

Single-cell transcriptome atlas reveals developmental trajectories and a novel metabolic pathway of catechin esters in tea leaves

Qiang Wang^{1,†}, Yi Wu^{1,†}, Anqi Peng¹, Jilai Cui^{1,2}, Mingyue Zhao¹, Yuting Pan¹, Mengting Zhang¹, Kai Tian³ , Wilfried Schwab^{1,4,*} and Chuankui Song^{1,*} 

¹State Key Laboratory of Tea Plant Biology and Utilization, International Joint Laboratory on Tea Chemistry and Health Effects, Anhui Agricultural University, Hefei, Anhui, China

²Key Laboratory of Tea Plant Biology of Henan Province, College of Life Science, Xinyang Normal University, Xinyang, Henan, China

³Key Laboratory of Ecological Security for Water Source Region of Mid-Line Project of South-To-North Diversion Project of Henan Province, School of Life Sciences and Agricultural Engineering, Nanyang Normal University, Nanyang, China

⁴Biotechnology of Natural Products, Technische Universität München, Freising, Germany

Received 26 May 2022;

revised 3 July 2022;

accepted 5 July 2022.

*Correspondence (Tel +86-551-65786065; Fax +86-551-65780360; email sckfriend@163.com (CS) and Tel +49 8161 71 2912; Fax +49 8161 71 2950; email wilfried.schwab@tum.de (WS))

†These authors contributed equally to this work.

Keywords: catechin, scRNA-seq, development, leaf, glycosyltransferase, tea plant.

Summary

The tea plant is an economically important woody beverage crop. The unique taste of tea is evoked by certain metabolites, especially catechin esters, whereas their precise formation mechanism in different cell types remains unclear. Here, a fast protoplast isolation method was established and the transcriptional profiles of 16 977 single cells from 1st and 3rd leaves were investigated. We first identified 79 marker genes based on six isolated tissues and constructed a transcriptome atlas, mapped developmental trajectories and further delineated the distribution of different cell types during leaf differentiation and genes associated with cell fate transformation. Interestingly, eight differently expressed genes were found to co-exist at four branch points. Genes involved in the biosynthesis of certain metabolites showed cell- and development-specific characteristics. An unexpected catechin ester glycosyltransferase was characterized for the first time in plants by a gene co-expression network in mesophyll cells. Thus, the first single-cell transcriptional landscape in woody crop leave was reported and a novel metabolism pathway of catechin esters in plants was discovered.

Introduction

Horticultural crops are domesticated to meet various human needs, including foods and beverages, energy, materials and ornamental plant (Xiong *et al.*, 2015). They also contain abundant natural compounds with high economic value, which are derived from plant secondary metabolism (Kabera, 2014). The tea plant (*Camellia sinensis* (L.) O. Ktze.), one of the world's most important woody cultivated crops, whose leaves are used to produce numerous kinds of tea, provides numerous secondary metabolites that account for its rich taste and health benefits (Xia *et al.*, 2019). Current studies have shown that the leaf development of tea tree is accompanied by regular changes of secondary metabolites, including compounds characteristic of tea such as polyphenols, caffeine and theanine (Li *et al.*, 2015; Xia *et al.*, 2019; Yu *et al.*, 2021; Zhao *et al.*, 2020). Although plant secondary metabolism is intimately related to development and environment, and the relationships between them at organ and tissue level have been studied extensively, their biosynthesis and distribution in different type cells was rarely reported (De Filippis, 2016).

Advances in single-cell RNA sequencing (scRNA-Seq) have now made it possible to sequence the transcriptome of rare cells with small amounts of starting material. Based on the specific unique molecular identification of single-cell, scRNA-seq enable monitoring of intracellular transcriptional activity at the single-cell level (Seyffert *et al.*, 2021). Currently, scRNA-seq of several plant

species, such as *Arabidopsis*, strawberry, rice and maize, peanuts, and one woody plant *populus* were reported (Bai *et al.*, 2022; Chen *et al.*, 2021; Kim *et al.*, 2021; Li *et al.*, 2021; Liu *et al.*, 2020, 2021; Marand *et al.*, 2021; Xie *et al.*, 2021; Zhang *et al.*, 2021a,b). Through these studies, a database of cell-specific markers and a single-cell atlas of various tissues and organs from different species has been established, providing new insights into organ development, cell differentiation and cell division (Jin *et al.*, 2021). However, gene function identification of non-model species often fails to reach the level of cell specificity, because of the lack of reliable specific marker genes for cell type identification. In addition, due to the bottleneck of cell sizes and protoplast isolation efficiency, its application in non-model plants is still limited.

Because of the thick cell wall and abundant secondary metabolism, woody plant tissues are hard to digest to release protoplasts (Barros *et al.*, 2015; Butt, 1985; Shen *et al.*, 2022). Moreover, the relatively long growth cycles and lack of effective transgenic technologies to validate the spatial and temporal expression profiles of genes have resulted in few cell type-specific marker genes being detected in the non-model woody plants. This exacerbates the difficulty of cell type identification and largely explains why most scRNA-seq analyses were done in model plants (Jin *et al.*, 2021; Liu *et al.*, 2021; Seyffert *et al.*, 2021; Zhang *et al.*, 2021a). Leaf is one of the principal vegetative organs of vascular plants for photosynthesis and transpiration. Until now, only a stem cell atlas of *populus* has

been reported for woody plants; scRNA-seq applications of woody leaves are not yet available (Chen *et al.*, 2021; Li *et al.*, 2021; Xie *et al.*, 2021). Compared with the leaf structure of herbaceous plants, tea leaves have a well-developed secondary cell wall, strong vascular bundles and a thick wax layer etc. Leaf growth and biosynthesis of characteristic secondary metabolites are critical for tea production and quality control and play an important role in plant protection (Lin *et al.*, 2022; Xia *et al.*, 2017; Xu *et al.*, 2016). Thus, tea leaves are a suitable model for studying secondary metabolites (Zeng *et al.*, 2020). Dissecting the developmental trajectories and the biosynthetic networks that give rise to the characteristic secondary metabolites could help discover new genes involved in their formation in woody plant (Falcone Ferreyra *et al.*, 2012; Sugiyama *et al.*, 2016; Xia *et al.*, 2019).

Here, we constructed the first single-cell level atlas of woody tea plant by protoplast isolation, single-cell sequencing and cell-specific tissue isolation. In addition, specific marker genes for each cluster and cell type of tea leaves were identified. Based on precise spatiotemporal transcripts, we delineate developmental trajectories of tea tender leaf cells and infer gene expression signatures associated with cell fate decisions. Interestingly, an unexpected uridine diphosphate-dependent glycosyltransferase (UGT) involved in the glycosylation of catechin esters, the major contributor of bitterness and astringency of tea, was identified *in vitro* and *in vivo* for the first time, reflecting the complexity of catechin ester metabolism in plants. Overall, we provide an example of the application of scRNA-seq in non-model woody plants and reveal insights into the complex developmental regulation of catechin esters metabolism in plants.

Results

Tea leaf structure and cell type identification

To assist the identification of cell types for scRNA-Seq, sections of tea plant leaves were examined using different approaches. Cross-cut slices of the main veins of the tea leaf were viewed under the microscope with visible light. Tea leaves show upper epidermis (UE) cells (epidermal pavement cells, Figure 1a–d, S1a, b, d) and lower epidermis (LE) cells (epidermal pavement cells, guard cells and non-glandular trichome (NGT) cells, Figure 1e, f, S1a–c), mesophyll cells (palisade (PM) and spongy mesophyll (SM) cells, Figure 1a–d, S1a, b); vascular bundle (VB) cells (includes xylem (XY), phloem (PH), vascular bundle sheath, parenchyma cells (PCC) and stone cells, Figure 1a–d, S1a). Only the lower epidermis has stomata and glandular hairs. While glandular hair density and stomata density are high and low, respectively, for 1 L, the opposite was observed for 3 L (Figure 1e, f). Besides, theanine, catechins (GCG and EGCG) and caffeine content in 1 L was higher than in 3 L (Figure 1g). In addition, section

observation provided many intuitive bases for cell type identification: trichome cells and stone cells will be filtered and removed by 40 µm sieve (Figure 1h, S1b). Epidermal cells and mature xylem cells have few green chloroplasts, but mesophyll cells have numerous chloroplasts that make them look dark green (Figure 1a–d). Xylem cells, cells of the vascular bundle sheath and epidermal cells all have thickened cell walls and are rich in lignin (Figure 1h).

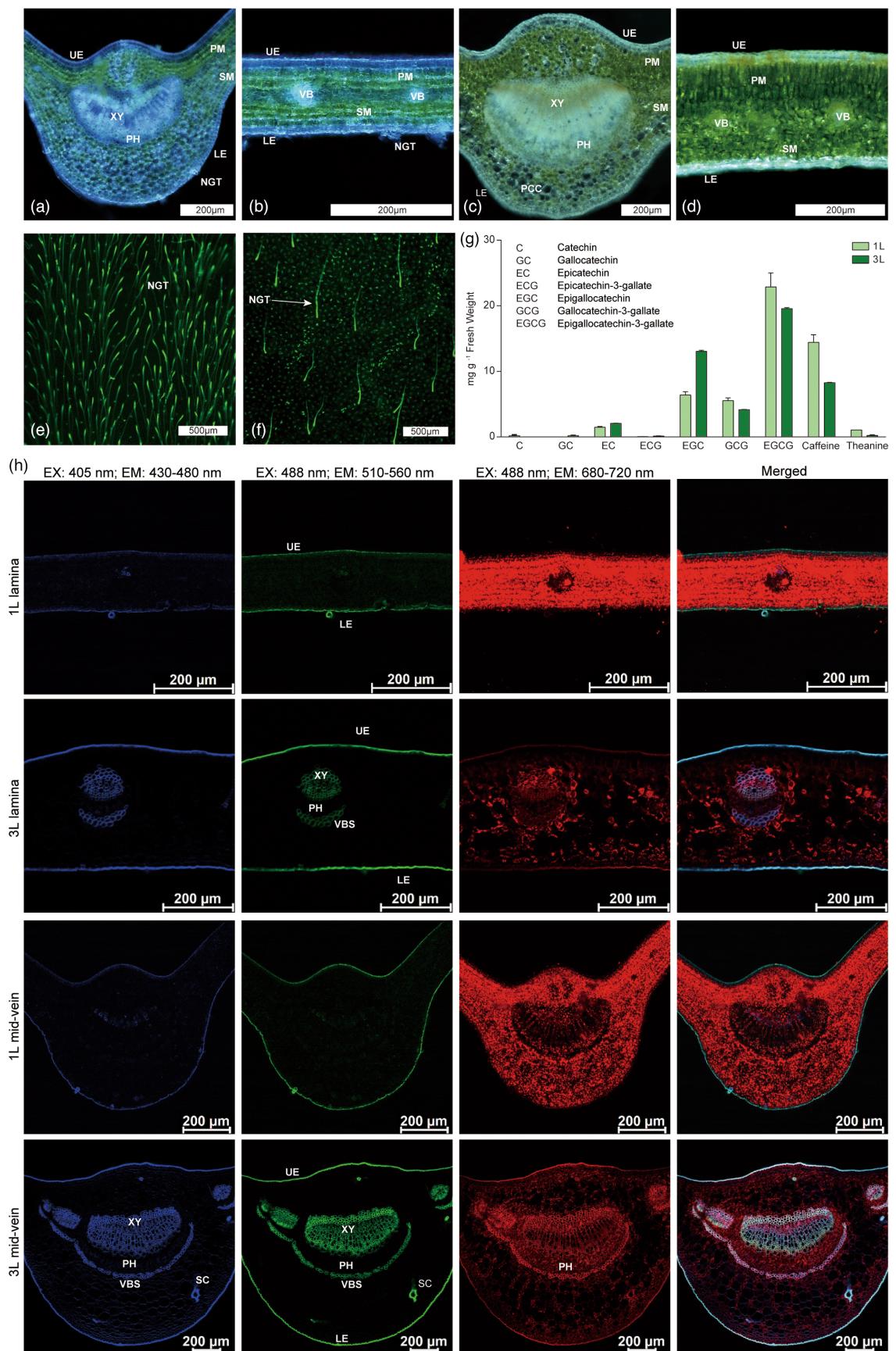
Protoplast fast isolation and single cell transcriptome sequencing

To reduce the effect of protoplasting treatment, we added snailase to shorten the digestion time from 16 to 4–6 h. The ratio of viable cells to total cells for each sample was from 73% to 82%. Approximately, $3.5\text{--}6.6 \times 10^6$ protoplasts per gram were isolated for each sample. Compared with the recently reported study on protoplasts isolation of tea cultivar 'Shuchazao', which maximal cell viability was 92.94% and the highest protoplast yield was 3.27×10^6 protoplasts per gram, our approach showed lower viability, higher cell yields and shorter digest time (Xu *et al.*, 2021). Microscopic examination revealed an abundance of cell types (Figure 2). The isolated protoplasts were introduced into a commercial 10× Chromium platform for scRNA-seq assays (Figure 2). Over 95% valid barcodes from a total of 18 395 cells were captured, a median 1476 (1 L) and 1696 (3 L) number of genes and a median 2530 (1 L) and 3568 (3 L) unique molecular identifier (UMI) counts per cell were detected. After filtering the cells, we generated a gene expression matrix consisted by 34 729 genes across 9458 filtered cells from the 1 L and 32 311 genes across 7519 filtered cells from the 3 L. Single cell transcriptome-related quality control parameters are detailed in a statistical table (Table S1).

Tea leaves cell atlas generation

Based on the results of cross-sectional observation, we drew the structure of a tea leaf to provide clues for the identification of cell type clusters (Figure 3a). We used standard computational pipelines to align the raw sequencing data to the tea plant genome, and after principal component analysis (PCA) and Shared Nearest Neighbour (SNN) clustering, a total of 16 977 cells were partitioned into 16 transcriptionally distinct clusters (Figure 3b, Table S2). Cluster 2, 4 and 7 were found only in 1 L, while cluster 5 and 10 were found in 3 L (Figure 3b, Figure S2, Table S2). To reveal local similarities and global structures of cell populations, the t-distributed stochastic neighbourhood embedding (tSNE) tool was used which reveals the cell distribution of clusters, samples and later cell types (Figure 3c, Figure S2). Then, we filtered out the protoplasting response genes according a published bulk RNA-Seq dataset (Jing *et al.*, 2019). The six-hour digestion affected the expression of over 730 genes, with clusters 0–3 being the most affected (Table S3). Some of

Figure 1 Cell characters of the 1st and 3rd leaves of *Camellia sinensis* var Shuchazao. (a) The mid-vein cross section of the 1st leaf. Bar = 200 µm. (b) The leaf lamina cross section of the 1st leaf. VB, vascular bundle. (c) The mid-vein cross section of the 3rd leaf. (d) The leaf lamina cross section of the 3rd leaf. (e) The lower epidermis of 3rd leaf. Bar = 500 µm. (f) The upper epidermis of 3rd leaf. (g) Contents of main catechins, caffeine and theanine in the 1st and 3rd leaf. (h) Autofluorescent images of the lamina and mid-vein section of 1st leaf and 3rd leaf. Abbreviation: EM, emission light wavelength; EX, excitation light wavelength LE, lower epidermis; NGT, non-glandular trichome; PH, phloem; PM, palisade mesophyll; SM, spongy mesophyll; UE, upper epidermis; VBS, vascular bundle sheath; XY, xylem.



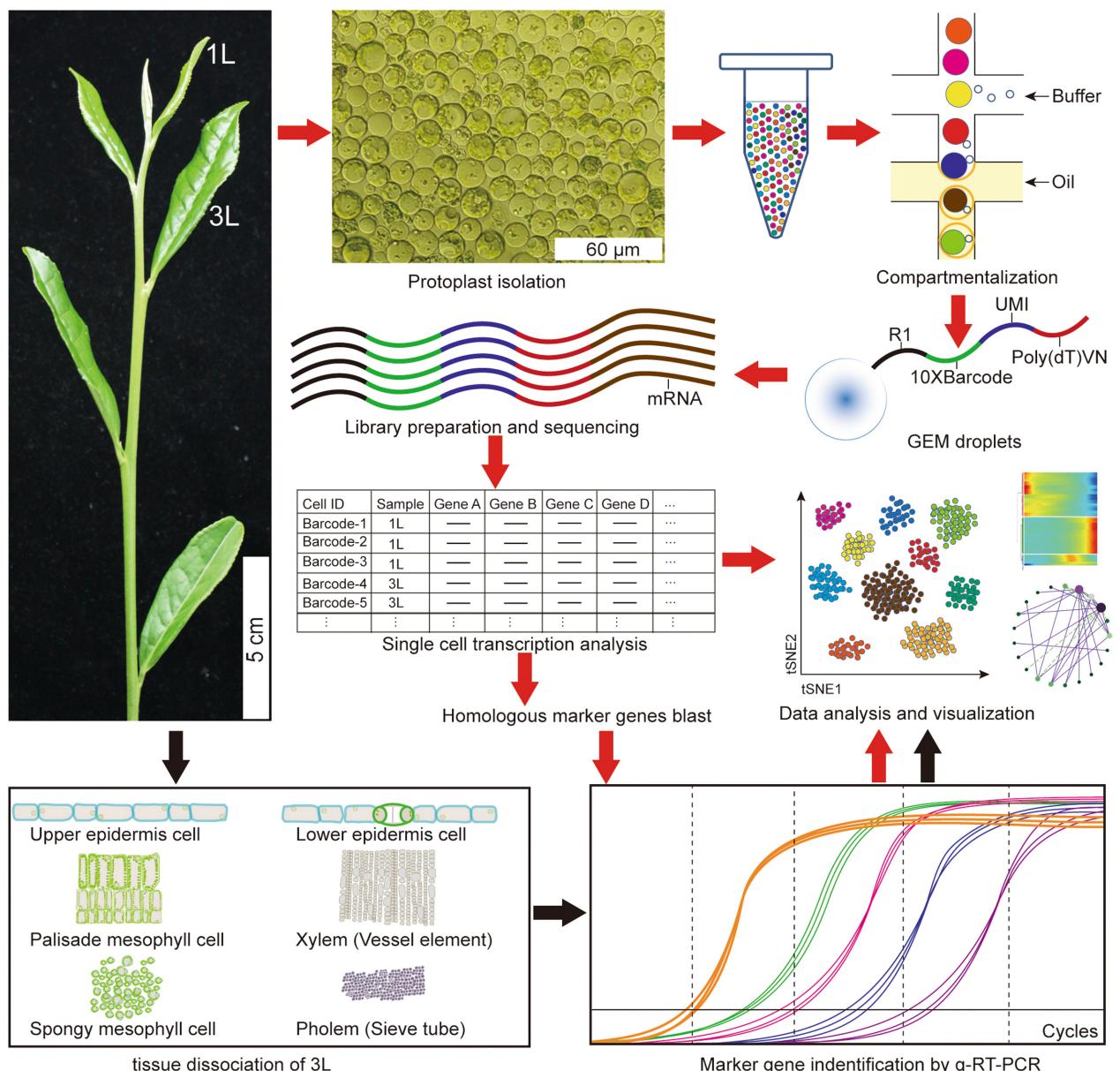


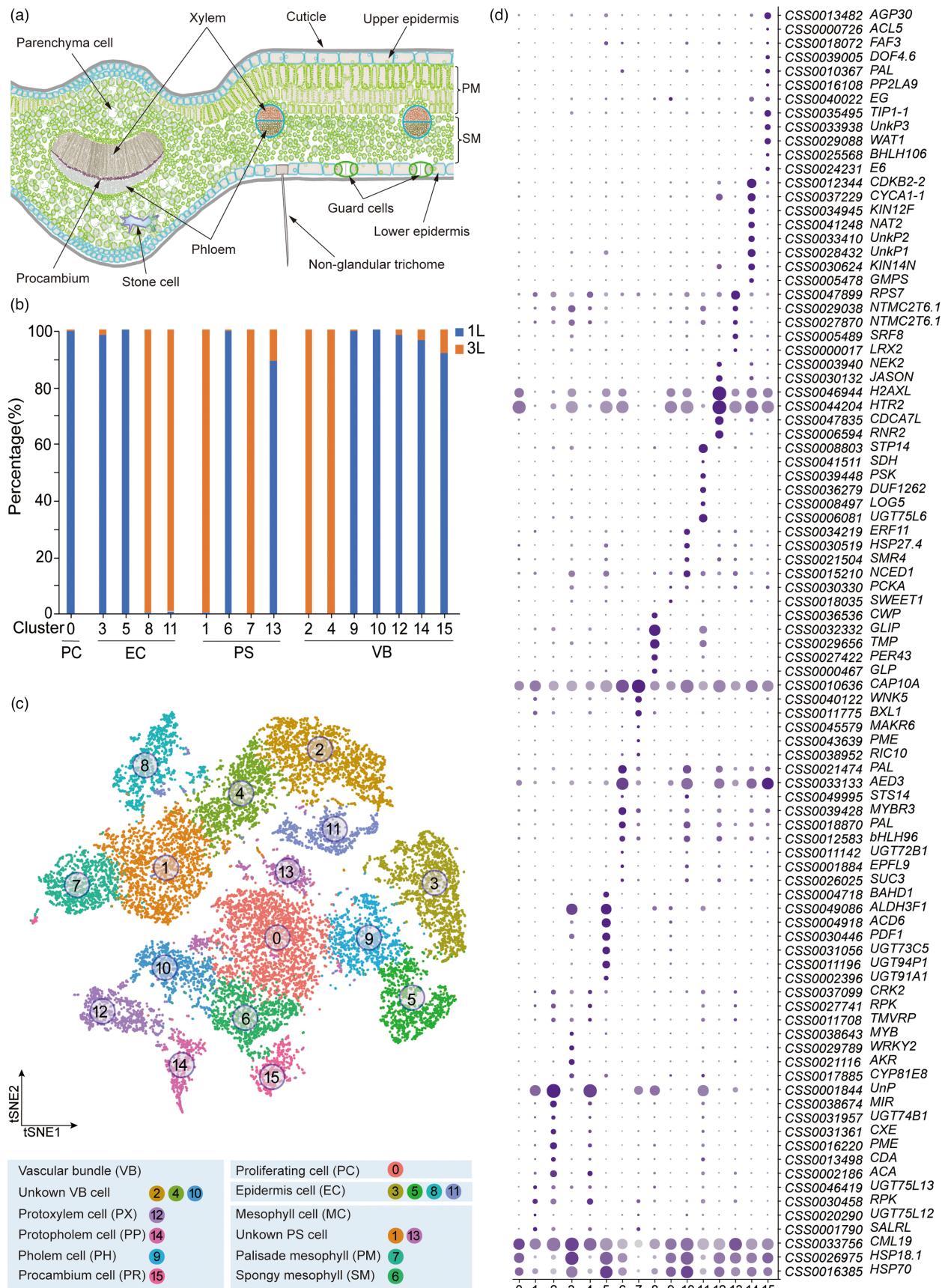
Figure 2 Brief flowchart of tea tender leaves scRNA-seq. The 1st and 3rd leaves of young shoot of *C. sinensis* var Shuchazao, representing 1 L and 3 L, respectively, were used for single-cell transcriptomics. Protoplasts were obtained by enzymatic hydrolysis. mRNA released by rupture of a single-cell suspension was combined with gel bead and emulsion to form GEMs [10× Genomics Chromium Single-Cell 3' kit (V3)]. The mRNA of the cell was independently reverse-transcribed in each GEM, and tagged cDNA was mixed and amplified for library construction. Libraries were sequenced on an Illumina NovaSeq 6000 sequencing system (paired-end multiplexing run, 150 bp) by LC-Bio Technology co. Ltd., (Hangzhou, China). Six tissues of 3 L were used for cDNA preparation to identify cluster-specific genes expression pattern by qRT-PCR. Red arrow: scRNA-seq analysis; black arrow: experiment validation.

them have already been detected in the tea plant and *Arabidopsis* when they respond to stress (Table S4) (Liu et al., 2020; Sun et al., 2021). Finally, we obtained 4,327 cluster-specific genes (CSGs) in 16 cell clusters (Table S5). Ninety-one highly and uniquely CSGs in each cluster were

displayed in a dot plot, including 12 homologous genes of *Arabidopsis* marker genes (Figure 3d, Table S8).

Because no cell-specific marker genes with specific biological functions or expression patterns have been well studied in tea plants at single-cell level, we used the following two strategies to

Figure 3 Overviews of the cell atlas of tender tea leaves. (a) Anatomy and cell types of tea 3 L. PM: palisade mesophyll cells; SM: spongy mesophyll cells. (b) Cell distribution in each cluster of 1 L and 3 L. (c) Visualization of 16 cell clusters using tSNE. Dots, individual cells; $n = 16\ 977$ cells; colour, cell clusters. (d) Expression pattern of representative cluster-specific marker genes. Dot diameter, proportion of cluster cells expressing a given gene. The detail information of selected genes is given in Table S8.



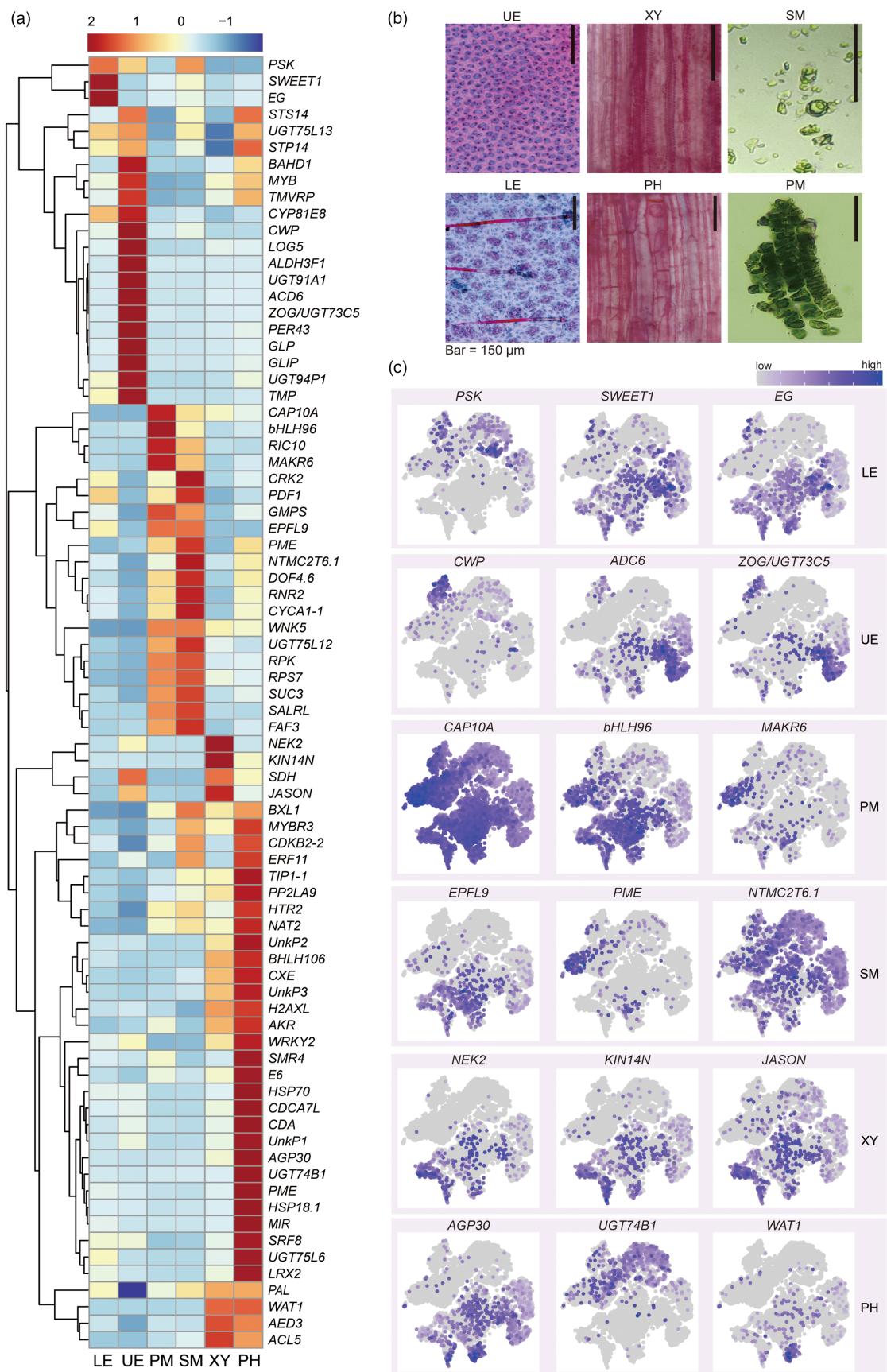


Figure 4 Marker genes identification by qRT-PCR of 6 isolated tissues from 3 L. (a) Heatmap of verified marker gene. The original expression data was given in the Table S9. (b) Identified cell types of 6 isolated tissues under the light microscope. Upper epidermis (UE), lower epidermis (LE), xylem (XY) and phloem and procambium (PH) were stained with safranin O/fast green, which stained lignified cells red and cytoplasm green. Spongy mesophyll cells (SM) and palisade mesophyll cells (PM) were isolated according to published methods and observed directly under the light microscope. (c) tSNE plots with the expression of selected marker genes for six tissues.

faithfully annotate cell clusters in the cell atlas of tea leaf. First, we used characteristic plant single-cell markers to search for homologous genes in the tea plant. Only few markers were found because of poor sequence similarity or no hits in CSGs of all clusters (Table S6) (Kim et al., 2021; Liu et al., 2020; Zhang et al., 2021b). Then we annotated the CSGs of each cluster with published scRNA-seq data in Arabidopsis and the KEGG pathway annotation (Table S7) (Jin et al., 2021). This provided us with a predicted cell type identification output. Second, we isolated six different types of cells from the 3 L of tea plants, which were upper epidermis (UE), lower epidermis (LE), xylem of main veins (XY), phloem and procambium of main veins (PH), palisade mesophyll (PM) and spongy mesophyll (SM) cells (Figure 4b). The above six cell types were subjected to the same digestion time to ensure identical protoplasting. Then we extracted the total RNA and after reverse transcription q-RT-PCR was performed for marker gene identification.

Sequencing the q-RT-PCR products to determine sequence specificity and filtering out genes with low expression level ($\Delta\Delta C_t < 0.0001$, relative to the housekeeping gene) identified 78 valid genes which were presented in a hierarchical clustering heatmap (Figure 4a, Table S9). Overall, the six tissues all have their own specific marker genes, and genes that are expressed simultaneously in two or more cell types (Figure 4a). A *phytosulfokines* (CSS0039448, PSK), *endoglucanase* (CSS0040022, EG) and a *bidirectional sugar transport SWEET* (CSS0018035, SWEET1) were highly expressed only in LE, indicating that they are probably involved in osmotic pressure regulation in guard cells (Daloiso et al., 2016). The tSNE plot showed that PSK was highly expressed in the 3 L, and the other two genes were highly transcribed in the 1 L. In UE, more than 10 markers were found, including three *UDP-glycosyltransferases* and a *predicted cytokinin riboside 5'-monophosphate phosphoribohydrolase LOG5*. Stress response genes and *GDSL esterase/lipase* (CSS0032332, GLIP) were highly expressed in epidermis cells. The marker genes in PM and SM were similar, but their expression levels were much higher than those in other cell types. The *CAP10A* marker gene showed a higher transcript abundance in PM than in SM (Figure 4a). The *Protodermal factor 1 like* gene is an epidermis cell marker of Arabidopsis (Zhang et al., 2021b). Its homologue gene (CSS0030446, PDF1) in tea leaves had the highest expression level in SM, followed by UE (Figure 4a). Two cell cycle regulated proteins, *serine/threonine-protein kinase Nek2* (CSS0003940, Nek2) and *Kinesin-like protein* (CSS0030624, KIN14N) were specially expressed in XY (Garcia-Hernandez et al., 2002; Quan et al., 2008). One oxidoreductase (CSS0041511, SDH) involved in lignin biosynthesis was predominantly expressed in XY and UE (Moinuddin et al., 2006; Xia et al., 2001). The thermospermone synthase is involved in Arabidopsis xylem specification through the prevention of premature cell death (Muñiz et al., 2008). Its homologue gene (CSS0000726, ACL5) was highly accumulated in XY and PH. In this study, the PH probably includes phloem, procambium and part of the vascular bundle sheath. *PHLOEM PROTEIN 2-like* (CSS0016108, PP2LA9) was highly expressed in PH, one of the

most abundant enigmatic proteins in sieve elements and companion cells (Dinant et al., 2003). In addition, several cell cycle-related proteins were also detected specially in PH, such as CSS0012344, CSS0021504 and CSS0047835 (Boudolf et al., 2001; Kumar et al., 2015; Vandepoele et al., 2002). Besides, heat shock proteins, histones and transcription factors also proved to be suitable marker genes for PH. The homologue gene of *WALLS ARE THIN LIKE* of Arabidopsis (CSS0029088, WAT1), associated with developing xylem vessels and fibres, was preferentially in both XY and PH (Ranocha et al., 2010). This is consistent with the xylem and vascular bundle sheath fibre observed in the vascular tissue of the 3 L (Figure 1h). Three marker genes of each tissue were selected for the tSNE plot to show the cell expression distribution (Figure 4c). The expression pattern of the epidermis marker genes showed the best agreement with the tSNE results (Figure 4c).

Based on the above results, we first grouped the 16 clusters into three primary cell types: epidermis cells (EC: cluster 3, 5, 8 and 11), mesophyll cells (MC: cluster 1, 6 and 7), vascular bundle cells (VB: cluster 2, 9, 10, 12, 14 and 15) (Figure 3c). Then by CSGs analysis of each cluster, we defined cluster 0 as proliferating cell (PR), because CSGs of this cluster were predominantly members of the histone superfamily, which is involved in cell division including *HISTONE4* (CSS0031807, HIS4) (Zhang et al., 2021b). CSGs in cluster 4 showed a similar pattern as in cluster 2, so we placed cluster 4 in the VB group (Figure 3d and S3, Table S5) (Vierling, 1991). According to gene expression in six tissues, we classified cluster 13 as unknown mesophyll population (Figure 3d, Figure 4a, Table S9). We applied two Arabidopsis marker gene systems and KEGG&GO annotation to annotate all CSGs of each cluster (Table S5) (Kim et al., 2021; Zhang et al., 2021b). Finally, we defined cluster 9 in combination with the cell character as phloem cell (PH), cluster 12 as protoxylem (PX), cluster 14 as protophloem cell (PP) and cluster 15 as procambium cell (PR) (Figure 1, S1, Table S2). In addition, we defined cluster 6 and 7 as spongy mesophyll (SM) and palisade mesophyll (PM) according the q-RT-PCR results of six tissues, respectively (Figure 4a). Our results show that the single-cell transcriptome atlas allows the identification of most of the major cell types of tea leaves (Figure 3a).

Differentiation trajectories of tea tender leaves

To verify the cell type and explore the continuous differentiation trajectory of tea leaves, we used cells from all clusters to conduct a pseudotime analysis (Figure 5). Because of the biological significance of the samples, we set early-differentiated cells like proliferating cell and procambium cell of the 1 L as the initial point (Figure 5a, b, d). The pseudotime analysis indicated that the pseudotime trajectory has four branch points that divide all cells into 9 states (Figure 5c). Cells capable of division of the 1 L were predominant at the starting point (Figure 5b). Epidermis cells were distributed over all nine states (Figure 5d). PM cells predominately occupied branch 4 (Figure 5c, d). PH, PR, PP, PC and SM of the 1 L exhibited a similar distribution pattern, mainly at state 1 (Figure 5b-d); Epidermis cells, unknown MS and

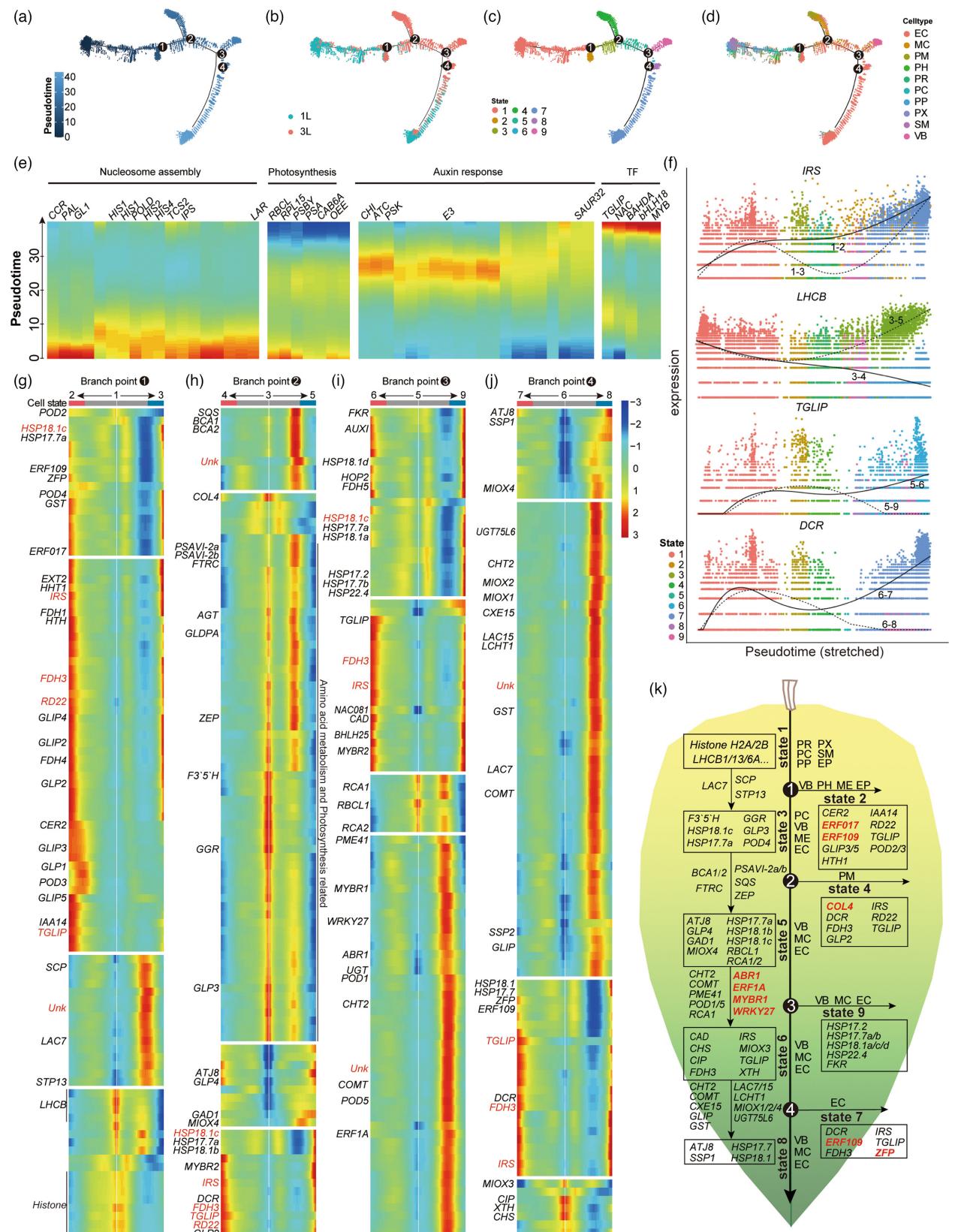


Figure 5 Differentiation trajectory and cell fate decision analysed by pseudotime analysis. (a–d) The cell ordering along the differentiation trajectory successively presented by pseudotime states, samples, branch states and cell types. (e) Heatmap of the top 50 significantly changed genes of 9 cell states. The description of these genes is given in Table S10. (f) Representative genes of four branch points were selected to show their expression trends before and after cell differentiation. (g–j) Heatmap of the top 100 significantly changed genes discovered by the Branched expression analysis modelling (BEAM) function from monocle in four branch points. The detail information of these genes is given in Table S11. (k) Brief chart of cell fate decision in the tea tender leaves.

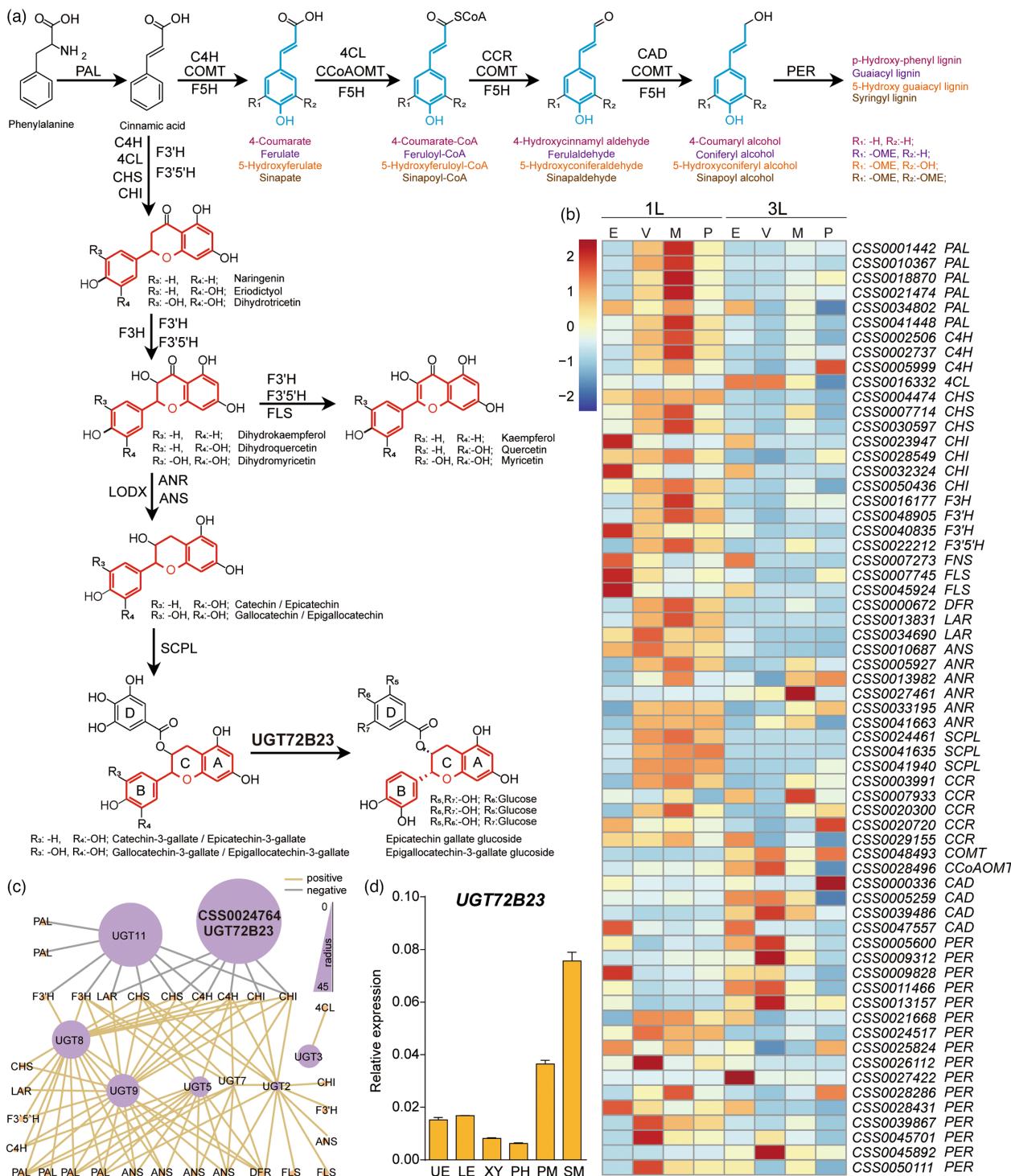


Figure 6 Cell-specific distribution of genes related to the biosynthesis of lignin and catechins, and discovery of novel glycosylated catechins. (a) Schematic diagram of the flavonoids and lignin biosynthesis pathway. Blue and red molecular structures represent backbones of lignin and catechins, respectively. (b) Heatmap of cell-specific genes showing of the lignin and catechin biosynthesis pathway. E: epidermis cell; V: vascular bundle cells; M: mesophyll cells; P: proliferating cell. The original data of the heatmap can be found in Table S12. (c) Co-expression network of UGTs and catechins biosynthesis genes in mesophyll cells. The purple dots present UGTs and the dot radius represents average gene expression level of scRNA-Seq of mesophyll cells. The yellow dots present catechins biosynthesis related genes. The symbols and spearman correlation coefficients data are given in Table S13. (d) q-RT-PCR of *UGT72B23* in 6 isolated tissues of 3 L. *CsUBI* (CSS0007748, ubiquitin-conjugating enzyme) and *CsACTIN* (CSS0008920, actin 7 isoform 1) were used as references.

unknown VB cells showed a similar pattern and were specifically distributed at branch 9 (Figure 5c, d). Further analysis was performed to find out the key genes involved in tea leaf growth

and development. Among the top 50 genes (Figure 5e, Table S9), were genes related to phenylpropanoids and flavonoids biosynthesis (Figure 5e, Figure 6a). BEAM result provided us a list of

significantly changed genes that probably determine or reflect cell fate before and after each of the four branch points (Figure 5f–j). Eight genes occurred during the four branch points, including genes involved in cutin biosynthetic, plant epidermis morphogenesis, wax and suberin biosynthesis and an uncharacterized protein (Figure 5f–j). One representative gene of each branch points was selected to show its expression in different cell states (Figure 5f). A heatmap of the significantly altered genes ($P < 0.01$) detected by the BEAM function of Monocle2 at four branch points shows gene expression patterns when cell fate decisions are made. Based on the above results, we mapped the cellular differentiation and development trajectory of the tender tea leaves and annotated the highly expressed genes of each state and branch point (Figure 5j).

In detail, highly expressed genes of state 1 were mainly histone family and photosynthesis-antenna proteins (Figure 5g). High-low-high fluctuations in the expression of photosynthesis-related genes were observed at both branch points 1 and 2 before differentiation (Figure 5g). This may explain why the parenchyma cells of the 1 L differentiate into palisade mesophyll cells and spongy mesophyll cells of the 3 L. From branch point 2 to branch 4, several genes were identified, including sesquiterpene synthase and zeaxanthin epoxidase (Figure 5h, k). PM cells differentiate specifically at stage 4. Genes highly expressed at this stage include a zinc finger protein involved in the plant circadian rhythm pathway, *FDH3*, *TGLIP*, *IRS*, *RD22* and *DCR* (Figure 5c, d, h, k). In epidermis cell-specific state 7, the ethylene-responsive transcription factor *ERF109-like* and zinc finger *CCCH domain-containing protein* were also highly expressed in addition to the genes in PM (Figure 5c, d, j, k). The corresponding branch point 4 indicates that three *inositol oxygenases* (*MIOX1/2/4*) and 2 *laccases* (*LAC7/15*) probably determine cell fate for specialization to epidermal cells (Figure 5j, k). At branch points 3 and 4, there were a number of highly expressed genes, including genes involved in secondary cell wall formation, such as *chitinase* (*CHT*), *laccase* (*LAC*), *XTH*, *peroxidase* (*POD*), a *pectin esterase* (*PME41*) (Figure 5i–k) (Bonawitz and Chapple, 2010). We also observed that a group of *LMW HSP* genes exhibited distinct expression patterns during the tea leaf development process (Figure 5g–k).

Cell specificity of secondary metabolites biosynthesis in tea leaves and their related genes during leaf development

It is known that the secondary metabolites of tea leaves change regularly during leaf development (Figure 1g) (Li et al., 2022b; Zhao et al., 2020). To further understand whether the biosynthesis of flavonoids, caffeine and theanine is cell-specific during leaf development, we analysed the expression of genes related to their biosynthesis in different cell types. The results showed that genes involved in the formation of flavonoids, caffeine and theanine showed cell specificity in both 1 L and 3 L (Figure 6, S4, Table S11). In particular, in flavonoid biosynthesis pathway, most

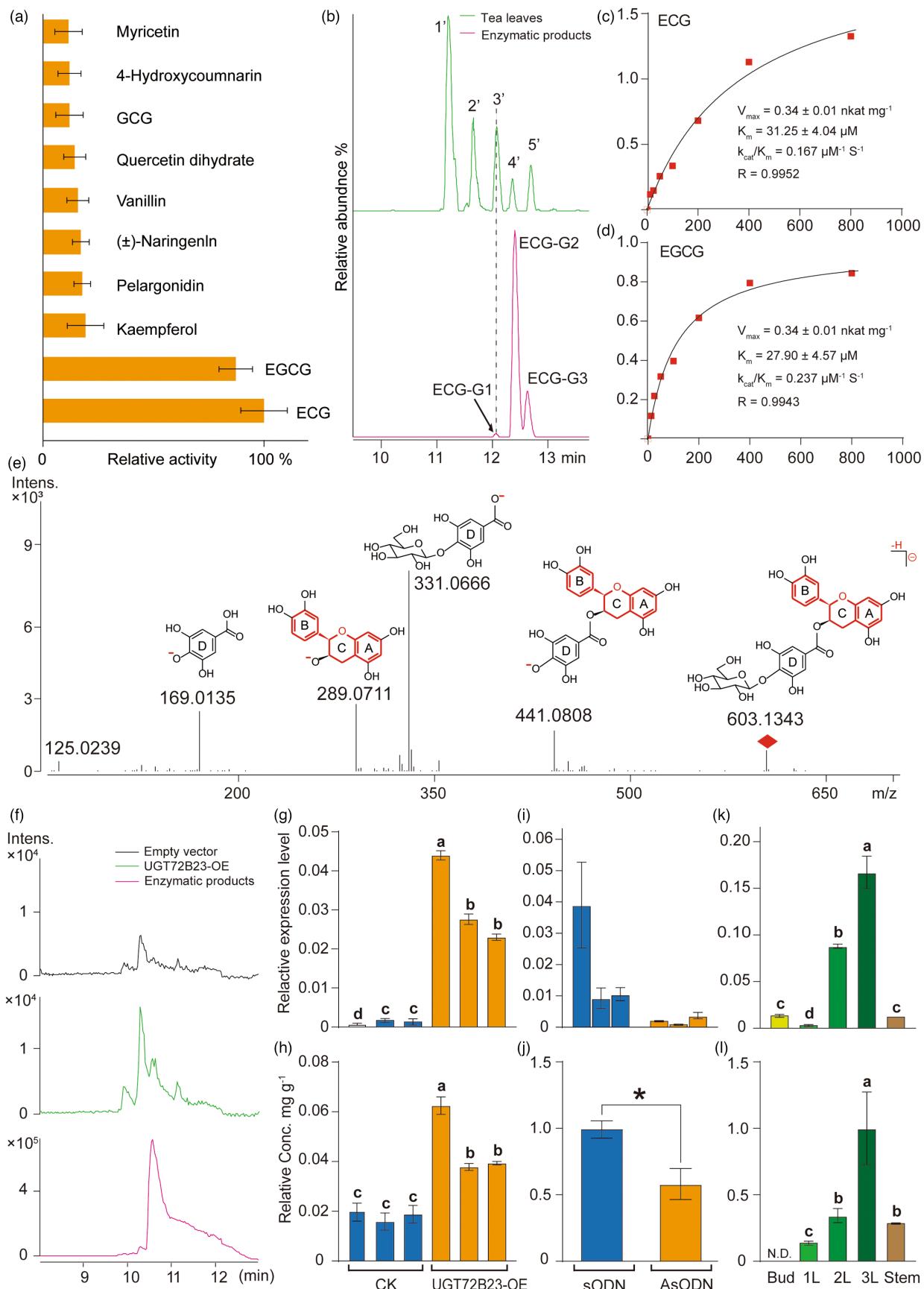
candidate gene were higher expressed in 1 L than that in the 3 L (Figure 6a, b). Five of 6 *PALs* (*phenylalanine ammonia-lyase*), 2 of 3 *C4Hs* (*cinnamate 4-hydroxylase*), 2 of 3 *CHSs* (*chalcone synthase*), *F3H* (*flavanone 3-hydroxylase*), 1 of 2 *F3'Hs* (*flavonoid 3'-hydroxylase*), *F3'5'H* (*flavonoid 3',5'-hydroxylase*), *DFR* (*dihydroflavonol 4-reductase*), 1 of 2 *LARs* (*leucoanthocyanidin reductase*) and 1 of 5 *ANRs* (*anthocyanidin reductase*) were highly expressed in mesophyll cells of 1 L (Figure 6b). Two of 3 *CHIs* (*chalcone isomerase*), 1 of 2 *F3'Hs*, *FNS* (*flavone synthase*), 2 of 2 *FLSs* (*flavonol synthase*) were predominantly expressed in epidermis cells of 1 L and then 3 L (Figure 6b). Transcription of three *Serine carboxypeptidase-like acyltransferases* (*SCPLs*) were rich in 1 L tissues except epidermis cells (Yao et al., 2022). In addition, there are other genes that are expressed cell-specifically in 3 L, such as *C4H* (*CSS0005999*) and *ANR* (*CSS0024761*) (Figure 6b). Genes of the theanine and caffeine biosynthetic pathways were also expressed in a cell-specific manner. Different homologous genes of an enzyme are transcribed in different cell types and at different stages of leaf development (Figure S4). *Glutamate/theanine synthase* (*GS/TS*), *CSS0022197* showed highest expressed in PC of 1 L, while *CSS0013560* showed highest expressed in VB, whereas *CSS0007758* and *CSS0050330* exhibited highest expressed in PC of 3 L (Figure S4). These results demonstrated, for the first time, the heterogeneity of cell types and the spatiotemporal expression of genes involved in secondary metabolism in developing tea plants.

As a woody plant, leaves of the tea plant are rich in lignin, which is one of the main components of secondary cell walls (Figure 1h, S1). To explore the lignin metabolism, the genes involved in lignin biosynthesis, which shares upstream substrates with the flavonoid biosynthesis was also analysed (Figure 6a, b). Results showed that homologous genes of *cinnamoyl-CoA reductase* were highly expressed in MC of 1 L and barely transcribed in 3 L (Figure 6a, b). *COMT* (*caffeinic acid O-methyltransferase*) and *CCoAOMT* (*caffeoyle-CoA-O-methyltransferase*) were highly expressed in VB of 3 L (Figure 6a, b). Four *CADs* (*cinnamyl alcohol dehydrogenase*) were hardly expressed in 1 L, but highly transcribed in EP, VB and PC of 3 L. *PER* (*Peroxidase*) is highly expressed in 3 L vascular and epidermis, but only slightly expressed in the corresponding 1 L cell types (Figure 6a, b). The high expression of these genes improves the mechanical strength of specific cell types, which coincides with the characteristics of the physical support of leaves (VB) and defence (EC). These results indicated that the expression of genes involved in lignin biosynthesis is spatiotemporally specific in different cell types during development.

Discovery of an unexpected catechin ester glucosyltransferase in plants

Glycosylation is an important modification of bioactive substances, affecting the storage, distribution, transport and activity of compounds. To explore the UGTs involved in the glycosylation

Figure 7 Characterization of UGT72B23 *in vivo* and *in vitro*. (a) Activity screening of recombinant CSS0024764 proteins with different substrates. The activity of ECG was set as 100%. Values are expressed as the mean \pm standard deviation of triplicate samples. (b) ECG glucoside identification. *m/z* 605 EIC of tea leaves, (c, d) Kinetic data of recombinant CSS0024764 for ECG and EGCG, respectively. (e) Mass spectral analysis of the formed product *rt* = 7.62 min, and ESI tandem mass product ions (*m/z*) of glycoside derivatives of epicatechin gallate following the glucose position in negative mode. (f) Mass spectral of *UGT72B23* overexpression in tobacco. (g, i, k) Relative expression level of *UGT72B23* in overexpressed tobacco lines, silenced in tea leaves by AsODN and different tissues of tea plant, respectively. (h, j, l) Relative concentration of ECG glucosides in overexpressed tobacco lines, silenced in tea leaves by AsODN and different tissues of tea plant, respectively. Duncan's multiple-range test was carried out, and statistical significance was calculated with one-way ANOVA using SPSS 20.0 ($P < 0.05$). CK, empty vector control; *UGT72B23-OE*, overexpression of *UGT72B23*.



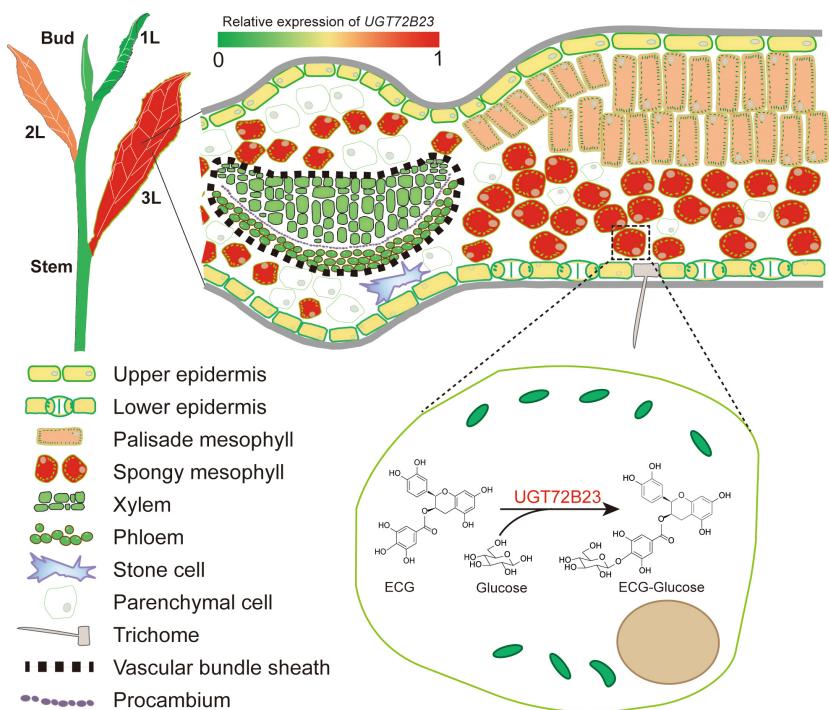


Figure 8 Model for spatiotemporal-specific UGT72B23-mediated ECG glucoside formation at single-cell resolution.

of catechins we constructed a genes-catechin co-expression network. The expression of eight *UGTs* correlated with the transcription of genes involved in catechin biosynthesis (absolute spearman correlation coefficients >0.4 , *P* value <0.05) (Figure 6c, Table S12). It should be noted that one gene CSS0024764 showed highest transcription level in mesophyll cells of 3 L (Figure 6c, Table S12) and was highly expressed in mesophyll cells and especially in SM (Figure 6d), suggesting that CSS0024764 (assigned UGT72B23 by the UGT Nomenclature Committee) might be a marker gene for SM.

We cloned and characterized UGT72B23 *in vitro* and *in vivo*. Unexpectedly, recombinant UGT72B23 could specifically catalyse the formation of catechin ester glucoside using epicatechin gallate (ECG) and epigallocatechin gallate (EGCG), two of the most important catechins contributing to bitterness and astringency in tea products, as the substrate (Figure 7a). We also searched for ECG/EGCG glucosides in tea leaves, but only ECG glucosides was found (Figure 7b). The kinetic parameters of UGT72B23 for ECG and EGCG were determined at pH 7.5 and 40 °C. The apparent K_m and K_{cat}/K_m ratios of UGT72B23 for ECG were 31.25 μM and 0.167 μM/s, and for EGCG were 27.9 μM and 0.237 μM/s, respectively (Figure 7c, d). The ECG glucoside formed by UGT72B23 was identified by LC–MS (Figure 7e). The presence of the fragment ion of *m/z* 331.0666 obtained by high resolution Q-TOF mass spectrometry suggests that the glucose is bound to the D ring although the precise structure needs to be further verified (Figure 7e). To further investigate whether UGT72B23 could glucosylate catechin ester *in planta*, the gene was transiently overexpressed in tobacco leaves and downregulated in tea leaves by gene-specific antisense oligodeoxynucleotide suppression (Jing et al., 2019). The content of ECG glucoside was markedly increased when UGT72B23 was overexpressed in tobacco, whereas EGCG glucoside was not detected in

tobacco (Figure 7f–h). The expression levels of UGT72B23 in individual tobacco leaves correlated very well with the content of ECG glucoside (Figure 7g, h). Silence of UGT72B23 in tea leaves led to a significantly reduction ECG glucoside determined by LC–MS compared with that in the control leaves (Figure 7i–j). Analysis in five different tissues of the tea plant showed that the expression of UGT72B23 correlated with the ECG glucoside content (Figure 7k–l). Thus, an unexpected catechin ester glucosyltransferase is involved in ECG glucoside formation in plants, reflecting the complexity of catechin ester metabolism in plants (Figure 8).

Discussion

Identification of 79 marker genes for tea leaves

Cell annotation is the key step and greatest conundrums to overcome in the application of single-cell sequencing in non-model plants. We found that the application of marker genes from *Arabidopsis* and other herbs is still limited in woody plants. A few of the model plant markers have homologous genes in the tea plant and showed different expression patterns than in the model plants. PH marker *SWEET* is highly expressed in LE of tea leaves and may be involved in the osmotic regulation of stomata guard cells or response to abiotic stresses (Kim et al., 2021; Wang et al., 2018a). This is due to the fact that gene expression changes dynamically during development. It varies with developmental status, tissue type and environment conditions (Seyfferth et al., 2021). In model plants, tissue-specific gene expression mainly relies on tractography techniques such as *in situ* hybridization and transgenic reporter lines (de Almeida Engler et al., 2001; Seyfferth et al., 2021; Zhang et al., 2021b). Non-fluorescent and fluorescent labels are commonly used when background noise is relatively low. However, there is no stable and efficient genetic

transformation system for tea plant and *in situ* hybridization protocols for tea plant results in highly autofluorescent background in very wide wavelength range because tea cells are rich in lignin and other secondary metabolites (Figure 1h) (Yang et al., 2020). As a result, there are few reports verifying spatiotemporal patterns of gene expression *in vivo* or *in situ*. Therefore, we tried an alternate approach and successfully isolated six tissue types from tea leaves according to published protocols of Arabidopsis and peanut (Endo et al., 2016; Liu et al., 2021). We obtained 79 reliable markers for UE, LE, PM, SM, XY and PH, including the novel UGT72B23 (Figures 4a, 6d). These markers were extremely specific for a single-cell type, such as the markers of UE and XY, and averagely specific for two or more cell types, such as the markers of ME (Figure 4a). Some markers for XY, probably due to mature xylem vessels of the third leaf, were characteristic of dead cells (You et al., 2019). Therefore, these specific markers provide a valuable reference for cell annotation of woody plant leaves. *In vivo* and *in situ* identification of cell types in non-model plants will also be an urgent problem to be solved in the future.

Generation of the first woody plant leave development cell atlas

By combining six tissue-specific markers and annotating all CSGs of each cluster by Arabidopsis markers (Tables S7, S9), the first cell atlas for leaf development in woody plants was created. (Figure 3c). The identification of cell types and the differentiation trajectories of the cells of tender tea leaf were presented in detail. Thus, we could observe the proportion of different cell types at different development stages and their theoretical positions in the development trajectories. Differential gene analysis and cell annotation of cells at different branch points and states can reveal key genes that determine or reflect the cell fate shift. By integrating the cell type and key genes involved in cell fate decision, we can present the developmental model of tender tea leaves (Figure 5k). This knowledge will facilitate future work on woody plant development at single-cell resolution and serve as a resource to understand cell fate determination during the leaf maturing process.

Gene expression related to biosynthesis of secondary metabolites at single-cell level during development

Based on the single-cell atlas and single-cell expression profiles, we were able to analyse the spatiotemporal expression patterns of specific cell types in the developmental process. Characteristic tea compounds and pathway genes related to the lignification of cell walls in woody plant cells were analysed. They showed spatiotemporal specific expression patterns in different cell types during development (Figures 6b, 8, S4). Flavonoid synthesis was concentrated and highly expressed in the mesophyll cells of young leaves, suggesting that flavonoids play important physiological functions in early leaf development (Falcone Ferreyra et al., 2012; Taylor and Grotewold, 2005). Lignin biosynthesis-related peroxidases were specifically expressed in vascular bundle, which is considered as mechanical support structure (Figures 6b, S4). In addition, we also identified several genes involved in the biosynthesis and regulation of characteristic tea compounds, such as the transcription factor *MYB*, and genes involved in theanine metabolism and transport (Dong et al., 2020; Li et al., 2022a; Lin et al., 2022; Yu et al., 2021). For example, *CsMYB139* and *CsMYB219* predicted to be involved in epidermal cell initiation is highly expressed in EP. *CsMYB2*, which is probably involved in

regulating shoot development, was highly expressed in PR, and an *amino acid permease* (*CSS0001846*) was specifically expressed in cluster 14, which was annotated as protophloem (Table S14). We also obtained more information about expression levels of these genes in other cell types, such as theanine biosynthesis and transport related genes which are mainly transcribed in ME and VB cells (Table S14). Genes involved in the same compound biosynthesis are not always co-expressed in the same cell type during development, indicating potential mechanisms of compound transport between different cell types.

Discovery of a novel catechin ester metabolic pathway by analysis of mesophyll cell expression matrix in tea leaves

Catechin derivatives are flavan-3-ols and thought to be associated with tea flavour, colour and quality. They account for 12–24% of the dry weight in tea leaves (Jin et al., 2014). Catechins mainly include non-ester catechins (e.g. EC and EGC) and ester type catechins (EGCG and ECG) in plants. Esterified catechin are the major components resulting in bitterness and astringency in tea. Using the transcriptomes of individual cell types, we were able to improve the prediction accuracy of the regulatory network and remove the background noise caused by the averaging of signals in bulk RNA-Seq. We detected for the first time, catechin ester glucosides in plant leaves, which might contribute to the high levels of catechins in SM during leaf development (Figure 8). Combined with the natural environment of wild tea plant, the glycosylation of catechin esters may be involved in leaf development and the response to environmental factors although its physiological functions in plants need further study.

Experimental procedures

Plant materials

Tender shoots of the current year of 6 years old *C. sinensis* var. *sinensis* cv. 'shuchazao' trees were collected from a tea plantation of Anhui Agriculture University (Hefei, China) in June. The complete 1st leaf (1 L) and 3rd leaf (3 L) blades were picked and surface sterilized with 75% alcohol for the following experiments. About 0.6 g of fresh leaf blades were cut and immediately put in a digestion solution for protoplast and tissue isolation. For chemical composition analyses, leaves were placed in liquid nitrogen and stored at –80 °C.

Leaf structure observation

To identify the cell types, fresh tea leaves were sliced transversely with a freshly sharpened blade. A part of sections was prepared and observed directly. The remaining sections with six different tissues were stained with safranin O/fast green to distinguish special cell types. The prepared temporary slices were imaged under visible light using the Zeiss Axio Zoom V16 microscope. To determine the thickness and degree of lignification of leaf cell, the untreated slices were also observed under the confocal laser scanning microscope (Lecia DMi8, Germany). The excitation light wavelength of the blue, green and red channel was 405 nm, 488 nm and 488 nm, respectively and the emission light wavelength range was 430–480 nm, 510–560 nm and 680–720 nm, respectively. Fluorescent images of each sample were taken under identical parameters. For the detection of non-glandular trichomes of 1 L and 3 L, green fluorescence of the lower epidermis was generated using a mercury arc lamp and excitation filter of

450–490 nm and an emission filter of 500–550 nm on Zeiss Axio Zoom V16 microscope.

Catechins, caffeine and theanine in leaves

Catechins, caffeine and theanine were extracted from 1 L and 3 L and quantified by HPLC according to previous methods (Li et al., 2019; Wang et al., 2018b) with minor modifications.

Tea leaf protoplast and tissue isolation

The digestion solutions for leaf protoplast and tissue were same and followed a protocol of Arabidopsis with modifications (Yoo et al., 2007). The detailed steps were as follows. 0.4 M D-mannitol (Sigma, Germany), 1.5% cellulase R10 (Yakult, Japan), 0.5% macerozyme R10 (Yakult, Tokyo, Japan) and 0.7% snailase (YEASEN, Shanghai, China) were dissolved in an initial solution which is 3/4 volume of final digestion solutions and contained 20 mM 4-morpholineethanesulfonic acid (pH = 5.5) and 20 mM KCl (Sigma, Germany). After cooling down to room temperature naturally, added 10 mM CaCl₂ (Sigma), 10 mM β-mercaptoethanol (Macklin, Shanghai, China) and 0.1% BSA (Sigma). 30 mL enzyme solutions were filtered in glass dishes (90 mm inner diameter) by 0.45 μm syringe filter. Rapid growing healthy leaves were selected as materials. Entire intact leaf blades were sterilized with 75% alcohol and cut into strips and immersed into digest solution in glass dish with a cover. The strips were evacuated for 5 min and then incubated statically in an incubator at 25 °C for 4 hrs in the darkness. The dished were rocked gently every 30 min during incubation period. The digestion mixture was filtered through a 40 μm nylon mesh sieve. Protoplasts were collected after gentle washing with 0.4 M D-mannitol solution twice. Centrifugation at 100 g for 2 min with a swinging bucket rotor of 3-18KS centrifuge (Sigma, Germany) with acceleration and brake settings of 3 and 3, respectively. The number of isolated protoplasts was measured with a haemocytometer counting chamber, and the viability of the protoplasts was assessed by 2 μg/mL fluorescein diacetate. The concentration of protoplasts was adjusted to 1500–2000 cells/μL by 0.4 M D-mannitol for loading onto the chromium controller of the 10× Genomics platform.

Cell-specific tissue isolation was performed according to the method described in Arabidopsis (Endo et al., 2016). Mid veins of 3 L were cut-off from blades, and the vein and lamella were digested separately in the enzyme solution for 3 h. The epidermis layers were removed using tweezers. The leaves without upper epidermis were digested for another 3 h to isolate the palisade mesophyll cells and the leaves without lower epidermis was used for the isolation of spongy mesophyll cells. The upper and lower epidermis was also removed from the digested main vein and then carefully dissected to separate the white xylem vessels and the phloem (vessels without xylem, including phloem, vascular bundle and procambium). All samples were washed twice with 0.4 M D-mannitol, then frozen in liquid nitrogen and stored at –80 °C for total mRNA extraction and reverse transcription.

scRNA-seq Library construction, sequencing and raw data quality control

The construction of the scRNA-seq libraries for tea leaves was performed strictly according to the manufacturer's instructions [10× Genomics Chromium Single-Cell 3' kit (V3)]. Single cells were sequenced with an Illumina NovaSeq 6000 sequencing system (paired-end multiplexing run, 150 bp) by LC-Bio Technology co. Ltd., (Hangzhou, China). Sequencing results were

demultiplexed and converted to FASTQ format using Illumina bcl2fastq software (version 2.20). Sample demultiplexing, barcode processing and single-cell 3' gene counting was done by using the Cell Ranger (version 5.0.1) and scRNA-seq data were aligned to the tea chromosome-level genome from the Tea Plant Information Archive (TPIA) (<http://tpia.teaplant.org>). Single cell captured from 1 L and 3 L were processed using 10× Genomics Chromium Single Cell 3' Solution. The Cell Ranger output was loaded into Seurat (version 3.1.1) which was used for dimensional reduction, clustering and analysis of scRNA-seq data. Overall, cells passed the quality control threshold when the following criteria were met. All genes expressed in less than one cell were removed, number of genes expressed per cell was >500 and <5000 as low and high cut-off, respectively, UMI counts were less than 500, the per cent of mitochondrial-DNA derived gene expression was <25%, doublets were removed by DoubletFinder with default settings (version 2.0.2).

Data visualization and clustering

To visualize the data, we further reduced the dimensionality of all filtered cells using Seurat and used tSNE to project the cells into 2D space (van der Maaten and Hinton, 2008). The steps include: (i) Using the LogNormalize method of the 'Normalization' function of the Seurat software to calculate the expression value of genes; (ii) PCA was performed using the normalized expression value, within all the principal components, the top 10 principal components were used to do clustering and tSNE analysis; (iii) To find clusters, a weighted graph-based clustering method (SNN-Shared Nearest Neighbour) is selected. Marker genes for each cluster were identified with the Wilcoxon rank-sum test (default parameters is 'bimod': Likelihood-ratio test) with default parameters via the FindAllMarkers function in Seurat. This selects marker genes that are expressed in more than 10% of the cells in a cluster and average log2-fold change (log2FC) >0.25.

To reduce the protoplasting effect, response genes were selected by comparison with a previously published transcriptome (Jing et al., 2019). Relative quantification of gene expression of single-cell and bulk RNA-Seq profiles was performed using actin *CSS0008920* as an endogenous reference gene for normalization. Subsequently, gene expression fold changes of 1 L and 3 L were calculated by log2FC. The log2FC threshold for protoplasting responsive genes was chosen abs(log2FC) >1, these genes were removed from the cluster-specific marker gene list and extracted for KEGG mapping.

Marker gene and cell type identification

To identify the cell type, we first performed orthologous gene alignments of the reported marker genes in Arabidopsis (Jin et al., 2021). The marker gene list was downloaded from the Plant Cell Marker DataBase (PCMDB, <http://www.tobaccodb.org/pcmdb/>) (Jin et al., 2021), and the protein sequences of tea plant and Arabidopsis were download from the TAIR and TPBA (Garcia-Hernandez et al., 2002; Xia et al., 2019). Using the Arabidopsis marker genes as the query sequences, the homologous genes of the tea plant were searched by TBtools (BLASTP: e-value: 1-e⁻⁵) (Chen et al., 2020a). The top score hits were selected and annotated as corresponding Arabidopsis cell type.

Then, we directly used the proved Arabidopsis marker genes as query to find out homologous genes in the tea plant. We then quantified their expression levels in six tissue types by quantitative

real-time reverse transcription PCR (q-RT-PCR). We also selected cluster-specific marker genes to identify tissue-specific genes with relative expression in these tissues by q-RT-PCR. Finally, we combined the *Arabidopsis* reference marker gene and q-RT-PCR data and assigned all clusters to specific cell types.

Pseudo-time trajectory analysis

Single-cell trajectories were constructed by Monocle (Version 2.0) (Qiu *et al.*, 2017). We used the expression matrix of all cells and epidermis subcluster cells to run separate pseudo-time analysis. The procambium cells of 1 L were set as starting point, to visualize the trajectory in the reduced dimensional space. Branched expression analysis modelling (BEAM) was used to further test for branch-dependent gene expression (Qiu *et al.*, 2017). Then we analysed key genes related to the development and differentiation process. To uncover the regulation network and functional genes, we also performed cluster analysis and KEGG/GO annotation using OmicStudio tools at <https://www.omicstudio.cn/tool/>.

Gene expression analysis

Total RNA from six dissociated leaf tissues was isolated using RNAiso-mate for Plant Tissue (Takara, Dalian, China) and RNAiso Plus (Takara) according to the manufacturer's instructions. The cDNA was synthesized by reverse transcription from total RNA using PrimeScriptRT Master Mix (Takara). Quantitative real-time PCR (q-RT-PCR) assays were performed on a CFX96 platform (Bio-Rad, CA), using the Hieff® qPCR SYBR Green Master Mix (YEASEN, Shanghai, China) according to the manufacturer's instructions and calculated by the comparative CT method (Livak and Schmittgen, 2001). Transcript abundances of *CsUBI* (CSS0007748, ubiquitin-conjugating enzyme) and *CsACTIN* (CSS0008920, actin 7 isoform 1) were determined as references. Table S15 gives all primer sequences.

Co-expression network analysis for flavonoid metabolic pathway

The CSGs expression matrix of 4297 mesophyll cells from both 1 L and 3 L were used for co-expression network analysis. The expression index of flavonoid biosynthesis pathway and glycosyltransferase (UGT) genes was selected from the scRNA-Seq file. Spearman correlation was applied to find related UGTs, (filter conditions: absolute spearman correlation coefficients >0.4, *P* value <0.05). Network graph was generated by OmicStudio tools and adjusted by Adobe Illustrator CC2019. Dot sizes represent the average gene expression levels in mesophyll cells as determined by scRNA-Seq.

RNA isolation, cDNA cloning and sequence analysis

Total RNA from leaves of *Camellia sinensis* var. *sinensis* cv. 'Shuchazao' was isolated using RNA isolator Total RNA Extraction Reagent (Vazyme Biotech Co., Ltd, Nanjing, China) according to the manufacturer's instructions. For qRT-PCR, the cDNA was generated by reverse transcription from total RNA using HiScript® III RT SuperMix (Vazyme Biotech Co. Ltd). The open reading frame sequences were amplified using Phusion® High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA). The PCR products, which were purified with a Gel Extraction Kit (CWBIOL, Jiangsu, China), were cloned into the pGEX-4 T1 vector and subsequently transformed into Trans1T1 competent cells.

Heterologous protein expression and purification

The complete open reading frame of *UGT72B23* was digested with *BamH*1 and *Sma*1, and the resulting gene fragments were subcloned into the expression vector pGEX-4 T-1. The recombinant plasmids with *UGT72B23* and the empty expression vector pGEX-4 T-1, which served as the negative control were then transformed into *E. coli* strain BL21 (DE3) pLysS cells. When the OD₆₀₀ of the culture reached between 0.6 and 0.8, isopropyl-β-D-thiogalactopyranoside was added to a final concentration of 1 mM to induce protein expression. The cultures were incubated at 16 °C with shaking at 200 rpm overnight. The next day, the cells were harvested by centrifugation at 4 °C (5000 rpm for 10 min). The recombinant proteins were purified using GST-binding resin (Novagen, Darmstadt, Germany) following the manufacturer's protocol. The protein concentration was determined by the Bradford method, and bovine serum albumin served as the standard protein. The purified *UGT72B23* recombinant protein was further analysed by SDS-PAGE.

Enzymatic activity assay

The enzyme assays were performed according to the method described previously (Jing *et al.*, 2019). For the initial screening, the enzyme reaction mixture (5 µL) consisted of 3.4 µL of Tris-HCl buffer (50 mM, pH 7.5, 10% glycerol and 10 mM 2-mercaptoethanol), 0.2 µL of 10 mM UDP-glucose, 0.2 µL of 20 µM substrate solution, 0.2 µL of 50 mM DL-dithiothreitol, and 1 µL of purified protein (0.5 µg per reaction). The recombinant enzyme assay was carried out using the UDP-Glo™ glycosyltransferase assay kit (Sheikh *et al.*, 2017). The optimum reaction temperature and pH was determined according to Chen (Chen *et al.*, 2020b). To determine kinetic parameters, at least eight different substrate concentrations covering the range from 0 to 800 µM were used in the aforementioned assay in triplicate at the optimized conditions (Zhao *et al.*, 2017).

Identification of the reaction products by LC-MS

The standard assays were scaled up to 500 µL, which included 5 mM UDP-glucose, 200 µM substrate, and 20 to 50 µg purified protein. The enzyme assays were incubated at 30 °C for 12 h. Five hundred µL of ethyl acetate was then added to stop the reaction and products were extracted twice. After ethyl acetate was vapourized, the residue was further dissolved in 500 µL methanol for LC-MS analysis. LC-MS was performed with a reverse-phase C18 column (1.8 µm, 100 × 2.1 mm) at 40 °C. A DIONEX Ultimate 3000 UHPLC system (Thermo Fisher Scientific, Waltham, MA) with an autosampler was utilized for all experiments. The solvents and LC parameters for product identification were selected according to the method described previously (Jing *et al.*, 2019). The products were identified by comparing their UV and MS spectra with those in the literature (Delcambre and Saucier, 2012).

Data analysis and graphing

Data were analysed using SPSS Statistics 20.0 (SPSS Inc., Chicago, IL) and presented as mean ± SD of at least triplicate measurements. Significance was determined at *P* < 0.05 by analysis of variance (ANOVA) followed by Duncan's least significant difference test. ScRNA-Seq analysis including tSNE map visualization for tea leaf cell types, marker gene dotplot and the hierarchical clustering heatmaps were produced to visualize the gene

expression level using the OmicStudio tools at <https://www.omicstudio.cn/tool/>.

For hierarchical clustering and visualization, all data were Z-score normalized by Euclidean distance.

Acknowledgements

This work was supported by grants from the National Natural Science Foundation of China (3196113030; 31870678; 32022076; 32102433), the postdoctoral fellowship from Anhui Agricultural University and German Research Foundation (DFG, SCHW 634/34-1). We thank Jun Fu and Xinxin Zhou from Hangzhou LC-BIO Co., Ltd for their great assistance with the RNA-seq data analysis. We sincerely appreciate professor Ru-duo Huang from school of life science of Anhui University and professor Ting Xue from school of life science of Anhui Agriculture University, for their encouragement and help.

Author contributions

C.S. and Q.W. conceptualized the initial study; C. S. and Q.W. were involved in the experimental layout; Q.W., Y.W., A. P., M. Z. and Y. P. performed the lab experiments, and Q.W. drafted the initial article; all authors discussed the results, reviewed the article and approved the final article.

Conflict of interests

The authors declare no competing interests.

References

- Bai, Y., Liu, H., Lyu, H., Su, L., Xiong, J. and Cheng, Z.M. (2022) Development of a single-cell atlas for woodland strawberry (*Fragaria vesca*) leaves during early *Botrytis cinerea* infection using single cell RNA-seq. *Hortic. Res.* **9**, uhab055.
- Barros, J., Serk, H., Granlund, I. and Pesquet, E. (2015) The cell biology of lignification in higher plants. *Ann. Bot.* **115**, 1053–1074.
- Bonawitz, N.D. and Chapple, C. (2010) The genetics of lignin biosynthesis: connecting genotype to phenotype. *Annu. Rev. Genet.* **44**, 337–363.
- Boudolf, V., Rombauts, S., Naudts, M., Inzé, D. and De Veylder, L. (2001) Identification of novel cyclin-dependent kinases interacting with the CKS1 protein of *Arabidopsis*. *J. Exp. Bot.* **52**, 1381–1382.
- Butt, A.D. (1985) A general method for the high-yield isolation of mesophyll protoplasts from deciduous tree species. *Plant Sci.* **42**, 55–59.
- Chen, C., Chen, H., Zhang, Y., Thomas, H.R., Frank, M.H., He, Y. and Xia, R. (2020a) TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* **13**, 1194–1202.
- Chen, Y., Guo, X., Gao, T., Zhang, N., Wan, X., Schwab, W. and Song, C. (2020b) UGT74AF3 enzymes specifically catalyze the glucosylation of 4-hydroxy-2,5-dimethylfuran-3(2H)-one, an important volatile compound in *Camellia sinensis*. *Hortic. Res.* **7**, 25.
- Chen, Y., Tong, S., Jiang, Y., Ai, F., Feng, Y., Zhang, J., Gong, J. et al. (2021) Transcriptional landscape of highly lignified poplar stems at single-cell resolution. *Genome Biol.* **22**, 319.
- Daloso, D.M., dos Anjos, L. and Fernie, A.R. (2016) Roles of sucrose in guard cell regulation. *N. Phytol.* **211**, 809–818.
- de Almeida Engler, J., De Groot, R., Van Montagu, M. and Engler, G. (2001) In situ hybridization to mRNA of *arabidopsis* tissue sections. *Methods* **23**, 325–334.
- De Filippis, L.F. (2016) Plant secondary metabolites. In *Plant-Environment Interaction*, (Azooz, M.M. and Ahmad, P., eds), pp. 263–299. Hoboken: Wiley online library.
- Delcambre, A. and Saucier, C. (2012) Identification of new flavan-3-ol monoglycosides by UHPLC-ESI-Q-TOF in grapes and wine. *J. Mass Spectrom.* **47**, 727–736.
- Dinant, S., Clark, A.M., Zhu, Y., Vilaine, F.O., Palauqui, J.-C., Kusiak, C. and Thompson, G.A. (2003) Diversity of the superfamily of phloem lectins (phloem protein 2) in angiosperms. *Plant Physiol.* **131**, 114–128.
- Dong, C., Li, F., Yang, T., Feng, L., Zhang, S., Li, F., Li, W. et al. (2020) Theanine transporters identified in tea plants (*Camellia sinensis* L.). *Plant J.* **101**, 57–70.
- Endo, M., Shimizu, H. and Araki, T. (2016) Rapid and simple isolation of vascular, epidermal and mesophyll cells from plant leaf tissue. *Nat. Protoc.* **11**, 1388–1395.
- Falcone Ferreyra, M.L., Rius, S. and Casati, P. (2012) Flavonoids: biosynthesis, biological functions, and biotechnological applications. *Front. Plant Sci.* **3**, 222.
- Garcia-Hernandez, M., Berardini, T., Chen, G., Crist, D., Doyle, A., Huala, E., Knee, E. et al. (2002) TAIR: a resource for integrated *Arabidopsis* data. *Funct. Integr. Genomic.* **2**, 239–253.
- Jin, J.-Q., Ma, J.-Q., Ma, C.-L., Yao, M.-Z. and Chen, L. (2014) Determination of catechin content in representative chinese tea germplasms. *J. Agr. Food Chem.* **62**, 9436–9441.
- Jin, J., Lu, P., Xu, Y., Tao, J., Li, Z., Wang, S., Yu, S. et al. (2021) PCMDB: a curated and comprehensive resource of plant cell markers. *Nucleic Acids Res.* **50**, D1448–D1455.
- Jing, T., Zhang, N., Gao, T., Zhao, M., Jin, J., Chen, Y., Xu, M. et al. (2019) Glucosylation of (Z)-3-hexenol informs intraspecies interactions in plants: a case study in *Camellia sinensis*. *Plant Cell Environ.* **42**, 1352–1367.
- Kabera, J. (2014) Plant secondary metabolites: biosynthesis, classification, function and pharmacological classification, function and pharmacological properties. *J. Pharm. Pharmacol.* **2**, 377–392.
- Kim, J.Y., Symeonidi, E., Pang, T.Y., Denyer, T., Weidauer, D., Bezrutczyk, M., Miras, M. et al. (2021) Distinct identities of leaf phloem cells revealed by single cell transcriptomics. *Plant Cell* **33**, 511–530.
- Kumar, N., Harashima, H., Kalve, S., Bramsiepe, J., Wang, K., Sizani, B.L., Bertrand, L.L. et al. (2015) Functional conservation in the SIAMESE-RELATED family of cyclin-dependent kinase inhibitors in land plants. *Plant Cell* **27**, 3065–3080.
- Li, C.F., Zhu, Y., Yu, Y., Zhao, Q.Y., Wang, S.J., Wang, X.C., Yao, M.Z. et al. (2015) Global transcriptome and gene regulation network for secondary metabolite biosynthesis of tea plant (*Camellia sinensis*). *BMC genomics* **16**, 560.
- Li, F., Dong, C., Yang, T., Ma, J., Zhang, S., Wei, C., Wan, X. et al. (2019) Seasonal theanine accumulation and related gene expression in the roots and leaf buds of tea plants (*Camellia Sinensis* L.). *Front. Plant Sci.* **10**, 1397.
- Li, H., Dai, X., Huang, X., Xu, M., Wang, Q., Yan, X., Sederoff, R.R. et al. (2021) Single-cell RNA sequencing reveals a high-resolution cell atlas of xylem in *Populus*. *J. Integr. Plant Bio.* **63**, 1906–1921.
- Li, P., Fu, J., Xu, Y., Shen, Y., Zhang, Y., Ye, Z., Tong, W. et al. (2022a) CsMYB1 integrates the regulation of trichome development and catechins biosynthesis in tea plant domestication. *N. Phytol.* **234**, 902–917.
- Li, P., Xia, E., Fu, J., Xu, Y., Zhao, X., Tong, W., Tang, Q. et al. (2022b) Diverse roles of MYB transcription factors in regulating secondary metabolite biosynthesis, shoot development, and stress responses in tea plants (*Camellia sinensis*). *Plant J.* **110**, 1144–1165.
- Lin, S., Chen, Z., Chen, T., Deng, W., Wan, X. and Zhang, Z. (2022) Theanine metabolism and transport in tea plants (*Camellia sinensis* L.): advances and perspectives. *Crit. Rev. Biotechnol.* **17**, 1–15.
- Liu, H., Hu, D., Du, P., Wang, L., Liang, X., Li, H., Lu, Q. et al. (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 Biotechnol. J.* **19**, 2261–2276.
- Liu, Z., Zhou, Y., Guo, J., Li, J., Tian, Z., Zhu, Z., Wang, J. et al. (2020) Global dynamic molecular profiling of stomatal lineage cell development by single-cell RNA sequencing. *Mol. Plant* **13**, 1178–1193.
- Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative pcr and the $2^{-\Delta\Delta C_T}$ method. *Methods* **25**, 402–408.
- Marand, A.P., Chen, Z., Gallavotti, A. and Schmitz, R.J. (2021) A cis-regulatory atlas in maize at single-cell resolution. *Cell* **184**, 3041–3055.
- Moinuddin, S.G.A., Youn, B., Bedgar, D.L., Costa, M.A., Helms, G.L., Kang, C., Davin, L.B. et al. (2006) Secoisolariciresinol dehydrogenase: mode of catalysis

- and stereospecificity of hydride transfer in *Podophyllum peltatum*. *Org. Biomol. Chem.* **4**, 808–816.
- Muñiz, L., Minguez, E.G., Singh, S.K., Pesquet, E., Vera-Sirera, F., Moreau-Courtois, C.L., Carbonell, J. et al. (2008) ACAULIS5 controls *Arabidopsis* xylem specification through the prevention of premature cell death. *Development* **135**, 2573–2582.
- Qiu, X., Hill, A., Packer, J., Lin, D., Ma, Y.A. and Trapnell, C. (2017) Single-cell mRNA quantification and differential analysis with Census. *Nat. Methods* **14**, 309–315.
- Quan, L., Xiao, R., Li, W., Oh, S.-A., Kong, H., Ambrose, J.C., Malcos, J.L. et al. (2008) Functional divergence of the duplicated *AtKIN14a* and *AtKIN14b* genes: critical roles in *Arabidopsis meiosis* and gametophyte development. *Plant J.* **53**, 1013–1026.
- Ranocha, P., Denancé, N., Vanholme, R., Freydier, A., Martinez, Y., Hoffmann, L., Köhler, L. et al. (2010) *Walls are thin 1 (WAT1)*, an *Arabidopsis* homolog of *Medicago truncatula NODULIN21*, is a tonoplast-localized protein required for secondary wall formation in fibers. *Plant J.* **63**, 469–483.
- Seyfferth, C., Renema, J., Wendrich, J.R., Eekhout, T., Seurinck, R., Vandamme, N., Blob, B. et al. (2021) Advances and opportunities in single-cell transcriptomics for plant research. *Annu. Rev. Plant Biol.* **72**, 847–866.
- Sheikh, M.O., Halmo, S.M., Patel, S., Middleton, D., Takeuchi, H., Schafer, C.M., West, C.M. et al. (2017) Rapid screening of sugar-nucleotide donor specificities of putative glycosyltransferases. *Glycobiology* **27**, 206–212.
- Shen, T., Xu, M., Qi, H., Feng, Y., Yang, Z. and Xu, M. (2022) Protoplast isolation and transcriptome analysis of developing xylem in *Pinus massoniana* (Pinaceae). *Mol. Bio. Rep.* **49**, 1857–1869.
- Sugiyama, A., Sano, C.M., Yazaki, K. and Sano, H. (2016) Caffeine fostering of mycoparasitic fungi against phytopathogens. *Plant Signal. Behav.* **11**, e1113362.
- Sun, Y., Zhou, J. and Guo, J. (2021) Advances in the knowledge of adaptive mechanisms mediating abiotic stress responses in *Camellia sinensis*. *Front. Biosci.* **26**, 1714–1722.
- Taylor, L.P. and Grotewold, E. (2005) Flavonoids as developmental regulators. *Curr. Opin. Plant Biol.* **8**, 317–323.
- van der Maaten, L. and Hinton, G. (2008) Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605.
- Vandepoele, K., Raes, J., De Veylder, L., Rouzé, P., Rombauts, S. and Inzé, D. (2002) Genome-wide analysis of core cell cycle genes in *Arabidopsis*. *Plant Cell* **14**, 903–916.
- Vierling, E. (1991) The roles of heat shock proteins in plants. *Annu. Rev. Plant Phys.* **42**, 579–620.
- Wang, L., Yao, L., Hao, X., Li, N., Qian, W., Yue, C., Ding, C. et al. (2018a) Tea plant SWEET transporters: expression profiling, sugar transport, and the involvement of CsSWEET16 in modifying cold tolerance in *Arabidopsis*. *Plant Mol. Biol.* **96**, 577–592.
- Wang, X., Guo, L., Gao, L., Shi, X., Zhao, X., Ma, X., Xia, T. et al. (2018b) Molecular evidence for catechin synthesis and accumulation in tea buds (*Camellia sinensis*). *J. Agr. Food Chem.* **66**, 63–69.
- Xia, E.-H., Zhang, H.-B., Sheng, J., Li, K., Zhang, Q.-J., Kim, C., Zhang, Y. et al. (2017) The tea tree genome provides insights into tea flavor and independent evolution of caffeine biosynthesis. *Molecular Plant* **10**, 866–877.
- Xia, E.H., Li, F.D., Tong, W., Li, P.H., Wu, Q., Zhao, H.J., Ge, R.H. et al. (2019) Tea Plant Information Archive: a comprehensive genomics and bioinformatics platform for tea plant. *Plant Biotechnol. J.* **17**, 1938–1953.
- Xia, Z.Q., Costa, M.A., Pelissier, H.C., Davin, L.B. and Lewis, N.G. (2001) Secoiridiciresinol dehydrogenase purification, cloning, and functional expression. Implications for human health protection. *J. Bio. Chem.* **276**, 12614–12623.
- Xie, J., Li, M., Zeng, J., Li, X. and Zhang, D. (2021) Single-cell RNA sequencing profiles of stem-differentiating xylem in poplar. *Plant Biotechnol. J.* **20**, 417–419.
- Xiong, J.-S., Ding, J. and Li, Y. (2015) Genome-editing technologies and their potential application in horticultural crop breeding. *Hortic. Res.* **2**, 15019.
- Xu, H., Wang, Y., Chen, Y., Zhang, P., Zhao, Y., Huang, Y., Wang, X. et al. (2016) Subcellular localization of galloylated catechins in tea plants [*Camellia sinensis* (L.) O. Kuntze] assessed via immunohistochemistry. *Front. Plant Sci.* **7**, 728.
- Xu, X.-f., Zhu, H.-y., Ren, Y.-f., Feng, C., Ye, Z.-h., Cai, H.-m., Wan, X.-c. et al. (2021) Efficient isolation and purification of tissue-specific protoplasts from tea plants (*Camellia sinensis* (L.) O. Kuntze). *Plant Methods* **17**, 84.
- Yang, T., Lu, X., Wang, Y., Xie, Y., Ma, J., Cheng, X., Xia, E. et al. (2020) HAK/KUP/KT family potassium transporter genes are involved in potassium deficiency and stress responses in tea plants (*Camellia sinensis* L.): expression and functional analysis. *BMC Genomics* **21**, 556.
- Yao, S., Liu, Y., Zhuang, J., Zhao, Y., Dai, X., Jiang, C., Wang, Z. et al. (2022) Insights into acylation mechanisms: co-expression of serine carboxypeptidase-like acyltransferases and their non-catalytic companion paralogs. *Plant J.* **111**, 117–133.
- Yoo, S.-D., Cho, Y.-H. and Sheen, J. (2007) *Arabidopsis* mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat. Protoc.* **2**, 1565–1572.
- You, Y., Sawikowska, A., Lee, J.E., Benstein, R.M., Neumann, M., Krajewski, P. and Schmid, M. (2019) Phloem companion cell-specific transcriptomic and epigenomic analyses identify MRF1, a regulator of flowering. *Plant Cell* **31**, 325–345.
- Yu, S., Li, P., Zhao, X., Tan, M., Ahmad, M.Z., Xu, Y., Tadege, M. et al. (2021) CsTCPs regulate shoot tip development and catechin biosynthesis in tea plant (*Camellia sinensis*). *Hortic. Res.* **8**, 104.
- Zeng, L., Zhou, X., Liao, Y. and Yang, Z. (2020) Roles of specialized metabolites in biological function and environmental adaptability of tea plant (*Camellia sinensis*) as a metabolite studying model. *J. Adv. Res.* **34**, 159–171.
- Zhang, T.Q., Chen, Y., Liu, Y., Lin, W.H. and Wang, J.W. (2021a) Single-cell transcriptome atlas and chromatin accessibility landscape reveal differentiation trajectories in the rice root. *Nat. Commun.* **12**, 2053.
- Zhang, T.Q., Chen, Y. and Wang, J.W. (2021b) A single-cell analysis of the *Arabidopsis* vegetative shoot apex. *Dev. Cell* **56**, 1056–1074.
- Zhao, S., Mi, X., Guo, R., Xia, X., Liu, L., An, Y., Yan, X. et al. (2020) The biosynthesis of main taste compounds is coordinately regulated by miRNAs and phytohormones in tea plant (*Camellia sinensis*). *J. Agr. Food Chem.* **68**, 6221–6236.
- Zhao, X., Wang, P., Li, M., Wang, Y., Jiang, X., Cui, L., Qian, Y. et al. (2017) Functional characterization of a new tea (*Camellia sinensis*) flavonoid glycosyltransferase. *J. Agr. Food Chem.* **65**, 2074–2083.

Supporting information

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

Figure S1 The 3rd leaf cross section and epidermis after staining with safranin O/fast green.

Figure S2 Visualization of 16 cell clusters using tSNE.

Figure S3 Heatmap shows the CSGs of 16 clusters with log2(FC) > 0.26.

Figure S4 Cell-specific distribution of genes related to the biosynthesis of theanine and caffeine.

Table S1 Statistical data of scRNA-seq.

Table S2 Statistical information of 16 cell clusters.

Table S3 Statistical information of protoplasting response genes in cluster.

Table S4 Protoplasting response gene list.

Table S5 Marker gene list of cell-clusters 0 to 15.

Table S6 Orthologous genes of characterized *Arabidopsis* cell-type specific marker gene retrieved from tea genome.

Table S7 Orthologous genes of *Arabidopsis* cell-type specific genes retrieved from tea genome.

Table S8 Marker gene for heatmap.

Table S9 Marker gene identified by qRT-PCR of 6 tissues of 3 L.

Table S10 Top 50 gene of pseudotime analysis.

Table S11 The top 100 significantly changed genes discovered by the BEAM of 4 differentiation trajectory branch points.

Table S12 Original data of heat map for tea characterized metabolites.

Table S13 Spearman correlation analysis between UGTs and functional gene of flavonoid biosynthesis pathway.

Table S14 Identify homolog of characterized gene that had reported.

Table S15 All primers for RT-PCR and gene cloning.