

**ORIGINAL RESEARCH**

# ScRNA-seq reveals the spatiotemporal distribution of camptothecin pathway and transposon activity in *Camptotheca acuminata* shoot apices and leaves

Shu Wang<sup>1</sup>  | Chuyi Zhang<sup>1</sup> | Ying Li<sup>1</sup> | Rucan Li<sup>1</sup> | Ke Du<sup>1</sup> |  
 Chao Sun<sup>1,2</sup> | Xiaofeng Shen<sup>1,2</sup> | Baolin Guo<sup>1,2</sup>

<sup>1</sup>Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

<sup>2</sup>State Key Laboratory for Quality Ensurance and Sustainable Use of Dao-di Herbs, Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing, China

**Correspondence**

Xiaofeng Shen  
 Email: [xfshen@implad.ac.cn](mailto:xfshen@implad.ac.cn)

Baolin Guo  
 Email: [blguo@implad.ac.cn](mailto:blguo@implad.ac.cn)

**Funding information**

the Chinese Academy of Medical Sciences Innovation Fund for Medical Sciences, Grant/Award Number: grant no. 2021-I2M-1-032

Edited by Y. Jiao

## Abstract

*Camptotheca acuminata* Decne., a significant natural source of the anticancer drug camptothecin (CPT), synthesizes CPT through the monoterpene indole alkaloid (MIA) pathway. In this study, we used single-cell RNA sequencing (scRNA-seq) to generate datasets encompassing over 60,000 cells from *C. acuminata* shoot apices and leaves. After cell clustering and annotation, we identified five major cell types in shoot apices and four in leaves. Analysis of MIA pathway gene expression revealed that most of them exhibited heightened expression in proliferating cells (PCs) and vascular cells (VCs). In contrast to MIA biosynthesis in *Catharanthus roseus*, CPT biosynthesis in *C. acuminata* did not exhibit multicellular compartmentalization. Some putative genes encoding enzymes and transcription factors (TFs) related to the biosynthesis of CPT and its derivatives were identified through co-expression analysis. These include 19 cytochrome P450 genes, 8 O-methyltransferase (OMT) genes, and 62 TFs. Additionally, these pathway genes exhibited dynamic expression patterns during VC and EC development. Furthermore, by integrating gene and transposable element (TE) expression data, we constructed novel single-cell transcriptome atlases for *C. acuminata*. This approach significantly facilitated the identification of rare cell types, including peripheral zone cells (PZs). Some TE families displayed cell type specific, tissue specific, or developmental stage-specific expression patterns, suggesting crucial roles for these TEs in cell differentiation and development. Overall, this study not only provides novel insights into CPT biosynthesis and spatial-temporal TE expression characteristics in *C. acuminata*, but also serves as a valuable resource for further comprehensive investigations into the development and physiology of this species.

## 1 | INTRODUCTION

*Camptotheca acuminata* Decne. is a major producer of the anticancer drug camptothecin (CPT) (Rayan et al., 2017), which has received widespread attention due to its remarkable antitumor effects through specific inhibition of type I DNA topoisomerase (G. L. Beretta

et al., 2013). In addition to the well-established clinical use of irinotecan and topotecan, an increasing number of CPT derivatives have been developed for the treatment of malignant tumours. Despite being classified as a quinolone alkaloid based on its chemical structure, CPT is synthesized through the monoterpene indole alkaloid (MIA) pathway (Figure S1). The MIA pathway emerges from the convergence of the indole pathway and the iridoid pathways. Playing a central role in CPT biosynthesis in *C. acuminata*, strictosidinic acid

Shu Wang and Chuyi Zhang contributed equally to this work.

synthase (STRAS) catalyzes the condensation of tryptamine and secologanic acid into strictosidinic acid. Tryptamine is synthesized through the indole biosynthetic pathway, whereas secologanic acid is generated via the iridoid biosynthetic pathway. In *C. acuminata*, at least five enzymes, namely, geraniol synthase (GES) (Fei Chen et al., 2016), 10-hydroxygeraniol oxidoreductase (10-HGO) (Alessio Valletta et al., 2010), iridodial synthase (IS) (Sadre et al., 2016), 7-deoxyloganic acid 7-hydroxylase (DL7H), and secologanic acid synthase (SLAS) (Yun Yang et al., 2019), have been characterized in the iridoid pathway. In the downstream pathway, strictosidinic acid is converted into CPT and its derivatives through a series of reactions, including oxidation, rearrangement, cyclization, deglycosylation, dehydrogenation, and hydroxylation processes. In *C. acuminata*, CYP71BE206 catalyzes the conversion of strictosamide to strictosamide epoxide (Xiang Pu et al., 2023). In addition, CPT 10-hydroxylase (CPT10H) and CPT 11-hydroxylase (CPT11H) can catalyze hydroxylation at C-10 and C-11, leading to the generation of 10-hydroxycamptotheclin and 11-hydroxycamptotheclin, respectively (Nguyen et al., 2021). Furthermore, 10-hydroxycamptotheclin can be further transformed into 10-methoxycamptotheclin by 10-hydroxycamptotheclin O-methyltransferase (Ca10OMT) (Salim et al., 2018). In addition, several transcription factors (TFs) involved in the MIA biosynthetic pathway have been discovered, such as OpMYB1 from MYB family, CrZCT1/2/3 from zinc finger protein family (ZNF), CrWRKY1 from WRKY family, CrBIS1/2 from basic helix-loop-helix (bHLH) family, and OpNAC1 from NAC (NAM, ATAF1/2, CUC1/2) family (Pauw et al., 2004; Suttipanta et al., 2011).

Single-cell RNA sequencing (scRNA-seq) has emerged as a promising approach for constructing transcriptome atlases at the single-cell level and has revolutionized the study of molecular and cellular biology (Chen et al., 2019). Previous studies have successfully identified rare cell types, traced cell lineages and determined cell fates. Remarkably, scRNA-seq has been used to improve the understanding of specialized metabolism in medicinal plants by illustrating the multicellular compartmentation of specialized metabolism and the discovery of key genes involved in metabolic pathways. For example, the spatial distribution of vinblastine biosynthesis pathway in *Catharanthus roseus* leaves was examined via scRNA-seq, and it was discovered that the pathway can be divided into three cell types: internal phloem-associated parenchyma (IPAP) cells, epidermal cells (ECs), and idioblast cells (ICs). In addition, a single-cell transcriptome atlas of tea leaves was constructed, and a novel catechin ester glycosyltransferase was identified using a gene co-expression network in mesophyll cells. (Wang et al., 2022). The rapid development of scRNA-seq has led to unprecedented opportunities for various applications in plants and led to great breakthroughs in life science (Kook Hui Ryu et al., 2021).

Transposable elements (TEs) are dispersed repetitive sequences in the genome capable of changing their position. They are prevalent in eukaryotes and constitute a substantial portion of the plant genome (R. Keith Slotkin and Robert Martienssen, 2007). Given their transcriptional activity, various tools have been used to quantify the expression of using bulk RNA-seq data (Sophie Lanciano and Gael Cristofari, 2020), leading to in-depth investigations into the

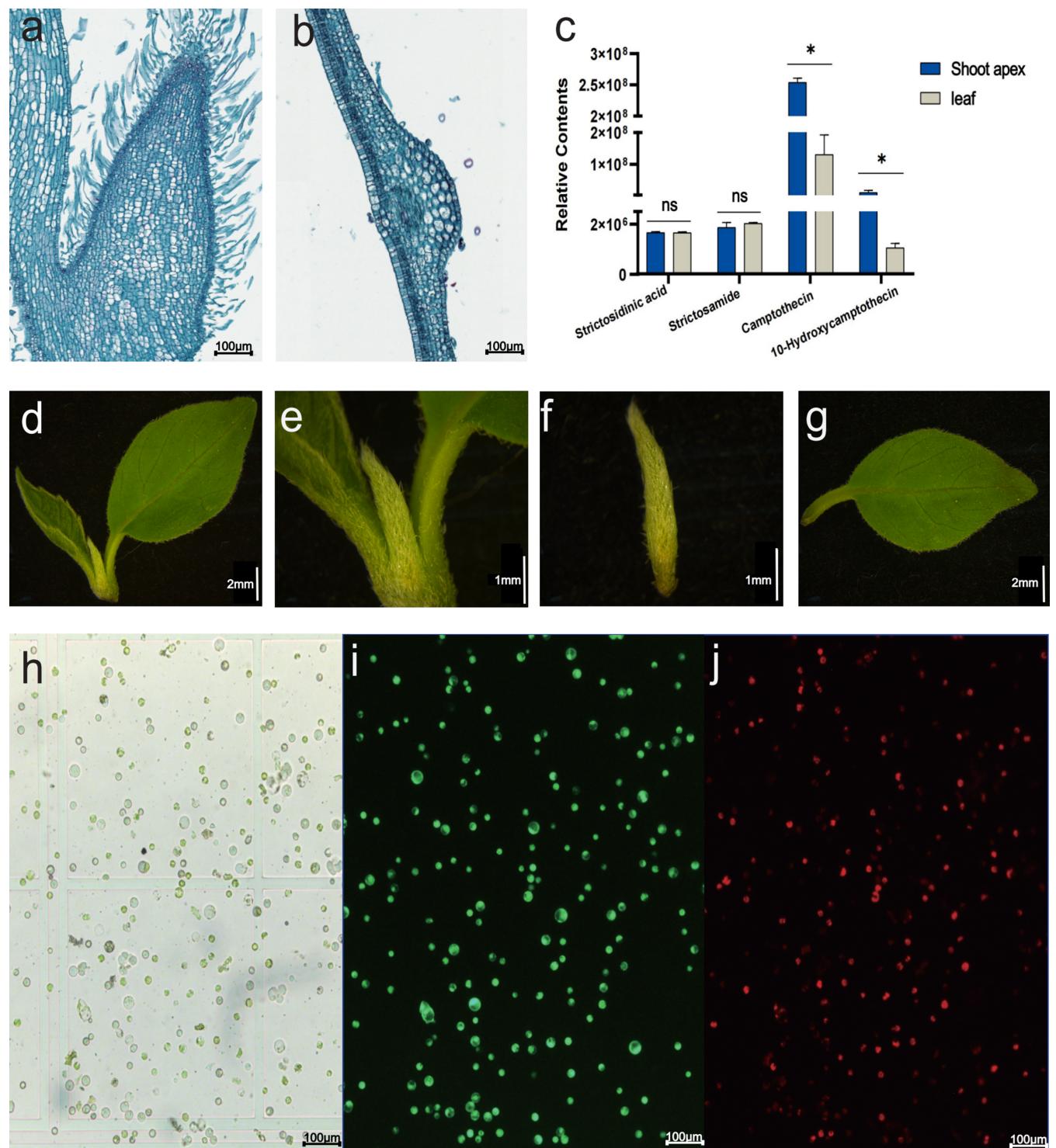
mechanisms underlying TE nonrandom transposition, spontaneous and nonspontaneous transcription, and their impact on gene expression (Guillaume Bourque et al., 2018). However, quantitative studies on TEs at single-cell resolution are rare. Recently, tools such as scTE (Jiangping He et al., 2021) and soloTE (Rocio Rodriguez-Quiroz and Braulio Valdebenito-Maturana, 2022) have been used to quantify the expression of all types of TEs in single-cell data, along with SCIFER (Emily C. Stow et al., 2022), a tool specifically designed for quantifying the expression of LINE/L1 elements. These methods have facilitated TE research at a single-cell level. Nonetheless, a notable dearth of detailed investigations that specifically focus on quantifying the expression of TEs at the single-cell level in plants results in an extreme scarcity of research on TE expression characteristics at the single-cell level.

In this study, we constructed high-quality single-cell transcriptome atlases of leaves and shoot apices from *C. acuminata* and investigated the spatiotemporal expression patterns of CPT biosynthesis-related genes at the single-cell level. The majority of genes involved in the CPT pathway exhibited a pronounced expression level in proliferating cells (PCs) and vascular cells (VCs). We reconstructed the developmental trajectory of ECs and VCs and explored the dynamic changes in expression patterns of CPT pathway genes during the development of VCs and ECs. Furthermore, we pioneered the construction of single-cell transcriptional maps based on gene and TE expression in plants and identified some TE with cell type specific, tissue specific, or developmental stage-specific expression patterns, suggesting crucial roles for these TEs in cell differentiation and development. The comprehensive dataset generated in this study serves as a valuable resource for investigating plant physiology and elucidating specialized metabolism, particularly CPT biosynthesis, at single-cell resolution.

## 2 | RESULTS

### 2.1 | Plant architecture observation and protoplasting

To enhance our comprehension of the structure and various cell types present in the shoot apices and leaves of *C. acuminata*, as well as to provide a valuable reference for cell type annotation in single-cell transcriptome analysis, we employed saffranine solid green staining. In *C. acuminata* leaves, anatomical and structural analyses revealed that the mesophyll comprises several layers of spongy parenchyma on the abaxial side and a single layer of elongated palisade parenchyma on the adaxial side. The leaf epidermis is composed of a single layer of thin-walled cells on both the adaxial and abaxial surfaces. The veins are vascular bundles that are arranged as rings in transverse sections and distributed in the cortex (Figure 1a). Compared with leaf tissues, shoot apices are mostly composed of parenchymal tissue and layers of epidermal cells with trichomes (Figure 1b). The angiosperm shoot apical meristem (SAM) consists of a small dome of cells with specific structural features, the central zone (CZ), the peripheral zone (PZ), and



**FIGURE 1** Anatomical features of *C. acuminata* leaves and isolation of protoplasts. Leaf cross-sections and shoot apex longitudinal sections stained with Safranin O and Fast Green (a-b). Comparison of the relative quantities of four CPT pathway metabolites in leaves and shoot apices (c). The shoot apices and leaves used to isolate the protoplasts (d-g). Protoplasts were subjected to staining with trypan blue (h, bright field) and fluorescein diacetate(FDA) (i, 488 nm) to discern their viability. Live cells cannot be stained by trypan blue and can emit green fluorescence after staining with the FDA. Protoplasts under bright light and UV excitation (360–380 nm). Chlorophyll emits red autofluorescence upon UV excitation (j, 368 nm).

the rib meristem (RM) (Bing Wang, et al., 2018). Metabolomic analyses were subsequently conducted on the shoot apices and leaves of *C. acuminata* (Figure 1d-g) to characterize and compare alkaloid

profiles between tissues. The analysis revealed that the camptothecin content in shoot apices was 1.9 times greater compared to that in leaves, and the 10-hydroxycamptothecin content in shoot apices was

found to be 9.7 times higher than in leaves. Four compounds related to CPT biosynthesis were identified, namely: strictosidinic acid, strictosamide, CPT, and 10-hydroxycamptotheclin. The quantities of the intermediates, strictosidinic acid and strictosamide, did not significantly differ between shoot apices and leaves (Figure 1c). However, the levels of the final products CPT and 10-hydroxycamptotheclin in the shoot apices were markedly higher than those in the leaves, which indicated that CPT and its derivatives preferred biosynthesis in young tissues during early development (Table S2).

To gain deeper insights into the spatiotemporal distribution of CPT biosynthesis, we generated single-cell transcriptome atlases of *C. acuminata* shoot apices and leaves. We systematically optimized the protoplast isolation protocols, resulting in a remarkable equilibrium between the yield and viability of the obtained protoplasts. Trypan blue and fluorescein diacetate staining were used to differentiate live and intact cells from diseased cells and cellular debris (Figure 1h–j). Ultimately, approximately 80,000 protoplasts with an average viability of 90.05% were used to generate the transcriptomic datasets.

## 2.2 | Single-cell transcriptome atlas of *C. acuminata* shoot apices and leaves

Four single-cell cDNA libraries were constructed, comprising two derived from leaf samples and two from shoot apex samples. These libraries were subjected to DNBSEQ-T7 high-throughput sequencing, generating a total of 360 GB of scRNA-seq data. Due to the occurrence of duplication events among plant specialized metabolic pathway genes and the limitations of the Cell Ranger pipeline in accurately quantifying genes with multiple copies, we employed Alevin to quantify the scRNA-seq datasets and identified 12,061 and 11,788 cells from leaf samples, characterized by median gene counts of 5,872 and 6,239 and median numbers of 18,758 and 21,237 unique molecular identifiers (UMIs) per cell, respectively. From the shoot apex samples, we identified 16,177 and 14,551 cells, with median gene counts of 5,247 and 5,927 and median numbers of 16,684 and 21,428 UMIs, respectively (Tables S3 and S4). Cells exhibiting abnormal gene numbers, UMI counts and high percentages of chloroplast genes (> 40%) were filtered out, leaving 11,177 and 11,788 high-quality cells in leaves, along with 14,143 and 15,349 in shoot apices for downstream analysis. Consequently, unsupervised clustering using Seurat generated a single-cell transcriptome atlas of leaves containing 14 distinct cell clusters and a map of shoot apices with 12 clusters.

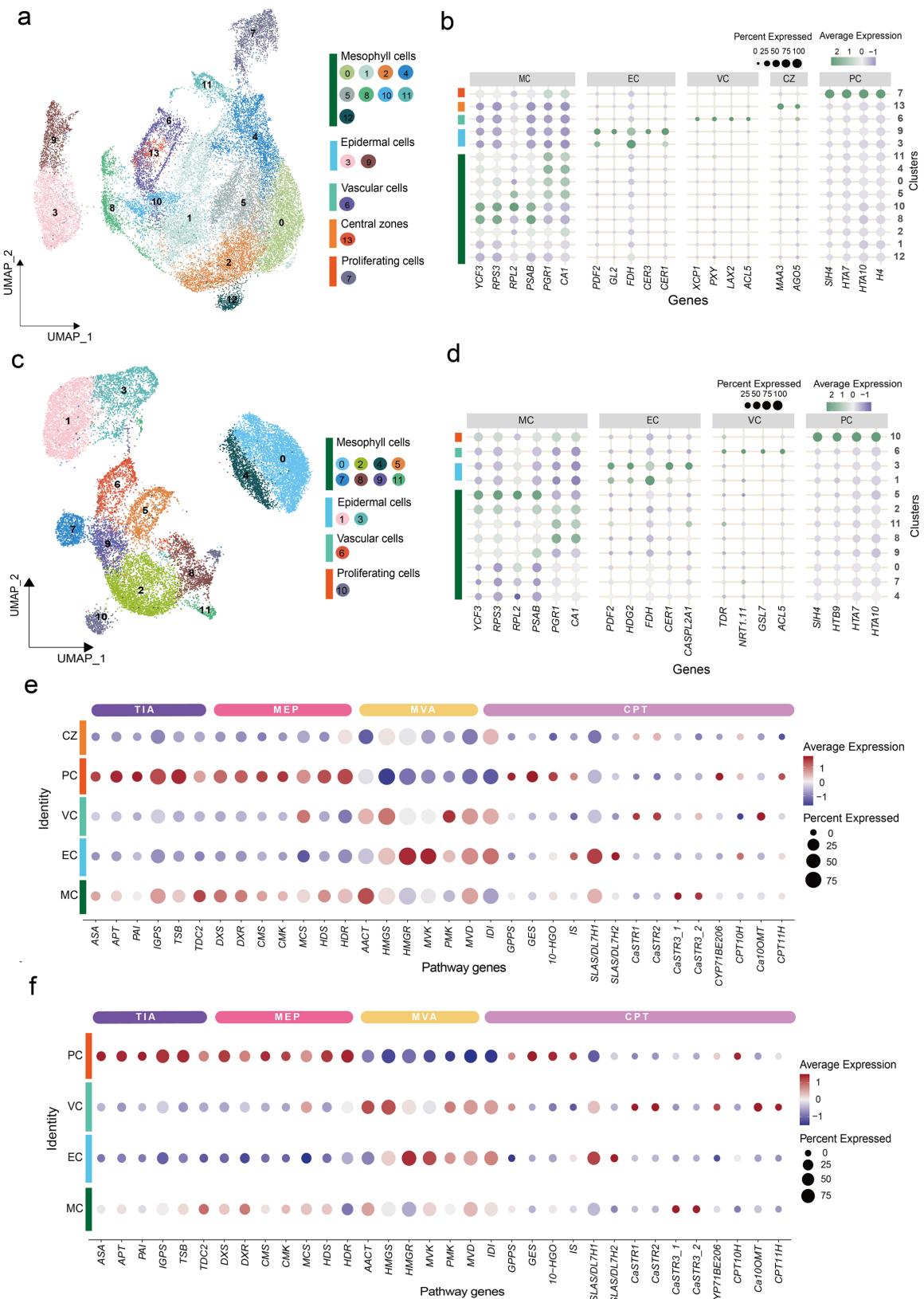
To precisely annotate cell clusters in *C. acuminata*, we applied the following two strategies: 1 choosing gene markers whose orthologs in other species have undergone spatial orientation research, such as RNA *in situ* hybridization; and 2 identifying a high correlation with the cluster in which cell identity has been identified by marker genes (Figure S2a, b and Table S5) (Sijie Sun et al., 2023). According to the above methods, we successfully annotated five broad populations of shoot apices: mesophyll cells (MCs), EC, PCs, VC, and CZs (Figure 2a, b and S2c). The MC population consisted of nine clusters (Clusters 0, 1, 2, 4, 5, 8, 10, 11, and 12). Some of these clusters (0, 2,

4, 5, 8, 10, and 11) expressed genes in the chloroplast genome, such as PHOTOSYSTEM I P700 Chlorophyll A APOPROTEIN A2 (PSAB) and PHOTOSYSTEM I ASSEMBLY PROTEIN YCF3 (YCF3) (S. Sato et al., 1999). The other clusters (Clusters 1 and 12) had a greater correlation with the groups that were identified as MCs. In Clusters 3 and 9, genes that have been experimentally shown to localize to the epidermis, such as PROTODERMAL FACTOR2 (PDF2) (Miyuki Nakamura et al., 2006) and FIDDLEHEAD (FDH) (Robert E. Pruitt et al., 2000), were detected; these genes were subsequently designated Clusters 3 and 9 as ECs. Cluster 6 was annotated as the VC population, in which several genes involved in xylem and phloem formation were highly expressed, such as PHLOEM INTERCALATED WITH XYLEM (PXY) (Kate Fisher and Simon Turner, 2007). We annotated Cluster 7 as a PC population because of the high proportion of cells expressing cell cycle-related genes, such as HISTONE H2A (HTA7) (K. S. Mysore et al., 2000) and HISTONE H4 (H4) (M. E. Chaboute et al., 1987). Cluster 13 was identified as the CZ population by the highly expressed genes, such as PROTEIN ARGONAUTE 5 (AGO5) (Tian-Qi Zhang et al., 2021). Cells identified in the leaf single-cell transcriptome atlas were also identified using the same strategies. The MC population consisted of eight clusters (Clusters 0, 2, 4, 5, 7, 8, 9, and 11). Clusters 1 and 3 were annotated as ECs. Clusters 6 and 10 were annotated as VC and PC, respectively (Figure 2c, d and Figure S2c,d).

## 2.3 | Cell type-specific distribution of MIA biosynthesis

To elucidate the spatiotemporal distribution of MIA biosynthesis within the shoot apices and leaves of *C. acuminata*, we subsequently analyzed the expression patterns of the characterized genes involved in the CPT biosynthetic pathway at single-cell resolution. IPP and its isomer, DMAPP, are fundamental constituents of terpenoid biosynthesis in higher plants. These building blocks are generated through the cytosolic mevalonate (MVA) pathway or the plastidic methylerythritol phosphate (MEP) pathway (Eva Vranová et al., 2013). Notably, in both the shoot apices and leaves of *C. acuminata*, nearly all the genes (Supplementary Table 6) associated with the MEP pathway exhibited the highest expression in the PCs, while *IDI* and the majority of the genes related to the MVA pathway showed the lowest expression in the PCs. This observation suggested that terpenoids were primarily synthesized via the MEP pathway in PCs. Furthermore, most genes involved in the indole pathway also exhibited high expression levels in the PCs of both tissues, indicating a strong demand for tryptophan and its derivatives in these cells.

Although the transcripts of characterized genes from the iridoid and downstream pathways enriched in distinct cell types within shoot apices or leaves, most of these genes exhibited analogous cell type-specific expression patterns in both tissues (Figure 2e and f). In leaves, GPPS, GES, 10-HGO, IS, and CPT10H were preferentially expressed in PCs, while STR1, STR2, CYP71BE206, 10OMT, and CPT11H showed a high expression preference in VC. Two copies of the STR3 gene were



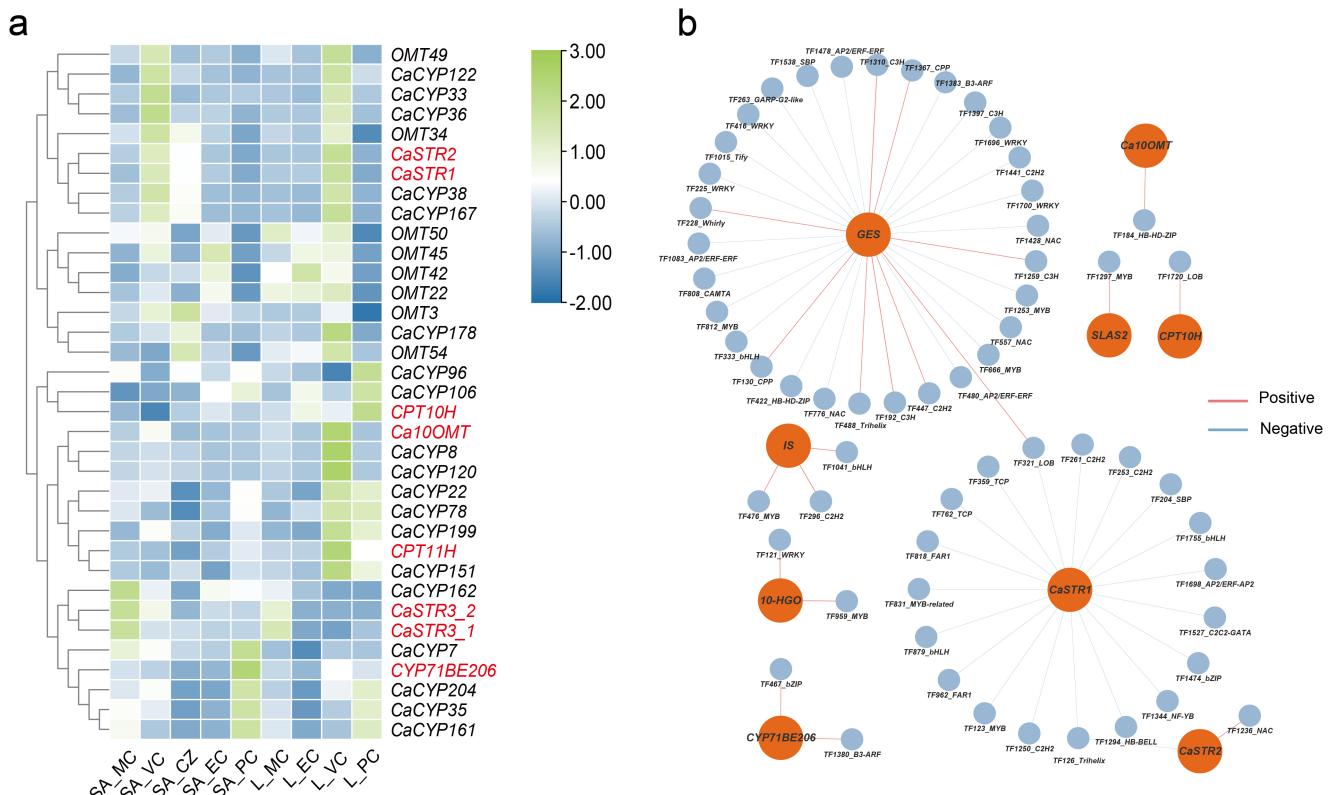
**FIGURE 2** Identification of cell clusters from shoot apices and leaves and the expression patterns of genes in the monoterpene indole alkaloid (MIA) biosynthetic pathways. Uniform manifold approximation and projections (UMAP) of shoot apex (a), and leaf cells (c), respectively. Each dot represents a cell, and the colours indicate the different clusters. The expression patterns of cell type-specific marker genes in shoot apices (b) and leaves (d), respectively. Detailed information about these marker genes is shown in Table S 6. The expression patterns of MIA pathway genes in shoot apices (e) and leaves (f), respectively.

highly expressed in MCs, and the *SLAS1* and *SLAS2* genes were highly expressed in ECs. In shoot apices, *CPT11H* transcripts were enriched in PCs, which exhibited a notably different expression pattern than that in leaves, while the expression patterns of the other genes remained similar. Taken together, our findings indicated that, except for *CPT11H*, all the other genes involved in CPT biosynthesis maintained their cell type-specific expression patterns in shoot apices and leaves. Most of the MIA pathway genes exhibited high expression in PCs and VCs, which did not align with the observations in *C. roseus*, where MIA biosynthesis primarily occurred in IPAP cells, ECs, and ICs. Furthermore, we found that the CPT biosynthetic pathways in the shoot apices and leaves were not obviously compartmentalized.

It has been reported that cytochrome P450s (CYP450s) play crucial roles in several reactions in the downstream synthesis pathway after strictosidinic acid. We screened potential candidate P450 enzymes through co-expression analysis at cell-type resolution (Figure 3a, and Figure S3a, b). The expression patterns of *CaCYP38*, *CaCYP167*, *CaCYP122*, *CaCYP33*, and *CaCYP36* were most similar to those of *CaSTR1* and *CaSTR2*, with the highest expression in the VCs of leaves and shoot apices. *CaCYP162* showed an expression pattern most similar to *CaSTR3*, with the highest expression in the MCs of shoot apices. *CaCYP8*, *CaCYP120*, and *CaCYP178* exhibited an expression pattern most similar to *Ca10OMT*, with the highest expression

observed in the VCs of the leaves. Similarly, the expression patterns of *CaCYP7*, *CaCYP204*, *CaCYP35*, and *CaCYP161* were most similar to *CYP71BE206*, which catalyzes strictosamide into strictosamide epoxide in the downstream synthesis pathway, and had the highest expression in PCs of shoot apices. The expression patterns of *CaCYP96* and *CaCYP106* were most similar to *CPT10H*, while the expression profiles of *CaCYP22*, *CaCYP78*, *CaCYP151* and *CaCYP199* were most similar to *CPT11H*, with the highest expression in PCs and VCs of leaves. Therefore, we inferred that these 19 CYP450s are candidate genes involved in downstream CPT biosynthesis (Figure 3a and Supplementary Table 7). Some methylated CPT derivatives, such as 11-methoxycamptothecin, 10-methoxycamptothecin, and 9-methoxycamptothecin, have been reported to have strong anticancer activity and have been isolated from *C. acuminata*. Thus, we also screened 8 candidate O-methyltransferase (OMT) enzymes that catalyze the conversion of CPT to methylated CPT derivatives in a similar manner (Figure 3a, Figure S3c and d).

TFs serve as pivotal molecular switches that regulate secondary metabolism in response to both internal and external signals. Through correlation analysis between the expression of predicted TF genes and CPT pathway genes, we identified 62 TFs from 23 different families that could potentially play roles in the biosynthesis of CPT and its derivatives (Figure 3b, Table S8). Among these, 20 TFs exhibited positive regulation, including 5 ZNFs, 4 MYBs, while 41 TFs are negative



**FIGURE 3** Co-expression analysis of putative genes encoding CYP450, OMT and TF. Cell type expression heatmap for the candidates of cytochrome P450s (CYP450s) and O-methyltransferase (OMT) genes. "SA" is the short name of the shoot apex, while the "L" represents the leaf. The identified CPT genes are marked in red (a). Co-expression network of CPT genes and TFs. Orange nodes represent the CPT genes, while the blue nodes represent TFs (b).

regulators, including 6 ZNFs, 5 MYBs, and 4 WRKYs. Interestingly, GES was regulated by the most TFs among all characterized pathway enzymes. Specifically, 31 TFs from 15 families were associated with GES, including 4 WRKYs, 6 ZNFs, 3 NACs, and 3 MYBs. Additionally, 19 TFs from 12 families were related to *STR1*, including 4 ZNFs, 2 bHLHs, and 2 TCPs. Notably, the MYB family demonstrated the most co-expression with pathway genes compared to other TF families. Our analysis revealed that the MYB family potentially regulates six pathway genes, suggesting its pivotal role in the biosynthesis of CPT. Moreover, some TFs have opposite regulatory functions on different genes. For instance, TF321 from the lateral organ boundaries (LOB) family is a negative regulator of *STR1* and a positive regulator of GES. This suggests that TF321 may play multiple important roles in regulating the biosynthesis of CPT.

## 2.4 | Dynamic expression of CPT biosynthetic pathway genes during EC and VC development

The ability of scRNA-seq to capture cells in different developmental states allows for the study of successive differentiation trajectories during tissue or cell development. To better analyze the development of ECs and VCs from the SAM, we merged the single-cell data of shoot apexes and leaves and obtained 20 distinct clusters after clustering (Figure 4a). All the clusters were annotated according to the above established strategies (Figure 4b and Figure S4).

The ECs were further divided into nine subclusters to reconstruct the EC developmental trajectory (Figure S5a, b). To further classify the cell types, the expression levels of the meristemoid (M) cells, guard mother cell (GMC) and EC markers *HOMEBOX-LEUCINE ZIPPER PROTEIN ROC2* (*HDG2*) (Miyuki Nakamura et al., 2006), *TRANSCRIPTION FACTOR FAMA* (*FAMA*) (Kyoko Ohashi-Ito and Dominique C. Bergmann, 2006), and *ECERIFERUM 3* (*CER3*) (Xinbo Chen et al., 2003) were analyzed (Figure S5c and Supplementary Table 9). The EC\_3 cluster primarily consisted of M cells expressing orthologs to the *Arabidopsis* identity marker *PDF2* and the transcription factor *SPEECHLESS* (*SPCH*), which assembled at the beginning of pseudo-time. In contrast, the EC\_6 and EC\_7 clusters were referred to as the mature epidermis due to the dominant expression of the marker gene *CER3*, which is located at the end of the trajectory (Figure 4c). It was clearly demonstrated that we captured cell populations that differentiated from M to ECs. Moreover, several key genes associated with the development of ECs, *PROTODERMAL FACTOR1* (*PDF1*), and *FIDDLEHEAD* (*FDH*) were selected to show the changes in expression during the development of EC. The expression of *PDF1* remained relatively stable throughout the process of EC differentiation, while the expression of *FDH* exhibited a gradual and consistent decrease (Figure 4d), which is consistent with the molecular function during the development of EC. The genes exhibiting significant expression changes along the branch were categorized into five distinct modules based on their expression profiles (Figure 4e). The Gene Ontology (GO) enrichment analysis provided a notable overrepresentation of genes associated with meristem growth regulation and cell growth, which was expected for early ECs (module 5). At the end of

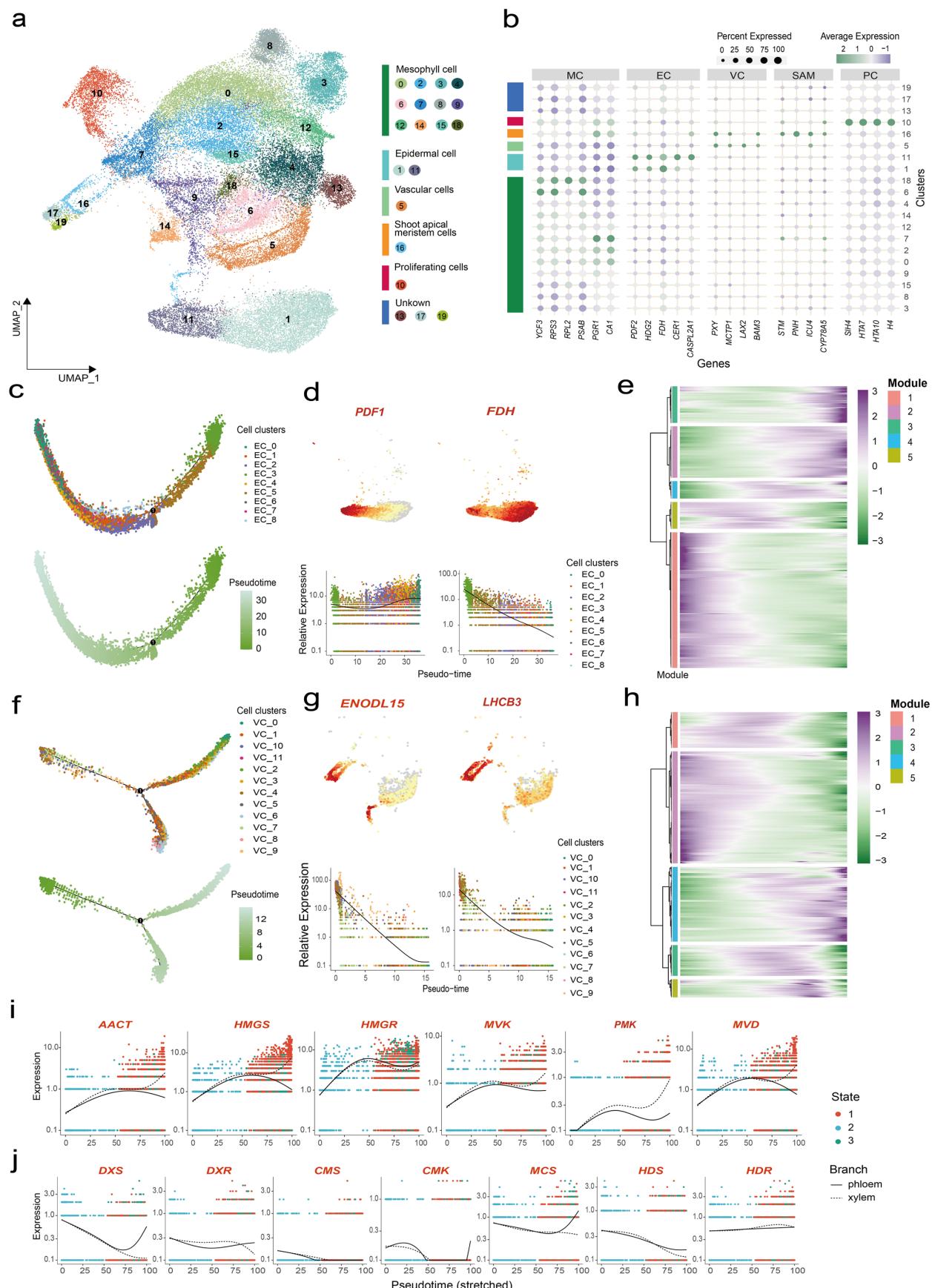
the trajectory, modules 2 and 3 included the expression of genes involved in the abiotic stress response, consistent with the hypersensitive nature of the mature epidermis to environmental stimuli. (Tables S10 and S11).

VCs were reassigned into 12 subclusters to reconstruct the VC developmental trajectory (Figure 4f, Figure S5d and e), among which the VC\_10 cluster assembled at the beginning of pseudotime axis mainly comprised procambium cells expressing orthologs to the *Arabidopsis* identity marker *TMM*. The VC\_0 and VC\_8 clusters were termed xylem cells and early phloem cells, respectively, because the marker genes *LOB DOMAIN-CONTAINING PROTEIN 15* (*LBD15*) and *WOODY ALTERED PHLOEM DEVELOPMENT* (*WDY*) were predominantly expressed (Figure S5f). In addition, a consistent decrease in the expression levels of *EARLY NODULIN-LIKE PROTEIN 15* (*ENOD15*) and *LIGHT HARVESTING CHLOROPHYLL-BINDING PROTEIN* (*LHCB3*) was observed (Figure 3g). GO enrichment analysis revealed that the enriched gene signals at the start of the pseudotime axis (module 2) corresponded to cell proliferation and the nucleosome assembly function of the procambium, whereas genes related to sugar uptake and ion transport were enriched at the end of the trajectory (module 5) (Figure 4h, Tables S12 and S13), which was consistent with the process of VCs development.

Subsequently, changes in pathway gene expression during EC and VC development were analyzed. During EC maturation, genes in the MVA pathway were increasingly expressed, whereas most genes in the MEP pathway demonstrated an inverse trend with development time (Figure S6). A similar expression pattern was observed during VC development (Figure S7), and further analysis revealed higher MVA pathway gene expression levels in mature xylem cells than in phloem cells (Figure 4i). Conversely, while MEP pathway gene expression decreased overall, phloem cells exhibited higher levels than xylem cells at late stages (Figure 4j). Additionally, the expression of other genes (Figure S6 and S7), such as *SLAS*, exhibited the opposite trend with development time along the EC trajectory but fluctuated in the VC samples. In addition, some genes showed specific enrichment in one type with negligible expression in the other. In summary, in addition to MVA and MEP exhibiting consistent expression patterns in both cell lineages, the trajectories of other genes differed between cell types. The dynamic expression of all pathway genes across pseudotime highlights the complexity of specialized metabolite biosynthesis during epidermis, xylem, and phloem cell differentiation. This provides insights into cell type-specific metabolic specialization in these critical epidermis cell and vascular cell populations.

## 2.5 | Differential expression of TEs among plant tissues and cells

To gain insights into whether TE has an impact on cell fate determination and development, we expanded our investigation beyond genes to TEs, another crucial component of the genome. Previous reports have highlighted the family-level specificity of TE expression in single-cell data, defining diverse cell identities (Jiangping He et al., 2021).

**FIGURE 4** Legend on next page.

Therefore, we annotated TEs in the genome of *C. acuminata*. Then, we employed scTE (v1.0.4) for gene and TE expression quantification, following the same analytical pipeline used for analysis based solely on gene expression. Consequently, a comprehensive single-cell expression atlas encompassing both gene and TE expression was generated (Figure 5a-c). Each cell cluster was effectively annotated using the same marker genes as previously identified (Figure S8). By comparing the cell identities in clusters generated based on the gene expression matrix and based on the gene/TE expression matrix, we observed that the majority of cells were consistently annotated to the same cell types via both approaches (Figure S9-S11). This indicated that the inclusion of TE expression does not hinder the accurate identification of cell identities, which is consistent with previous research (Jiangping He et al., 2021). Additionally, we found that analysis based on the gene/TE expression matrix more readily isolated CZ and PZ cells than that generated solely from the gene expression matrix. This suggested that the inclusion of TE expression serves to augment intercellular heterogeneity and facilitate the discovery of rarer cell types.

According to the scRNA-seq data of *C. acuminata*, we detected a total of 7,829 families of TEs expressed in the leaves and 7,991 in the shoot apices (Tables S14 and S15). Among them, 344 families were specific to leaves, and 506 were specific to shoot apices. Among these expressed and unique TE families, the LTR superfamily was the most common. While the overall proportions of each TE family in the shoot apices and leaves were quite similar, the shoot apices exhibited a higher proportion of tissue-specific TEs from the Gypsy superfamily compared to the leaves. (Figure 5d). Moreover, TEs also exhibited differential expression across different clusters. Some TE families in leaf cells, such as rnd-1-family-17 and rnd-1-family-20, were highly expressed in Cluster 9, while other families, such as CA12:12859547..12861064-LTR and rnd-1-family-15, were almost exclusively expressed in Cluster 11. Cluster-specific TE families were also observed in shoot apex cells (Figure 5e, Figures S12, S13, Tables S16 and S17). All of these findings indicated the presence of cell type-specific and tissue-specific expression of TEs and highlighted the potential of TEs as cellular identity markers. Interestingly, whether in shoot apex cells or leaf cells, highly specific expression of several TE families was detected in the PC cells (Figure 5e), suggesting that some TEs may be involved in cell proliferation during these stages.

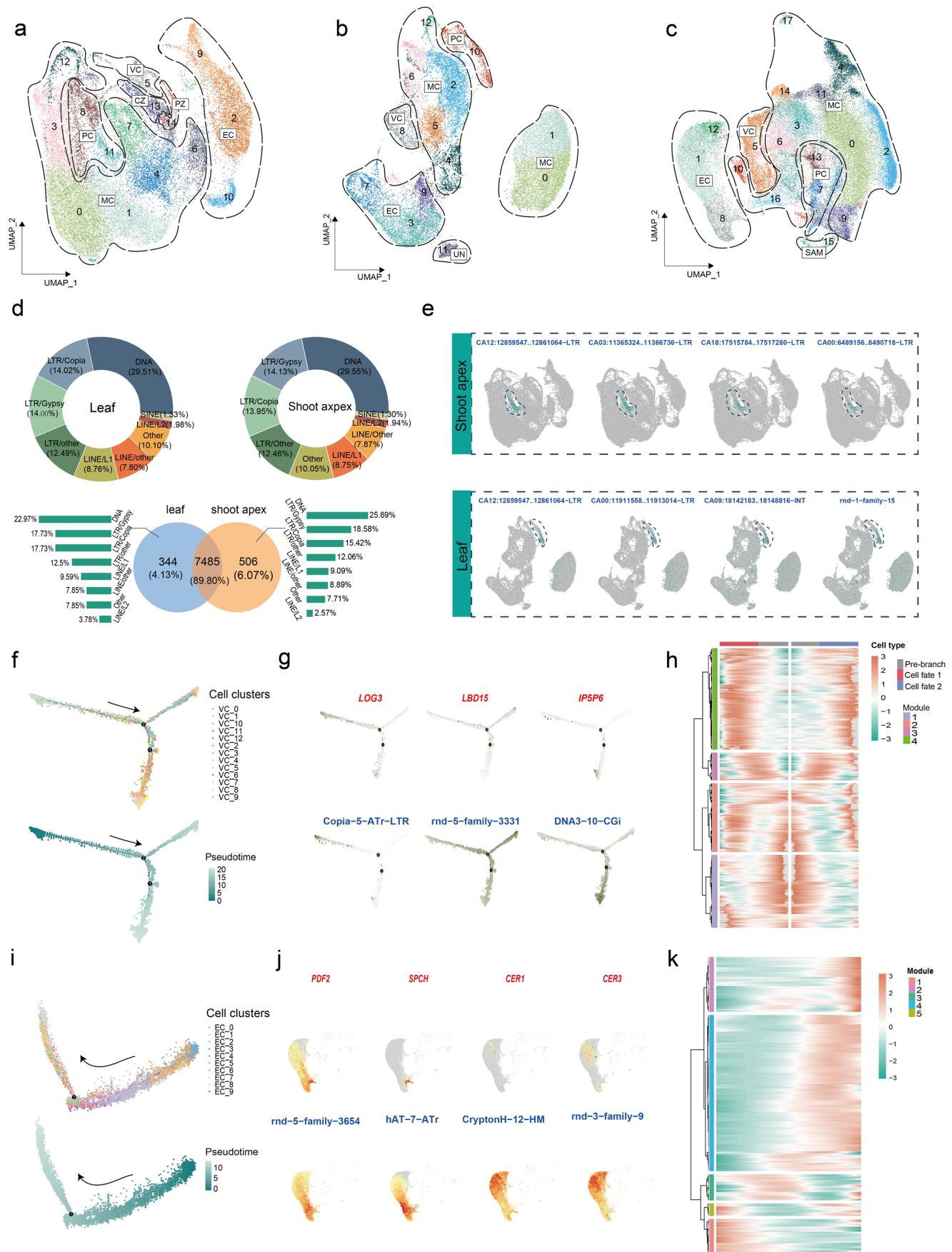
TE expression is regulated by complex mechanisms such as small RNAs and chromatin modifications. Recent studies have also reported that certain TFs have regulatory effects on TE expression (Guillaume Bourque et al., 2018). Therefore, we calculated the correlation between the expression of predicted TFs and TE families, which showed cell cluster-specific expression in shoot apices and leaves, respectively. High-confidence results were filtered, and the top 100 highly positively correlated pairs of TE families and TFs were displayed separately for shoot apices and leaves (Tables S18, S19 and Figure S14). Among these, co-expression relationships of 13 TE families with 46 TFs were displayed in shoot apices, while 20 TE families with 38 TFs in leaves. Notably, certain TE families demonstrated a high number of co-expression relationships with TFs. For instance, the hAT-N48C-DR TE family exhibited the highest number of co-expression relationships with 27 TFs in shoot apices, while the MicOch-5.428b TE family showed co-expression with 14 TFs in leaves. However, only 26 pairs (containing 7 TFs and 6 TE families) exhibited highly correlated expression in both samples (Supplementary Table 20), suggesting specific relationships existing in the two tissues, which is consistent with the tissue-specific expression pattern of TE family mentioned above. Among these seven TFs, four belong to the ZNF family, including the C2C2, C2H2 and C3H subfamilies, while the remaining three belong to the bHLH, Whirly, and NAC families. These findings exhibited pairs of TFs and TE families with possible underlying connections, providing valuable insights for further investigations into the regulatory relationships between TEs and TFs.

## 2.6 | The state-specific expression of TEs in VC and EC cell development

Given the developmental relationship between certain cell types, the heterogeneity of TE expression at the single-cell level may be associated with cell development. Therefore, we performed pseudotime trajectory analysis based on gene and TE expression based on the VC and EC populations. In the VC developmental trajectory, we defined State 4 as developmental initiation using the marker gene CYTOKININ RIBOSIDE 5'-MONIPHOSPHATE PHOSPHORIBOHYDROLASE (LOG3) for progenitor cells. The

State 3, representing xylem cells, was defined by HISTIDINE KINASE 3 (HK3) and LBD15, while the remaining branch endpoint,

**FIGURE 4** Reconstruction of the developmental trajectories of epidermal cells (ECs) and vascular cells (VCs). UMAP visualization of cells after the integration of shoot apex and leaf scRNA-seq data. Each dot represents a cell, and the colours indicate the different clusters (a). The expression patterns of cell type-specific marker genes in the integrated data (b). Detailed information about these marker genes is shown in Table S 7. The pseudotime trajectory of ECs. Each dot represents a cell, and distinct cell clusters of ECs are shown. The depth of green indicates the pseudotime score (c). UMAP visualization of transcript accumulation and gene expression kinetics during pseudotime progression for PROTODERMAL FACTOR1 (PDF1), FIDDLEHEAD (FDH), EARLY NODULIN-LIKE PROTEIN 15 (ENODL15) and LIGHT HARVESTING CHLOROPHYLL-BINDING PROTEIN (LHCB3). The colour intensity represents the relative transcript expression level (d and g). The expression heatmap of pseudotime-dependent genes among ECs over the developmental trajectory (e). The pseudotime trajectory of VCs. Each dot represents a cell, and distinct cell clusters of VCs are shown. The depth of green indicates the pseudotime score (f). The expression heatmap of pseudotime-dependent genes among VCs over the developmental trajectory (h). Gene expression kinetics during the VC pseudotime trajectory for the cytosolic mevalonate (MVA) pathway genes and the plastidic methylerythritol phosphate (MEP) pathway genes, respectively (i and j).

**FIGURE 5** Legend on next page.

representing phloem cells, was defined by PROTEIN NRT1/PTR FAMILY 1.2 (PTR6) and TYPE IV INOSITOL POLYPHOSPHATE 5-PHOSPHATASE 6 (IP5P6) (Figure 5f, g, Figure S15a, b and Figure S16). During this process, certain TE families, such as Tx1-55-DR and Copia-5-ATr-LTR, exhibited specific high expression patterns during early development, while DNA3-10-GCi exhibited specific high expression patterns at the endpoint of phloem cell development. Additionally, during the development of xylem cells, TE families such as the rnd-5-family-3331 and rnd-5-family-3637 exhibited expression patterns similar to those of the xylem marker gene *LBD15* (Figure 5g and Figure S16). Next, the TE families detected by branch expression analysis modelling (BEAM) were selected for generating the expression heatmap (Figure 5h). As the cells progressed from the starting point towards the two branch endpoints, the TE families clustered into four distinct clusters, demonstrating differential expression patterns according to the development stage. During the process of early epidermal cell development to mature wax-secreting epidermal cells, State 3 was defined as developmental initiation by *PDF2* and *SPCH*, while State 1 was defined as the developmental endpoint using *VERY-LONG-CHAIN ALDEHYDE DECARBOXYLASE* (*CER1*) and *CER3* (Figure 5i, j, Figure S15c and d). In this trajectory, we observed some TE families exhibiting similar landscapes to those in the VC trajectory (Figure 5h, j and k). These findings revealed the characteristics of TE families with stage-specific expression during development. In addition, based on annotation information, we found that TE families located near and within trajectory marker genes are mainly DNA transposons, in which some exhibited similar expression trends as highly expressed marker genes (Figure S17, Tables S21 and S22). These findings suggested that these TE families may have potential effects on key genes involved in cell development and differentiations. Therefore, our study revealed that certain TE families exhibit cell stage-specific expression and may play a role in cell fate determination in *C. acuminata*.

### 3 | DISCUSSION

Single-cell RNA-seq has transformed plant science by providing unprecedented insights into cellular diversity and functional genomics at single-cell resolution (Carolin Seyfferth et al., 2021). Although the utilization of single-cell transcriptome atlases in crop and model plant

species has elucidated cell development and agronomic gene regulation (Hao Liu et al., 2021; Rahul Shaw et al., 2021), related research in medicinal plants has been relatively limited, with less focus on specialized metabolism and TE transcription. This study aimed to generate transcriptome atlases of *C. acuminata* to investigate the spatiotemporal distribution of genes associated with CPT synthesis pathways at the single-cell level. Furthermore, we integrated TE expression to construct a gene/TE expression atlas and revealed the presence of spatio-temporal heterogeneity in TE expression in plants and the potential roles in cell development and differentiation.

The loss of some cell types is commonly observed in single-cell transcriptome studies (Lin et al., 2023). For example, S cells and glandular hair cells were not identified in *Arabidopsis thaliana* (Tenorio Berrio et al., 2022), and a similar limitation was encountered in our study. This loss of cell types can be attributed to two potential factors. First, the inherent structural characteristics of these cells pose significant challenges during protoplasting. For example, trichomes possess a rigid cell wall structure that imparts exceptional resistance to protoplasting procedures (Peina Zhou et al., 2022). The utilization of single-nucleus RNA-seq provides a promising alternative approach to capture additional cell types through nuclear separation rather than protoplasting (Kaimeng Wang et al., 2023). This method ensures that nucleus preparation is less influenced by cellular features or spatial location within tissues. Moreover, due to the absence of marker genes for certain cell types, such as ICs, it remains uncertain whether they were present within the dataset but remained undetectable. Given this challenge, the application of spatial transcriptomics proves valuable in acquiring *in situ* transcript localization information, offering crucial insights for optimizing cell annotation (Anjali Rao et al., 2021).

The multicellular localization patterns of several metabolic pathways, including the glucosinolate pathway in *A. thaliana*, the benzylisoquinoline alkaloid (BIA) metabolism pathway in *Opium poppy* (Natali Ozber and Peter J. Facchini, 2022), taxol metabolism pathway in *Taxus* (Yu et al., 2023) and the MIA pathway in *C. roseus*, have been extensively studied. Based on the given context, it is evident that certain metabolic pathways exhibit apparent compartmentalization. For example, in *A. thaliana*, the biosynthesis and accumulation of glucosinolates involve three distinct cell types: xylem parenchyma cells, phloem cells, and ICs. Similarly, in *Opium poppy*, sieve elements and laticifers are involved in the BIA pathway. In *C. roseus*, the MIA

**FIGURE 5** Single-cell transcriptome atlas constructed based on gene/transposable elements (TEs) expression and spatiotemporal TE-specific expression. UMAP visualizations based on gene/TE expression of shoot apex cells, leaf cells, and cells from integrated datasets of both tissues. Each dot represents a cell, and the colors indicate the different clusters (a-c). Classification and proportions of the expressed TE families in shoot apices and leaves. The Venn diagram shows the numbers of unique and shared TE families in the two tissues, while the bar plots show the classification and proportions of the tissue-specific TE families (d). Visualization of several cell type-specific TE families by UMAP of leaf and shoot apex cells. The colour intensity represents the relative transcript expression level (e). The pseudotime trajectory of VCs and ECs is based on gene/TE expression, respectively. Each dot represents a cell, and distinct cell clusters of VCs and ECs are shown. The depth of blue indicates the pseudotime score (f and i). The expression patterns of stage-specific marker genes and TE families along the EC developmental trajectory. The colour intensity represents the relative transcript expression level (g). The expression heatmap of TE families related to the VC developmental trajectory branches was identified by the BEAM function in Monocle. Cell fate 1 represents phloem cells, and cell fate 2 represents xylem cells (h). UMAP plot of the selected stage-specific marker genes and TE families in ECs (j). The expression heatmap demonstrates the pseudotime-dependent TE families across the course of EC development (k).

pathway is partitioned among three cell types: internal phloem-associated IPAP cells, ECs, and ICs. However, it is notable that not all specialized metabolic pathways exist in compartmentalization. For instance, most taxol biosynthesis genes are predominantly expressed in MCs. In this study, we employed single-cell RNA-seq to explore, for the first time, the localization of identified enzymes in the CPT pathway. According to the results, we did not observe an apparent compartmentalization of CPT biosynthesis. Regrettably, the absence of key cell types, particularly ICs, which serve as the primary accumulation site of CPT, and the incomplete understanding of the CPT pathway may have imposed limitations on our analysis of the intricate higher-order organization of CPT biosynthesis. Overall, further exploration is required to investigate whether the compartmentalization of CPT biosynthesis existed within *C. acuminata* leaves and shoot apices.

Co-expression analyses using bulk mRNA-seq have been used to discover genes involved in the CPT biosynthetic pathway and related TFs (Hu et al., 2020; Zhang et al., 2023). However, expression abundances derived from whole tissues are averaged over all cells in the organ. Thus, if a gene is expressed in a rare cell type, the power of co-expression analyses will be limited. By using transcriptomes from individual cell types, we improved prediction accuracy and reduced background noise caused by the averaging of expression abundances in bulk RNA-Seq. Here, for the first time, we identified candidate genes involved in the CPT downstream pathway through single-cell level co-expression analysis, including 19 candidate CYP450s and 8 candidate OMTs. Furthermore, our single-cell data predicted that 11 ZNFs, 8 MYBs, 4 bHLHs, and 5 WRKYs are involved in regulating CPT biosynthesis in leaves and shoot apices. Previous studies identified one MYB family member and its target genes, where overexpression of *OpMYB1* in hairy roots of *Ophiorrhiza pumila* resulted in reduced production of camptothecin and decreased expression of *OpTDC*. We suggested several new MYB members strongly associated with *IS* and *STR2*. Additionally, we identified new TF families potentially involved in CPT biosynthesis, such as the C3H family and the LOB family. Interestingly, we also found some TEs co-expressed with TFs. While reciprocal regulation between TEs and TFs has been discovered in animals, further investigations are needed to elucidate this phenomenon in plants.

TEs have attracted increasing research interest in recent years due to their emerging roles in chromosomal structural rearrangement and gene regulation. Using scTE, we generated single-cell transcriptome atlases based on gene and transposable element expression profiles. Within these atlases, we observed tissue specific, cell-specific, and developmental stage-specific expression patterns of TE families. Intriguingly, we observed extremely specific expression patterns of several TE families in PCs, suggesting a potential shared transcriptional mode or activation machinery among members of these highly expressed TE families in PCs. Previous studies have shown that extensive mitosis can weaken methylation maintenance, reactivating TE transposition (Sarah E. Johnstone et al., 2022). Therefore, we speculate that the specific transcriptional activities of these TE families in PCs may be related to low methylation resulting from extensive

proliferation. Additionally, we observed that most genes involved in the CPT biosynthetic pathway also exhibited high expression, specifically in PCs. However, further investigations are needed to determine whether the expression of these TEs is involved in the regulation of pathway genes.

## 4 | MATERIALS AND METHODS

### 4.1 | Plant material

*C. acuminata* plants were planted at 25°C in a growth chamber under a 16/8 h photoperiod. The leaves (1.8–2.0 cm long) and shoot apices of one-year-old plants were used for single-cell transcriptome sequencing and metabolome analysis.

### 4.2 | UPLC-MS/MS analysis

Young and healthy leaves as well as shoot apices were collected and immediately frozen in liquid nitrogen. The samples were ground with a ball mill, after which 70% methanol–water extraction solution was added. After centrifuging, the supernatant was filtered for UPLC-MS/MS analysis (AB SCIEX, ExionLC™ AD, Applied Biosystems 4500 QTRAP). Column, Agilent SB-C18 (1.8 μm, 2.1 × 100 mm). The mobile phase consisted of solvent A, pure water with 0.1% formic acid; solvent B, acetonitrile with 0.1% formic acid. Sample measurements were performed with a gradient program that employed the following starting conditions: 0–9 min, 95% A, 5% B; 9–10 min, 5% A, 95% B, 10–14 min, 95% A, 5.0% B; flow velocity, 0.35 mL/minute; column temperature, 40°C; and injection volume, 2 μL.

### 4.3 | Protoplast isolation

In brief, young healthy leaves and shoot apices were collected and processed in an enzyme solution (2% cellulase R-10, 0.3% macerozyme R-10, 0.5% pectinase, 0.5 M mannitol, and 20 mM MES, pH 5.6–5.8) for 2.5 h at 25°C. The protoplasts were centrifuged at 100 g for 6 min after passing through 70 and 40 μm strainers and were subsequently washed once with protoplasting solution without enzymes. Trypan blue and fluorescein diacetate staining were used to test protoplast viability, and cells were counted using a hemocytometer. The final suspension volume was adjusted to achieve a cell density of 1500–1800 cells/μL. The protoplasts were kept cold until further processing.

### 4.4 | scRNA-seq library construction and sequencing

The scRNA-seq libraries were constructed using the Chromium Single-cell 3' Gel Beads-in-emulsion (GEM) Library and Gel Bead Kit v.3 (16 rxns PN-1000268, 10x Genomics) according to the user's

manual supplied with the kit. Approximately 20000 cells for one sample were loaded on a single-cell A chip. Libraries were sequenced on DNBSEQ-T7 according to the manufacturer's instructions.

#### 4.5 | Preprocessing of scRNA-seq data

Raw sequencing reads from high-throughput single-cell RNA sequencing of leaf and shoot apex samples were processed using the Alevin pipeline (Avi Srivastava et al., 2019) (Salmon v1.9.0) to generate unique molecular identifier (UMI) counts and digital expression matrices. These matrices were imported into Seurat (Yuhan Hao et al., 2021) (v4.3.0) to perform integrated analysis. Quality control was conducted to exclude low-quality cells based on the number of detected genes, UMI counts, and proportions of chloroplast genes. Potential doublets were also removed using DoubletFinder (Christopher S. McGinnis et al., 2019) (v2.0.3). Cell cycle scores were computed using cell cycle-related genes identified by homology alignment with known marker genes (Table S1). Normalization was performed via the SCTransform (vst.flavor = "v1") to eliminate the effects of chloroplast DNA, UMI counts and the cell cycle. Principal component analysis identified major sources of variation from 3000 highly variable genes selected by RunPCA. Batch effects were mitigated by applying Harmony (Ilya Korsunsky et al., 2019) to the normalized data.

#### 4.6 | Cell clustering and annotation

Graph-based clustering was performed using the FindNeighbors and FindClusters functions in Seurat. Nonlinear dimensionality reduction was applied using RunUMAP function. Differentially expressed genes between clusters were identified using FindAllMarkers with thresholds of  $pct.2 < 0.1$  and  $\logfc.threshold = 0.25$ . FeaturePlot was used to visualize the cell distributions and expression of marker genes in the UMAP space. The average expression was calculated for each cluster. Correlations between cluster averages are depicted in a heatmap created with pheatmap (v1.0.12).

#### 4.7 | Gene co-expression analysis

Based on existing databases such as PlantTFDB and SwissProt, potential genes encoding TFs, CYP450s, and OMTs were predicted in the *C. acuminata* genome by iTAK and HMMsearch. Single-cell data from shoot apices and leaves were utilized for further analysis. The "rcorr()" function from the R package Hmisc (v5.1-2) was employed to calculate Spearman correlation coefficients between these potential genes and the identified CPT pathway genes. The filtering criteria for TF genes were an absolute correlation coefficient  $>0.75$  and a  $p$ -value  $<0.05$ , and for enzyme genes, a correlation coefficient  $>0.5$  and a  $p$ -value  $<0.05$ . Subsequently, we generated a network graph using Cytoscape (v3.10.2) to visualize the known CPT biosynthetic enzyme genes along with their first-degree network neighbors. Nodes connected to CPT

synthesis genes were selected to construct a heatmap demonstrating gene expression. Based on the expression patterns, candidate enzymes and TFs related to CPT synthesis were inferred.

#### 4.8 | Pseudotime analysis

Pseudotime trajectory analysis was conducted using the R package Monocle (Xiaojie Qiu et al., 2017) (v2.26.0). The log-normalized data derived from the Seurat object were imported into Monocle object via the "as.CellDataSet" function. The cells were ordered along the trajectory and visualized in a dimensionally reduced space. The determination of the trajectory root was identified by developmental marker genes. The differentialGeneTest function was applied to identify genes exhibiting significant changes along pseudotime with the  $q$ -value  $<0.1$ . The identified genes were then clustered using the plot\_pseudotime\_heatmap function with default parameters. The plot\_genes\_in\_pseudotime and plot\_genes\_branches\_pseudotime functions were used to demonstrate dynamic expression patterns of genes throughout pseudotime.

#### 4.9 | Analysis of scRNA-seq data using scTE

We employed RepeatModeler (Jullien M. Flynn et al., 2020) (v2.0.5), TEclass (v2.1.3d) (György Abrusán et al., 2009), and EDTA (v2.2.0) (Shujun Ou et al., 2019) tools to perform initial TE annotation for the target species. Subsequently, we integrated the obtained TE annotations with the RepBase database to eliminate redundancies and create a comprehensive TE database specific to *C. acuminata*. RepeatMasker (v4.1.5) was utilized to final annotate TEs in the genome based on the integrated TE database. The annotation files of genes and TEs were integrated to construct index through the commander 'scTE\_build' of the scTE (Jiangping He et al., 2021) (v1.0.4) pipeline. The BAM files required for the pipeline were quantitatively generated by CellRanger (v7.0.1) from the single-cell data sets. Subsequently, the gene/TE expression matrices were generated through the commander 'scTE' with the arguments '-x, -i, -o, and -expect-cells' and were used in Seurat (v4.3.1) for further analysis via the same process as that used for the analysis based on gene expression alone. For the co-expression analysis, differentially expressed TE families across clusters were screened with a  $\logfc.threshold > 1$  and an adjusted  $p$ -value  $<0.05$ . The calculation and visualization of correlations followed the methods described above. High-confidence results were obtained by filtering for correlations greater than 0.75 and  $p$ -values less than 0.05.

## 5 | CONCLUSIONS

In the present study, we utilized single-cell RNA-seq to generate extensive datasets derived from both the shoot apices and leaves of *C. acuminata*. Our analysis revealed that most genes involved in MIA biosynthesis pathways exhibited remarkable expression in PCs and

VCs. As a result, it became apparent that CPT biosynthesis did not exist distinct multicellular compartmentation. The expression of pathway genes changes dynamically during the development of VCs and ECs. In addition, a total of 19 CYP450s, 8 OMTs, and 62 TFs were identified as candidates related to the biosynthesis of CPT and its derivatives through co-expression analysis. Furthermore, in the realm of plant research, we made pioneering strides by constructing single-cell transcriptomic atlases based on the expression of both genes and TEs. Through meticulous examination, we observed cell type-specific, developmental stage-specific, and tissue-specific expression patterns of TE families in *C. acuminata*. In conclusion, the comprehensive and invaluable dataset generated through this study serves as a potent resource for exploring plant physiology, unravelling the intricacies of specialized metabolism, and specifically, shedding light on the biosynthesis of CPT at the single-cell level.

## AUTHOR CONTRIBUTIONS

X.S. and B. G. conceived and designed the project. X.S. and S.W. performed the experiments, and Y.L., C.Z., R.L., K.D. analyzed the data. S.W. and C.Z. wrote the manuscript draft. S.W., X.S., C.Z., and S.C. revised the manuscript. All authors have read and agreed to the published version of the manuscript.

## FUNDING INFORMATION

This research was supported by the Chinese Academy of Medical Sciences Innovation Fund for Medical Sciences (grant no. 2021-I2M-1-032). The funders had no role in the design of the study; in the collection, analyzes or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

## DATA AVAILABILITY STATEMENT

The raw scRNA-seq data sets generated in this study have been deposited in CNCB (China National Center for Bioinformation) under accession number PRJCA023356.

## ORCID

Shu Wang  <https://orcid.org/0000-0002-3992-1830>

## REFERENCES

- Alessio Valletta, Livio Trainotti, Anna Rita Santamaria, and Gabriella Pasqua (2010) ‘Cell-specific expression of tryptophan decarboxylase and 10-hydroxygeraniol oxidoreductase, key genes involved in camptothecin biosynthesis in *Camptotheca acuminata* Decne (Nyssaceae)’. *BMC Plant Biol* 10, 69. Available at: <https://doi.org/10.1186/1471-2229-10-69>
- Anjali Rao, Dalia Barkley, Gustavo S. França, and Itai Yanai (2021) ‘Exploring tissue architecture using spatial transcriptomics’, *Nature*, 596(7871), pp. 211–220. Available at: <https://doi.org/10.1038/s41586-021-03634-9>.
- Avi Srivastava, Laraib Malik, Tom Smith, Ian Sudbery, and Rob Patro (2019) ‘Alevin efficiently estimates accurate gene abundances from dscRNA-seq data’, *Genome Biology*, 20(1), p. 65. Available at: <https://doi.org/10.1186/s13059-019-1670-y>.
- Bing Wang, Steven M. Smith, and Jiayang Li (2018) ‘Genetic Regulation of Shoot Architecture’, *Annual Review of Plant Biology*, 69, pp. 437–468. Available at: <https://doi.org/10.1146/annurev-arplant-042817-040422>.
- Carolin Seyfferth, Jim Renema, Jos R. Wendrich, Thomas Eekhout, Ruth Seurinck, Niels Vandamme, Bernhard Blob, Yvan Saeys, Yrjo Helariutta, Kenneth D. Birnbaum, and Bert De Rybel (2021) ‘Advances and Opportunities in Single-Cell Transcriptomics for Plant Research’, *Annual Review of Plant Biology*, 72, pp. 847–866. Available at: <https://doi.org/10.1146/annurev-arplant-081720-010120>.
- Chen, H., Ye, F. and Guo, G. (2019) ‘Revolutionizing immunology with single-cell RNA sequencing’, *Cellular & Molecular Immunology*, 16(3), pp. 242–249. Available at: <https://doi.org/10.1038/s41423-019-0214-4>.
- Christopher S. McGinnis, Lyndsay M. Murrow, and Zev J. Gartner (2019) ‘DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors’, *Cell Systems*, 8(4), pp. 329–337.e4. Available at: <https://doi.org/10.1016/j.cels.2019.03.003>.
- Emily C. Stow, Melody Baddoo, Alexis J. LaRosa, Dawn LaCoste, Prescott Deininger, and Victoria Belancio (2022) ‘SCIFER: approach for analysis of LINE-1 mRNA expression in single cells at a single locus resolution’, *Mobile DNA*, 13(1), p. 21. Available at: <https://doi.org/10.1186/s13100-022-00276-0>.
- Eva Vranová, Diana Coman, and Wilhelm Gruissem (2013) ‘Network analysis of the MVA and MEP pathways for isoprenoid synthesis’, *Annual Review of Plant Biology*, 64, pp. 665–700. Available at: <https://doi.org/10.1146/annurev-arplant-050312-120116>.
- Fei Chen, Wei Li, Liangzhen Jiang, Xiang Pu, Yun Yang, Guolin Zhang, and Yinggang Luo (2016) ‘Functional characterization of a geraniol synthase-encoding gene from *Camptotheca acuminata* and its application in production of geraniol in *Escherichia coli*’, *Journal of Industrial Microbiology & Biotechnology*, 43(9), pp. 1281–1292. Available at: <https://doi.org/10.1007/s10295-016-1802-2>.
- G. L. Beretta, L. Gatti, P. Perego, and N. Zaffaroni (2013) ‘Camptothecin resistance in cancer: insights into the molecular mechanisms of a DNA-damaging drug’, *Current Medicinal Chemistry*, 20(12), pp. 1541–1565. Available at: <https://doi.org/10.2174/0929867311320120006>.
- Guillaume Bourque, Kathleen H. Burns, Mary Gehring, Vera Gorbunova, Andrei Seluanov, Molly Hammell, Michaël Imbeault, Zsuzsanna Izsvák, Henry L. Levin, Todd S. Macfarlan, Dixie L. Mager, and Cédric Feschotte (2018) ‘Ten things you should know about transposable elements’, *Genome Biology*, 19(1), p. 199. Available at: <https://doi.org/10.1186/s13059-018-1577-z>.
- György Abrusán, Norbert Grundmann, Luc DeMester, and Wojciech Makalowski (2009) ‘TEclass-a tool for automated classification of unknown eukaryotic transposable elements’, *Bioinformatics (Oxford, England)*, 25(10), pp. 1329–1330. Available at: <https://doi.org/10.1093/bioinformatics/btp084>.
- Hao Liu, Dongxiu Hu, Puxuan Du, Liping Wang, Xuanqiang Liang, Haifeng Li, Qing Lu, Shaoxiong Li, Haiyan Liu, Xiaoping Chen, Rajeev K. Varshney, and Yanbin Hong (2021) ‘Single-cell RNA-seq describes the transcriptome landscape and identifies critical transcription factors in the leaf blade of the allotetraploid peanut (*Arachis hypogaea* L.)’, *Plant Biotechnology Journal*, 19(11), pp. 2261–2276. Available at: <https://doi.org/10.1111/pbi.13656>.
- Hu, Y.-T., Xu, Z.-C., Tian, Y., Gao, R.-R., Ji, A.-J., Pu, X.-D., Wang, Y., Liu, X., and Song, J.-Y. (2020) ‘Genome-wide identification and analysis of AP2/ERF transcription factors related to camptothecin biosynthesis in *Camptotheca acuminata*’, *Chinese Journal of Natural Medicines*, 18(8), pp. 582–593. Available at: [https://doi.org/10.1016/S1875-5364\(20\)30070-4](https://doi.org/10.1016/S1875-5364(20)30070-4).
- Ilya Korsunsky, Nghia Millard, Jean Fan, Kamil Slowikowski, Fan Zhang, Kevin Wei, Yuriy Baglaenko, Michael Brenner, Po-Ru Loh, and Soumya Raychaudhuri (2019) ‘Fast, sensitive and accurate integration of

- single-cell data with Harmony', *Nature Methods*, 16(12), pp. 1289–1296. Available at: <https://doi.org/10.1038/s41592-019-0619-0>.
- Jiangping He, Isaac A. Babarinde, Li Sun, Shuyang Xu, Ruhai Chen, Junjie Shi, Yuanjie Wei, Yuhao Li, Gang Ma, Qiang Zhuang, Andrew P. Hutchins, and Jiekai Chen (2021) 'Identifying transposable element expression dynamics and heterogeneity during development at the single-cell level with a processing pipeline scTE', *Nature Communications*, 12(1), p. 1456. Available at: <https://doi.org/10.1038/s41467-021-21808-x>.
- Jullien M. Flynn, Robert Hubley, Clément Goubert, Jeb Rosen, Andrew G. Clark, Cédric Feschotte, and Arian F. Smit (2020) 'RepeatModeler2 for automated genomic discovery of transposable element families', *Proceedings of the National Academy of Sciences of the United States of America*, 117(17), pp. 9451–9457. Available at: <https://doi.org/10.1073/pnas.1921046117>.
- K. S. Mysore, J. Nam, and S. B. Gelvin (2000) 'An *Arabidopsis* histone H2A mutant is deficient in Agrobacterium T-DNA integration', *Proceedings of the National Academy of Sciences of the United States of America*, 97(2), pp. 948–953. Available at: <https://doi.org/10.1073/pnas.97.2.948>.
- Kaimeng Wang, Caiyao Zhao, Sunhuan Xiang, Kunyu Duan, Xiaoli Chen, Xing Guo, and Sunil Kumar Sahu (2023) 'An optimized FACS-free single-nucleus RNA sequencing (snRNA-seq) method for plant science research', *Plant Science*, 326, p. 111535. Available at: <https://doi.org/10.1016/j.plantsci.2022.111535>.
- Kate Fisher and Simon Turner (2007) 'PXY, a Receptor-like Kinase Essential for Maintaining Polarity during Plant Vascular-Tissue Development', *Current Biology*, 17(12), pp. 1061–1066. Available at: <https://doi.org/10.1016/j.cub.2007.05.049>.
- Kook Hui Ryu, Yan Zhu, and John Schiefelbein (2021) 'Plant Cell Identity in the Era of Single-Cell Transcriptomics', *Annual Review of Genetics*, 55, pp. 479–496. Available at: <https://doi.org/10.1146/annurev-genet-071719-020453>.
- Kyoko Ohashi-Ito and Dominique C. Bergmann (2006) 'Arabidopsis FAMA controls the final proliferation/differentiation switch during stomatal development', *The Plant Cell*, 18(10), pp. 2493–2505. Available at: <https://doi.org/10.1105/tpc.106.046136>.
- Lin, J.-L., Chen, L., Wu, W.-K., Guo, X.-X., Yu, C.-H., Xu, M., Nie, G.-B., Dun, J.-L., Li, Y., Xu, B., Wang, L.-J., Chen, X.-Y., Gao, W. and Huang, J.-Q. (2023) 'Single-cell RNA sequencing reveals a hierarchical transcriptional regulatory network of terpenoid biosynthesis in cotton secretory glandular cells', *Molecular Plant*, 16(12), pp. 1990–2003. Available at: <https://doi.org/10.1016/j.molp.2023.10.008>.
- M. E. Chaboute, N. Chaubet, G. Philipps, M. Ehling, and C. Gigot (1987) 'Genomic organization and nucleotide sequences of two histone H3 and two histone H4 genes of *Arabidopsis thaliana*', *Plant Molecular Biology*, 8(2), pp. 179–191. Available at: <https://doi.org/10.1007/BF00025329>.
- Miyuki Nakamura, Hiroshi Katsumata, Mitsutomo Abe, Naoto Yabe, Yoshibumi Komeda, Kotaro T. Yamamoto, and Taku Takahashi (2006) 'Characterization of the class IV homeodomain-Leucine Zipper gene family in *Arabidopsis*', *Plant Physiology*, 141(4), pp. 1363–1375. Available at: <https://doi.org/10.1104/pp.106.077388>.
- Natali Ozber and Peter J. Facchini (2022) 'Phloem-specific localization of benzylisoquinoline alkaloid metabolism in opium poppy', *Journal of Plant Physiology*, 271, p. 153641. Available at: <https://doi.org/10.1016/j.jplph.2022.153641>.
- Nguyen, T.-A.M., Nguyen, T.-D., Leung, Y.Y., McConnachie, M., Sannikov, O., Xia, Z. and Dang, T.-T.T. (2021) 'Discovering and harnessing oxidative enzymes for chemoenzymatic synthesis and diversification of anticancer camptothecin analogues', *Communications Chemistry*, 4(1), p. 177. Available at: <https://doi.org/10.1038/s42004-021-00602-2>.
- Pauw, B., Hilliou, F.A.O., Martin, V.S., Chatel, G., de Wolf, C.J.F., Champion, A., Pré, M., van Duijn, B., Kijne, J.W., van der Fits, L. and Memelink, J. (2004) 'Zinc finger proteins act as transcriptional repressors of alkaloid biosynthesis genes in *Catharanthus roseus*', *The Journal of Biological Chemistry*, 279(51), pp. 52940–52948. Available at: <https://doi.org/10.1074/jbc.M404391200>.
- Peina Zhou, Hongyu Chen, Jingjie Dang, Zunrui Shi, Yongfang Shao, Chanchan Liu, Longjiang Fan, and Qinan Wu (2022) 'Single-cell transcriptome of *Nepeta tenuifolia* leaves reveal differentiation trajectories in glandular trichomes', *Frontiers in Plant Science*, 13, p. 988594. Available at: <https://doi.org/10.3389/fpls.2022.988594>.
- R. Keith Slotkin and Robert Martienssen (2007) 'Transposable elements and the epigenetic regulation of the genome', *Nature Reviews. Genetics*, 8(4), pp. 272–285. Available at: <https://doi.org/10.1038/nrg2072>.
- Rahul Shaw, Xin Tian, and Jian Xu (2021) 'Single-Cell Transcriptome Analysis in Plants: Advances and Challenges', *Molecular Plant*, 14(1), pp. 115–126. Available at: <https://doi.org/10.1016/j.molp.2020.10.012>.
- Rayan, A., Raiyn, J. and Falah, M. (2017) 'Nature is the best source of anti-cancer drugs: Indexing natural products for their anticancer bioactivity', *PLoS One*, 12(11), p. e0187925. Available at: <https://doi.org/10.1371/journal.pone.0187925>.
- Robert E. Pruitt, Jean-Philippe Vielle-Calzada, Sara E. Ploense, Ueli Grossniklaus, and Susan J. Lolle (2000) 'FIDDLEHEAD, a gene required to suppress epidermal cell interactions in *Arabidopsis*, encodes a putative lipid biosynthetic enzyme', *Proceedings of the National Academy of Sciences*, 97(3), pp. 1311–1316. Available at: <https://doi.org/10.1073/pnas.97.3.1311>.
- Rocío Rodríguez-Quiroz and Braulio Valdebenito-Maturana (2022) 'SoloTE for improved analysis of transposable elements in single-cell RNA-Seq data using locus-specific expression', *Communications Biology*, 5(1), p. 1063. Available at: <https://doi.org/10.1038/s42003-022-04020-5>.
- S. Sato, Y. Nakamura, T. Kaneko, E. Asamizu, and S. Tabata (1999) 'Complete structure of the chloroplast genome of *Arabidopsis thaliana*', *DNA research: an international journal for rapid publication of reports on genes and genomes*, 6(5), pp. 283–290. Available at: <https://doi.org/10.1093/dnare/6.5.283>.
- Sadre, R., Magallanes-Lundback, M., Pradhan, S., Salim, V., Mesberg, A., Jones, A.D. and DellaPenna, D. (2016) 'Metabolite Diversity in Alkaloid Biosynthesis: A Multilane (Diastereomer) Highway for Camptothecin Synthesis in *Camptotheca acuminata*', *The Plant Cell*, 28(8), pp. 1926–1944. Available at: <https://doi.org/10.1105/tpc.16.00193>.
- Salim, V., Jones, A.D. and DellaPenna, D. (2018) 'Camptotheca acuminata 10-hydroxycamptothecin O-methyltransferase: an alkaloid biosynthetic enzyme co-opted from flavonoid metabolism', *The Plant Journal: For Cell and Molecular Biology*, 95(1), pp. 112–125. Available at: <https://doi.org/10.1111/tpj.13936>.
- Sarah E. Johnstone, Vadim N. Gladyshev, Martin J. Aryee, and Bradley E. Bernstein (2022) 'Epigenetic clocks, aging, and cancer', *Science (New York, N.Y.)*, 378(6626), pp. 1276–1277. Available at: <https://doi.org/10.1126/science.abn4009>.
- Shujun Ou, Weija Su, Yi Liao, Kapeel Chougule, Jireh R. A. Agda, Adam J. Hellenga, Carlos Santiago Blanco Lugo, Tyler A. Elliott, Doreen Ware, Thomas Peterson, Ning Jiang, Candice N. Hirsch, and Matthew B. Hufford (2019) 'Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline', *Genome Biology*, 20(1), p. 275. Available at: <https://doi.org/10.1186/s13059-019-1905-y>.
- Sijie Sun, Xiaofeng Shen, Yi Li, Ying Li, Shu Wang, Rucan Li, Huibo Zhang, Guoan Shen, Baolin Guo, Jianhe Wei, Jiang Xu, Benoit St-Pierre, Shilin Chen, and Chao Sun (2023) 'Single-cell RNA sequencing provides a high-resolution roadmap for understanding the multicellular compartmentation of specialized metabolism', *Nature Plants*, 9(1), pp. 179–190. Available at: <https://doi.org/10.1038/s41477-022-01291-y>.
- Sophie Lanciano and Gael Cristofari (2020) 'Measuring and interpreting transposable element expression', *Nature Reviews. Genetics*, 21(12), pp. 721–736. Available at: <https://doi.org/10.1038/s41576-020-0251-y>.
- Suppitanta, N., Pattanaik, S., Kulshrestha, M., Patra, B., Singh, S.K. and Yuan, L. (2011) 'The transcription factor CrWRKY1 positively regulates the terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*', *Physiologia Plantarum*, 151(1), pp. 103–113.

- Plant Physiology, 157(4), pp. 2081–2093. Available at: <https://doi.org/10.1104/pp.111.181834>.
- Tenorio Berrio, R., Verstaen, K., Vandamme, N., Pevernagie, J., Achon, I., Van Duyse, J., Van Isterdael, G., Saeys, Y., De Veylder, L., Inzé, D. and Dubois, M. (2022) ‘Single-cell transcriptomics sheds light on the identity and metabolism of developing leaf cells’, *Plant Physiology*, 188(2), pp. 898–918. Available at: <https://doi.org/10.1093/plphys/kiab489>.
- Tian-Qi Zhang, Yu Chen, and Jia-Wei Wang (2021) ‘A single-cell analysis of the *Arabidopsis* vegetative shoot apex’, *Developmental Cell*, 56(7), pp. 1056–1074.e8. Available at: <https://doi.org/10.1016/j.devcel.2021.02.021>.
- Wang, Q., Wu, Y., Peng, A., Cui, J., Zhao, M., Pan, Y., Zhang, M., Tian, K., Schwab, W. and Song, C. (2022) ‘Single-cell transcriptome atlas reveals developmental trajectories and a novel metabolic pathway of catechin esters in tea leaves’, *Plant Biotechnology Journal*, 20(11), pp. 2089–2106. Available at: <https://doi.org/10.1111/pbi.13891>.
- Xiang Pu, Minji Wang, Menghan Chen, Xinyu Lin, Ming Lei, Jiahua Zhang, Shengnan Yang, Hanguang Wang, Jinqiu Liao, Li Zhang, and Qianming Huang (2023) ‘Proteomics-Guided Mining and Characterization of Epoxidase Involved in Camptothecin Biosynthesis from *Camptotheca acuminata*’, *ACS chemical biology*, 18(8), pp. 1772–1785. Available at: <https://doi.org/10.1021/acschembio.3c00222>.
- Xiaojie Qiu, Qi Mao, Ying Tang, Li Wang, Raghav Chawla, Hannah A. Pliner, and Cole Trapnell (2017) ‘Reversed graph embedding resolves complex single-cell trajectories’, *Nature Methods*, 14(10), pp. 979–982. Available at: <https://doi.org/10.1038/nmeth.4402>.
- Xinbo Chen, S. Mark Goodwin, Virginia L. Boroff, Xionglun Liu, and Matthew A. Jenks (2003) ‘Cloning and characterization of the WAX2 gene of *Arabidopsis* involved in cuticle membrane and wax production’, *The Plant Cell*, 15(5), pp. 1170–1185. Available at: <https://doi.org/10.1105/tpc.010926>.
- Yu, C., Hou, K., Zhang, H., Liang, X., Chen, C., Wang, Z., Wu, Q., Chen, G., He, J., Bai, E., Li, X., Du, T., Wang, Y., Wang, M., Feng, S., Wang, H. and Shen, C. (2023) ‘Integrated mass spectrometry imaging and single-cell transcriptome atlas strategies provide novel insights into taxoid biosynthesis and transport in *Taxus mairei* stems’, *The Plant Journal: For Cell and Molecular Biology* [Preprint]. Available at: <https://doi.org/10.1111/tpj.16315>.
- Yuhan Hao, Stephanie Hao, Erica Andersen-Nissen, William M. Mauck, Shiwei Zheng, Andrew Butler, Maddie J. Lee, Aaron J. Wilk, Charlotte Darby, Michael Zager, Paul Hoffman, Marlon Stoeckius, Efthymia Papalexi, Eleni P. Mimitou, Jaison Jain, Avi Srivastava, Tim Stuart, Lamar M. Fleming, Bertrand Yeung, Angela J. Rogers, Juliana M. McElrath, Catherine A. Blish, Raphael Gottardo, Peter Smibert, and Rahul Satija (2021) ‘Integrated analysis of multimodal single-cell data’, *Cell*, 184(13), pp. 3573–3587.e29. Available at: <https://doi.org/10.1016/j.cell.2021.04.048>.
- Yun Yang, Wei Li, Jing Pang, Liangzhen Jiang, Xixing Qu, Xiang Pu, Guolin Zhang, and Yinggang Luo (2019) ‘Bifunctional Cytochrome P450 Enzymes Involved in Camptothecin Biosynthesis’, *ACS chemical biology*, 14(6), pp. 1091–1096. Available at: <https://doi.org/10.1021/acschembio.8b01124>.
- Zhang, H., Shen, X., Sun, S., Li, Y., Wang, S., Wei, J., Guo, B. and Sun, C. (2023) ‘Integrated transcriptome and proteome analysis provides new insights into camptothecin biosynthesis and regulation in *Camptotheca acuminata*’, *Physiologia Plantarum*, 175(3), p. e13916. Available at: <https://doi.org/10.1111/ppl.13916>.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Wang, S., Zhang, C., Li, Y., Li, R., Du, K., Sun, C. et al. (2024) ScRNA-seq reveals the spatiotemporal distribution of camptothecin pathway and transposon activity in *Camptotheca acuminata* shoot apices and leaves. *Physiologia Plantarum*, 176(5), e14508. Available from: <https://doi.org/10.1111/ppl.14508>