes associated with each cell cluster was constructed (Fig. 4b). Correlation analysis revealed a strong relationship between the scRNA-seq and ST data (Supplementary Fig. 12d–f).

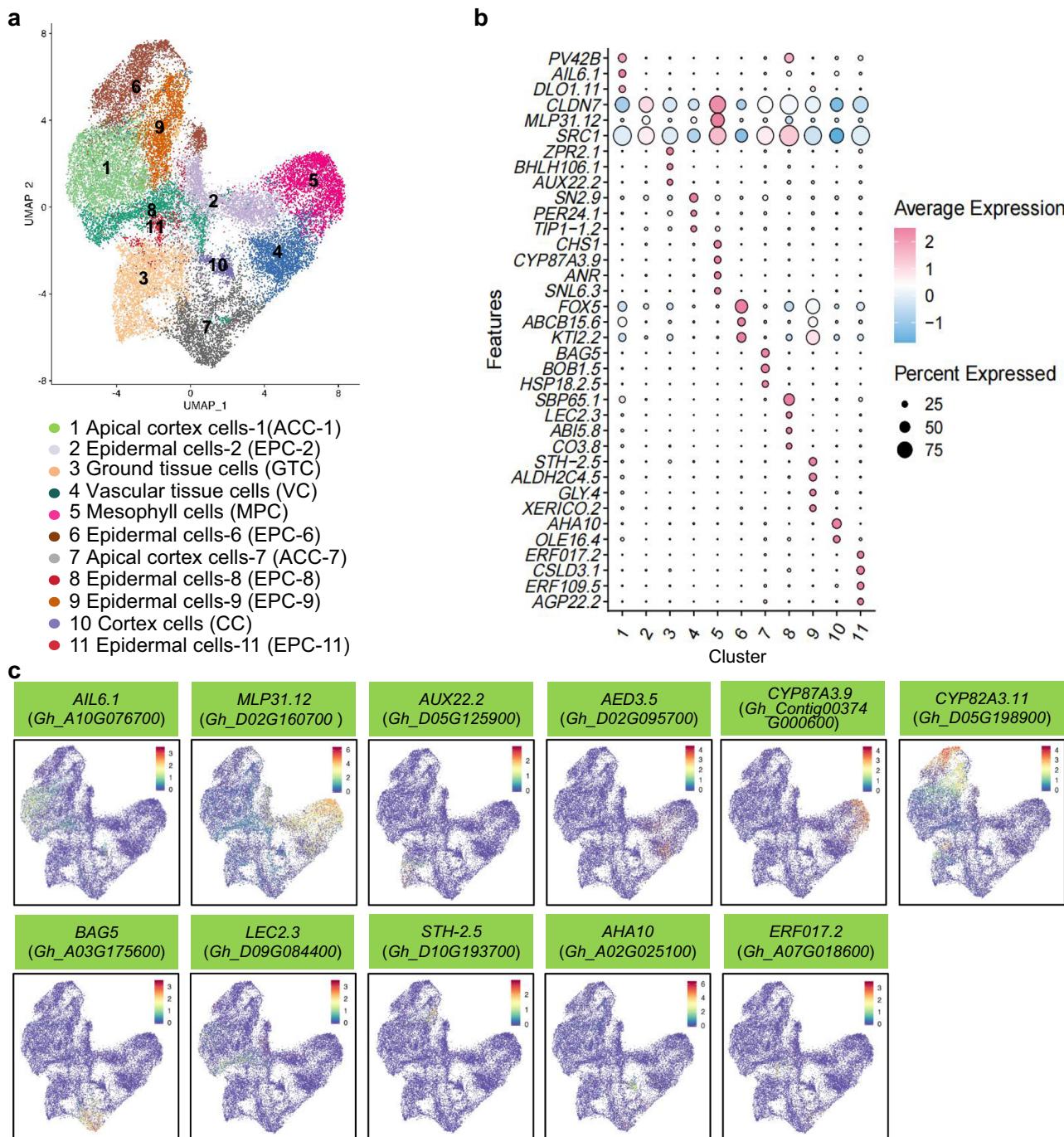


Fig. 4 | Single-cell analysis of somatic embryo development in cotton.

a Reduction and clustering of single-cells in globular, torpedo, and cotyledonary embryos. Each point represents one cell, and different clusters of cells are represented by different colors. From cluster 1 to cluster 11, the number of cells contained in each cluster is 3800, 3356, 3118, 3008, 2807, 2771, 2537, 2284, 2175, 565, and 489, respectively. **b** Dotplot shows the expression of the representative marker

genes in each 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). **c** Expression of the representative marker genes in different cell clusters. The UMAP plot of transcript accumulation for the marker genes in embryos are shown. The color intensity indicates the relative transcript level of each marker gene. Source data are provided as a Source Data file.

Each cell type was annotated with marker genes identified by ST, and DEGs were subsequently identified in each cluster (Supplementary Data 16 and Supplementary Table 12). For example, *Gh_A05G087300*, which was highly expressed in cluster 2, is a homolog of *mitogen-activated protein kinase 3 (MPK3)*, which plays an important role in epidermal cells and embryonic development in *A. thaliana*⁶¹. Thus, cluster 2 was classified as ‘epidermal cells_2’. Homologs of *auxin22 (AUX22)* (*Gh_D05G125900*) and *cytochrome P450 87A3 (CYP87A3)* (*Gh_Contig00374G000600*) were highly expressed in ‘ground tissue

*cells_3’ (clusters 3) and ‘mesophyll cells’ (cluster 5), respectively (Fig. 4c). *AUX22* is involved in the progression from embryogenic callus to somatic embryo⁶². *CYP87A3* is an auxin-induced gene originally isolated from rice coleoptiles during the early stages of growth⁶³. The leaf development-associated gene *bcl-2-associated athanogene family molecular chaperone regulator 5 (BAG5)* (*Gh_D02G193400*)⁶⁴ was highly expressed in cluster 7 (Fig. 4c). Accordingly, cluster 7 was annotated as ‘apical cortex cells_7’. A homolog of *LOB domain-containing protein 12 (LBD12)* (*Gh_A03G037200*), which is associated with the formation of*

shoot apical meristem in rice⁶⁵, was highly expressed in 'apical cortex cells_1', 'epidermal cells_9', and 'cortex cells' (clusters 1, 9, and 10).

Next, we used pseudo-time analysis to construct a developmental trajectory for cell differentiation (Fig. 5a). Embryonic development proceeds from the globular to cotyledonary embryos, so we can determine the start and end points according to the proportion of the cells in the embryos at different developmental stages in each pseudo-time state (Supplementary Fig. 12g), and infer the early and late stages of cell differentiation in the pseudo-time trajectory. Different cell clusters were apparent at different branching positions along the pseudo-time trajectory (Fig. 5b). Such observations can be used to make an initial determination of the relationship between different cell types and developmental stages⁶⁶. For example, similar distribution patterns were observed in the developmental trajectories of 'epidermal cells_2' (cluster 2) and 'ground tissue cells' (cluster 3), indicating that these clusters share similar developmental stages during the process of cell differentiation (Fig. 5b).

Gene expression was analyzed at different pseudo-time stages for each single cell cluster (Fig. 5c), and the data were divided into three categories. In the first category, marker gene expression was the highest during the early pseudo-time stage and then decreased gradually. Marker genes in this category included *ethylene-responsive transcription factor ERF026 (ERF026)* (*Gh_A1IG008000*) in 'epidermal cells_11' (cluster 11), *pathogenesis-related protein STH-2 (STH-2.5)* (*Gh_D10G193700*) in 'epidermal cells_9' (cluster 9), *bZIP transcription factor TGA10 (TGA10.2)* (*Gh_D01G215500*) in 'epidermal cells_8' (cluster 8), and *ATPase 10, plasma membrane-type (AHA10)* (*Gh_A02G025100*) in 'cortex cells' (cluster 10). In the second category, marker gene expression increased gradually over pseudo-time and was the highest during the late pseudo-time stage. Marker genes in this category included *peroxidase 24 (PER24.1)* (*Gh_A1OG093300*) in 'vascular tissue cells' (cluster 4), *cytochrome P450 87A3 (CYP87A3.9)* (*Gh_Contig00374G000600*) in 'mesophyll cells' (cluster 5), and *gibberellin-regulated protein 4 (GASA4.1)* (*Gh_A09G018000*) in 'epidermal cells_2' (cluster 2). In the third category, marker gene expression was the highest during the middle pseudo-time stage and then decreased gradually. Marker genes in this category included *seed biotin-containing protein SBP65 (SBP65.1)* (*Gh_D11G065500*) in 'apical cortex cells_1' (cluster 1) and *expansin-like B1 (EXLB1.4)* (*Gh_A12G236400*) in 'ground tissue cells' (cluster 3). The gene *STH-2* is associated with early germination in the seeds of *Abrus cantoniensis*⁶⁷ and *GASA4* is expressed in the roots and buds of mature *A. thaliana*⁶⁸. These results suggest that genes expressed during the early pseudo-time stage also play key roles in early embryonic development. Overall, the expression pattern of each marker gene matches its role at specific developmental stages.

Construction of a gene expression map for somatic embryo development

As mentioned previously, the initial onset of embryogenesis is critical for the transformation of non-embryogenic callus cells into embryogenic cells. According to the ST data, several highly expressed genes were associated with 'non-embryogenic callus' (cluster 13), 'pro-embryogenic cell' (cluster 6), and 'globular embryo cell_10' (cluster 10) (Supplementary Fig. 13). In 'pro-embryogenic cell' (cluster 6), *Gh_A01G086600* is a homolog of *indole-3-acetate o-methyltransferase 1 (IAMT1)* (Supplementary Fig. 13a), which is involved in auxin metabolism in *A. thaliana*⁶⁹. This suggests that *Gh_A01G086600* may regulate the initial formation of globular embryos through auxin-related metabolic processes. Similarly, *Gh_A12G174900*, also in 'pro-embryogenic cell' (cluster 6), is a homolog of *receptor-like transmembrane kinase 1 (TMK1)* (Supplementary Fig. 13l), which is involved in auxin signal transduction in *A. thaliana*⁷⁰, suggesting that this gene may regulate the formation of globular embryos by participating in auxin signal transduction. *In situ* hybridization (ISH) revealed that *ABC*

transporter G family member 11 (ABCG11) (Gh_D11G232500) was primarily expressed in the endodermis of globular and torpedo embryos (Fig. 6b-d and Supplementary Fig. 14a-f). *ABCG11* encodes an epidermal lipid transporter in *A. thaliana*⁷¹.

The establishment and maintenance of embryonic epidermis are crucial for the normal development of embryos. In higher plants, the epidermis of shoot organs originates from the embryonic epidermis, which protects the embryo from biotic and abiotic stressors^{72,73}. Using the ST dataset, we analyzed key genes expressed in 'epidermal cell' (cluster 3), 'cortex cell_1' (cluster 1), and 'cortex cell_12' (cluster 12) to identify genes that may regulate the development of cortical and epidermal cells. *Cysteine-rich receptor-like protein kinase 29 (CRK29)* belongs to a class of receptor-like kinases which encode cysteine-rich protein kinases⁷⁴. In cotton, *CRK29 (Gh_D10G017200)* was expressed in 'epidermal cell' (cluster 3) as well as the epidermis of early globular and torpedo embryos (Supplementary Fig. 13d). ISH revealed that *Gh_A13G249900*, a homolog of *histidine-containing phosphotransmitter 3 (AHP3)*, was expressed in the cortex of globular embryos, and distributed in the cortex of both torpedo and cotyledonary embryos (Supplementary Fig. 14g-l). In *A. thaliana*, *AHP3* encodes a histidine phosphate transfer protein involved in the cytokinin signaling pathway⁷⁵. ISH revealed that *LBD12* was primarily expressed in the cortex of torpedo embryos (Supplementary Fig. 14m-r), which further verifies the reliability of our cell cluster classification.

AATP1 and *DOX2* negatively regulate cotton somatic embryo development

Utilizing the ST dataset, we found that *AAA-atpase1 (AATP1)* was expressed in 'globular embryo_10' (cluster 10), while *DOX2* was in 'pro-embryogenic cells' (cluster 6) (Fig. 2f). Both clusters are key cell types associated with somatic embryo regeneration in cotton⁷⁶. *Gh_D09G025900* is a homolog of *AATP1*, which plays a key role in inducing signaling to increase systemic resistance to pathogen pressure in potato (*Solanum tuberosum*) tubers⁷⁷. In tomatoes, *DOX2* encodes an α -dioxygenase that participates in α -linoleic acid metabolism and plays a role in protecting tissues from oxidative damage and apoptosis⁷⁸.

Because both *AATP1 (Gh_D09G025900)* and *DOX2 (Gh_A09G193100)* were found to be highly expressed during the early stages of cotton SE, we generated three overexpression (OE) and three CRISPR/Cas gene editing lines to evaluate the function of each gene (Fig. 7 and Supplementary Fig. 15). The expression levels of *AATP1* and *DOX2* in OE lines significantly increased compared with that in wild-type (WT) (Fig. 7a, b). Both *AATP1-CAS* and *DOX2-CAS* transgenic calli (Cs) proliferated faster and exhibited a larger volume at the end of the first 30 days of induction than WT Cs. In contrast, the proliferation rates and volumes of *AATP1-OE* and *DOX2-OE* Cs were lower than that of the WT control (Fig. 7c, d and Supplementary Fig. 15). By the end of the 90-day period, both *AATP1-CAS* and *DOX2-CAS* Cs were friable and could be easily recognized as embryogenic callus (EC). In contrast, *AATP1-OE* and *DOX2-OE* Cs proliferated more slowly during the initial stage of callus induction than the WT control and CAS Cs, and their volumes were also lower. Consistently, the differentiation rates of *AATP1-CAS* and *DOX2-CAS* Cs increased by the end of 90 days (Fig. 7e). However, no significant differences were observed in either volume or weight of OE Cs relative to WT control. After 90 days of calli induction, the *AATP1-OE* and *DOX2-OE* Cs were hard and exhibited no sign of EC development (Fig. 7c). Both *AATP1-CAS* and *DOX2-CAS* Cs exhibited higher induction and differentiation rates than WT control (Fig. 7e).

At day 30th, the cells of *AATP1-CAS* and *DOX2-CAS* Cs were primarily round or oval, indicating a transformation toward EC cells. In contrast, the cells of OE Cs appear elongated, indicating that these cells were in a non-embryogenic state (Supplementary Fig. 16). Taken together, these results suggest that both *AATP1* and *DOX2* negatively regulate callus initiation and EC formation.

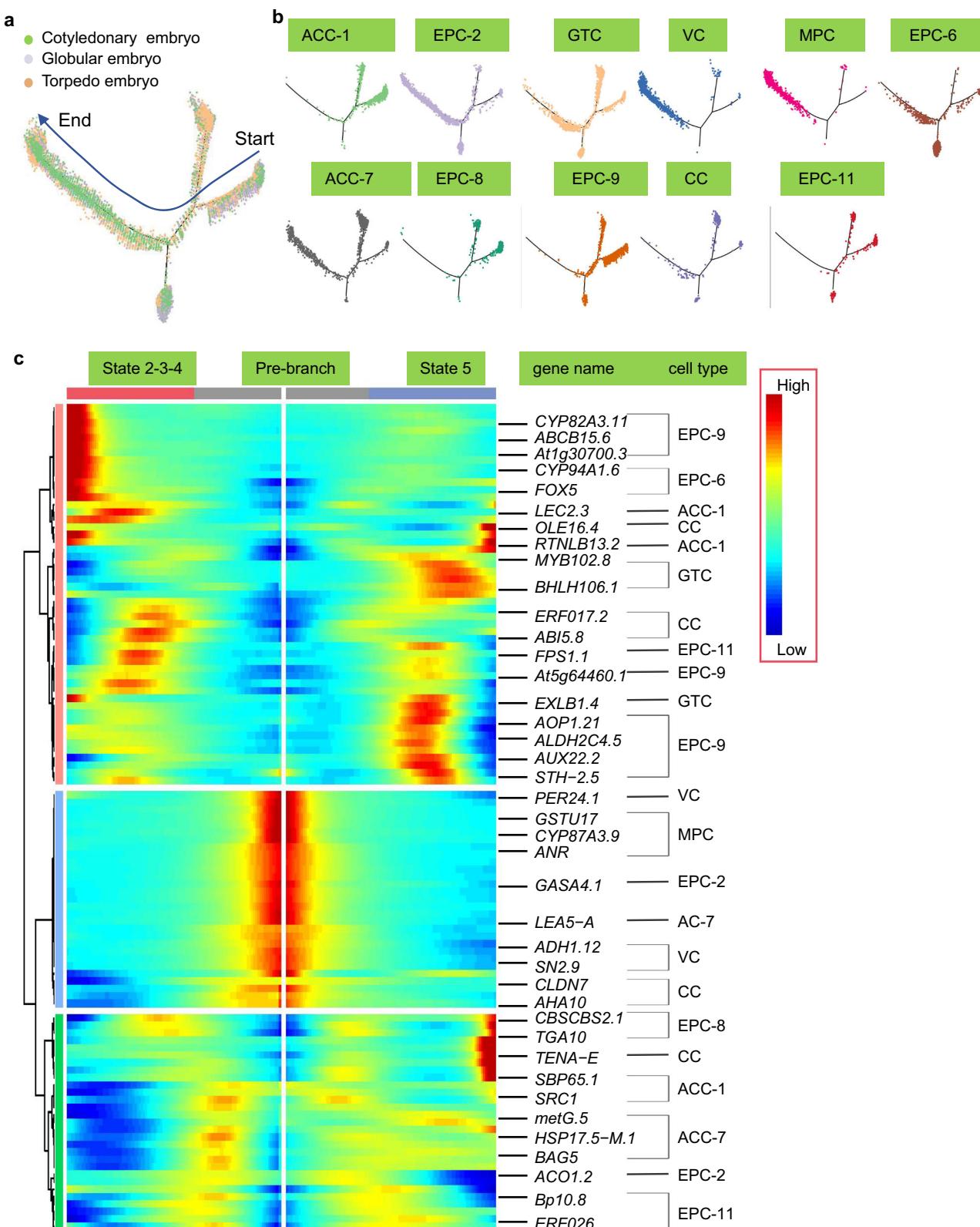


Fig. 5 | Pseudo-time analysis of cotton somatic embryos. **a** Distribution of cell clusters across the three stages of embryo development on pseudo-time trajectories. **b** Pseudo-time trajectories are colored according to cell clusters. **c** Expression dynamics and clustering of marker genes over pseudo-time. The

names of genes and cell types determining the developmental trajectories are shown on the right side of the heatmap. The color key ranging from blue (low) to red (high) indicates the expression levels.

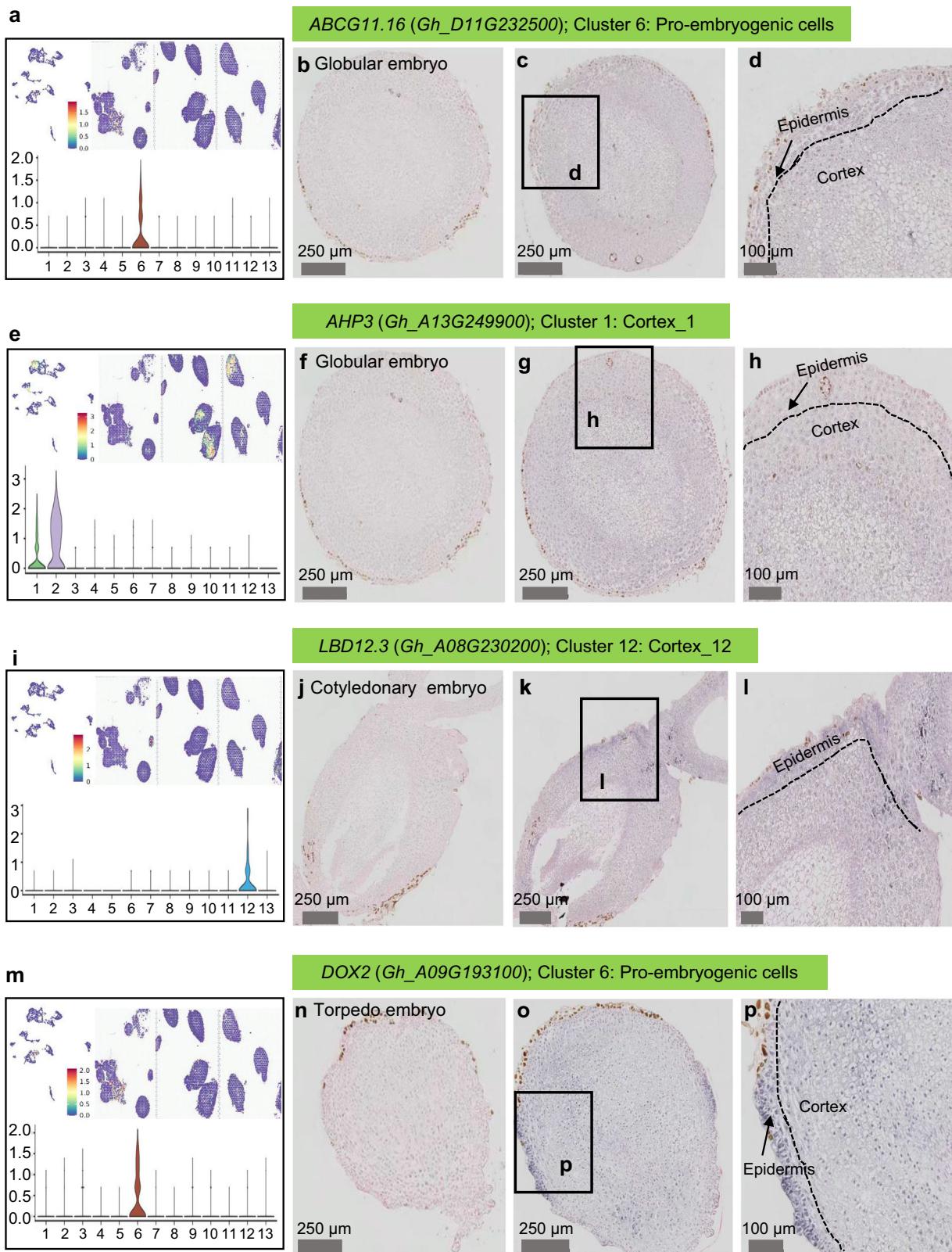
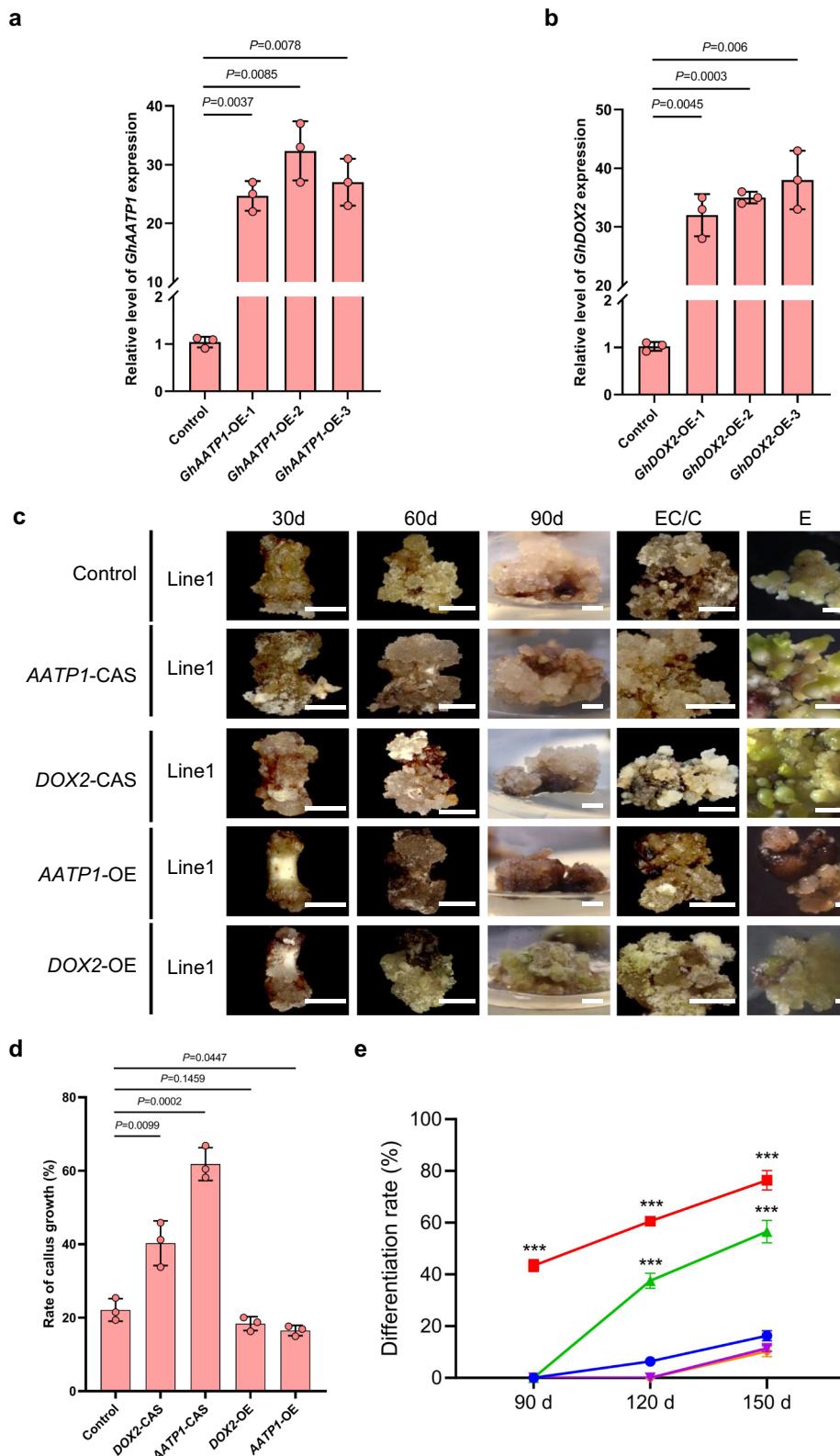


Fig. 6 | Spatiotemporal analysis of specific genes regulating SE in cotton. **a, e, i, m** Spatial location maps and violin plots of the expression of marker genes *ABCG11.16* (*Gh_D11G232500*), *AHP3* (*Gh_A13G249900*), *LBD12.3* (*Gh_A08G230200*), and *DOX2* (*Gh_A09G193100*) in different clusters. The color intensity indicates the relative transcript level of each marker gene. **b, f, j, n** Negative ISH controls. Data shown are representative images of three independent experiments with similar results. **c, g, k, o** ISH localization of marker genes in cotton somatic embryos. Data shown are representative images of three independent experiments with similar results. **d, h, l, p** Enlarged views of the rectangular sections indicated in **c, g, k, o**. Data shown are representative images of three independent experiments with similar results. Scale bars (black line) = 250 μ m (**b, c, f, g, j, k, n, o**) and 100 μ m (**d, h, l, p**).

results. **c, g, k, o** ISH localization of marker genes in cotton somatic embryos. Data shown are representative images of three independent experiments with similar results. **d, h, l, p** Enlarged views of the rectangular sections indicated in **c, g, k, o**. Data shown are representative images of three independent experiments with similar results. Scale bars (black line) = 250 μ m (**b, c, f, g, j, k, n, o**) and 100 μ m (**d, h, l, p**).



MSI analysis of metabolites during cotton somatic embryo development

We utilized a newly-developed air flow-assisted desorption electrospray ionization mass spectrometry imaging (FAADESI-MSI)⁷⁹ technique to examine dynamic metabolic changes during the development of cotton somatic embryos (Fig. 8 and Supplementary Fig. 17a). We detected the spatial distribution of metabolites by combining histological sections with the results of spatial shrunken centroids

clustering of adjacent tissues (Supplementary Fig. 17b–d). To identify marker metabolites associated with somatic embryo development, we compared the metabolite profiles of tissue sections collected from different developmental stages, i.e., (1) cortex vs. cotyledon, (2) cortex vs. vascular tissue, (3) cotyledon vs. vascular tissue, and (4) hypocotyl callus vs. hypocotyl. We compared the metabolite profiles of the different regions using unsupervised principal component analysis (PCA) (Supplementary Fig. 18a). To identify differentially

Fig. 7 | Functional analysis of representative marker genes for the development of callus tissue in cotton. **a, b** Relative gene expression of *GhAATP1* (a) and *GhDOX2* (b) in OE lines and Control materials detected by qRT-PCR. *GhHIS* was used as an internal reference gene. Data are means \pm SD from three independent repetitions ($n = 3$). Statistical analysis was assessed by unpaired two-tailed *t* test and significant differences were indicated with *p*-values at confidence intervals 0.95. **c** The developmental patterns of line 1 each of *AATP1*-CAS, *DOX2*-CAS, Control, *AATP1*-OE, and *DOX2*-OE materials after 30 d, 60 d, and 90 d of initial callus induction, and at the embryogenic callus/callus (EC/C) stage and embryo (E) stage. Data shown are representative images of three independent experiments with similar results. Other lines are shown in Supplementary Fig. 15. Scale bar = 5 mm. **d** Rate of callus growth of different materials 30 d after initial callus induction. The

rate of callus growth is calculated as the number of callus cells/hypocotyls. Data are means \pm SD from three independent repetitions ($n = 3$). Statistical analysis was assessed by unpaired two-tailed *t* test and significant differences were indicated with *p*-values at confidence intervals 0.95. **e** Line graph of differentiation rate of different materials 90 d, 120 d, and 150 d after initial callus induction. The differentiation rate is the rate at which callus tissue transforms into embryogenic callus and is calculated as the number of embryogenic callus cells/callus. Data are means \pm SD from three independent repetitions ($n = 3$). Statistical analysis was assessed by unpaired two-tailed *t* test and significant differences were indicated with *p*-values at confidence intervals 0.95. Source data are provided as a Source Data file.

produced metabolites (DPMs), we used orthogonal projections to latent structures-discriminant analysis (OPLS-DA) to distinguish metabolite profiles between groups (Supplementary Fig. 18b). The signal intensities of metabolites in different replicate samples were evaluated by statistic analysis (Supplementary Fig. 18c, d). Supplementary Figs. 19–27 shows the imaging of metabolites in some key metabolic pathways based on AFAADES-MSI imaging, including lipid metabolism ('linoleic acid metabolism'), carbohydrate metabolism ('citrate cycle (TCA cycle)'), amino acid metabolism ('arginine and proline metabolism', 'glycine, serine and threonine metabolism', 'alanine, aspartate and glutamate metabolism', 'valine, leucine and iso-leucine biosynthesis', and 'histidine metabolism'), and nucleotide metabolism ('purine metabolism', and 'pyrimidine metabolism'). It was worth noting that most metabolites showed highly heterogeneous spatial characteristics, for example, L-arginine was specifically expressed in the cotyledon region, and linoleic acid and 12(13)-EpOME were specifically expressed in the cortex region. The above results indicated that AFAADES-MSI technique could indeed perform imaging analysis of metabolites in cotton somatic embryos with high sensitivity, wide coverage and strong specificity.

The metabolic pathways associated with DPMs were determined according to the exact mass number, SmetDB database, and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis. Among these DPMs detected during the process of somatic embryo development (Supplementary Data 17 and 18), 74 metabolites were assigned to 26 plant development-associated metabolic pathways (Fig. 8b). For example, 'histidine metabolism' is necessary for embryonic development⁸⁰ and 'pyrimidine and purine metabolism' supports nucleotide biosynthesis during embryonic development⁸¹. 'Arginine and proline metabolism' is essential for embryogenesis and organogenesis because arginine is both a precursor for proline synthesis and also plays a role in the synthesis of polyamines^{82,83} (Supplementary Fig. 21), which are important for the transformation of cotton embryogenic callus into somatic embryos³⁵.

In the cortex vs. cotyledon (group 1) DPM comparison (Supplementary Fig. 28), several metabolites exhibited significantly different abundances in different tissues. For example, dihydrozeatin (C02029) was highly abundant in cotyledons, whereas linoleic acid (C01595), (S)- β -aminoisobutyrate (C03284), and D-malic acid (C00497) were highly abundant in the cortex (Supplementary Data 17). Of these, the abundance of linoleic acid (C01595) was particularly high in the cortex (Supplementary Fig. 28d).

In the cortex vs. vascular tissue (group 2) DPM comparison (Supplementary Fig. 29), many of the identified metabolites participate in several metabolic pathways. For example, pyroglutamic acid (C01879), succinic acid (C00042), and linoleic acid (C01595) were highly abundant in the cortex, whereas spermidine (C00315), and isoeugenol (C10469) were highly abundant in the vascular tissue.

In the cotyledon vs. vascular tissue (group 3) DPM comparison (Supplementary Fig. 30), N-carbamoylputrescine (C00436), L-arginine (C00062), and indole-3-acetaldehyde (C00637) were highly abundant in cotyledons (Supplementary Data 17). The high expression of related

differential metabolites in cotyledons indicates that these metabolites may play a role in cotyledon formation.

In the callus vs. hypocotyl (group 4) DPM comparison (Supplementary Fig. 31), dihydrozeatin (C02029), and dTDP (C00363) were highly abundant in hypocotyls. These metabolites are involved in 'zeatin biosynthesis' and 'purine metabolism'. L-aspartic acid (C00049), L-histidine (C00135), and galabiose (C00760) were highly abundant in callus tissues (Supplementary Fig. 31d, Supplementary Data 17). Both L-aspartic acid and L-histidine are utilized in 'histidine metabolism', which is essential for embryonic development⁸⁰.

Further, to verify the accuracy of the metabolites detected via SM, we analyzed the samples at the same development stages using liquid chromatography-tandem mass spectrometry (LC-MS)/MS also and conducted referential target metabolites analysis on some representative metabolites, further proving the metabolites identified in the SM analysis (Supplementary Fig. 32). The annotation of representative target analyte obtained from spatial metabolomics were further identified by MS/MS spectrum (Supplementary Figs. 33–62).

Gene-metabolite joint analysis during somatic embryo development

Finally, a joint analysis of SM and ST data was conducted. Based on the ST data, we produced a heatmap, spatial and UMAP visualization of the expression of genes involved in the metabolic pathways associated with the identified DPMs (Supplementary Fig. 63, Supplementary Data 19). In addition, quantitative strength analysis of correlation was performed on the key DPMs-related gene dataset obtained from the integrated analysis of ST and SM in Supplementary Data 19 (Supplementary Fig. 64). The majority of genes related to the synthesis of polyamines (e.g., spermine and spermidine), including *spermidine synthase 1 (SPDSYNI)* (*Gh_A03G164400*) and *spermine synthase (SPMS)* (*Gh_A08G106900*), were highly expressed in globular embryos, which is consistent with the higher abundance of these polyamines in these embryos (Supplementary Fig. 63a). A homologous gene of *putative 12-oxophytodienoate reductase 11 (OPR11)* (*Gh_Contig00833G000600*) was highly expressed in cotyledony embryos (Supplementary Data 19). The consistency between the ST and SM results suggests that spermine and spermidine may represent marker metabolites for globular embryo development, while jasmonic acid may be a marker metabolite for cotyledony embryo development.

KEGG analysis revealed that *DOX2* participates in 'alpha-linolenic acid metabolism' as α -dioxygenase and plays a role in the generation of 2(R)-HPOT from alpha-linolenic acid (Supplementary Fig. 65a), indicating that *DOX2* negatively regulates the synthesis of alpha-linolenic acid. Combined with the SM data, we found that the abundance of alpha-linolenic acid was the highest at the torpedo embryo stage (Supplementary Fig. 65b). Alpha-linolenic acid is a type of fatty acid (FA), and reduced FA biosynthesis is linked to embryonic developmental defects^{84,85}. These results suggest that *DOX2* may negatively regulate cotton SE by negatively regulating the synthesis of alpha-linolenic acid.

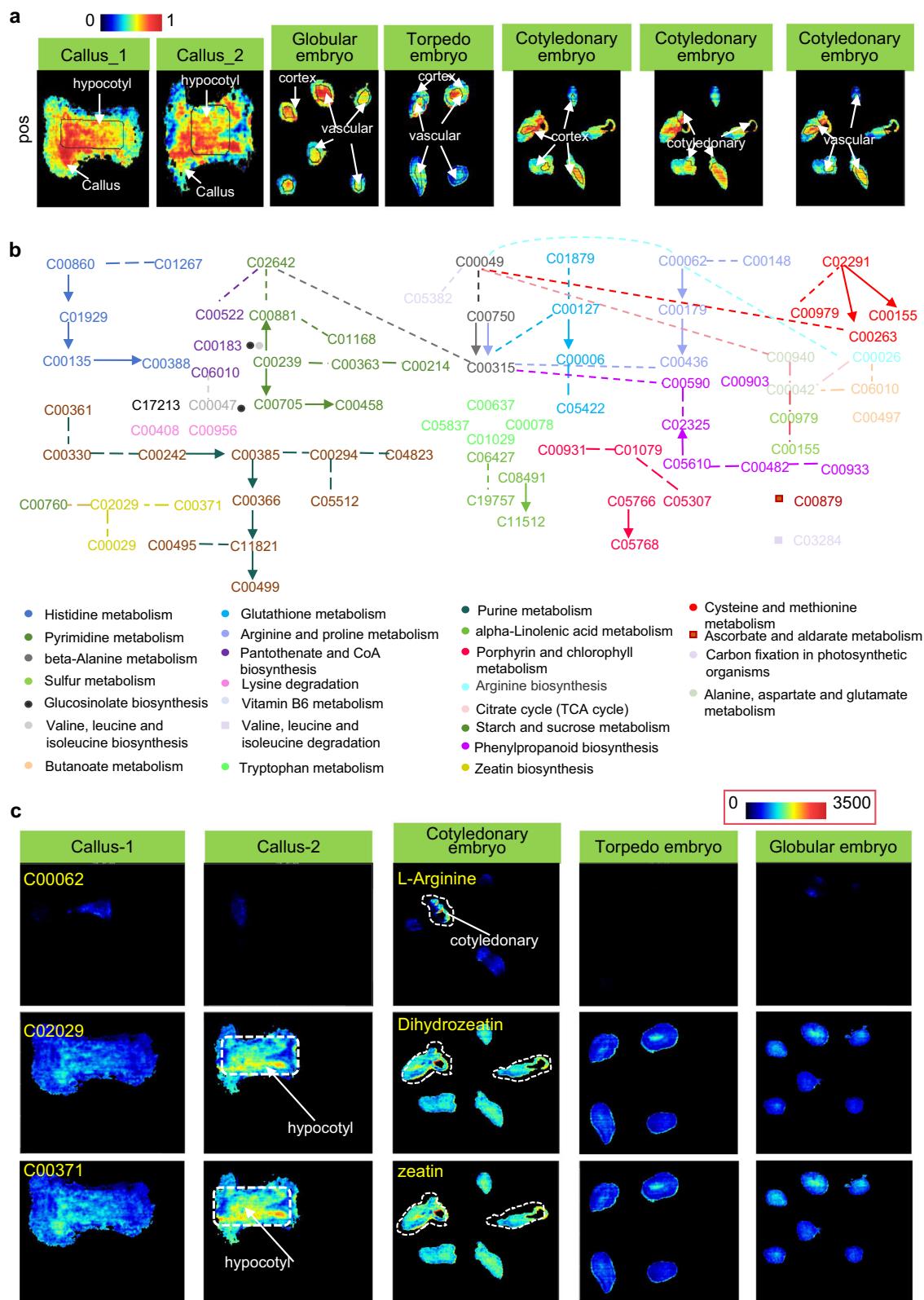


Fig. 8 | Spatial analysis of metabolites in developing cotton somatic embryos. **a** Some MSIs in positive ion scanning mode were selected to label tissue regions in samples at different developmental stages. These MSIs do not represent specific metabolites. **b** Metabolic network of cotton somatic embryos. The metabolic network was constructed based on the DPMs associated with different stages of

embryo development. The numbers represent KEGG annotations, with colors corresponding to the associated metabolic pathways. Solid lines represent one-step reactions and dotted line represent two- or more-step reactions. **c** MSI of L-arginine, dihydrozeatin, and zeatin in somatic embryos.

Discussion

One of the bottlenecks in cotton breeding is the low efficiency of genetic transformation. Primarily, the genetic transformation of cotton involves tissue culture technology. Unlike rice and other plants, cotton goes through a long developmental stage during tissue culture and regeneration. Besides, the callus in cotton does not directly differentiate into regenerative buds but through somatic embryo regeneration². Callus induction is possible in many cotton genotypes, but only a few successfully transform from callus to somatic embryos and eventually form regenerated plants. Several studies based on forward and reverse genetics have identified the regulatory genes associated with specific developmental stages and modes in cotton but have not analyzed the entire transcriptome and gene regulatory network^{11,86,87}. Therefore, we speculated that an in-depth and systematic study of the dynamic changes of transcriptome and metabolome at the single-cell level during the development of cotton somatic embryo based on an integrated spatial transcriptomic and metabolomic approach might help analyze and understand the fate of different cell types during the transformation from callus to somatic embryo. Based on the transcriptome data, we created a visual and easily searchable database that provides information on the expression patterns and dynamics of genes associated with cotton somatic embryo development. This information will help save time and cost, accelerate somatic embryo development, and provide an essential reference for promoting genetic transformation and breeding of cotton.

In this study, we investigated tissue sections of cotton somatic embryos at different developmental stages and found an array of dynamic changes across cell types and structures over the course of development (Supplementary Fig. 1). At the callus stage of cotton, cells are generally scattered, loose, and irregular, while the cellular arrangement becomes closer and tighter at the embryogenic callus stage^{88,89}. When the initial globular embryo is formed, the cells begin to redifferentiate and the epidermis of the somatic embryo begins to form and enclose the interior of the embryo, and the formation of the cortex and cambium also becomes apparent^{90,91}. During the torpedo embryo stage, the polarity of the embryo is gradually established, the shape of the vascular bundles becomes clearer, and the developmental patterns at the top and base of the vascular bundles become defined⁹². In the cotyledonary embryo, the vascular tissue is more developed and exhibits a 'V-shaped' appearance⁴¹. At this time, the embryo begins to develop cotyledons⁹.

Both ST and scRNA-seq have been utilized to study the process of mammalian cell regeneration in vitro, including gastrulation in mice⁹³. We utilized a combination of ST, scRNA-seq, and SM methodologies to explore transcriptomic and metabolic changes associated with specific tissues and cell types during the entire course of somatic embryo development (Fig. 1). We identified and annotated key cell types and marker genes at different developmental stages during SE. Based on the spatial distribution of representative marker genes in specific tissues and cell types, we characterized the specific cell types identified through scRNA-seq. This overcomes the shortcoming associated with ST at single-cell resolution, and allows for a comprehensive analysis of heterogeneity within the same cell type. The expression characteristics of certain representative marker genes were further verified by ISH (Fig. 6, Supplementary Fig. 14).

We identified key marker genes associated with each stage of embryonic development, including the early embryonic marker gene *WYKY23* (*Gh_D08G138200*) in callus⁹⁴, the procambium marker gene *ATHB-8* in globular embryos⁹⁵, the embryonic vascular tissue and radicle initiation marker gene *ARF5* in torpedo embryos, and the shoot apical meristem and cotyledon formation marker gene *HAT2* in cotyledonary embryos⁹⁶. Our ST data allowed us to visualize gene expression, identify cell types and developmental stages, discover tissue-specific DEGs, and predict the function of these DEGs (Figs. 2 and 3). For example, cluster 9 participates in 'shoot system development' and

cluster 3 is related to the regulation of 'response to stimulus' and 'defense response to bacterium'. These results are consistent with the biological functions associated with cells in these clusters, further supporting the credibility of our ST data. These results will aid our understanding of the roles of cell type-specific DEGs in regulating the development of specific tissues. In addition to the well-known *MP/ARFs* and *REV*, we also identified several marker genes in different cell clusters, such as *ABCG11* and *TMK1* in 'pro-embryogenic cell' (cluster 6), *ATHB-20* and *ATHB-13* in 'ground tissue cell' (section 7: cluster 1), and *WOX13* in 'non-embryogenic callus cell' (section 1: cluster 1) (Supplementary Data 13 and 14). These genes may play a role in the development of the embryo into the next stage; for example, *WOX13* may play a regulatory role in the development of callus into embryogenic callus. These newly identified marker genes can be used as references in future studies.

And in contrast with ST, scRNA-seq allows studies to be conducted in single cells at high resolution^{97,98}. We identified many similarities between the two datasets. For example, common cell types (e.g., epidermal, vascular, and cortex cells) were identified by both ST and scRNA-seq. In addition, we also used a multimodal intersection analysis to demonstrate the correspondence between ST and scRNA-seq clusters (Supplementary Fig. 66). The results show that 'Epidermal cell' (cluster 3) in ST corresponds to 'Apical cortex cells_1' (cluster 1), 'Epidermal cells_6' (cluster 6), 'Apical cortex cells_7' (cluster 7), 'Epidermal cells_9' (cluster 9) and 'Epidermal cells_11' (cluster 11) in scRNA-seq, while the names of these cell clusters in ST and scRNA-seq are similar (Supplementary Fig. 66). In addition, certain common genes, such as *LBD12*, which was found to be highly expressed in 'cortex cell_12' (cluster 12) by ST, was consistently found to be expressed in 'apical cortex cells_1', 'epidermal cells_9', and 'cortex cells' (clusters 1, 9, and 10) by scRNA-seq (Supplementary Fig. 13i, Supplementary Data 16). Thus, we combined the ST and scRNA-seq datasets to perform high-resolution spatial studies of different cell types in cotton somatic embryos.

Because many metabolites function as feedback regulators of gene expression^{99,100}, we utilized SM to analyze and identify DPMs in samples from different stages of SE (Fig. 8 and Supplementary Data 17). We found that both L-histidine and L-valine were highly enriched in callus tissues (Supplementary Data 17). A previous study suggests that an increasing content of total amino acids in embryonic tissues is related to the process of embryo formation¹⁰¹. Spermine and spermidine were highly enriched in globular embryos, and these polyamine metabolites are related to the differentiation of embryogenic calli into early embryos³⁵. Sinapine and sinapylalcohol were highly enriched in torpedo embryos (Supplementary Data 18), and these metabolites participate in phenylpropanoid biosynthesis and play an important protective role during plant development¹⁰².

N-carbamoylputrescine, indole-3-acetaldehyde, and dihydrozeatin were highly abundant in cotyledons (Supplementary Data 17). KEGG analysis revealed that these metabolites participate in several metabolic pathways, including 'tryptophan metabolism', and 'pyrimidine and purine metabolism' (Supplementary Figs. 28-31). 'Tryptophan metabolism' and 'zeatin biosynthesis' are auxin-related and cytokinin-related pathways, respectively^{103,104}, suggesting that the phytohormones cytokinin and auxin may play an important role in the induction of cotyledon embryos into seedlings. Additionally, 'pyrimidine and purine metabolism' is involved in chromatin synthesis¹⁰⁵, which may be required for cell division during the process of cotyledonary embryo development. The differential abundance of related metabolites in cotyledons suggests that these DPMs play a role in cotyledon formation and development (Supplementary Data 17). Both linoleic acid and pyroglutamic acid were highly abundant in the cortex (Supplementary Data 17). The unsaturated fatty acid linoleic acid is involved in 'linolenic acid metabolism', and is related to stress defense in plants¹⁰⁶, including in cortex tissues¹⁰⁷. Through *in situ* imaging of

metabolites in embryonic tissues, we identified many key DPMs which may regulate the development of somatic embryos. These results bring a metabolomic perspective to the study of somatic embryo regeneration in cotton.

We conducted a joint network analysis of ST and SM data, and the results show consistency (Supplementary Figs. 63–64, Supplementary Data 19). In particular, among the studied metabolites, polyamines have been shown to play key roles in the formation of somatic embryos³⁵. According to both ST and SM data, it appears that DOX2 and AATP1 may be involved in the regulation of the initiation of SE in cotton. Both DOX2 and alpha-linolenic acid were associated with normal embryonic development (Supplementary Fig. 65)^{8,48,35}. Functional analysis confirmed the potential roles of AATP1 and DOX2 in the regulation of callus induction and differentiation and the formation of embryos (Fig. 7).

Finally, based on our ST and scRNA-seq data, we constructed a website (<https://cotton.cricaas.com.cn/somaticembryo/>) on the spatial characteristics of gene expression during SE in cotton (Supplementary Fig. 67). This website will facilitate the application of transcriptome data obtained from this work in future studies of callus and somatic embryo development. Overall, our multi-omics approach, which combined ST, scRNA-seq, and SM analyses, proved to be highly suitable for the elucidation of SE in cotton. We identified key regulatory factors by identifying DEGs and DPMs during different developmental stages, at different spatial organizations, and in different cell types (Supplementary Fig. 68). Our multi-omics approach presents an efficient strategy for solving the problems associated with in vitro cotton regeneration and will enable us to further study the process of plant development.

Methods

Sample preparation of 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)”.

To perform ST, sections were collected from 2-3 embryos each at the globular, torpedo, and cotyledonary embryonic stages. Tissue samples were prepared and processed based on the User Guide of Visium Spatial Gene Expression. Briefly, fresh samples of globular embryos, torpedo embryos, cotyledonary embryos, embryogenic callus tissue, and non-embryogenic callus tissue were directly embedded in optimum cutting temperature (OCT) compound (OE Biotech, Shanghai, Item No. 4583) in a histology plastic cassette, and cut into 10 µm serial frozen sections at -20 °C on a cryostat microtome (Leica, CM 1950). Subsequently, RNA was extracted and assessed for quality (Supplementary Data 20). The eligible sections with RNA integrity number (RIN) greater than 7 was 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× Genomics Visium array slides were fixed, stained, and permeabilized for 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 were 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. 24 minutes was chosen as the optimal permeabilization time, because the fluorescence signal is the strongest and the signal diffusion is the weakest (Supplementary Fig. 69).

Tissue permeabilization, cDNA synthesis, library construction and sequencing

Fixed and stained tissue sections were permeabilized for 24 minutes 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. 70).

Quality control of sequencing data and gene quantification

Sequencing of high-throughput data yielded reads in fastq format. The 10× Genomics Space Ranger software (version 1.2.0) was used to process the Visium ST sequencing data and bright field microscopy images, detect the captured tissue area on the chip, and distinguish reads based on spatial barcode information. Quality control metrics, including the total number of spots, the number of reads per spot, the number of detected genes, and the number of UMIs were calculated (Supplementary Fig. 71). STAR software (version 2.7.1a) was used to integrate the cotton reference genome for comparative analysis and to create a gene-spot matrix for gene expression analysis.

Quantitative quality control and data preprocessing

Based on the preliminary quality control results obtained from the Space Ranger software (version 1.2.0), additional quality control and processing was carried out using the Seurat4 software package. Since the sequencing data of all samples were of good quality (Supplementary Fig. 2) 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 spot. Seurat’s sctransform function was employed to normalize data, construct a regularized negative binomial gene expression model, detect high variance features, and store data in SCT for further analysis. Highly variable genes were selected using the FindVariableGenes function (mean.function = ExpMean, dispersion.function = LogVMR) in Seurat4.

Dimensionality reduction and clustering analysis

To reduce the batch effects inherent to data, mutual nearest neighbors (MNN)¹⁰⁸ approach was performed with the R package (version 1.3.4) batchelor. Graph-based clustering was employed to group cells according to their gene expression profiles using Seurat’s FindClusters function. Seurat’s “RunUMAP” function was employed to visualize cells using a UMAP algorithm.

Differential gene expression and enrichment analysis

Differentially expressed genes (DEGs) were identified using the “FindMarkers” function of Seurat4 package. Statistical significance was assessed according to the following thresholds: $p\text{-value} < 0.05$ and $|\log_2\text{foldchange}| > 0.58$. Then, GO enrichment analyses of DEGs were performed with the annotated gene name of DEGs (<https://pantherdb.org/>). The reference species is *G. hirsutum*.

Preparation of single-cell suspensions of cotton somatic embryos

We used an improved reagent formula and treatment protocol for the dissociation of cotton protoplasts. Specifically, the cracking fluid was reformulated as 20 mM MES (pH 5.7), 3% (wt/vol) cellulase R10, 1% (wt/vol)

vol) Macerozyme R10, 1% (wt/vol) snailase, 1% (wt/vol) cutinase, 0.4 M mannitol, and 20 mM KCl. The treatment conditions were also improved, including the use of dark conditions, a temperature of 28 °C, and 20 RPM horizontal shaking. Each sample was soaked in the enzymolysis solution and then vacuum infiltrated for 15 mins to ensure thorough infiltration. The dissociated cells were filtered with a 40 µm cell strainer to remove large tissue blocks.

Construction of cDNA library

The RIN of single-cell suspensions of cotton somatic embryos was evaluated (Supplementary Data 20), RIN > 7 qualified samples can be used for single-cell transcriptome sequencing. The scRNA-seq library was prepared using a Chromium Single Cell 3' Gel Beads-in-Emulsion (GEM) Library & Gel Bead Kit v3, according to the manufacturer's instructions (10× Genomics, California, USA). During the library preparation, the cDNA lengths we obtained were distributed within the standard range (Supplementary Fig. 72).

ScRNA-seq data preprocessing

The Cell Ranger pipeline (version 5.0.0) provided by 10× Genomics was used to demultiplex cellular barcodes and map reads to the cotton reference genome. The STAR aligner was used to quantify transcripts. The UMI count matrix was produced with the R package Seurat4. To filter out low-quality cells and multiple captures, cells with UMIs or gene numbers outside the limit of the mean value ± 2 standard deviations (SD) were removed. For this process, a Gaussian distribution of UMIs and gene numbers was assumed. Subsequently, potential doublets were identified using the DoubletFinder package¹⁰⁹ (version 2.0.2). Finally, normalized counts were obtained with Seurat's NormalizeData function.

Dimensionality reduction and clustering analysis

Highly variable genes were selected using the FindVariableGenes function (mean.function = ExpMean, dispersion.function = LogVMR) in Seurat4. Utilizing MNN approach, any batch effects inherent in the scRNA-seq data were corrected with the R package batchelor¹⁰⁸. Both PCA and UMAP algorithms were utilized to reduce the dimensionality of the single cell transcriptome. Based on the dimensionality reduction results of PCA, UMAP was used to visualize cell clusters according to their gene expression profiles with the "FindClusters" and "RunUMAP" functions in Seurat4.

Differential gene expression and enrichment analysis

DEGs were identified using the "FindMarkers" function (test.use = MAST) in Seurat4. Statistical significance was assessed according to the following thresholds: *p*-value < 0.05 and |log₂ foldchange| > 0.58. Both GO and KEGG pathway enrichment analyses were performed using R based on the hypergeometric distribution.

Pseudo-time analysis

The development of pseudo-time was determined with the Monocle2 package (v2.9.0). The raw count was first converted from a Seurat4 object to a Cell Data Set object using the "importCDS" function in Monocle2. The "differentialGeneTest" function of the Monocle2 package was used to select ordering genes (*q*-val < 0.01). Ordering genes allow cells to be ordered along the pseudo-time trajectory. The dimensional reduction clustering analysis was performed using the "reduceDimension" function, followed by trajectory inference with the "orderCells" function, using default parameters. To evaluate changes over pseudo-time, gene expression was graphed using the "plot_genes_in_pseudotime" function.

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 were summed for each sample. Second, "1" was added to the summation value to take log₁₀ (i.e., log₁₀ [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 ggpqr package, the correlation value and *p*-value were calculated by Pearson analysis.

Multimodal intersection analysis

Marker genes of scRNA-seq and ST were obtained by "FindAllmarkers". The parameters are set as follows (test.use = "t", logfc.threshold = 0.25, min.pct = 0.25, return.thresh = 1e-2/1e-5). Hypergeometric distribution analysis is realized by phyer function in R language. Use the below formula 1 to calculate the significance *p*-value.

$$P = 1 - \sum_{i=0}^{m-1} \frac{\binom{M}{i} \binom{N-M}{n-i}}{\binom{N}{n}} \quad (1)$$

Parameter i: Both cell markers of scRNA-seq and spot markers of ST; M: cell markers for scRNA-seq; N: scRNA-seq and ST express genes simultaneously; n: spot markers for ST.

Preparation of materials for in situ hybridization

First, cotton somatic embryo tissue samples were placed into a glass bottle containing 4–5 mL of freshly prepared 4% paraformaldehyde fixative (1× PBS 100 mL, 1 M NaOH 0.5 mL, 4 g paraformaldehyde, pH 7.0). The fixative was then replaced and the samples were incubated overnight (≤ 16 h) at 4 °C. Subsequently, the tissue samples were dehydrated in a graded series of ethanol and xylene. Finally, the samples were embedded in paraffin using premade forms and stored at 4 °C for further analysis.

In situ hybridization

The alkaline phosphatase (AP) (Boster, Wuhan) chromogenic method: the embedded samples were sliced to 5 µm-thick sections, baked at 60 °C for 45 mins, and submerged in xylene for deparaffinization. Fragments of mRNA were exposed by adding a freshly-diluted solution of 3% citric acid and pepsin to the sections and incubating at 37 °C for 10 mins. Each slide was washed three times with 0.5 M phosphate buffered solution (PBS) for 5 mins. Samples were pre-hybridized by adding a pre-hybrid solution to each section and incubating at 37 °C for 2–4 hours. Samples were subsequently hybridized by dropping a probe onto each section before incubating in a thermostat wet box overnight at 60 °C. After hybridization, each section was washed twice with 2 × Saline Sodium Citrate (SSC) for 5 mins at 37 °C, once with 0.5 × SSC for 15 mins, and once with 0.2 × SSC for 15 mins at 37 °C. The washed sections were incubated for 30 mins in a 37 °C wet box. The incubated sections were subsequently incubated again with anti-digoxin of biotinylated mice for 1 hour, followed by four 5 mins washes with 0.5 M PBS. The washed sections were incubated with SABC-AP for 40 mins in a 37 °C wet box, followed by four 5 mins washes with 0.5 M PBS. Finally, the BCIP/NBT chromogenic reaction was performed by 1:20 diluting BCIP/NBT (x 20) with 0.01 M TBS (pH=9.5) and adding the chromogenic solution to the specimens. The reaction was carried out for 20–30 mins at a temperature of between 30–37 °C, with blue-purple staining indicating a positive result. Primers or probes used in this study can be found in Supplementary Data 21.

Construction of the CRISPR/Cas9 gene editing vectors and overexpression vectors

The basic pCambia2300 vector was utilized for the construction of the OE vectors, using the *Xba*I and *Bam*H I cleavage sites. The amplified genes were driven by either the cauliflower mosaic virus (CaMV) 35S promoter or other promoters. Positive cells were selected with the *neomycin phosphotransferase II* gene (*NPTII*), which confers kanamycin resistance. The pCambia2300 vector was also utilized for the

construction of CRISPR/Cas9 gene editing vectors. NLS-Cas9-NLS was inserted to construct the pCambia2300-p35S-NLS-Cas9-NLS-tNOS (pCambia2300-Cas9) vector. pAtU6-sgRNA-grNA-U6t was amplified and inserted upstream of 35S-NLS-Cas9-NLS with the *Hind III* and *Sbf I* restriction endonucleases. All primers used in this study can be found in Supplementary Data 22. Vector maps are shown in Supplementary Fig. 73.

Sterile cotton seedling culture

First, cotton seeds Zhongmiansuo24 (ZM24) were delinted with sulfuric acid. After delinting, the cotton seeds were dried in a 40 °C oven. After 48 h, the delinted, dried cotton seeds were cleaned with 75% alcohol. Seeds were sterilized by soaking them in 3% H₂O₂ for 20 h, followed by three rinses with sterile water. Sterilized seeds were germinated in the sterile seedling medium for 5 days under the following conditions: 16 h light/8 h dark, 28 °C day/25 °C night, 65% relative humidity, and 200 μmol m⁻² s⁻¹ photo-synthetically active radiation (PAR) intensity.

Preparation of medium

Murashige and Skoog (MS) medium with B5 vitamins (Phytotech, USA) was used for all experiments. MS medium was used to produce callus induction medium, embryogenic callus induction medium, embryoid induction medium, and regeneration seedling induction medium.

Agrobacterium infection and co-culture

Etiolated 5-day-old seedlings were collected and the hypocotyls were cut into 6–8 0.8 cm sections on an ultra-clean table. The hypocotyl sections were soaked in an *Agrobacterium* solution containing 50 mg/L acetosyringone (AS) for 3–5 min, and subsequently placed in a dish with sterile filter paper. After drying for approximately 10 min on ultra-clean workbench, the cut hypocotyl segments were placed in co-culture medium containing 50 mg/L AS, and cultured in a light incubator for 48 h (dark, 23 °C, 65% relative humidity).

Screening of transgenic callus

After two days of co-culture, the hypocotyl segments were cultured in callus induction medium containing selected antibiotics to induce callus initiation and proliferation. The segments were sub-cultured once a month and cultivated 2–3 times until the emergence of enlarged callus tissue. The culture condition was as follows: 16 h light/8 h dark, 28 °C day/25 °C night, 65% relative humidity, and 200 μmol m⁻² s⁻¹ PAR intensity. Callus tissue exhibiting a relatively large size was considered a positive result.

Induction of embryogenic callus

Positive callus tissue was cultured in embryogenic callus induction medium, sub-cultured once a month, and cultivated 2–3 times until the emergence of granular, millet-sized, yellow-green embryogenic callus. The culture condition was as follows: 16 h light/8 h dark, 28 °C day/25 °C night, 65% relative humidity, and 200 μmol m⁻² s⁻¹ PAR intensity. Positive embryogenic callus was further propagated.

Induction of embryoids

The positive embryogenic callus was cultured in embryoid induction medium, sub-cultured once every 2 weeks, and cultivated 2–3 times until the emergence of embryoids. The culture condition was as follows: 16 h light/8 h dark, 28 °C day/25 °C night, 65% relative humidity, and 200 μmol m⁻² s⁻¹ PAR intensity.

Seedling regeneration

Developmentally normal somatic embryos were collected and cultured in a seedling induction medium, sub-cultured once every 2 weeks, and cultivated 2–3 times until regenerated seedlings appeared. The culture condition was as follows: 16 h light/8 h dark,

28 °C day/25 °C night, 65% relative humidity, and 200 μmol m⁻² s⁻¹ PAR intensity.

Seedling cultivation

Rooted regenerated seedlings were placed in a nutrient solution and transplanted into a greenhouse for cultivation under the following conditions: 25 °C, 14 h light/10 h dark, and 300 μmol m⁻² s⁻¹ PAR intensity. The seeds of the T0 generation were self-crossed and harvested.

Detection of overexpressing plants by qRT-PCR

Sterile *AATPI-OE* and *DOX2-OE* seedlings were cultured in solid MS medium. To determine gene expression levels, the hypocotyls of 5-day-old seedlings were collected for fluorescence quantitative real-time PCR (qRT-PCR) analysis with specific primers (Supplementary Data 22). 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.

Statistical analysis

$$\text{Rate of callus growth}(\%) = \frac{(\text{tissue weight on: the 30th day of callus induction} - \text{initial weight of in vitro explant})}{\text{initial weight of in vitro explant}} \quad (2)$$

Three biological replicates were performed in each period.

$$\text{Differentiation rate of embryonic callus}(\%) = \frac{\text{number of embryonic callus}}{\text{total number of all normally grown explants}} \quad (3)$$

The differentiation rate of embryonic callus was counted from the 90th day of callus induction, and the differentiation rate was counted every 30 days. A total of 150 callus were collected from 50 callus of each line, and 3 biological replicates were performed in each period (Supplementary Data 23). Statistical analysis was assessed by unpaired two-tailed *t* test and significant differences were indicated with *p*-values at confidence intervals 0.95.

Spatial metabolic sample preparation

The fresh globular embryos, torpedo embryos, cotyledonary embryos, embryogenic callus tissues, and non-embryogenic callus tissues were embedded directly in the Leica Cryo-Gel (Item No. 14020108926). After embedding, we preserved samples under –80 °C, later sliced them in 10 μm serial sagittal slices (*n* = 10) with the cryostat microtome (Leica CM 1950, Leica Microsystem, Germany), and thaw-mounted them onto the positive-charge desorption plate (Thermo Scientific, U.S.A.). Later, we preserved sections under –80 °C prior to subsequent analyses, desiccated then for a 1-h duration under –20 °C and later for a 2-h duration under ambient temperature prior to MSI analyses.

AFADESI-MSI analysis

The AFADESI-MSI analysis was performed using an AFADESI ion source and a Q-OT-qIT hybrid mass spectrometer (MS) (Orbitrap Fusion Lumos; Thermo Fisher Scientific, MA, USA). For positive ion scanning mode, acetonitrile: water (80:20, V/V) containing 0.1% formic acid was used as the spray solvent. For negative ion scanning mode, acetonitrile:water (80:20, V/V) was used. The analysis was carried out at an x-axis scanning speed of 0.2 mm/s, with a 0.1 mm distance between two lines, an x-axis scanning length of 10 mm, and a y-axis scanning length of 10 mm. MSI was performed between a mass to charge ratio (*m/z*) of 70 to 1000, with an accuracy of <5 ppm mass error and a resolution of 70,000. Dinitrogen (0.6 Mpa) was used as the inert gas, with a capillary temperature of 350 °C. An Xcalibur data acquisition and processing system was used to collect the data, after which the

platform parameters and data collection sequences were adjusted according to sample size, step spacing, and scanning speed.

AFADESI-MSI data preprocessing

The raw data (.raw format) were converted to .imzML files using imzMLConverter¹¹⁰. The Cardinal¹¹¹ software package, an open-source R package for data transformation, visualization, and analysis, was used for background deduction, peak alignment, peak screening, normalization, and smoothing. Peak detection and peak recognition were performed using the MSI data. A series of processing steps, including peak alignment and interpolation of missing values, were performed to obtain the final quantitative data. All MSI images were normalized using total ion count normalization (TIC) for each pixel¹¹². 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 analysis results of spatial metabolomics, we conducted LC-MS/MS analysis on globular embryos, torpedo embryos, cotyledonary embryos samples from the same batch used for SM. The metabolomics assay was performed by Shanghai Luming Biotechnology Co., Ltd. (Shanghai, China). Three biological replicates were utilized for each group of samples. Briefly, 50 mg samples were added to 1.5 mL Eppendorf tubes containing 600 µL of methanol-water solution (methanol:water = 7:3, v/v), with L-2-chlorophenylalanine (4 µg/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 µL) from each tube were filtered using 0.22 µm 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 µm, 2.1 × 100 mm) 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 µL. For MS, the mass range was from 100 to 1,200 *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 compounds detected by AFADESI-MSI were annotated using the pySM pipeline¹¹³ and an in-house SmetDB database (a database containing about 20000 endogenous metabolites for plants constructed by Shanghai Luming Biotechnology Co., Ltd., Shanghai, China). Briefly, unsupervised PCA was performed to determine the overall distribution among samples and the stability of the analytical process. OPLS-DA was used to identify differences in the metabolic profiles and DPMs between groups. The importance of each feature for group separation was determined by retrieving the Variable Importance in Projection (VIP) of each compound for each model. A two-tailed Student's *t* test was utilized to ascertain statistically

significant differences in DPMs between groups. Statistically significant DPMs were determined when VIP > 1.0 and *p*-value < 0.05. 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% product ion threshold. Compounds were identified using the Metlin and LuMet-plant (a mass spectral database containing approximately 1000 standard substances constructed by Shanghai Luming Biotechnology Co., Ltd.) databases, based on 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% in groups, by replacing the zero value by 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 (Supplementary Data 24). 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 analyte were demonstrated in Supplementary Figs. 32–62.

Standard solutions and calibration curves

In the pursuit of targeted metabolite analysis, we employed L-aspartic acid (B24126-100 mg, Shanghai Yuanye Bio-Technology Co., Ltd), and L-arginine (B21920-200 mg, Shanghai Yuanye Bio-Technology Co., Ltd) as the subjects of our study. The stock solutions of 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, 0.08 ng/mL. For quality control purpose, 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 µL of ice-cold methanol-water (7:3, V/V), which included Succinic acid-d4 and Aspartic acid (3-13 C) 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 µL of the supernatant was decanted into sample vials. This supernatant was subjected to drying under a nitrogen stream and subsequently re-dissolved in 300 µL 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 g) for 5 min. Finally, 150 µL 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.

Gene-metabolite joint analysis- multimodal intersection analysis

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 somatic embryos were selected to construct transcriptome-metabolome association networks. The addmodulescore function in Seurat4 package was scored to visualize the expression of genes related to metabolite synthesis in spatial organization.

Database website

The R package Shiny was used to construct an interactive website (<https://cotton.cricaas.com.cn/>) containing a database of single-cell transcriptomics data and ST data for cotton. The shinythemes package was used to set custom themes and design the user interface (UI). Several Seurat4 functions, including DimPlot, VlnPlot, FeaturePlot, and SpatialPlot, were used to graph and display data and provide output content to the UI. The website is divided into three primary modules (three tabs are displayed at the top of the webpage): cotton single-cell transcriptomic data (scRNA-seq data), and two modules containing cotton ST data (Spatial Transcriptomics data1 and Spatial Transcriptomics data2). Each label can be clicked to enter pages containing detailed results displays. The function parameter area is located on the left side of the webpage, and the result display area is located on the right side. Above the result display area is a cluster display plots/gene ViolinPlot/FeaturePlot. To adjust the display result, one can select the point size, dimensionality reduction method (t-SNE/UMAP), group (cluster or sample), and the specified gene. A list of the top 10 identified marker genes 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 have been deposited in NCBI under BioProject [PRJNA800600](#). Spatial metabolomics were deposited in [METASPACE](#). LC-MS metabolomics data and

targeted metabolites data are available in Open Archive for Miscellaneous Data (OMIX) of National Genomics Data Center (NGDC) under accession [OMIX04900](#) and [OMIX005016](#).

Code availability

Customized code can be found at Github [<https://github.com/superly917/Cell-Atlas-of-Developing-Somatic-Embryos-in-Cotton>].

References

- Cui, Y. et al. Rapid mining of candidate genes for verticillium wilt resistance in cotton based on BSA-seq analysis. *Front. Plant Sci.* **12**, 703011 (2021).
- Kumar, S. et al. Comparative transcriptomics of non-embryogenic and embryogenic callus in semi-recalcitrant and non-recalcitrant upland cotton lines. *Plants-Basel* **10**, 1775 (2021).
- Elhiti et al. Molecular regulation of plant somatic embryogenesis. *VITRO CELL DEV-PL* **49**, 631–642 (2013).
- Salaün, C., Lepiniec, L. & Dubreucq, B. Genetic and molecular control of somatic embryogenesis. *Plants (Basel, Switz.)* **10**, 1467 (2021).
- Wu, X., Liu, C., Zhang, C. & Li, F. Progress of somatic embryogenesis in cotton. *Chinese Bulletin of Botany* **25**, 469–475 (2008).
- Souter, M. & Lindsey, K. Polarity and signaling in plant embryogenesis. *J. Exp. Bot.* **51**, 971–983 (2000).
- Hai-hong, S., Chuan-liang, L. I. U., Chao-jun, Z., Zhi-xia, W. U. & Fu-guang, L. I. Progress in mechanisms of cotton somatic embryogenesis. *Acta Botanica Boreali.-Occidentalia Sin.* **29**, 637–642 (2009).
- Hussain, S. S., Rao, A. Q., Husnain, T. & Riazuddin, S. Cotton somatic embryo morphology affects its conversion to plant. *Biol. Plant.* **53**, 307–311 (2009).
- Guo, H. et al. Dynamic transcriptome analysis reveals uncharacterized complex regulatory pathway underlying genotype-recalcitrant somatic embryogenesis transdifferentiation in cotton. *Genes* **11**, 519 (2020).
- Xiao, Y. et al. Effects of GhWUS from upland cotton (*Gossypium hirsutum* L.) on somatic embryogenesis and shoot regeneration. *Plant Sci.* **270**, 157–165 (2018).
- Jin, F. Y. et al. Comparative transcriptome analysis between somatic embryos (SEs) and zygotic embryos in cotton: evidence for stress response functions in SE development. *Plant Biotechnol. J.* **12**, 161–173 (2014).
- West, M. LEAFY COTYLEDON1 is an essential regulator of late embryogenesis and cotyledon identity in *Arabidopsis*. *Plant Cell* **6**, 1731–1745 (1994).
- Parcy, F., Valon, C., Kohara, A., Miséra, S. & Giraudat, J. The ABSCISIC ACID-INSENSITIVE3, FUSCA3, and LEAFY COTYLEDON1 loci act in concert to control multiple aspects of *Arabidopsis* seed development. *Plant Cell* **9**, 1265–1277 (1997).
- Zheng et al. AtWuschel promotes formation of the embryogenic callus in *Gossypium hirsutum*. *PLoS One* **9**, e87502 (2014).
- Nodine, M. & Bartel, D. P. MicroRNAs prevent precocious gene expression and enable pattern formation during plant embryogenesis. *Genes Dev.* **24**, 2678–2692 (2010).
- Ding, M. et al. Transcriptomic analysis reveals somatic embryogenesis-associated signaling pathways and gene expression regulation in maize (*Zea mays* L.). *Plant Mol. Biol.* **104**, 647–663 (2020).
- Pais, M. S. Somatic embryogenesis induction in woody species: the future after omics data assessment. *Front. Plant Sci.* **10**, 240 (2019).
- Magnani, E., Jiménez-Gómez, J. M., Soubigou-Taconnat, L., Lepiniec, L. & Fiume, E. Profiling the onset of somatic embryogenesis in *Arabidopsis*. *BMC Genomics* **18**, 998 (2017).

19. Ji, L. et al. Genome-wide reinforcement of DNA methylation occurs during somatic embryogenesis in Soybean. *Plant Cell* **31**, 2315–2331 (2019).
20. Li, C., Hu, F., Chen, H. & Zhao, J. Transcriptome characteristics during cell wall formation of endosperm cellularization and embryo differentiation in *Arabidopsis*. *Front. Plant Sci.* **13**, 998664 (2022).
21. Yi, F. et al. High temporal-resolution transcriptome landscape of early maize seed development. *Plant Cell* **31**, 974–992 (2019).
22. Liu, Z. et al. Global dynamic molecular profiling of stomatal lineage cell development by single-cell RNA sequencing. *Mol. Plant* **13**, 1178–1193 (2020).
23. 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).
24. Farmer, A., Thibivilliers, S., Ryu, K. H., Schiefelbein, 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).
25. Denyer, T. et al. Spatiotemporal developmental trajectories in the *Arabidopsis* root revealed using high-throughput single-cell RNA sequencing. *Dev. Cell* **48**, 840–852 (2019).
26. Xie, Y. et al. Single-cell RNA sequencing efficiently predicts transcription factor targets in plants. *Front. Plant Sci.* **11**, 603302 (2020).
27. Giacomello, S. A new era for plant science: spatial single-cell transcriptomics. *Curr. Opin. Plant Biol.* **60**, 102041 (2021).
28. Caprioli, R. M. Imaging mass spectrometry: molecular microscopy for enabling a new age of discovery. *Proteomics* **14**, 807–809 (2014).
29. 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).
30. 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).
31. Dong, Y. et al. High mass resolution, spatial metabolite mapping enhances the current plant gene and pathway discovery toolbox. *N. Phytol.* **228**, 1986–2002 (2020).
32. Barbosa, E. A., Junior, M. C. B. Jr, Rocha, T. L. & Engler, J. Imaging mass spectrometry of endogenous polypeptides and secondary metabolites from galls induced by root-knot nematodes in tomato roots. *Mol. Plant Microbe Interact.* **31**, 1048–1059 (2018).
33. Taira, S. et al. Mass spectrometric imaging of ginsenosides localization in panax ginseng root. *Am. J. Chin. Med.* **38**, 485–493 (2010).
34. Zhang, T. et al. Chemical imaging reveals diverse functions of tricarboxylic acid metabolites in root growth and development. *Nat. Commun.* **14**, 2567 (2023).
35. Cheng, W. H. et al. Polyamine and its metabolite H₂O₂ play a key role in the conversion of embryogenic callus into somatic embryos in upland cotton (*Gossypium hirsutum* L.). *Front. Plant Sci.* **6**, 1063 (2015).
36. Ha, S., Vankova, R., Yamaguchi-Shinozaki, K., Shinozaki, K. & Tran, L. S. Cytokinins: metabolism and function in plant adaptation to environmental stresses. *Trends Plant Sci.* **17**, 172–179 (2012).
37. Zhang, B. H., Wang, Q. L., Liu, F., Wang, K. B. & Frazier, T. P. Highly efficient plant regeneration through somatic embryogenesis in 20 elite commercial cotton (*Gossypium hirsutum* L.) cultivars. *Plant Omics* **2**, 259–268 (2009).
38. Tvorogova, V. E. et al. The WUSCHEL-related homeobox transcription factor MtWOX9-1 stimulates somatic embryogenesis in *Medicago truncatula*. *Plant Cell Tiss. Org.* **138**, 517–527 (2019).
39. Robles et al. Differential contributions of ribosomal protein genes to *Arabidopsis thaliana* leaf development. *Plant J.* **65**, 724–736 (2011).
40. Cheng, Y., Dai, X. & Zhao, Y. Auxin synthesized by the YUCCA flavin Monooxygenases is essential for embryogenesis and leaf formation in *Arabidopsis*. *Plant Cell* **19**, 2430–2439 (2007).
41. Alvarez-Venegas, R., Clelia, D. L. P. A. & Casas-Mollano, J. A. Epigenetic advances on somatic embryogenesis of agronomical and important crops. Springer International Publishing <https://doi.org/10.1007/978-3-319-07971-4>, 91–109 (2014).
42. Zhai, N. & Xu, L. Pluripotency acquisition in the middle cell layer of callus is required for organ regeneration. *Nat. Plants* **7**, 1453–1460 (2021).
43. Smetana, O. et al. High levels of auxin signalling define the stem-cell organizer of the vascular cambium. *Nature* **565**, 485–489 (2019).
44. Sakakibara, K. et al. WOX13-like genes are required for reprogramming of leaf and protoplast cells into stem cells in the moss *Physcomitrella patens*. *Dev* **141**, 1660–1670 (2014).
45. Baima, S. The *Arabidopsis* ATHB-8 HD-Zip protein acts as a differentiation-promoting transcription factor of the vascular meristems. *Plant Physiol.* **126**, 643–655 (2001).
46. Ji, J. et al. WOX4 promotes procambial development. *Plant Physiol.* **152**, 1346–1356 (2010).
47. Ding, A. et al. ERF4 and MYB52 transcription factors play antagonistic roles in regulating homogalacturonan demethylesterification in *Arabidopsis* seed coat mucilage. *Plant Cell* **33**, 381–403 (2021).
48. Gulzar, B. et al. Genes, proteins and other networks regulating somatic embryogenesis in plants. *J. Genet. Eng. Biotechnol.* **18**, 31 (2020).
49. Baima, S. et al. The expression of the Athb-8 homeobox gene is restricted to provascular cells in *Arabidopsis thaliana*. *Dev* **121**, 4171–4182 (1995).
50. Haecker, A. et al. Expression dynamics of WOX genes mark cell fate decisions during early embryonic patterning in *Arabidopsis thaliana*. *Dev* **131**, 657–668 (2004).
51. Wang, Y. et al. Transcriptomes analysis reveals novel insight into the molecular mechanisms of somatic embryogenesis in *Hevea brasiliensis*. *BMC Genom.* **22**, 183 (2021).
52. Möller, B. K. et al. Auxin response cell-autonomously controls ground tissue initiation in the early *Arabidopsis* embryo. *Proc. Natl Acad. Sci. USA* **114**, E2533–e2539 (2017).
53. El Ouakfaoui, S. et al. Control of somatic embryogenesis and embryo development by AP2 transcription factors. *Plant Mol. Biol.* **74**, 313–326 (2010).
54. Warghat, A. R., Thakur, K. & Sood, A. Plant stem cells: what we know and what is anticipated. *Mol. Biol. Rep.* **45**, 2897–2905 (2018).
55. Capron, A., Chatfield, S., Provart, N. & Berleth, T. Embryogenesis: pattern formation from a single cell. *Arabidopsis Book* **7**, e0126 (2011).
56. Scheres & Ben. Stem-cell niches: nursery rhymes across kingdoms. *Nat. Rev. Mol. Cell Biol.* **8**, 345–354 (2007).
57. Helariutta, Y. et al. The SHORT-ROOT gene controls radial patterning of the *Arabidopsis* root through radial signaling. *Cell* **101**, 555–567 (2000).
58. Prigge, M. J. et al. Class III homeodomain-leucine zipper gene family members have overlapping, antagonistic, and distinct roles in *Arabidopsis* development. *Plant Cell* **17**, 61–76 (2005).
59. Schlereth, A. et al. MONOPTEROS controls embryonic root initiation by regulating a mobile transcription factor. *Nature* **464**, 913–U128 (2010).
60. Itoh, J.-I., Hibara, K.-I., Sato, Y. & Nagato, Y. Developmental role and auxin responsiveness of class III homeodomain leucine zipper gene family members in rice. *Plant Physiol.* **147**, 1960–1975 (2008).
61. Wang, H. C., Ngwenyama, N., Liu, Y. D., Walker, J. C. & Zhang, S. Q. Stomatal development and patterning are regulated by

- environmentally responsive mitogen-activated protein kinases in *Arabidopsis*. *Plant Cell* **19**, 63–73 (2007).
62. Qi, S. et al. Global transcriptome and coexpression network analyses reveal new insights into somatic embryogenesis in *Hybrid Sweetgum* (*Liquidambar styraciflua* × *Liquidambar formosana*). *Front. Plant Sci.* **12**, 751866 (2021).
63. Chaban, C., Waller, F., Furuya, M. & Nick, P. Auxin responsiveness of a novel cytochrome p450 in rice coleoptiles. *Plant Physiol.* **133**, 2000–2009 (2003).
64. Li, L. et al. CaM/BAG5/Hsc70 signaling complex dynamically regulates leaf senescence. *Sci. Rep.* **6**, 31889 (2016).
65. Ma, W. et al. The LBD12-1 transcription factor suppresses apical meristem size by repressing argonaute 10 expression. *Plant Physiol.* **173**, 801–811 (2017).
66. Haghverdi, L., Büttner, M., Wolf, F. A., Buettner, F. & Theis, F. J. Diffusion pseudotime robustly reconstructs lineage branching. *Nat. Methods* **13**, 845–848 (2016).
67. Yanxia, Z., Jianping, J., Yanfen, H., Qingsong, D. & Kunhua, W. Comparative transcriptome analysis of the effects of friction and exogenous gibberellin on germination in *Abrus cantoniensis*. *Plant Signal. Behav.* **17**, 2149113 (2022).
68. Herzog, M., Dorne, A. M. & Grellet, F. GASA, a gibberellin-regulated gene family from *Arabidopsis thaliana* related to the tomato GAST1 gene. *Plant Mol. Biol.* **27**, 743–752 (1995).
69. Takubo, E. et al. Role of *Arabidopsis* INDOLE-3-ACETIC ACID CARBOXYL METHYLTRANSFERASE 1 in auxin metabolism. *Biochem. Biophys. Res. Co.* **527**, 1033–1038 (2020).
70. Cao, M. et al. TMK1-mediated auxin signalling regulates differential growth of the apical hook. *Nature* **568**, 240–243 (2019).
71. McFarlane, H. E., Shin, J. J. H., Bird, D. A. & Samuels, A. L. *Arabidopsis* ABCG transporters, which are required for export of diverse cuticular lipids, dimerize in different combinations. *Plant Cell* **22**, 3066–3075 (2010).
72. Ogawa, E. et al. ATML1 and PDF2 play a redundant and essential role in *Arabidopsis* embryo development. *Plant Cell Physiol.* **56**, 1183–1192 (2015).
73. Benková, E. et al. Local, efflux-dependent auxin gradients as a common module for plant organ formation. *Cell* **115**, 591–602 (2003).
74. Yadeta, K. A. et al. A Cysteine-Rich protein kinase associates with a membrane immune complex and the cysteine residues are required for cell death. *Plant Physiol.* **173**, 771–787 (2017).
75. Nishiyama, R. et al. *Arabidopsis* AHP2, AHP3, and AHP5 histidine phosphotransfer proteins function as redundant negative regulators of drought stress response. *Proc. Natl Acad. Sci. USA* **110**, 4840–4845 (2013).
76. Guillou, C. et al. Indirect somatic embryogenesis of *Theobroma cacao* L. in liquid medium and improvement of embryo-to-plantlet conversion rate. *Vitr. Cell. Dev. Biol. Plant* **54**, 377–391 (2018).
77. Conrath, U. et al. Enhanced resistance to *Phytophthora infestans* and *Alternaria solani* in leaves and tubers, respectively, of potato plants with decreased activity of the plastidic ATP/ADP transporter. *Planta* **217**, 75–83 (2003).
78. Tirajoh, A., Aung, T. S. T., McKay, A. B. & Plant, A. L. Stress-responsive alpha-dioxygenase expression in tomato roots. *J. Exp. Bot.* **56**, 713–723 (2005).
79. He, J. et al. A sensitive and wide coverage ambient mass spectrometry imaging method for functional metabolites based molecular histology. *Adv. Sci.* **5**, 1800250 (2018).
80. Stepansky, A. & Leustek, T. Histidine biosynthesis in plants. *Amino Acids* **30**, 127–142 (2006).
81. Deltour, R. Nuclear activation during early germination of the higher plant embryo. *J. Cell Sci.* **75**, 43–83 (1985).
82. Pathak, M. R., Teixeira da Silva, J. A. & Wani, S. H. Polyamines in response to abiotic stress tolerance through transgenic approaches. *GM Crops Food* **5**, 87–96 (2014).
83. Bagni, N. & Tassoni, A. Biosynthesis, oxidation and conjugation of aliphatic polyamines in higher plants. *Amino Acids* **20**, 301–317 (2001).
84. Bates, P. D. et al. Fatty acid synthesis is inhibited by inefficient utilization of unusual fatty acids for glycerolipid assembly. *Proc. Natl Acad. Sci. USA* **111**, 1204–1209 (2014).
85. Yin, Y. et al. Ultra-high α-linolenic acid accumulating developmental defective embryo was rescued by lysophosphatidic acid acyltransferase 2. *Plant J.* **103**, 2151–2167 (2020).
86. Sun, Y., Zhang, X., Chao, H., Guo, X. & Nie, Y. Somatic embryogenesis and plant regeneration from different wild diploid cotton (*Gossypium*) species. *Plant Cell Rep.* **25**, 289–296 (2006).
87. Jin, S. et al. Identification of a novel elite genotype for in vitro culture and genetic transformation of cotton. *Bio. Plant.* **50**, 519–524 (2006).
88. Zimmerman, J. L. Somatic embryogenesis: a model for early development in higher plants. *Plant Cell* **5**, 1411–1423 (1993).
89. Fehér, A. Callus, dedifferentiation, totipotency, somatic embryogenesis: what these terms mean in the era of molecular plant biology? *Front. Plant Sci.* **10**, 536 (2019).
90. Rathore, K. S., Campbell, L. M., Sherwood, S. & Nunes, E. Cotton (*Gossypium hirsutum* L.). *Methods Mol. Biol.* **1224**, 11–23 (2015).
91. Zhang, B. H. et al. Plant regeneration via somatic embryogenesis in cotton. *Plant Cell Tiss. Org.* **60**, 89–94 (2000).
92. Yang, X. Y. & Zhang, X. L. Regulation of somatic embryogenesis in higher plants. *Crit. Rev. Plant Sci.* **29**, 36–57 (2010).
93. Brink, S., Alemany, A., Batenburg, V. V., Moris, N. & Oudenaarden, A. V. Single-cell and spatial transcriptomics reveal somitogenesis in gastruloids. *Nature* **579**, 1–1 (2020).
94. Grunewald, W. et al. Tightly controlled WRKY23 expression mediates *Arabidopsis* embryo development. *EMBO Rep.* **14**, 1136–1142 (2013).
95. Gardiner, J., Donner, T. J. & Scarpella, E. Simultaneous activation of SHR and ATHB8 expression defines switch to preprocambial cell state in *Arabidopsis* leaf development. *Dev. Dyn.* **240**, 261–270 (2011).
96. Turchi, L. et al. *Arabidopsis* HD-Zip II transcription factors control apical embryo development and meristem function. *Dev* **140**, 2118–2129 (2013).
97. Crosetto, N., Bienko, M. & van Oudenaarden, A. Spatially resolved transcriptomics and beyond. *Nat. Rev. Genet.* **16**, 57–66 (2015).
98. Kolodziejczyk, A., Kim, J. K., Svensson, V., Marioni, J. & Teichmann, S. The technology and biology of single-cell RNA sequencing. *Mol. Cell* **58**, 610–620 (2015).
99. Geng, Z. et al. Recent advances in genes involved in secondary metabolite synthesis, hyphal development, energy metabolism and pathogenicity in *Fusarium graminearum* (teleomorph *Gibberella zeae*). *Biotechnol. Adv.* **32**, 390–402 (2014).
100. Bhamhani, S. et al. Transcriptome and metabolite analyses in *Azadirachta indica*: identification of genes involved in biosynthesis of bioactive triterpenoids. *Sci. Rep.* **7**, 5043 (2017).
101. Hiroshi, K. & Hiroshi, H. Changes in endogenous amino acid compositions during somatic embryogenesis in *Daucus carota* L. *Plant Cell Physiol.* **1**, 27–38 (1984).
102. Vanholme, R., De Meester, B., Ralph, J. & Boerjan, W. Lignin biosynthesis and its integration into metabolism. *Curr. Opin. Biotech.* **56**, 230–239 (2019).
103. Mano, Y. & Nemoto, K. The pathway of auxin biosynthesis in plants. *J. Exp. Bot.* **63**, 2853–2872 (2012).
104. Keshishian, E. A. & Rashotte, A. M. Plant cytokinin signalling. *Essays Biochem* **58**, 13–27 (2015).
105. Stasolla, C., Katahira, R., Thorpe, T. A. & Ashihara, H. Purine and pyrimidine nucleotide metabolism in higher plants. *J. Plant Physiol.* **160**, 1271–1295 (2003).

106. Abdelrahman, M. et al. Defective cytokinin signaling reprograms lipid and flavonoid gene-to-metabolite networks to mitigate high salinity in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **118**, e2105021118 (2021).
107. Javelle, M., Vernoud, V., Rogowsky, P. M. & Ingram, G. C. Epidermis: the formation and functions of a fundamental plant tissue. *N. Phytol.* **189**, 17–39 (2011).
108. 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).
109. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* **8**, 329–337.e324 (2019).
110. Race, A. M., Styles, I. B. & Bunch, J. Inclusive sharing of mass spectrometry imaging data requires a converter for all. *J. Proteom.* **75**, 5111–5112 (2012).
111. Bemis, K. D. et al. Cardinal: an R package for statistical analysis of mass spectrometry-based imaging experiments. *Bioinformatics* **31**, 2418–2420 (2015).
112. Wang, G. et al. Analyzing cell-type-specific dynamics of metabolism in kidney repair. *Nat. Metab.* **4**, 1109–1118 (2022).
113. Palmer, A. et al. FDR-controlled metabolite annotation for high-resolution imaging mass spectrometry. *Nat. Methods* **14**, 57–60 (2017).

Acknowledgements

This study was funded by support from the Innovation Program of Chinese Academy of Agricultural Sciences (CAAS-CSIAF-202402), National Natural Science Foundation of China (31621005), National Key Research and Development Program (2021YFF1000102 and 2022YFD1200300), National Natural Science Foundation of China (31670233), the Innovation Scientists and Technicians Troop Construction Projects of Henan Province, the Agricultural Science and the Technology Innovation Program of Chinese Academy of Agricultural Sciences.

Author contributions

Conceptualization, X.S. and F.L.; Methodology, X.Y., Z.L., J.Y. (Jiachen Yuan), A.Q., Y.W., and Y.C.; Formal Analysis, X.S., X.Y., Z.L., J.Y. (Jiachen Yuan) and A.Q.; Investigation, X.G., X.Y., Z.L., J.Y. (Jiachen Yuan), A.Q., W.Q., Y.W., Y.L. (Yao Lu), Y.C., Y.L., Y.Z. (Yaping Zhou), J.Y. (Jincheng Yang), Y.Z. (Yixin Zhang), X.L., P.W., Z.Z., M.H., S.S., and H.L.; Writing X.S., X.Y., and A.Q.; Writing – Review & Editing, X.S., A.Q., L.H.E., L.S.P.T., and F.L.; Supervision, X.S. and F.L.; Project Administration, X.S. and F.L.; Funding Acquisition, X.S. and F.L.

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/s41467-025-55870-6>.

Correspondence and requests for materials should be addressed to Xuwu Sun or Fuguang Li.

Peer review information *Nature Communications* thanks the anonymous reviewers 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