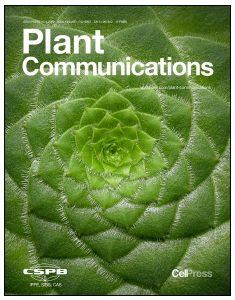
Journal Pre-proof



Mass spectrometry imaging and single-cell transcriptional profiling reveal the tissue-specific regulation of bioactive ingredient biosynthesis in *Taxus* leaves

Xiaori Zhan, Tian Qiu, Hongshan Zhang, Hou Kailin, Xueshuang Liang, Cheng Chen, Zhijing Wang, Qicong Wu, Xiaojia Wang, Xiao-lin Li, Mingshuang Wang, Shangguo Feng, Houqing Zeng, Chunna Yu, Huizhong Wang, Chenjia Shen

PII: S2590-3462(23)00147-5

DOI: https://doi.org/10.1016/j.xplc.2023.100630

Reference: XPLC 100630

To appear in: *PLANT COMMUNICATIONS*

Received Date: 29 December 2022

Revised Date: 31 March 2023

Accepted Date: 22 May 2023

Please cite this article as: Zhan, X., Qiu, T., Zhang, H., Kailin, H., Liang, X., Chen, C., Wang, Z., Wu, Q., Wang, X., Li, X.-l., Wang, M., Feng, S., Zeng, H., Yu, C., Wang, H., Shen, C., Mass spectrometry imaging and single-cell transcriptional profiling reveal the tissue-specific regulation of bioactive ingredient biosynthesis in *Taxus* leaves, *PLANT COMMUNICATIONS* (2023), doi: https://doi.org/10.1016/j.xplc.2023.100630.

Mass spectrometry imaging and single-cell transcriptional profiling reveal the tissue-specific regulation of bioactive ingredient biosynthesis in *Taxus* leaves

Running title: Single-cell transcriptome atlas of *Taxus* leaf

Xiaori Zhan[1,2], Tian Qiu[1,2], Hongshan Zhang[1,2,3], Hou Kailin[1,2], Xueshuang Liang[1,2], Cheng Chen[4], Zhijing Wang[4], Qicong Wu[1,2], Xiaojia Wang[1], Xiao-lin Li[5], Mingshuang Wang[1,2], Shangguo Feng[1,2], Houqing Zeng[1,3], Chunna Yu[1,2], Huizhong Wang[1,2*], Chenjia Shen[1,2,3 *]

[1] College of Life and Environmental Sciences, Hangzhou Normal University, Hangzhou 311121, China;

[2] Zhejiang Provincial Key Laboratory for Genetic Improvement and Quality Control of Medicinal Plants, Hangzhou Normal University, Hangzhou 311121, China;

[3] Kharkiv Institute, Hangzhou Normal University, Hangzhou 311121, China;

[4] College of Pharmacy, Hangzhou Normal University, Hangzhou 311121, China.

[5] State Key Laboratory Breeding Base of Dao-di Herbs, National Resource Center for Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China;

[*]Authors for correspondence:

Chenjia Shen (shencj@hznu.edu.cn), Tel: +86-571-28865198; Fax: +86-571-28865198;

Huizhong Wang (whz62@163.com), Tel: +86-571-28865198; Fax: +86-571-28865198.

Short Summary

MALDI-IMS analysis visualized various secondary metabolites in *T. mairei* leaf sections. *Taxus* leaf single-cell sequencing showed that most of the taxol biosynthesis

genes are expressed mainly in leaf mesophyll cells. Furthermore, a number of novel and cell specific transcription factors were identified, providing valuable resources for studying the cell type specific regulation of secondary metabolism.

Abstract

Taxus leaves provide the raw industrial materials for taxol, a natural antineoplastic drug widely used in the treatment of various cancers. However, the precise distribution, biosynthesis, and transcriptional regulation of taxoids and other active components in *Taxus* leaves remain unknown. Matrix-Assisted Laser Desorption/Ionization-Mass Spectrometry Imaging (MALDI-IMS) analysis was used to visualize various secondary metabolites in *T. mairei* leaf sections, confirming the tissue-specific accumulation of different active metabolites. Single-cell sequencing was applied to produce expression profiles of 8846 cells, with a median of 2,352 genes per cell. Based on a series of cluster-specific markers, cells were grouped into 15 clusters, suggesting a high degree of cell heterogeneity in the *T. mairei* leaves. Our data created the first *Taxus* leaf metabolic single-cell atlas and reveal spatial and temporal expression patterns of several secondary metabolic pathways. According to the cell type annotation, most of the taxol biosynthesis genes are expressed mainly in leaf mesophyll cells, phenolic acid and flavonoid biosynthesis genes are expressed highly in leaf epidermis cells (including stomatal complex and guard cells), and terpenoid and steroid biosynthesis genes are expressed specifically in leaf mesophyll cells. Furthermore, a number of novel and cell specific transcription factors involved in secondary metabolite biosynthesis were identified, including MYB17, WRKY12, WRKY31, ERF13, GT_2, and bHLH46. Our research established the transcriptional landscape of the major cell types in *T. mairei* leaves at single-cell resolution and provided valuable resources for studying the basic principles of cell type specific regulation of secondary metabolism.

Keywords:

ScRNA-seq, taxol biosynthesis, *Taxus mairei*, secondary metabolism, transcription regulation

Introduction

Taxol, a tetracyclic diterpenoid first isolated from *Taxus brevifolia*, is reported to be one of the most effective anti-tumor drugs (Wani et al., 1971). However, the slow growth of wild yew and low taxol contents in plant tissues cause shortages of taxol (Dang et al., 2017). Although the highest levels of taxoids are located in the stem bark, stripping bark is a destructive production method. As the commercial value of taxol increases, so does the destruction of *Taxus* trees (Yu et al., 2018). *Taxus wallichiana* var. *mairei* (*T. mairei*) is an evergreen tree from the family Taxaceae and is widely distributed in Southeast Asia (Wang et al., 2019b). Leaves and twigs of *T. mairei* are used for the industrial extraction of taxol, because they have sufficient biomass and are renewable.

The production of taxol involves a complex, 20-step biosynthetic pathway that begins with a diterpenoid precursor (Croteau et al., 2006; Yu *et al.*, 2018). The first discrete process in the biogenesis of taxol is the construction of the taxane skeleton by taxadiene synthase (TS), which is a key, but not rate-limiting, enzyme in the pathway (Kui et al., 2022). Baccatin III is produced during the second process in the pathway and involves a series of cytochrome P450 taxoid hydroxylases, including the 2*α*-, 5*α*-, 7*β*-, 9*α*-, 10*β*-, 13*α*-, and 14*β*-hydroxylases, and 10-deacetylbaccatin III-10-*O*-acetyltransferase (DBAT) (Kaspera and Croteau, 2006). Then, baccatin III-3-amino,3-phenylpropanoyltransferase (BAPT) catalyzes the connection of a phenylalanine-derived C13-side chain to baccatin III, producing 3′-*N*-debenzoyl-2′-deoxytaxol. Finally, 3′-*N*-debenzoyl-2′-deoxytaxol is converted to paclitaxel by 3′-*N*-debenzoyl-2′-deoxytaxol-*N*-benzoyl transferase (DBTNBT) in the last step of taxol biosynthesis (Long et al., 2008).

Transcription factors (TFs) play important roles in the transcriptional regulation of taxol biosynthesis. Several MYB (TmMYB39, TcMYB29a and TmMYB3), WRKY (TcWKRY33, TcWKRY8 and TcWKRY47), ERF (TcERF12 and TcERF15), and MYC (TcMYC2a) family members and their taxol biosynthesis-related target genes have been functionally identified in various *Taxus* species (Cao et al., 2022; Chen et al., 2021;

Yu et al., 2022a; Yu et al., 2020; Zhang et al., 2018; Zhang et al., 2015). Mining novel TFs is, therefore, an effective way to accelerate the genetic improvement of *Taxus* trees.

During the past decade, a number of taxoids have been isolated and identified from *Taxus* leaves. New taxane diterpenoids, such as 9,13-diacetyltaxumairol W, 10,13-dibenzoyltaxacustin, 7,13-diacetylwallifoliol, 7,13-dibenzoylwallifoliol, and 7,9-dibenzoyltaxumairol P, have been isolated from *T. sumatrana* leaves (Shen et al., 2002). Three novel taxanes, namely baccatin VIII, baccatin IX, and baccatin X, were isolated from the leaves of *T. yunnanensis* (Hai et al., 2014). Seven taxanes, taxinine, taxinine A, taxinine B, 2-deacetoxytaxinine B, taxacin, taxchinin B, and taxol, were isolated as antiplatelet components from the leaves of *T. cuspidata* (Kim and Yun-Choi, 2010).

Recently, the spatial distribution of active components in medicinal plants has been a research focus. For example, the spatial distribution of the alkaloid components of sacred lotus leaves has been investigated (Liu et al., 2021b). In *Arabidopsis*, the distributions of flavonol glycosides, fatty acids, fatty acid esters, galactolipids, and glycosphingolipids vary significantly between the different substructures of leaves (Hieta et al., 2021). The spatial distribution of active alkaloids, coumarins, sugars, and organic acids has been visualized in *Clausena lansium* leaves (Tang et al., 2021). Data on the spatial distribution of secondary metabolites provide guidance for sampling and help to reveal the roles of metabolites in plant leaves.

In vascular plants, leaves, the primary sites of light acquisition and carbon fixation, harbor a number of cell types organized in different tissues (Kalve et al., 2014). Since different cell types carry nearly identical genetic information, variations in metabolism are controlled by transcriptional regulation during the process of cell differentiation (Wang et al., 2021). Newly developed single-cell RNA sequencing (ScRNA-seq) techniques enable the transcriptome profiles of individual cells to be characterized. In plants, the application of ScRNA-seq has been limited to a few plant species, depending on the protoplast preparation and availability of the genome sequence (Liu et al., 2022). ScRNA-seq provides a transcriptomic atlas, defining various developmental processes, including inflorescence development, mesophyll cell development, and xylem

formation (Li et al., 2021a; Tao et al., 2022; Zong et al., 2022). Recently, ScRNA-seq techniques have been applied to uncover novel secondary metabolite biosynthetic pathways. For examples, scRNA-seq revealed the biosynthetic pathway of a floral scent in *Nicotiana attenuata* corolla cells and a novel metabolic pathway of catechin esters in tea leaves (Kang et al., 2022; Wang et al., 2022a). Thus, scRNA-seq can be used extensively to reveal distinct gene expression patterns within heterogeneous cell populations.

To clarify the cell type specificity of secondary metabolite biosynthesis and accumulation, MS imaging, and ScRNA-seq were performed in this study. Matrix-assisted laser desorption/ionization (MALDI) imaging mass spectrometry (IMS) analysis is used to visualize various secondary metabolites in leaf sections, confirming the tissue-specific accumulation of different active metabolites. ScRNA-seq analysis can provide a metabolic profile of *T. mairei* leaves at the single-cell level. The identification of cell type specific TFs may strengthen the understanding of the transcriptional regulation of secondary metabolism in *T. mairei* leaves. Our study provides novel insights into the tissue specific accumulation and regulation of secondary metabolites in medicinal plants.

Results

Overview of leaf data from timsTOF fleX MALDI-2

To detect the accumulation pattern of active ingredients in the leaves of *T. mairei*, timsTOF fleX MALDI-2 was performed (**Fig. 1a** and **Fig.S1**). The raw MS imaging data were imported into SCiLS Lab workstation to process the MS features, including baseline deduction, peak alignment, data smoothing, and normalization (**Table S1**). Segmentation analysis showed that all detected features were located on a colored leaf map with 3510 points. The map separated the features into different locations, suggesting that the active ingredients accumulated in tissue-specific patterns (**Fig. 1b**). All detected features were further separated into four principal components (PCs), referring to different known leaf tissues, including phloem and spongy mesophyll cells

(SM, PC1), epidermal cells (EP, PC2), bundle sheath cells (BS, PC3), and palisade mesophyll cells (PM, PC4) (**Fig. 1c**).

The loading intensities of all features were displayed using a heatmap (**Fig. 1d** and **Table S2**). MeV clustering analysis grouped all features into six clusters. In detail, Cluster I contained 974 features highly accumulated in PM cells; Clusters II and III contained 717 and 485 features significantly accumulated in SM cells, respectively; Cluster IV contained 411 features mainly accumulated in EP cells; Cluster V contained 636 features predominantly accumulated in BS cells; and Cluster VI contained the 283 features highly accumulated in both EP and BS cells (**Fig. 1e**).

Visualization of active ingredient locations in *T. mairei* leaf

Our data detected numerous active ingredients, including heterocycles, steroids, coumarins, phenolics, terpenes, glycosides, and taxoids, in the leaves of *T. mairei*. Four typical active compounds belonging to each major class were visualized by color maps (**Fig. 2a**). The flavonoids ginkgetin, quercetin, and rutin were mainly distributed in the EP cells, and (R)-glabridin uniformly accumulated in four tissue groups. The steroids crustecdysone, cortisol, and ponasterone A were mainly distributed in the BS cells, and schottenol accumulated to a high level in the BS and PM cells. Four selected coumarin metabolites, including bergaptol, coumestan, coumesterol, and dicumarol, were detected at high levels in the EP cells, but rarely in the BS and PM cells. The phenolics (-)-gossypol and theaflavin were only detected in the BS cells, while procyanidin accumulated at high levels in the SM and BS cells, and xenognosin A accumulated to significant levels in the SM cells. The terpenes carnosol, kahweol, and lactaroviolin were mainly distributed in the SM cells, whereas squalene was mostly distributed in the PM cells. The glycosides chrysoeriol 7-*O*-neohesperidoside an glyvenol accumulated at high levels in the PM cells, and digitalin and pseudojervine evenly accumulated in the BS, SM, and PM cells. The taxoids 10-deacetyl cephalomannine, 10-deacetyl paclitaxel, baccatin III and 10-deacetyl baccatin III accumulated mainly in the SM cells (**Fig. 2b**). The detailed location information of selected active ingredients in *T. mairei*

leaf was listed in **Table S3**.

ScRNA-seq of *T. mairei* leaves

Leaf protoplasts were isolated to produce single cells (**Fig. 3a**). Approximately, 1500 protoplasts/μL were isolated from the leaves of *T. mairei*. The basic experimental procedures for library construction and scRNA-seq are shown in **Fig. 3b**. The purified protoplasts were processed using a commercial 10 × Chromium platform. The number of effective cells was evaluated using barcodes and UMI counts (**Fig. 3c**). ScRNA sequencing data were produced from a pool of 8846 cells with, on average, 35,874 reads and 2352 genes per cell (**Fig. 3d**). *t*-SNE projections of cells are shown as colored UMI counts (**Fig. 3e)**. The expression data of 36,808 genes were imported into Seurat loupe software, identifying a total of 15,552 cluster-enriched genes (**Table S4**). Based on the cluster-enriched genes, all cells were grouped into 15 clusters with uneven cell numbers (**Fig. S2a**). Cluster 0 contained the largest number of cells (1180 cells) and Cluster 14 contained the smallest number of cells (90 cells) (**Fig. S2b**). The *t*-SNE projection of cells belonging to each cluster are shown (**Fig. S2c)**. Three representative cluster-enriched genes for each cell cluster are shown in **Fig. 3f**.

Cell type marker and cluster annotation of leaves cells

A number of cell type marker genes for leaf have been identified in model plants (Liu *et al.*, 2022). By sequence alignment, several *Arabidopsis* leaf cell type marker genes were used to annotate the cell types (**Table S5**). Here, 11 cell type marker genes were applied to annotate the *T. mairei* leaf cell types: *1-AMINO-CYCLOPROPANE-1-CARBOXYLIC ACID OXIDASE 3* (*ACO3*, ctg2007-gene.6), a BS and vein cell-specific expressed gene, was enriched in the Cluster 3; *LIGHT HARVESTING COMPLEX PHOTOSYSTEM II SUBUNIT 6* (*LHCB6*, ctg18492_gene.1) and *CHLOROPHYLL A/B BINDING PROTEIN 3* (*CAB3*, ctg8753_gene.1), two leaf mesophyll cell-specific expressed genes, were enriched in Clusters 6 and 7; *FOUR LIPS* (*FLP*, ctg9019_gene.2) and *GLABRA 2* (*GL2*, ctg9664_gene.1), two leaf stomatal complex cell-specific

expressed genes, were enriched in Clusters 8 and 12; *PHLOEM INTERCALATED WITH XYLEM* (*PXY*, ctg3593_gene.5), a vascular/procambium cell-specific expressed gene, was enriched in Cluster 9; *ALTERED PHLOEM DEVELOPMENT* (*APL*, ctg8447_gene.3), a phloem cell-specific expressed gene, was enriched in Cluster 14; *MAP KINASE 9* (*MAPK9*, ctg4435_gene.11), a leaf guard cell-specific expressed gene, was enriched in Clusters 8 and 12; *SMALLER TRICHOMES WITH VARIABLE BRANCHES 2* (*SVB2*, ctg4709_gene.1) and *ECERIFERUM 3* (*CER3*, ctg560_gene.1), two leaf epidermis cell-specific expressed genes, were enriched in Cluster 12; and *RIBULOSE BISPHOSPHATE CARBOXYLASE SMALL CHAIN 1A* (*RBCS1A*, ctg596-gene.10), a leaf pavement cell-specific expressed gene, was enriched in Cluster 13 (**Fig. 3g**). Our data indicated the high degree of cell heterogeneity in *T. mairei* leaves.

Based on the cell type marker genes, most of the cell clusters were annotated. In detail, Cluster 3 contained BS cells and vein cells; Clusters 6 and 7 contained leaf mesophyll cells; Clusters 8 and 12 contained leaf stomatal complex cells, guard cells, and epidermal cells; Cluster 9 contained vascular cells and procambium cells; and Cluster 13 contained leaf pavement cells (**Fig. 3h**).

Cell type specificity of taxol biosynthesis in *T. mairei* leaves

A series of key enzymes are reported to be involved in the taxol biosynthesis pathway (Croteau *et al.*, 2006). The scRNA sequencing analysis detected many taxol biosynthesis-related genes, including three *TS* genes, four *T5OH* genes, four *T13OH* genes, seven *TAT* genes, four *T10OH* genes, five *TBT* genes, six *DBBT* genes, two *BAPT* genes, one *DBAT* gene, three *DBTNBT* genes, three *PAL* genes, three *T7OH* genes, and one *T14OH* gene (**Fig. 4a**). The cell coverage of gene expression was calculated to identify the major genes. Of the *TS* genes, ctg6088_gene.1 displayed the largest cell coverage (7.85%); of the *T5OH* genes, ctg7747_gene.2 was expressed in more than 5% of cells (8.17%); of the *T13OH* genes, the expression of ctg593_gene.6 occupied the largest cell coverage (5.37%); of the *TAT* genes, the expression pf ctg195_gene.31 showed the largest cell coverage (13.48%); of the *T10OH* genes, ctg2120_gene.5

displayed the largest cell coverage (5.51%); of the *TBT* genes, ctg6463_gene.2 showed the largest cell coverage (6.67%); of the *DBBT* genes, ctg12512_gene.2 occupied the large cell coverage (12.47%); of the *BAPT* genes, ctg5183_gene.4 displayed the largest cell coverage (3.96%); of the *DBTNBT* genes, ctg887_gene.16 showed the largest cell coverage (22.48%); of the *PAL* genes, ctg965_gene.9 showed the largest cell coverage (15.0%); and of the *T7OH* genes, ctg2120_gene.2 showed the largest cell coverage (4.79%) (**Fig. 4b**).

Expression analysis showed that most of the taxol biosynthesis-related genes were expressed in the leaf mesophyll cells (Clusters 6 and 7). However, the expressions of *T10OH* (ctg2120_gene.5), *TBT* (ctg6463_gene.2), *DBBT* (ctg4356_gene.1), and *DBTNB* (ctg887_gene.16) were detected in the BS and vein cells (Cluster 3) (**Fig. 4c**). UMAP visualization of the expression patterns of 12 major taxol-related genes are shown in **Fig. S3**. The successive differentiation trajectory was calculated to reveal the gene expression pattern during the development of leaf mesophyll tissue. The pseudotime trajectory has five branch points, dividing all cells into six branches (**Fig. 4d**). Our data suggested that most of the taxol biosynthesis pathway-related genes, such as *DBAT* (ctg10533_gene.2), *T10OH* (ctg2120_gene.10 and ctg5026_gene.32), *TAT* (ctg195_gene.31), and *TS* (ctg6088_gene.1), were expressed in the mature leaf mesophyll cells. Only one gene, *T14OH* (ctg8816_gene.2), was expressed in the young leaf mesophyll cells (**Fig. 4e**).

Main population of cells involved in taxol biosynthesis

Considering that most taxol biosynthesis-related genes were expressed in the leaf mesophyll cells (Clusters 6 and 7), based on the specific expressed marker genes, all of the leaf mesophyll cells were re-clustered into four subclasses: A to D (**Fig. S4a**). There were 168 cells in subclass A, 109 in subclass B, 103 in subclass C, and 79 in subclass D (**Fig. S4b**). UMAP visualization of the expression patterns of subclass A–D genes are shown in **Fig. S4c**. Interestingly, most of the taxol biosynthesis pathway genes were expressed in the cells belonging to subclasses C and D (**Fig. 4f**). The cells in subclasses

A and B were associated with young leaf mesophyll tissue and those in subclasses C and D with mature leaf mesophyll tissue. ScRNA data indicated that taxol biosynthesis occurred in a mature population of mesophyll cells, subclasses C and D, in Cluster 6/7.

Enrichment analysis grouped the Cluster 6/7-specific expressed genes into different GO and KEGG terms, respectively. Two taxol biosynthesis-related terms, 'paclitaxel metabolic process' (GO:0042616, P = 1.75E-34) and 'taxoid 14-beta-hydroxylase activity' (GO:0036203, P = 3.04E-32), were identified as significantly enriched GO terms (**Fig. S4d**). One terpenoid precursor biosynthesis-related term, 'Terpenoid backbone biosynthesis' (map00900, P = 2.80E-7), was identified as a significantly enriched KEGG term (**Fig. S4e**).

Cell specific expression pattern of the CYP super family

As a part of the classic taxol biosynthesis pathway, CYP-mediated oxygenation is required for the modification of the taxane skeleton (Jennewein et al., 2003). Based on the public *T. mairei* genome, 487 CYP genes, including 74 CYP725 subfamily genes were identified (**Table S6)**. Notably, the CYP725 subfamily genes were enriched in Cluster 6/7 (**Fig. 5a**). Re-clustering analysis showed that most of the CYP725 subfamily genes were expressed in cells belonging to subclasses C and D of Cluster 6/7. The expression patten of CYP725 subfamily genes showed high similarity to the pattern exhibited by taxol biosynthesis-related genes at the single cell level (**Fig. 5b**). For example, ctg7747_gene.4, ctg7598_gene.6, ctg7598_gene.5, ctg11883_gene.1, ctg6028_gene.1, and ctg7747_gene.3 were Cluster 7-specifically expressed CYP725 genes, providing several candidate genes to improve the taxol synthesis pathway (**Fig. 5c** and **d**). A pseudotime trajectory analysis showed that most of the CYP725 genes were expressed in the later stage of leaf mesophyll tissue development (**Fig. 5e**).

Cell type specificity for terpenoid and steroid metabolism

For terpenoid and steroid metabolism, three *DXS* genes, two *DXR* genes, one *CMK* gene, three *MDS* genes, two *HDS* genes, four *GPPS* genes, six *GGPPS* genes, one

319 *SQLE* gene, two *CYP51* genes, one *EBP* gene, and three *DWF1* genes were identified
320 by scRNA sequencing (**Fig. S5a** and **Table S7**). The expressions of ctg7293_gene.1
321 (*DXS*), ctg895_gene.1 (*DXR*), ctg10290_gene.1 and ctg739_gene.2 (*MDS*s),
322 ctg1206_gene.2 (*GPPS*), ctg1119_gene.5 (*CYP51*), and ctg5554_gene.1 (*DWF1*) were
323 detected in more than 10% of cells (**Fig. S5b**). Most of the terpenoid and steroid
324 biosynthesis-related genes were detected in Cluster 7 cells (**Fig. S5c**).
325 Expression analysis showed that most of the terpenoid and steroid biosynthesis
326 pathway genes were expressed in cells belonging to Clusters 6 and 7. According to the
327 annotation information, terpenoid and steroid biosynthesis-related genes might be
328 expressed in leaf mesophyll cells. The successive differentiation trajectory during the
329 development of leaf mesophyll tissue is shown in **Fig. S5d-e**. Some terpenoid and
330 steroid biosynthesis-related genes, such as ctg10290_gene.1 and ctg1119_gene.7, were
331 expressed in the young leaf mesophyll tissue, while other genes, such as
332 ctg2545_gene.5 and ctg6176_gene.1, were expressed in the mature leaf mesophyll
333 tissue (**Fig. S5f**).
334
335 **Cell type specificity for phenolic acid and flavonoid metabolism**
336 For phenolic acid metabolism, scRNA sequencing detected several genes encoding
337 three key enzymes, including three *C4H* genes, nine *C3H* genes, and six *COMT* genes
338 (**Table S8**). The cell coverage was calculated to identify the major genes involved in
339 phenolic acid metabolism. Of the *C4H* genes, ctg2905_gene.4 displayed the largest cell
340 coverage (14.84%); of the *C3H* genes, ctg6560_gene.3 showed the largest cell coverage
341 (2.78%); and of the *COMT* genes, ctg15240_gene.2 showed the largest cell coverage
342 (8.70%). For flavonoid metabolism, scRNA sequencing detected three *4CL* genes, eight
343 *CHS* genes, one *CHI* gene, two *F3H* genes, and six *FLS* genes. Of the *CHS* genes,
344 ctg3517_gene.14 occupied the largest cell coverage (4.66%); of the *F3H* genes, the
345 ctg17092_gene.1 displayed the largest cell coverage (4.51%); and of the *FLS* genes,
346 ctg2147_gene.1 showed the largest cell coverage (6.01%) (**Fig. S6a**).
347 Expression analysis showed that most of the phenolic acid and flavonoid

biosynthesis pathway genes were expressed in cells from Clusters 8, 11, and 12. According to their annotation information, phenolic acid and flavonoid biosynthesis-related genes may be expressed in the leaf EP cells, stomatal complex cells, and guard cells (**Fig. S6b**).

Discovery of cell specific TFs involved in the regulation of secondary metabolism

Although a number of TFs have been identified in *Taxus* species, information on the transcriptional regulation of secondary metabolism at the single cell level is limited. The expression levels of all TF genes were listed in **Table S9**. Considering that mesophyll tissue and leaf epidermal tissue are the main sites of secondary metabolism, several cell cluster-specific expressed TFs, including six MYB genes, five bHLH genes, four GT_2 genes, five ERF genes, and seven WRKY genes, were selected basing on the criteria log2(target cluster/other clusters) > 0.26 and $P < 0.01$. (**Table S10**). The expression patterns of the above TF genes in all annotated Clusters are shown in **Fig. 6a**. To identify the potential TFs involved in secondary metabolism, the 2000 bp promoter regions of 24 taxoid biosynthesis-related genes, 6 phenolic acid biosynthesis-related genes, 6 flavonoid metabolism-related genes, 13 terpenoid biosynthesis-related genes, and 5 steroid biosynthesis-related genes were extracted from the public *T. mairei* genome. The distribution of various *cis*-elements, such as MBE, GT element, W-box, ABRE, and GCC-box, in all selected promoters is shown in **Fig. 6b**.

Two criteria, similar expression pattern and matched binding elements, were applied to screen the cell specific TFs. For taxol biosynthesis, two *MYB* genes (ctg21726_gene.1 and ctg2891_gene.5), four *GT_2* genes (ctg2396_gene.1, ctg3435_gene.2, ctg3885_gene.7, and ctg21853_gene.2), three *bHLH* genes (ctg18566_gene.1, ctg458_gene.1, and ctg183_gene.1), five *WRKY* genes (ctg3845_gene.3, ctg8684_gene.3, ctg3000_gene.1, and ctg1880_gene.5), and three *ERF* genes (ctg996_gene.4, ctg4461_gene.2, and ctg505_gene.15) were identified as potential regulators. Furthermore, a number of TFs were predicted to be involved in various secondary metabolism pathways and the regulation network is shown in **Fig.6c**.

Verification of the binding of TFs to their potential targets

Six TFs, including MYB17 (ctg21726_gene.1), WRKY12 (ctg8684_gene.3), WRKY31 (ctg3845_gene.3), ERF13 (ctg996_gene.4), GT_2 (ctg3435_gene.2), and bHLH46 (ctg18566_gene.1), were randomly selected to verify the binding of cell specific expressed TFs to their target sites. The full-length sequences of these six TF genes were cloned and inserted into vectors. The EMSA results showed that MYB17 bound directly to the MBE element in the *TS* promoter, WRKY12 bound directly to the W-box in the *DBAT* promoter, WRKY31 bound directly to the W-box in the *SQE* promoter, ERF13 bound directly to the GCC-box in the *C3H* promoter, GT_2 bound directly to the GT element in the *CHS* promoter, and bHLH46 bound directly to the ABRE element in the *GGPPS* promoter (**Fig. 7a-f**).

To determine the transcriptional activities of MYB17, WRKY12, WRKY31, ERF13, GT_2, and bHLH46, a dual-LUC reporter assay was performed using a tobacco transient expression system (**Fig. 7g**). The results showed that WRKY12, WRKY31, GT_2, and ERF13 significantly activated the expression of their downstream target genes (*DBAT*, *SQE*, *CHS*, and *C3H*), respectively, whereas, MYB17 and bHLH46 significantly inhibited the expression of their downstream target genes (*TS* and *GGPPS*), respectively (**Fig. 7h**).

Discussion

The complexity of biochemical processes in plant tissue is not only programed by genetic information but is also affected by the chemical's localization. Thousands of secondary metabolites, including flavonoids, taxoids, polysaccharides, and abietane diterpenoids, have been detected in the *T. mairei* leaves (Hao et al., 2017; Monacelli et al., 2002; Wang et al., 2019a; Yang et al., 2016). However, the precise locations of different types of secondary metabolites in *T. mairei* leaves are largely unknown.

High-throughput MALDI-2 imaging is a recently developed tool used to explore the spatial distribution patterns of chemicals in plant tissues (Bien et al., 2022). Mesophyll

406 tissue is the main site of photosynthesis, which provides energy and raw materials for
407 the biosynthesis of primary and secondary metabolites (Allahverdiyeva et al., 2015).
408 Statistical results showed that the largest number of metabolites were enriched in SM
409 (1202 metabolites) and PM (947 metabolites), suggesting that *T. mairei* mesophyll cell-
410 related tissues are the main sites of active compound biosynthesis. Due to their different
411 biological functions, various types of metabolites preferentially accumulated in specific
412 tissues. For example, three classic flavonoids, ginkgetin, quercetin, and rutin,
413 preferentially accumulated in the EP cells, matching their roles in protecting plants
414 against exposure to UV radiation and high light conditions (Ferreyra et al., 2021).
415 MALDI-2 imaging further showed that several coumarins, including bergaptol and
416 coumestan, accumulated at high levels in the leaf epidermal cells. EP cell-accumulated
417 coumarins play an important role in the response to nutrient-deficient conditions (Robe
418 et al., 2021). Furthermore, most of the detected steroids, phenolics, and terpenes
419 accumulated at high levels in the SM cells, and most of the detected glycosides and
420 taxoids accumulated in the SM and PM cells. Our data provide detailed and precise
421 location information on different types of active compounds to promote the
422 comprehensive utilization of *Taxus* leaves.

423 *T. mairei* leaves provide an important, reproducible resource for the industrial
424 extraction of active ingredients, especially taxoids (Li et al., 2021b). MALDI-2 imaging
425 results confirmed a distinctive accumulation pattern of active ingredients in different *T.*
426 *mairei* leaf cell types. The revolutionary application of 10× Genomics technology in
427 plant scRNA-seq enhanced the ability to analyze the heterogeneity of gene expression
428 at the single-cell level (Liu *et al.*, 2022). To date, leaf single-cell transcriptional profiles
429 of several plants have been reported, including those of *Nepeta tenuifolia*, *Arabidopsis*,
430 rice, peanut, *Brassica rapa*, and tea (Liu et al., 2021a; Lopez-Anido et al., 2021; Wang
431 *et al.*, 2022a; Wang *et al.*, 2021; Zhou et al., 2022a). In these studies, although the
432 number of captured cells varied in different experiments, the median number of genes
433 per cell was less than 2000. In the present study, the median number of genes per cell
434 was 2352, suggesting sufficient sequencing depth to screen valuable genes.

All *T. mairei* leaf cells were grouped into a number of cell clusters. Due to the lack of marker genes, cell cluster annotation is difficult to overcome in medicinal plants. In model plants, many cell type marker genes have been confirmed by *in situ* hybridization and reporter systems. For examples, *in situ* hybridization confirmed that *OsDEF7* (LOC_Os02g41904) is highly expressed in inner mesophyll cells and *OsENOD93* (LOC_Os06g05000) is highly expressed in mature mesophyll cells (Wang *et al.*, 2021); a reporter system confirmed that *PbZIP9* is expressed specifically in leaf vasculature cells (Kim et al., 2021); and a GFP fluorescence system confirmed that *AtMYS1* (AT2G38300) is specifically expressed in guard cells (Zhang et al., 2021). According to known markers (PCMDB database), 11 cluster specific expressed genes were used to annotate the cell types covering the major leaf tissues in *T. mairei* leaves. Due to the lack of reliable marker genes, several big clusters could not be annotated with known marker genes. Establishing a rapid genetic transformation system and *in situ* hybridization system will be essential to improving cell annotation in *T. mairei* in the future.

The biosynthesis of endogenous active components involves a number of rate-limiting enzymes, which are encoded by multiple genes (Zhou et al., 2022b). In the present study, differences in the cell converge of different encoding genes were carefully evaluated. TS catalyzes the formation of taxadiene, the first committed precursor to taxol (Croteau *et al.*, 2006). Among the three *TS* genes, the expression of ctg5306_gene.4 and ctg7747_gene.1 is silent in most leaf cells. As an important taxol rate-limiting enzyme, BAPT is encoded by two genes (ctg5183_gene.3 and ctg5183_gene.4). Interestingly, the cell coverage of ctg5183_gene.4 is about 50-fold than that of ctg5183_gene.3. TBT, which transforms the debenzoyl-intermediate to 10-DAB, is another important rate-limiting enzyme involved in the chemical modification of the taxoid core structure (Walker and Croteau, 2000). Among the five *TBT* genes, only ctg6463_gene.2 was expressed in more than 5% of leaf cells. Similarly, numerous silenced genes involved in the terpenoid-, steroid-, phenolic acid-, and flavonoid-metabolism pathways were identified. Elevation of the cell coverage of the silenced

464 genes is potentially a way to promote efficiency in active compound biosynthesis.

465 Genetic improvement of medicinal plants requires the spatial screening of key
466 metabolic genes (Wang et al., 2022b). Most taxol biosynthesis-related genes were
467 expressed in the leaf mesophyll cells, and re-clustering analysis further focused them
468 into subclasses C and D of Cluster6/7, suggesting that there is a specific cell group for
469 taxol biosynthesis within mesophyll cells. The terpenoid biosynthesis pathway provides
470 the precursors, GGPP, for synthesizing taxol in *Taxus* species (Croteau *et al.*, 2006).
471 The main site of terpenoid biosynthesis pathway gene expression is also the leaf
472 mesophyll cells, guaranteeing a supply of precursors for taxol synthesis (Soliman and
473 Tang, 2015). Carbon flux to steroid biosynthesis is a competitive pathway to terpenoid
474 biosynthesis (Syklowska-Baranek et al., 2022). Clustering analysis showed that the
475 main expression sites for steroid biosynthesis-related genes was similar to that for
476 terpenoid biosynthesis genes. Down-regulation of the terpenoid biosynthesis genes has
477 been shown to be a useful approach to improving the production of taxol (Oksman-
478 Caldentey and Inze, 2004). Flavonoids and phenolic acid derivatives are important
479 photo-protective compounds and antioxidants (Casanova et al., 2020). Significantly,
480 phenolic acid and flavonoid metabolism-related genes are specifically expressed in leaf
481 EP cells and stomatal complex cells. The accumulation of phenolic acids and flavonoids
482 in leaf EP cells might play a role in protection against the invasion of pathogenic
483 microorganisms and damaging UV light on the front line (Schafer and Wink, 2009).

484 Plant secondary metabolism is conducted by a complex TF network (Chen et al.,
485 2020; Feng et al., 2023; Yu et al., 2022b; Yu *et al.*, 2020; Yu et al., 2021; Zhan et al.,
486 2022). To date, several TFs, including stimulus responsive TFs (such as *TcERF12* and
487 *TcERF15*), tissue-specific expressed TFs (such as *TmMYB3*), and species specific
488 expressed TFs (such as *TmMYB39*), have been functionally identified in various *Taxus*
489 species (Yu *et al.*, 2022a; Yu *et al.*, 2020; Zhang *et al.*, 2015; Zhou et al., 2019). After
490 searching the *T. mairei* genome, a large number of TFs were annotated in the ScRNA-
491 seq data set (Xiong et al., 2021). To screen novel TFs at single-cell resolution, we
492 proposed two screening criteria: *cis*-element based potential physical bonding

capability and matched spatio-temporal expression patterns between the TF and its targets.

Based on these criteria, six MYBs, five bHLHs, four GT_2s, five ERFs, and seven WRKYs, were predicted to be involved in a regulation network of secondary metabolism in *T. mairei* leaves. Our previous studies have identified two MYB family members and their target genes. *TmMYB39*, a female-predominantly expressed TF, regulates the taxol biosynthesis pathway by targeting *T10OH*, *T13OH*, and *TBT* genes (Yu *et al.*, 2022a). *TmMYB3*, a phloem-specific expressed TF, enhances taxol biosynthesis by binding to the promoters of *TBT*, *DBTNBT*, and *TS* (Yu *et al.*, 2020). Our results suggested a new TcMYB17-TS pair, enriching the understanding of the role of the MYB family in taxol biosynthesis. In *T. mairei*, jasmonate-induced WRKY26 and salicylic acid-induced WRKY33 activated the expression of a *DBAT* gene by binding to the W-box in its promoter region (Chen et al., 2022; Chen *et al.*, 2021). In our study, WRKY12 was also identified as a regulator of *DBAT* expression, indicating that WRKY12 might be an important target of phytohormone-induced taxol biosynthesis.

In *Taxus* species, TFs involved in secondary metabolism, except taxol biosynthesis, are largely unknown. SQE acts as a rate-limiting enzyme in the biosynthesis of steroidal compounds. Recently, two *Asparagus racemosus* bZIP family TFs (TGA1 and TGA2) and two *Panax notoginseng* AP2/ERF family TFs (ERF2 and ERF3) were considered to be involved in the regulation of *SQE* expression (Lin et al., 2020; Upadhyay et al., 2020). In *T. mairei*, *SQE* (ctg680_gene.8) is highly expressed in mesophyll and stomatal complex cells, and its promoter contains multiple *cis*-elements, such as MBE, ABRE, and W-box. Thus, mesophyll and stomatal complex cell-specific expressed *MYB*, *bHLH*, and *WRKY* family TFs are potential regulators of the steroid biosynthesis pathway. In bean, the silencer region of the *CHS15* promoter contains multiple binding sites for SBF1, a GT_1 like factor (Harrison et al., 1991). In *T. mairei*, the promoter of the *CHS* gene (ctg13079_gene.1) contains multiple GT_2 binding sites. *In vitro* and *in vivo* assays confirmed the role of epidermis-specific expressed GT_2 in the activation of

CHS expression, leading to the accumulation of flavonoids in EPs. The gene expression patterns were largely consistent with the location of compounds from each category.

Our research establishes a transcriptional landscape of the major cell types in *T. mairei* leaves at single-cell resolution and provides a valuable resource for studying the basic principles of cell type-specific regulation of secondary metabolism (**Fig. 7i-k**).

Materials and Methods

Plant materials and genome data

Five-year-old *T. mairei* trees were grown in an open field in the 'Cangqian' campus of Hangzhou Normal University, Hangzhou, China. Leaves of young twigs were harvested for microsection and protoplasting. The reference *T. mairei* genome was downloaded from the Genome Sequence Archive at the National Genomics Data Center (Xiong *et al.*, 2021).

Sample preparation and MS imaging

The young leaves of *T. mairei* were frozen and cut into 20 μm-thick sections using Leica CM1950 freezing microtome. The resulting sections were transferred onto the MALDI-2 special conductive glass slide. Then, the slice was vacuum dried under a stable stream of nitrogen and transferred to -80℃ for sealed storage. MALDI-2 special matrix solution containing 2,5-dihydroxybenzoic acid and 2,5-dihydroxyacetophenone was evenly sprayed on the slide with leaf sections.

The coated conductive slide was put on the target plate of TimsTOF flex MALDI-2 system, and the detection area was screened using Bruker Data imaging software. After setting the imaging resolution, the detection area was divided into several two-dimensional points according to the section size. The leaf slice was scanned by laser radiation to detect the ionized released molecules at the target site. Details of the experimental setup and data processing are provided as follows: detection range of metabolite charge ratio (150-1200 Da); ion detection mode (positive); MADLI-2 (on); ion mobility (on); rise time (300 ms); 1/K0 range (0.6-1.7 V.s/cm2); imaging resolution

(20 μm); laser frequency (HZ 1000); laser energy (44%); and acquisition times per pixel (50).

MS imaging data processing

The MS data was loaded to SCiLS Lab MVS software (ver.2021a Pro, Bruker Daltonics) to generate MS images. A reduced mass list of all the peaks was stored in imzML format for further processing. The imaging pixels in the target area were clustered using unsupervised spatial clustering analysis function of SCiLS software. The regions with similar accumulation pattern were marked with the same color to intuitively classify the slices from the molecular level. For metabolite annotation, MS spectrum data was extracted and qualitied by Bruker Metaboscape TM workstation. The resulting data was searched against the Bruker Library MS Metabase 3.0 database (https://www.bruker.com/en/products-and-solutions/mass-spectrometry/ms-software/metabolomics-spectral-libraries.html). Molecular mass error was set <10 ppm. Principal component analysis (PCA) is used to analyze the distribution of ion features on the imaging.

Preparation of *Taxus* leaf protoplasts

About 20 tender leaves of *T. mairei* were harvested and cut into small pieces. The leaf samples were treated with enzymolysis solution (cellulase R-10, isolation enzyme P, mannitol, KCl and MES, pH = 5.7) in a triangular flask and vacuum infiltrated for 10 min. The leaf samples were incubated in darkness for 4–5 h at 28℃ to isolate protoplasts. After reaching a certain number of protoplasts, the enzymolysis solution was removed. The isolated protoplasts were washed with 8% mannitol solution three times and filtered with a 40-μm cell sifter. Protoplast cell activity was detected by trypan blue staining. The dead cells were removed using a Miltenyi® Dead Cell Removal Kit (MACS 130-090-101).

ScRNA-seq library construction and raw data processing

For library construction, single-cell suspensions were loaded to 10 × Chromium according to the manufacturer's instructions of a 10 × Genomics Chromium Single-Cell 3 kit (V3). Libraries were sequenced on an Illumina NovaSeq 6000 sequencing system by LC-BioTechnology Co. Ltd., (Hangzhou, China) at a minimum depth of 20,000 reads per cell. The paired-end multiplexing run was set at 150 bp.

Sequencing results were downloaded and converted to FASTQ format using Illumina bcl2fastq software (v. 2.20). Sample demultiplexing, barcode processing, and single-cell 3′ gene counting were carried out using the Cell Ranger pipeline (v. 6.1.1). scRNA-seq data were aligned to the *T. mairei* reference genome. The quality control parameters, such as Q30, median unique molecular identifier (UMI), and gene number per cell, were determined using CellRanger (v. 3.0.0) software and Seurat software package (v. 2.3.4). To remove the low-quality cells, the data were filtered according to the average number of genes detected in a single cell. Specifically, cells with detected genes < 500 and > 5000, mitochondrial reads > 5%, and mitochondrial reads > 20% were considered to be doublets and were removed by DoubletFinder software (v. 2.0.2).

Cell visualization and clustering

To visualize the data, the dimensionality of all resulting cells were reduced into 2D-apace using Seurat software. Th expression levels of genes in each cell were calculated using the 'Normalization' function in Seurat software. The normalized expression values were used for principal component analysis (PCA) and the top 10 PCs were used for clustering and t-SNE analysis with a weighted graph-based clustering method. The marker genes for each cell cluster were selected using the 'FindAllMarkers' function in Seurat software with Wilcoxon rank-sum test and Likelihood-ratio test.

Tissue-specific marker genes and leaf cell type identification

To identify the cell type specific markers, the orthologous gene alignments of the reported leaf-related marker genes in model plant Arabidopsis. The detailed sequences of Arabidopsis marker genes were downloaded from the Plant Cell Marker DataBase

(http://www.tobaccodb.org/pcmdb/) (Jin et al., 2022). Using the Arabidopsis marker genes as the query sequences, the homologous genes of the *T. mairei* were searched by BLASTP function in TBtools software. The top score hits were selected and annotated as corresponding *T. mairei* cell types.

Pseudo-time trajectory analysis

To visualize the trajectory in a reduced dimensional space, all cells were ordered using Monocle software (v. 3.0) with a matrix of cells and gene expression profiles. For pseudo-time analysis, the sequence count was first converted into the 'CellDataSet' with the 'importCDS' function in Monocle 2. Then, we used the 'estimateSizeFactors' and 'estimateDispersions' functions to pre-calculate several parameters. 'Differential GeneTest' module in Monocle software was applied to screen the differentially expressed genes between different groups of cells. Dimensional reduction clustering analysis was carried out using the 'reduceDinension' function with default parameters. Then, gene expression was plotted to track changes during pseudo-time according to a previously published method (Liu *et al.*, 2022).

Electrophoretic Mobility Shift Assay (EMSA)

The 2000 bp promoter sequences of various secondary metabolism-related genes were extracted from the *T. mairei* genome (Xiong *et al.*, 2021). The promoter sequences were uploaded and scanned using the PlantCARE program (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/).

The cDNAs of *MYB17*, *WRKY12*, *WRKY31*, *ERF13*, *GT_2*, and *bHLH46* were cloned and inserted into pGEX4-T vector. The resulting vector was transformed into *Escherichia coli* cells (DE3) to produce recombinant TF-His tag protein. The IPTG (1mM) induced recombinant protein was purified by His60 Ni Superflow Resin (Clontech, Beijing, China). The purified His fusion proteins were isolated by 12% SDS-PAGE electrophoresis and used for further EMSA experiment.

EMSA experiment was performed using the Light Shift Chemiluminescent EMSA

kit (Beyotime, Shanghai, China) according to its protocol (Yu *et al.*, 2020). In detail, two MYB binding elements (CAGTTG and CAACTG), two bHLH binding elements (CACGT and CACGTG), one WKRY binding element (TTGACC), two GT_2 binding elements (GGTAATT and GGTAAAT), and two ERF binding elements (GCCGCC and CCGAC), were used to design probes. Potential TF binding elements derived from the downstream promoters were labeled with 5'6-FAM fluorescent dye. Unlabeled sequences were used as competition probes and sequences with 'CCCGGG' were used as mutation probes.

Dual-Luciferase reporter (Dual-LUC) assays

The cDNAs of several selected TF genes were cloned into the pGreenII62-SK vector as effectors and the downstream promoter sequences were cloned into the pGreenII0800-LUC vector as reporters, respectively. The completed constructs were co-expressed in tobacco leaves via *Agrobacterium tumefaciens* GV3101 strain-mediated transient transformation technology. The binding activities of TFs to the selected promoters were calculated from the LUC/REN ratio, which was determined using a dual luciferase assay kit (Promega, Beijing, China).

Bioinformatics and statistical analysis

The enrichment of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) for the cell-specific expressed genes were analyzed by the Blast2GO program (https://www.blast2go.com/). The statistical analysis was performed using SPSS software ver.19.0 (SPSS Inc., Chicago, IL). Significance was determined at $P < 0.05$ by analysis of variance (ANOVA) followed by Duncan's least significant difference test. For hierarchical clustering and visualization, all data were Z-score normalized using MeV software by Euclidean distance. For the MS imaging analysis, two biological repeats were applied. For the EMSA experiment, three technician repeats were applied. For the Dual-LUC assay, three biological repeats were applied.

Data availability

The *T. mairei* reference genome was downloaded from NCBI database (ID: PRJNA730337). scRNA-seq data are available at the NCBI database (BioProject ID PRJNA909435). The datasets generated and analysed during the current study are available in the 'Baidu Netdisk' (https://pan.baidu.com/s/1hXL8X9bCF5ZLi4UGKXRlHQ; (password: tjhg)).

Supplemental Information

Figure S1 The secondary biological repeat for timsTOF fleX MALDI-2 analysis.

Figure S2 The transcriptomic profiles of the genes with high variation in expression level.

Figure S3 UMAP visualization of expression patterns of 12 major taxol-related genes.

Figure S4 Re-clustering of the leaf mesophyll cells.

Figure S5 Cell type specificity for terpenoid and steroid metabolism.

Figure S6 Cell type specificity of the phenolic acid and flavonoid metabolisms.

Table S1 The raw data of timsTOF fleX MALDI-2 analysis.

Table S2 The loading intensities of all features in four PCs.

Table S3 The detailed information of selected active ingredient locations in *T. mairei* leaf.

Table S4 The detail information of the cluster-enriched genes.

Table S5 Gene list used for cell-type identification in this study.

Table S6 The ID list of CYP superfamily genes in *T. mairei* genome.

Table S7 The genes involved in the terpenoid and steroid metabolism.

Table S8 The genes involved in the phenolic acid and flavonoid metabolism.

Table S9 The expression levels of all TF genes in different cell clusters.

Table S10 The related expression level of selected TF genes in different cell clusters.

Fundings

This work was funded by National Natural Science Foundation of China (32271905

and 32270382); Zhejiang Provincial Natural Science Foundation of China under Grant No. LY23C160001, LY18C050005, LY19C150005, LY19C160001; Opening Project of Zhejiang Provincial Key Laboratory of Forest Aromatic Plants-based Healthcare Functions (2022E10008); Open Foundation of State Key Laboratory of Subtropical Silviculture, Zhejiang A & F University (KF201708); Major Increase or Decrease Program in The Central Finance Level (Grant No. 2060302); and Zhejiang Provincial Key Research & Development Project Grants (2017C02011, 2018C02030).

Acknowledgments

We are grateful to LC Sciences company (Hangzhou, China) for transcriptomic analysis. We are also grateful to Professor Jianbin Yan (Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China) for genomic analysis.

Conflict of Interest Statement

The authors have no relevant financial or non-financial interests to disclose.

Authors' Contributions

Z.X. S.F. H.Z.(Houqing Zeng) and C.S. conceptualized the initial study; C.S., H.Z. (Hongshan Zhang) M.W. and Q.W. were involved in the experimental layout; H.Z.(Hongshan Zhang), Q.W., T.Q., H.K., Z.W., C.C. X.W., X.L. (Xueshuang Liang) and X.L. (Xiao-lin Li) performed the lab experiments, S.F., H.W, C.Y. and C.S. drafted the initial article; all authors discussed the results, reviewed the article, and approved the final article.

References

Allahverdiyeva, Y., Suorsa, M., Tikkanen, M., and Aro, E.M. (2015). Photoprotection of photosystems in fluctuating light intensities. Journal of experimental botany **66**:2427-2436. 10.1093/jxb/eru463.

Bien, T., Koerfer, K., Schwenzfeier, J., Dreisewerd, K., and Soltwisch, J. (2022). Mass spectrometry imaging to explore molecular heterogeneity in cell culture. Proceedings of the National Academy of Sciences of the United States of America **119**:e2114365119. 10.1073/pnas.2114365119.

Cao, X., Xu, L., Li, L., Wan, W., and Jiang, J. (2022). TcMYB29a, an ABA-responsive R2R3-MYB transcriptional factor, upregulates Taxol biosynthesis in *Taxus chinensis*. Front Plant Sci **13**:804593. 10.3389/fpls.2022.804593.

Casanova, J.M., Dos Santos Nascimento, L.B., Casanova, L.M., Leal-Costa, M.V., Costa, S.S., and Tavares, E.S. (2020). Differential distribution of flavonoids and phenolic acids in leaves of kalanchoe delagoensis Ecklon & Zeyher (Crassulaceae). Microsc Microanal **26**:1061-1068. 10.1017/S1431927620024344.

Chen, L., Wu, L., Yang, L., Yu, H., Huang, P., Wang, Y., Yao, R., and Zhang, M. (2022). TcJAV3-TcWRKY26 cascade is a missing link in the jasmonate-activated expression of Taxol biosynthesis gene DBAT in *Taxus chinensis*. Int J Mol Sci **23**10.3390/ijms232113194.

Chen, Y., Zhang, M., Jin, X., Tao, H., Wang, Y., Peng, B., Fu, C., and Yu, L. (2020). Transcriptional reprogramming strategies and miRNA-mediated regulation networks of *Taxus media* induced into callus cells from tissues. BMC genomics **21**:168. 10.1186/s12864-020-6576-2.

Chen, Y., Zhang, H., Zhang, M., Zhang, W., Ou, Z., Peng, Z., Fu, C., Zhao, C., and Yu, L. (2021). Salicylic acid-responsive factor TcWRKY33 positively regulates Taxol biosynthesis in *Taxus chinensis* in direct and indirect ways. Front Plant Sci **12**:697476. 10.3389/fpls.2021.697476.

Croteau, R., Ketchum, R.E., Long, R.M., Kaspera, R., and Wildung, M.R. (2006). Taxol biosynthesis and molecular genetics. Phytochemistry reviews : proceedings of the Phytochemical Society of Europe **5**:75-97. 10.1007/s11101-005-3748-2.

Dang, P.H., Nguyen, H.X., Duong, T.T.T., Tran, T.K.T., Nguyen, P.T., Vu, T.K.T., Vuong, H.C., Phan, N.H.T., Nguyen, M.T.T., Nguyen, N.T., et al. (2017). alpha-Glucosidase inhibitory and cytotoxic taxane diterpenoids from the stem bark of *Taxus wallichiana*. J Nat Prod **80**:1087-1095. 10.1021/acs.jnatprod.7b00006.

Feng, S., Kailin, H., Zhang, H., Chen, C., Huang, J., Wu, Q., Zhang, Z., Gao, Y., Wu, X., Wang, H., et al. (2023). Investigation of the role of TmMYB16/123 and their targets (TmMTP1/11) in the tolerance of *Taxus media* to cadmium. Tree Physiol 10.1093/treephys/tpad019.

Ferreyra, M.L.F., Serra, P., and Casati, P. (2021). Recent advances on the roles of flavonoids as plant protective molecules after UV and high light exposure. Physiol Plant **173**:736-749. 10.1111/ppl.13543.

Hai, P., Wen, S.Z., Li, Y., Gao, Y., Jiang, X.J., and Wang, F. (2014). New taxane diterpenoids from *Taxus yunnanensis*. Nat Prod Bioprospect **4**:47-51. 10.1007/s13659-

014-0003-9.

Hao, J., Guo, H., Shi, X., Wang, Y., Wan, Q., Song, Y.B., Zhang, L., Dong, M., and Shen, C. (2017). Comparative proteomic analyses of two *Taxus* species (*Taxus* × *media* and *Taxus mairei*) reveals variations in the metabolisms associated with paclitaxel and other metabolites. Plant & cell physiology **58**:1878-1890. 10.1093/pcp/pcx128.

Harrison, M.J., Lawton, M.A., Lamb, C.J., and Dixon, R.A. (1991). Characterization of a nuclear protein that binds to three elements within the silencer region of a bean chalcone synthase gene promoter. Proceedings of the National Academy of Sciences of the United States of America **88**:2515-2519. 10.1073/pnas.88.6.2515.

Hieta, J.P., Sipari, N., Raikkonen, H., Keinanen, M., and Kostiainen, R. (2021). Mass spectrometry imaging of *Arabidopsis thaliana* leaves at the single-cell level by infrared laser ablation atmospheric pressure photoionization (LAAPPI). J Am Soc Mass Spectrom **32**:2895-2903. 10.1021/jasms.1c00295.

Jennewein, S., Rithner, C.D., Williams, R.M., and Croteau, R. (2003). Taxoid metabolism: Taxoid 14beta-hydroxylase is a cytochrome P450-dependent monooxygenase. Archives of biochemistry and biophysics **413**:262-270. 10.1016/s0003-9861(03)00090-0.

Jin, J., Lu, P., Xu, Y., Tao, J., Li, Z., Wang, S., Yu, S., Wang, C., Xie, X., Gao, J., et al. (2022). PCMDB: a curated and comprehensive resource of plant cell markers. Nucleic acids research **50**:D1448-D1455. 10.1093/nar/gkab949.

Kalve, S., De Vos, D., and Beemster, G.T. (2014). Leaf development: a cellular perspective. Front Plant Sci **5**:362. 10.3389/fpls.2014.00362.

Kang, M., Choi, Y., Kim, H., and Kim, S.G. (2022). Single-cell RNA-sequencing of *Nicotiana attenuata* corolla cells reveals the biosynthetic pathway of a floral scent. The New phytologist **234**:527-544. 10.1111/nph.17992.

Kaspera, R., and Croteau, R. (2006). Cytochrome P450 oxygenases of Taxol biosynthesis. Phytochemistry reviews : proceedings of the Phytochemical Society of Europe **5**:433-444. 10.1007/s11101-006-9006-4.

Kim, J.Y., Symeonidi, E., Pang, T.Y., Denyer, T., Weidauer, D., Bezrutczyk, M., Miras, M., Zollner, N., Hartwig, T., Wudick, M.M., et al. (2021). Distinct identities of leaf phloem cells revealed by single cell transcriptomics. Plant Cell **33**:511-530. 10.1093/plcell/koaa060.

Kim, S.Y., and Yun-Choi, H.S. (2010). A comparative optical aggregometry study of antiplatelet activity of taxanes from *Taxus cuspidata*. Thromb Res **125**:e281-284. 10.1016/j.thromres.2009.12.024.

Kui, L., Majeed, A., and Dong, Y. (2022). Reference-grade *Taxus* genome unleashes its pharmacological potential. Trends Plant Sci **27**:10-12. 10.1016/j.tplants.2021.10.010.

Li, H., Dai, X., Huang, X., Xu, M., Wang, Q., Yan, X., Sederoff, R.R., and Li, Q. (2021a). Single-cell RNA sequencing reveals a high-resolution cell atlas of xylem in Populus. J Integr Plant Biol **63**:1906-1921. 10.1111/jipb.13159.

Li, L., Chen, Y., Ma, Y., Wang, Z., Wang, T., and Xie, Y. (2021b). Optimization of Taxol extraction process using response surface methodology and investigation of temporal and spatial distribution of Taxol in *Taxus mairei*. Molecules

2610.3390/molecules26185485.

Lin, T., Du, J., Zheng, X., Zhou, P., Li, P., and Lu, X. (2020). Comparative transcriptome analysis of MeJA-responsive AP2/ERF transcription factors involved in notoginsenosides biosynthesis. 3 Biotech **10**:290. 10.1007/s13205-020-02246-w.

Liu, H., Hu, D., Du, P., Wang, L., Liang, X., Li, H., Lu, Q., Li, S., Liu, H., Chen, X., et al. (2021a). 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. 10.1111/pbi.13656.

Liu, J., Shi, K., Shi, J., Feng, Y., Hao, C., Peng, J., and Chen, S. (2021b). A simple strategy to monitor the temporal and spatial distribution of alkaloids in sacred lotus leaves. Bioscience, biotechnology, and biochemistry **85**:1332-1340. 10.1093/bbb/zbab038.

Liu, Z., Yu, X., Qin, A., Zhao, Z., Liu, Y., Sun, S., Liu, H., Guo, C., Wu, R., Yang, J., et al. (2022). Research strategies for single-cell transcriptome analysis in plant leaves. The Plant journal : for cell and molecular biology **112**:27-37. 10.1111/tpj.15927.

Long, R.M., Lagisetti, C., Coates, R.M., and Croteau, R.B. (2008). Specificity of the *N*-benzoyl transferase responsible for the last step of Taxol biosynthesis. Archives of biochemistry and biophysics **477**:384-389. 10.1016/j.abb.2008.06.021.

Lopez-Anido, C.B., Vaten, A., Smoot, N.K., Sharma, N., Guo, V., Gong, Y., Anleu Gil, M.X., Weimer, A.K., and Bergmann, D.C. (2021). Single-cell resolution of lineage trajectories in the Arabidopsis stomatal lineage and developing leaf. Dev Cell **56**:1043-1055 e1044. 10.1016/j.devcel.2021.03.014.

Monacelli, B., Pasqua, G., Botta, B., Vinciguerra, V., Gacs-Baitz, E., and Monache, G.D. (2002). Abietane diterpenoids from callus cultures of *Taxus baccata*. Planta medica **68**:764-766. 10.1055/s-2002-33802.

Oksman-Caldentey, K.M., and Inze, D. (2004). Plant cell factories in the post-genomic era: new ways to produce designer secondary metabolites. Trends Plant Sci **9**:433-440. 10.1016/j.tplants.2004.07.006.

Robe, K., Conejero, G., Gao, F., Lefebvre-Legendre, L., Sylvestre-Gonon, E., Rofidal, V., Hem, S., Rouhier, N., Barberon, M., Hecker, A., et al. (2021). Coumarin accumulation and trafficking in *Arabidopsis thaliana*: a complex and dynamic process. The New phytologist **229**:2062-2079. 10.1111/nph.17090.

Schafer, H., and Wink, M. (2009). Medicinally important secondary metabolites in recombinant microorganisms or plants: progress in alkaloid biosynthesis. Biotechnol J **4**:1684-1703. 10.1002/biot.200900229.

Shen, Y.C., Wang, S.S., Pan, Y.L., Lo, K.L., Chakraborty, R., Chien, C.T., Kuo, Y.H., and Lin, Y.C. (2002). New taxane diterpenoids from the leaves and twigs of *Taxus sumatrana*. J Nat Prod **65**:1848-1852. 10.1021/np0202273.

Soliman, S., and Tang, Y. (2015). Natural and engineered production of taxadiene with taxadiene synthase. Biotechnology and bioengineering **112**:229-235. 10.1002/bit.25468.

Syklowska-Baranek, K., Kaminska, M., Paczkowski, C., Pietrosiuk, A., and Szakiel, A. (2022). Metabolic modifications in terpenoid and steroid pathways triggered by methyl jasmonate in *Taxus* × *media* hairy roots. Plants (Basel) **11**10.3390/plants11091120.

Tang, X., Zhao, M., Chen, Z., Huang, J., Chen, Y., Wang, F., and Wan, K. (2021). Visualizing the spatial distribution of metabolites in *Clausena lansium* (Lour.) skeels using matrix-assisted laser desorption/ionization mass spectrometry imaging. Phytochemistry **192**:112930. 10.1016/j.phytochem.2021.112930.

Tao, S., Liu, P., Shi, Y., Feng, Y., Gao, J., Chen, L., Zhang, A., Cheng, X., Wei, H., Zhang, T., et al. (2022). Single-cell transcriptome and network analyses unveil key transcription factors regulating mesophyll cell development in Maize. Genes (Basel) **13**10.3390/genes13020374.

Upadhyay, S., Jeena, G.S., Kumar, S., and Shukla, R.K. (2020). *Asparagus racemosus* bZIP transcription factor-regulated squalene epoxidase (ArSQE) promotes germination and abiotic stress tolerance in transgenic tobacco. Plant science : an international journal of experimental plant biology **290**:110291. 10.1016/j.plantsci.2019.110291.

Walker, K., and Croteau, R. (2000). Taxol biosynthesis: molecular cloning of a benzoyl-CoA:taxane 2alpha-O-benzoyltransferase cDNA from *Taxus* and functional expression in *Escherichia coli*. Proceedings of the National Academy of Sciences of the United States of America **97**:13591-13596. 10.1073/pnas.250491997.

Wang, Q., Wu, Y., Peng, A., Cui, J., Zhao, M., Pan, Y., Zhang, M., Tian, K., Schwab, W., and Song, C. (2022a). Single-cell transcriptome atlas reveals developmental trajectories and a novel metabolic pathway of catechin esters in tea leaves. Plant Biotechnol J 10.1111/pbi.13891.

Wang, T., Zhang, F., Zhuang, W., Shu, X., and Wang, Z. (2019a). Metabolic variations of flavonoids in leaves of *T. media* and *T. mairei* obtained by UPLC-ESI-MS/MS. Molecules **24**10.3390/molecules24183323.

Wang, T., Chen, Y., Zhuang, W., Zhang, F., Shu, X., Wang, Z., and Yang, Q. (2019b). Transcriptome sequencing reveals regulatory mechanisms of Taxol synthesis in *Taxus wallichiana* var. *mairei*. Int J Genomics **2019**:1596895. 10.1155/2019/1596895.

Wang, Y., Huan, Q., Li, K., and Qian, W. (2021). Single-cell transcriptome atlas of the leaf and root of rice seedlings. J Genet Genomics **48**:881-898. 10.1016/j.jgg.2021.06.001.

Wang, Y., Gui, C., Wu, J., Gao, X., Huang, T., Cui, F., Liu, H., and Sethupathy, S. (2022b). Spatio-temporal modification of lignin biosynthesis in plants: a promising strategy for lignocellulose improvement and lignin valorization. Front Bioeng Biotechnol **10**:917459. 10.3389/fbioe.2022.917459.

Wani, M.C., Taylor, H.L., Wall, M.E., Coggon, P., and McPhail, A.T. (1971). Plant antitumor agents. VI. The isolation and structure of taxol, a novel antileukemic and antitumor agent from *Taxus brevifolia*. J Am Chem Soc **93**:2325-2327. 10.1021/ja00738a045.

Xiong, X., Gou, J., Liao, Q., Li, Y., Zhou, Q., Bi, G., Li, C., Du, R., Wang, X., Sun, T., et al. (2021). The *Taxus* genome provides insights into paclitaxel biosynthesis. Nat Plants **7**:1026-1036. 10.1038/s41477-021-00963-5.

Yang, L., Zheng, Z.S., Cheng, F., Ruan, X., Jiang, D.A., Pan, C.D., and Wang, Q. (2016). Seasonal dynamics of metabolites in needles of *Taxus wallichiana* var. *mairei*. Molecules **21**10.3390/molecules21101403.

Yu, C., Luo, X., Zhan, X., Hao, J., Zhang, L., YB, L.S., Shen, C., and Dong, M. (2018). Comparative metabolomics reveals the metabolic variations between two endangered *Taxus* species (*T. fuana* and *T. yunnanensis*) in the Himalayas. BMC plant biology **18**:197. 10.1186/s12870-018-1412-4.

Yu, C., Zhang, C., Xu, X., Huang, J., Chen, Y., Luo, X., Wang, H., and Shen, C. (2021). Omic analysis of the endangered Taxaceae species *Pseudotaxus chienii* revealed the differences in taxol biosynthesis pathway between *Pseudotaxus* and *Taxus yunnanensis* trees. BMC plant biology **21**:104. 10.1186/s12870-021-02883-0.

Yu, C., Huang, J., Wu, Q., Zhang, C., Li, X.L., Xu, X., Feng, S., Zhan, X., Chen, Z., Wang, H., et al. (2022a). Role of female-predominant MYB39-bHLH13 complex in sexually dimorphic accumulation of taxol in *Taxus media*. Hortic Res **9**:uhac062. 10.1093/hr/uhac062.

Yu, C., Luo, X., Zhang, C., Xu, X., Huang, J., Chen, Y., Feng, S., Zhan, X., Zhang, L., Yuan, H., et al. (2020). Tissue-specific study across the stem of *Taxus media* identifies a phloem-specific TmMYB3 involved in the transcriptional regulation of paclitaxel biosynthesis. The Plant journal : for cell and molecular biology **103**:95-110. 10.1111/tpj.14710.

Zhan, X., Chen, Z., Chen, R., and Shen, C. (2022). Environmental and genetic factors involved in plant protection-associated secondary metabolite biosynthesis pathways. Front Plant Sci **13**:877304. 10.3389/fpls.2022.877304.

Zhang, M., Li, S., Nie, L., Chen, Q., Xu, X., Yu, L., and Fu, C. (2015). Two jasmonate-responsive factors, TcERF12 and TcERF15, respectively act as repressor and activator of tasy gene of taxol biosynthesis in *Taxus chinensis*. Plant Mol Biol **89**:463-473. 10.1007/s11103-015-0382-2.

Zhang, M., Jin, X., Chen, Y., Wei, M., Liao, W., Zhao, S., Fu, C., and Yu, L. (2018). TcMYC2a, a basic helix-loop-helix transcription factor, transduces JA-signals and regulates taxol biosynthesis in *Taxus chinensis*. Front Plant Sci **9**:863. 10.3389/fpls.2018.00863.

Zhang, T.Q., Chen, Y., and Wang, J.W. (2021). A single-cell analysis of the Arabidopsis vegetative shoot apex. Dev Cell **56**:1056-1074 e1058. 10.1016/j.devcel.2021.02.021.

Zhou, P., Chen, H., Dang, J., Shi, Z., Shao, Y., Liu, C., Fan, L., and Wu, Q. (2022a). Single-cell transcriptome of *Nepeta tenuifolia* leaves reveal differentiation trajectories in glandular trichomes. Front Plant Sci **13**:988594. 10.3389/fpls.2022.988594.

Zhou, T., Luo, X., Yu, C., Zhang, C., Zhang, L., Song, Y.B., Dong, M., and Shen, C. (2019). Transcriptome analyses provide insights into the expression pattern and sequence similarity of several taxol biosynthesis-related genes in three *Taxus* species. BMC plant biology **19**:33. 10.1186/s12870-019-1645-x.

Zhou, X., Rao, S., Wrightstone, E., Sun, T., Lui, A.C.W., Welsch, R., and Li, L. (2022b). Phytoene synthase: the key rate-limiting enzyme of carotenoid biosynthesis in plants. Front Plant Sci **13**:884720. 10.3389/fpls.2022.884720.

Zong, J., Wang, L., Zhu, L., Bian, L., Zhang, B., Chen, X., Huang, G., Zhang, X., Fan, J., Cao, L., et al. (2022). A rice single cell transcriptomic atlas defines the developmental trajectories of rice floret and inflorescence meristems. The New

phytologist **234**:494-512. 10.1111/nph.18008.

Figure Legends

Figure 1 **Analysis of the data from timsTOF fleX MALDI-2.** (a) Sampling and sectioning strategy for spatial metabolomic analysis. (b) Segmentation analysis outputs a colored map with different data points. (c) PCA of all the detected metabolites in *T. mairei* leaves. Four PCs indicated different accumulation patterns of detected metabolites in leaves. (d) The heatmap showed the relative accumulation levels of each detected metabolite in four PCs. The heatmap scale ranges from -2 to +2 on a $\log_2$ scale. (e) Clustering analysis grouped all the detected MS features into six clusters.

Figure 2 **Visualization of taxoid accumulation in *T. mairei* leaves.** (a) The tissue specific distribution of various typic metabolites, including four flavonoids, four steroids, four coumarins, four phenolics, four terpenes, four glycosides, and four taxoids, were detected by MALDI-IMS analysis. The color scale ranges from 0% to 100%. (b) The proportion of various typic metabolites in different leaf tissues, including spongy mesophyll cells, epidermis cells, bundle sheath cells, and palisade mesophyll cells.

Figure 3 **ScRNA-seq and cell clustering of *T. mairei* leaves.** (a) Isolation of protoplasts from leaf tissues. (b) Workflow of the scRNA-seq of *T. mairei* leaves. Protoplasts isolated the leaves were loaded into a 10x Genomics Chromium Controller. (c) The number of effective cells was quantified by barcodes and UMI counts. (d) The number of cells, reads and genes. (e) T-SNE projection of cells colored by UMI counts. (f) The expression pattern of three representative cluster-enriched genes for each cluster. Dot diameter, proportion of cluster cells expressing a given gene. (g) The expression pattern of 11 leaf tissue-specific expressed marker genes in 15 different clusters. (h) Visualization of 15 cell clusters using tSNE method. Dots indicated individual cells and colour indicated different cell clusters.

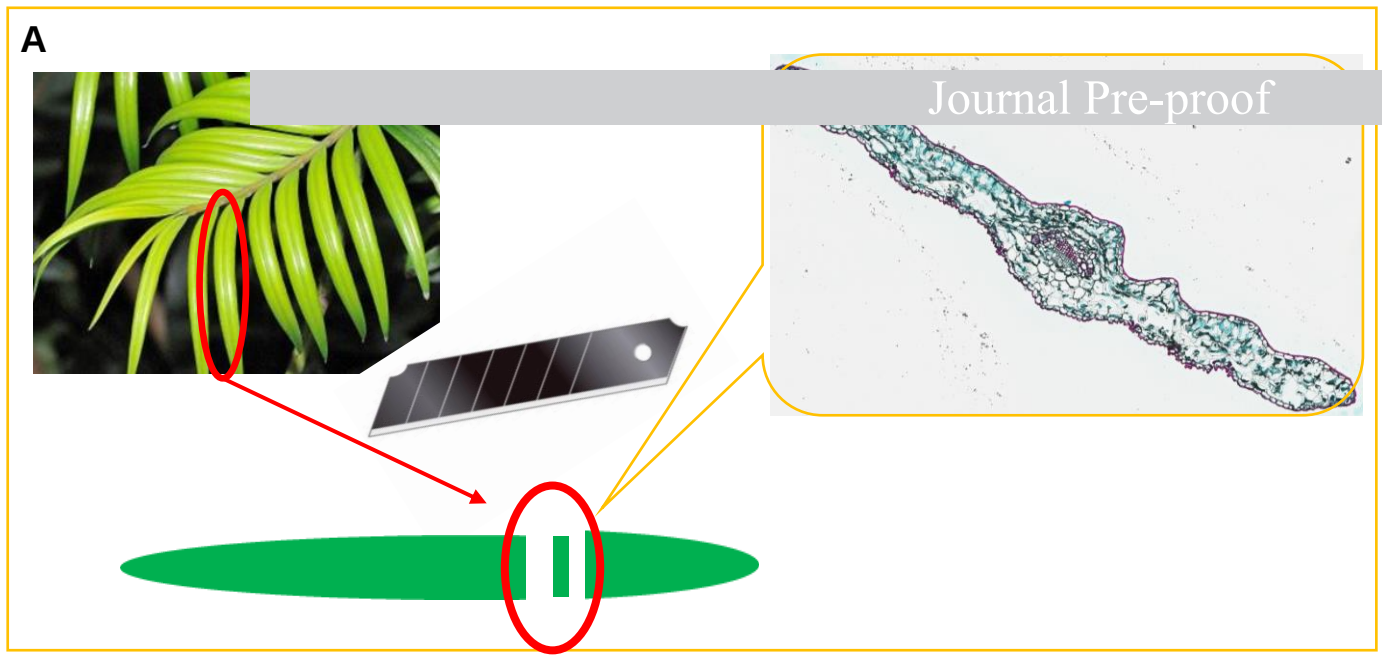
Figure 4 **Cell type specificity of Taxol biosynthesis in *T. mairei* leaves.** (a) Overview

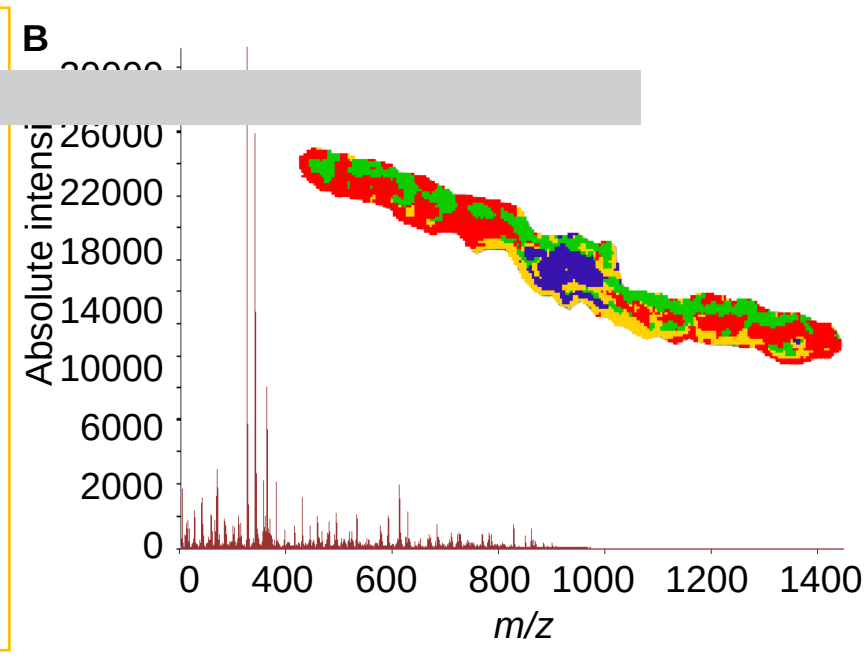
of the paclitaxel biosynthesis pathway, including enzymes and metabolites. (b) The cell coverage of Taxol biosynthesis pathway-related genes. Each key enzyme was encoding by at least one gene. Red stars indicated the encoding gene with the largest cell coverage. (c) Cell specific expression analysis of the taxol biosynthesis-related genes. Clusters 6 and 7 refer mesophyll cells. (d) The successive differentiation trajectory during the mesophyll tissue development process. (e) Heatmap showing expression levels of Taxol-related genes under mesophyll tissue development process. The heatmap scale ranges from +3 to -3 on a $\log_2$ scale. (f) All the leaf mesophyll cells were re-clustered into four subclasses: A to D. The red color indicated expression levels and the dot size indicated cell percent.

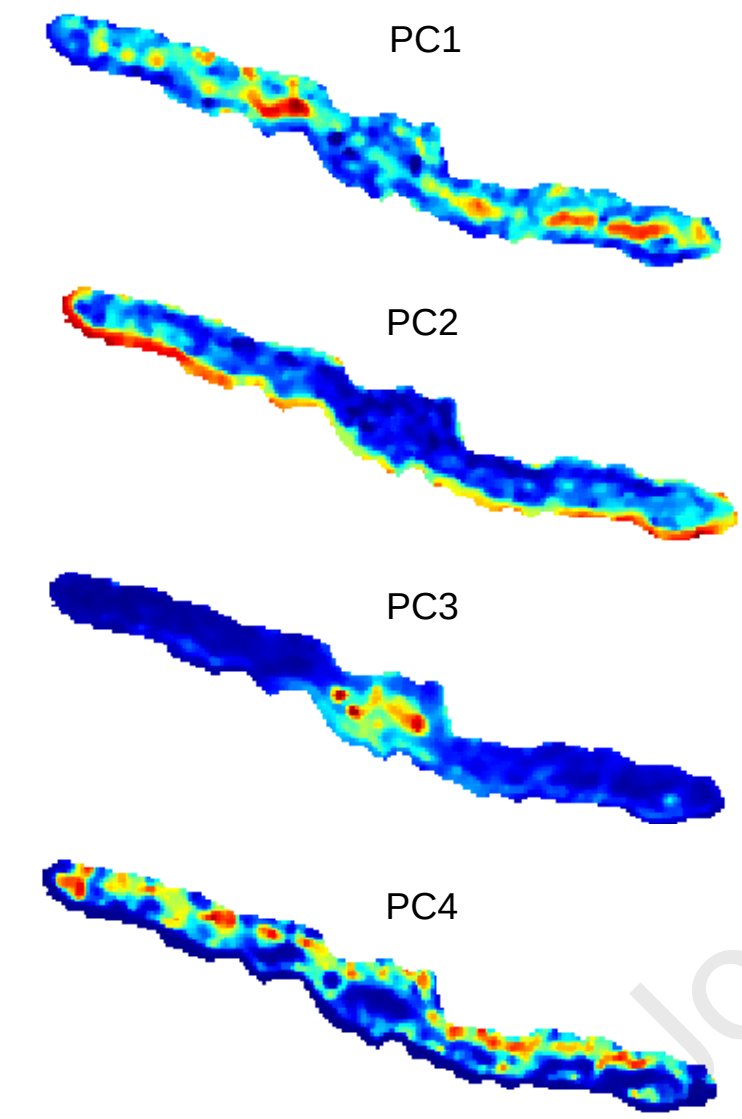
Figure 5 **Expression pattern of CYP super family.** (a) Cell specific expression analysis of CYP family at single cell level by ScRNA-seq. Blue box indicated that the CYP725 subfamily members were enriched in the cells belonging to Cluster 6 and 7. (b) All the leaf mesophyll cells were re-clustered into four subclasses: A to D. The red color indicated expression levels and the dot size indicated cell percent. (c) UMAP visualization of expression patterns of selected six CYP725 family members in *T. mairei* leaves. (d) Violin Plot analysis of expression patterns of selected six CYP725 family members in *T. mairei* leaves. (e) KEGG enrichment analysis of the genes significantly expressed in the Cluster 6 and 7.

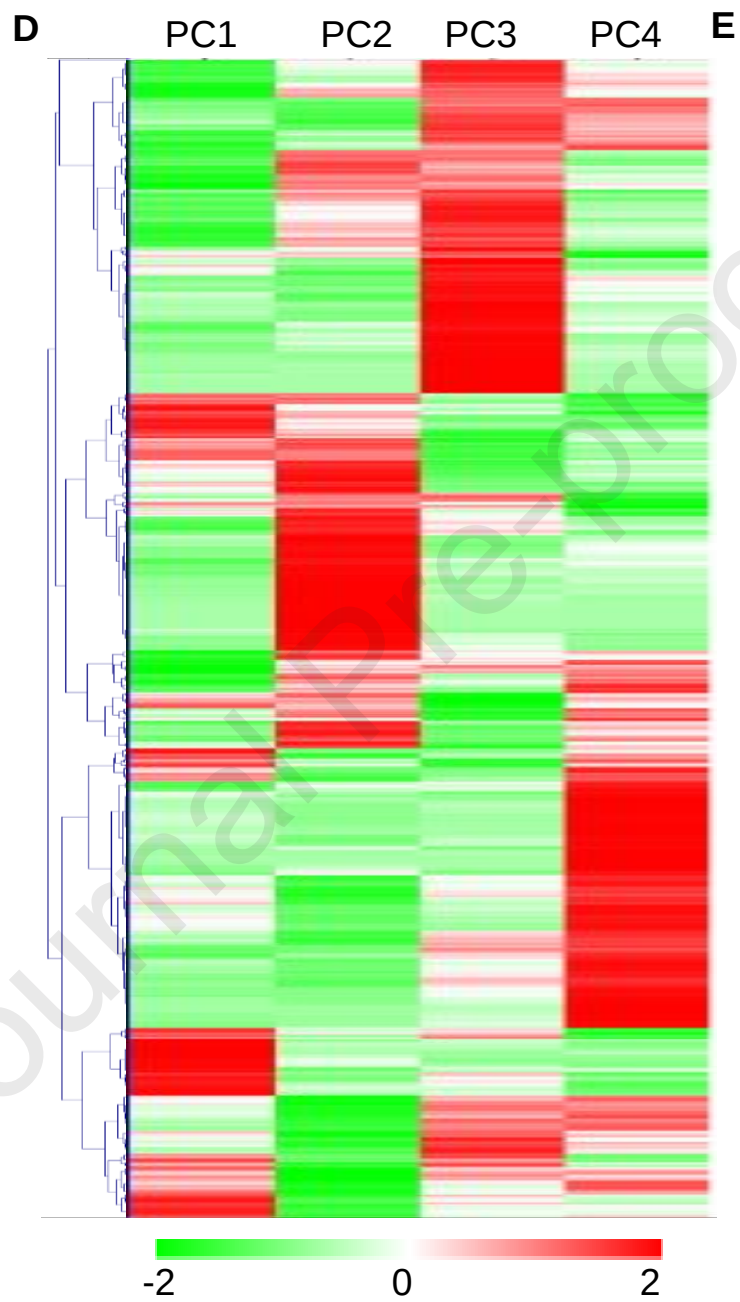
Figure 6 **Discovery of cell specific TFs involved in secondary metabolism in *T. mairei* leaves.** (a) Cell specific expression analysis of TFs at single cell level by ScRNA-seq. The proportion of the expression level of each TF in eight metabolism-related Clusters. (b) Screening of TF-binding elements in the promoter regions of secondary metabolism-related genes. Crimson triangles indicated the different TF-binding elements, including MBE (light blue), GT-element (pink), GCC-box (red), W-box (yellow), and ABRE (purple). (c) Prediction of transcription regulation networks according to the cell type specific expression pattern.

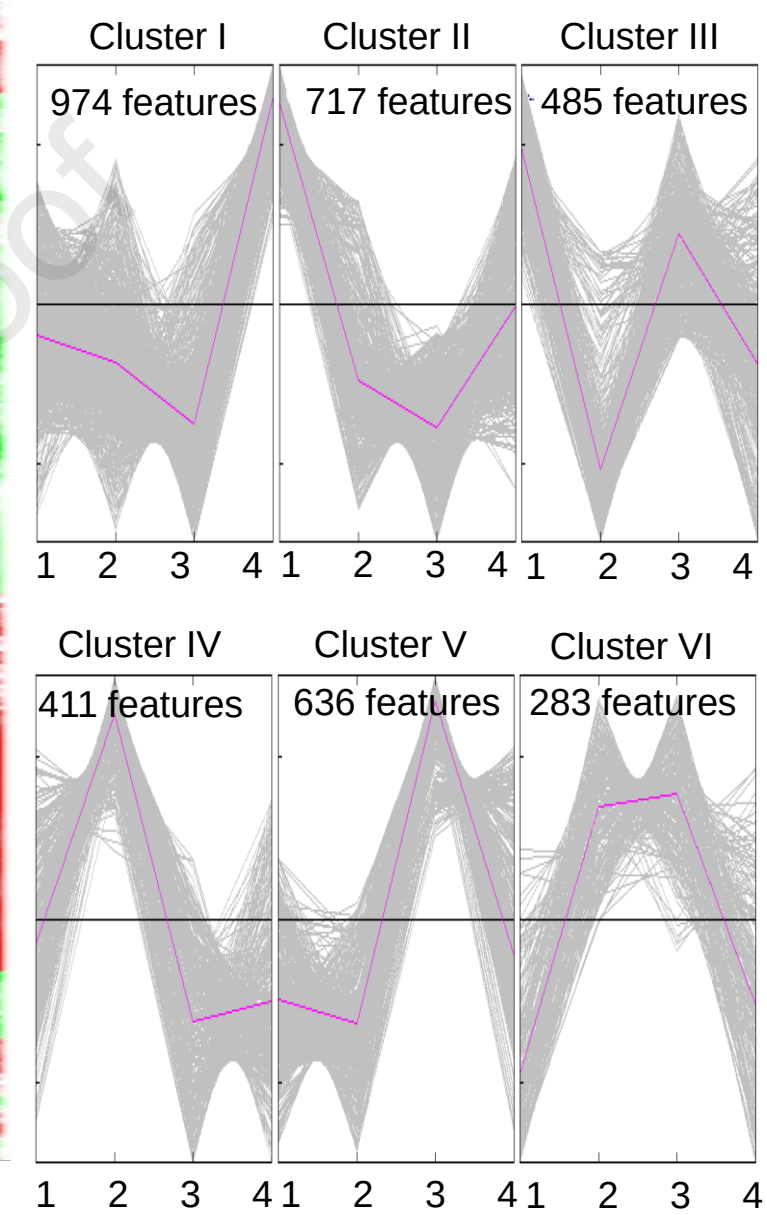
Figure 7 **Verification of the binding of TFs to their potential targets.** (a-f) The GST only or TTF-GST fusion protein was incubated with probes containing the binding element derived from the promoters. - and + represent absence and presence, respectively, and 20× or 200× show increasing amounts of probes for competition. (g) Overview of constructs prepared for dual-luciferase reporter assays. The promoter fragments were ligated with the pGreenII 0800-LUC vector to produce the reporters. The effectors were produced by inserting the TF genes into the pGreenII-62-SK vector. (h) The dual luciferase assays in tobacco leaves showed that co-transformation of TFs affected the target promoters. (i) The cell type locations in *T. mairei* leaves. (j) The location of taxol biosynthesis-related genes. (l) The location of terpenoid and steroid metabolism-related genes. (k) The location of phenolic acid and flavonoid metabolism-related genes.











A

Flavonoids: Ginkgetin; Quercetin; Rutin; (R)-Glabridin

Steroids: Crustecdysone; Cortisol; Ponasterone A; Schottenol

Coumarins: Bergaptol; Coumestan; Coumesterol; Dicumarol

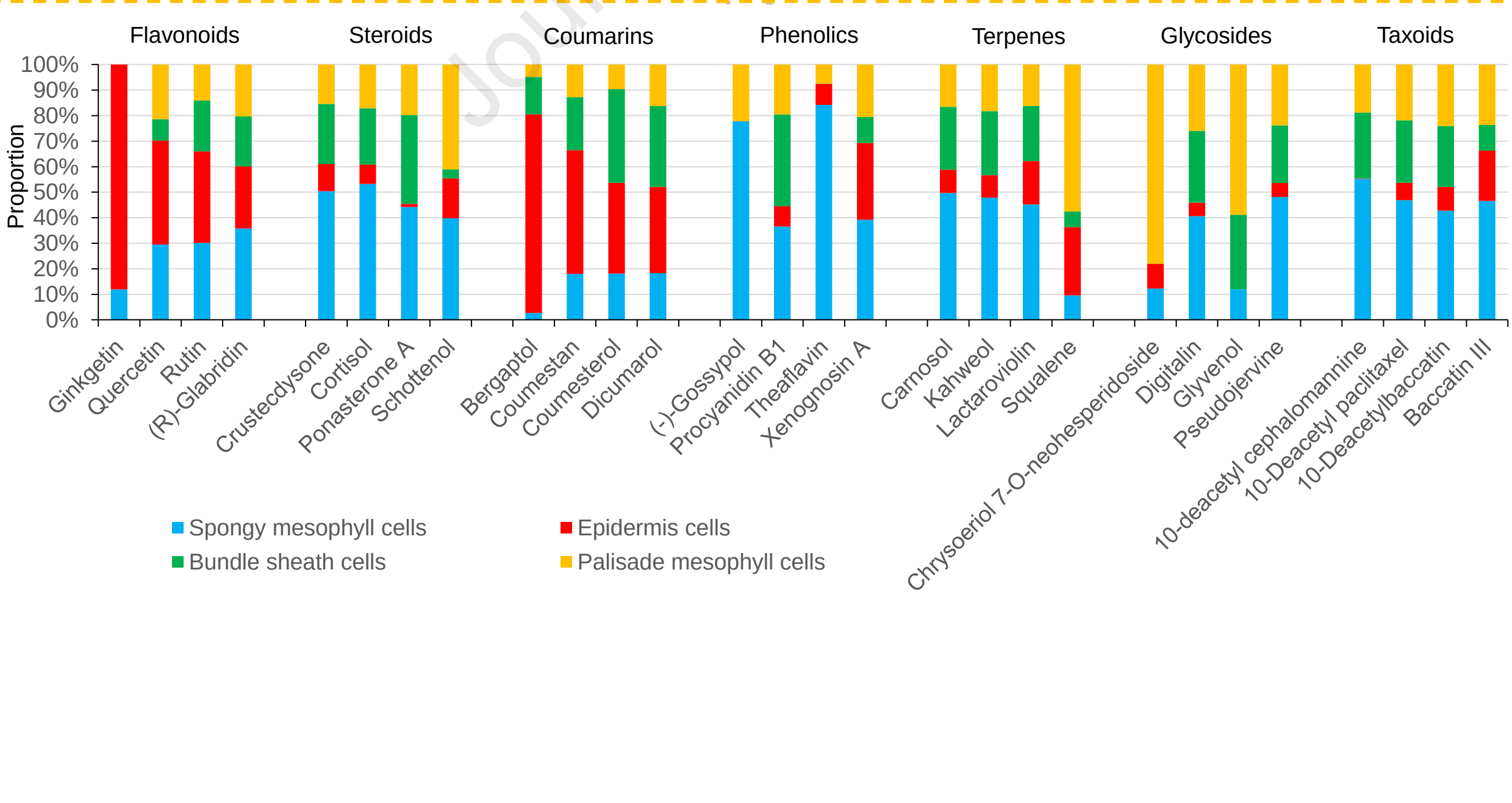
Phenolics: (-)-Gossypol; Procyanidin; Theaflavin; Xenognosin A

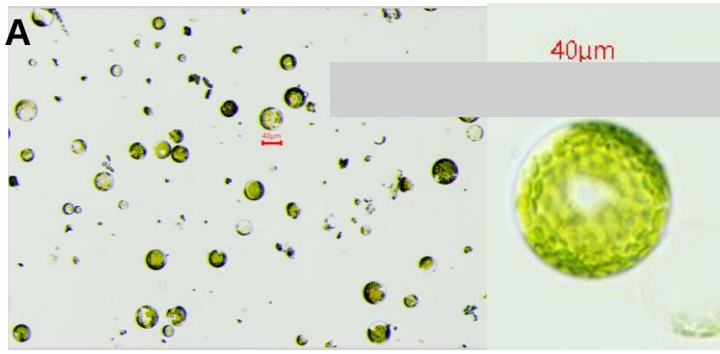
Terpenes: Carnosol; Kahweol; Lactaroviolin; Squalene

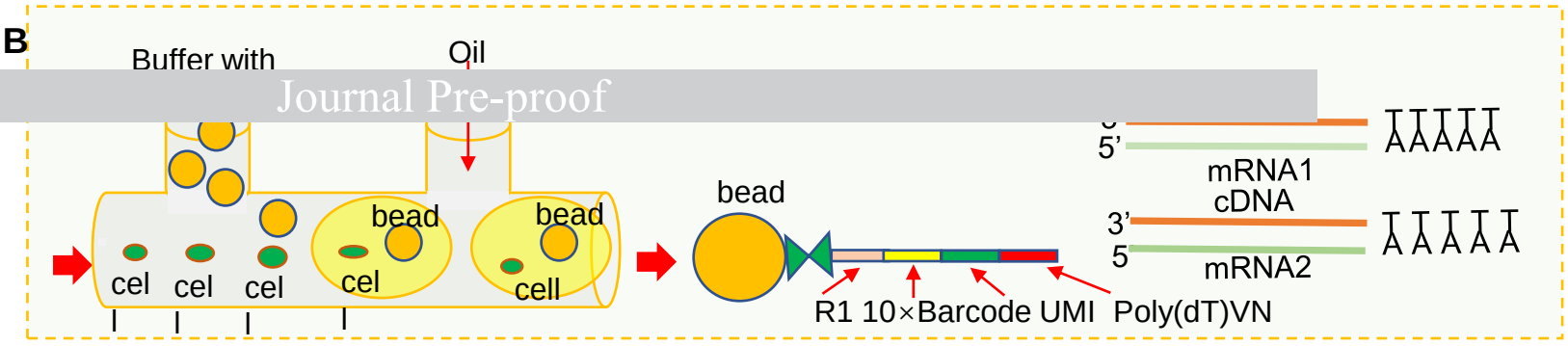
Glycosides: Chrysoeriol 7-O-neohesperidoside; Digitalin; Glyvenol; Pseudojervine

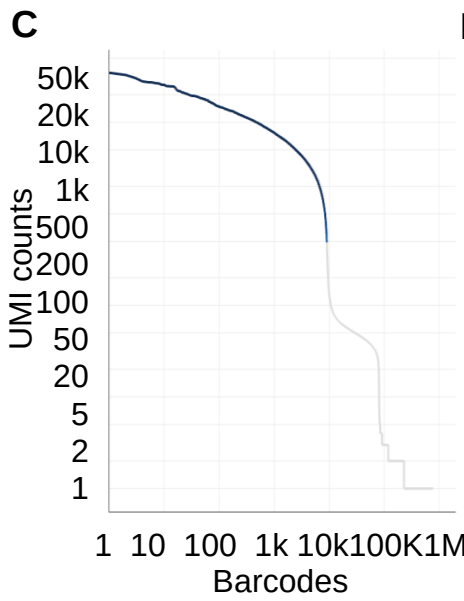
Taxoids: 10-Deacetyl cephalomannine; 10-Deacetyl paclitaxel; 10-Deacetyl baccatin III; Baccatin III

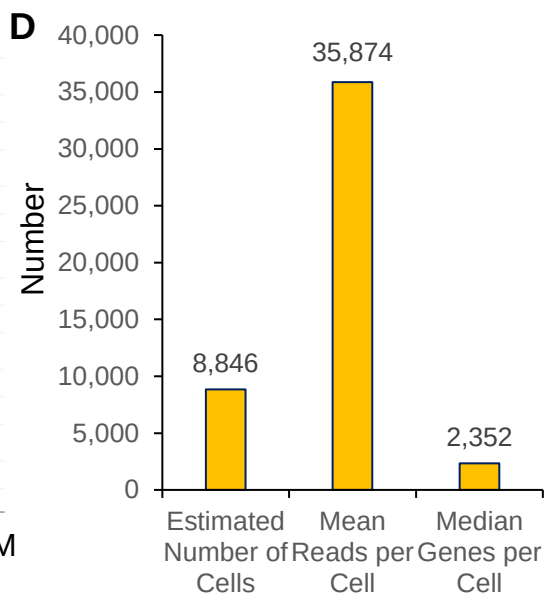
B

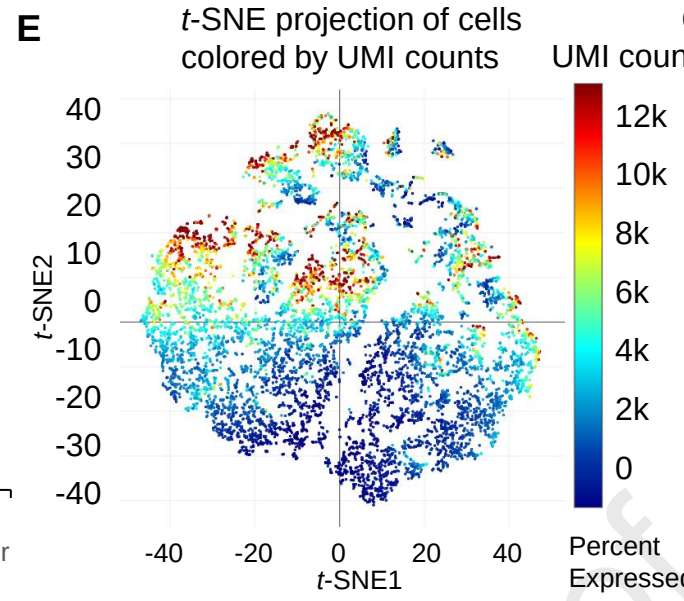


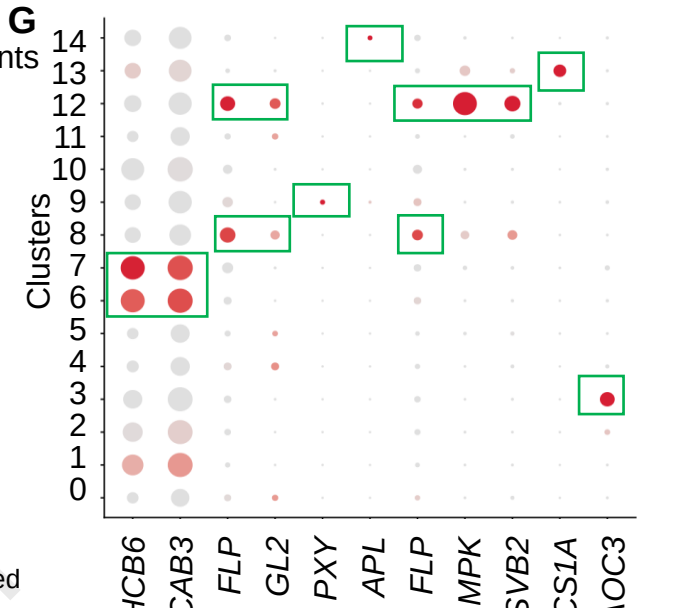


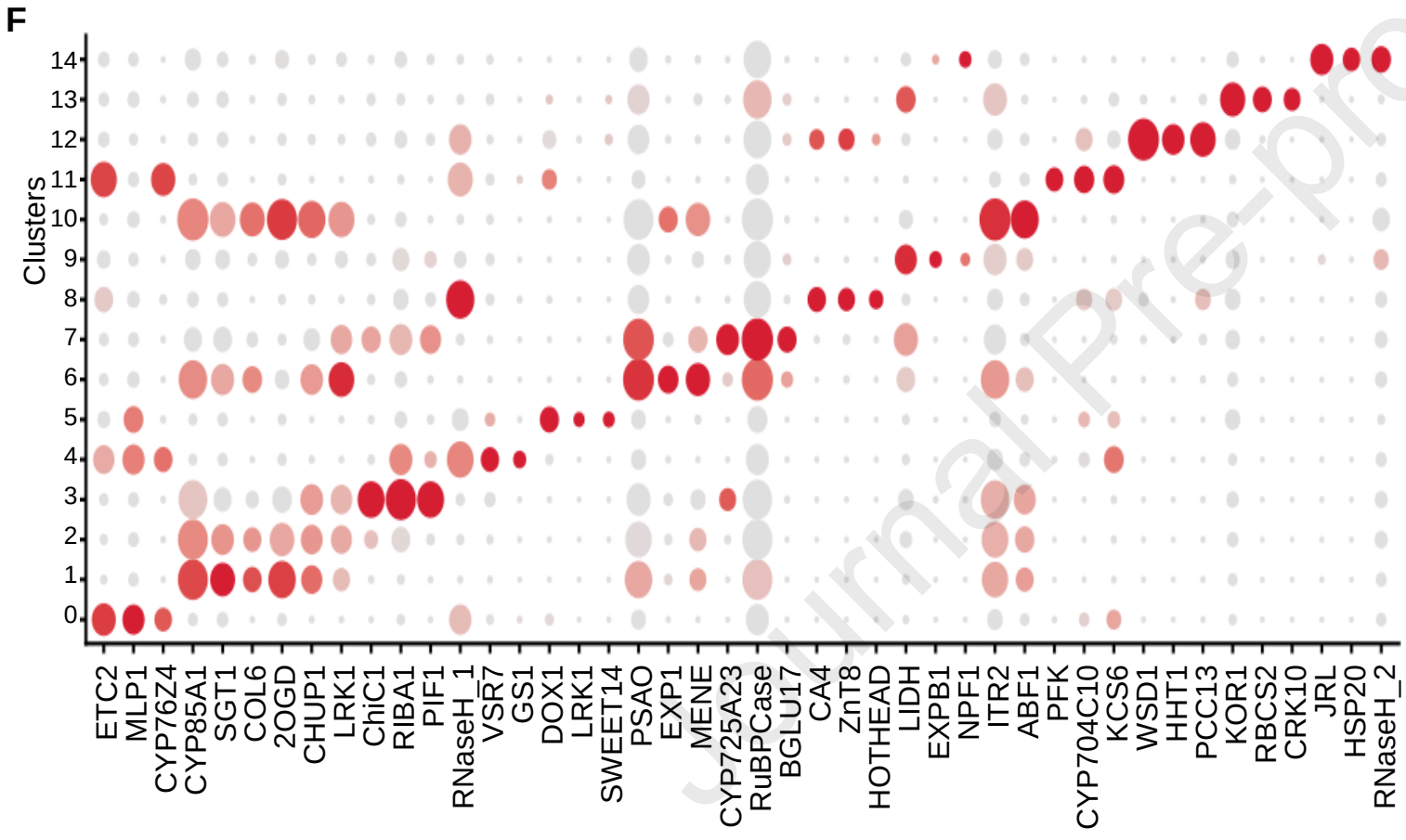


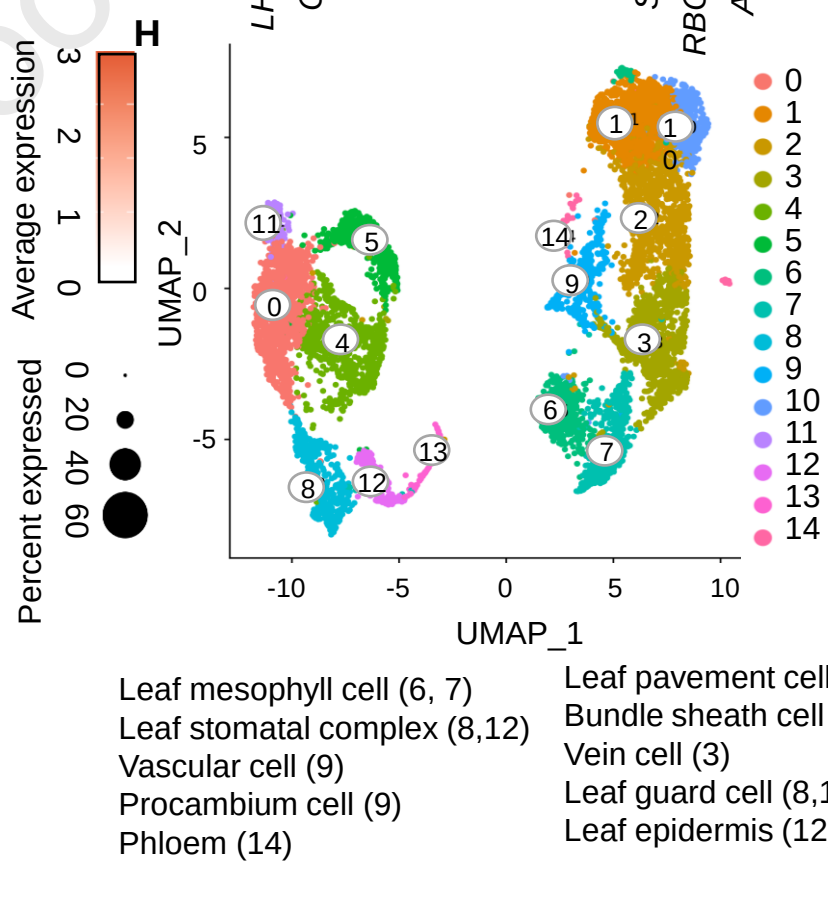








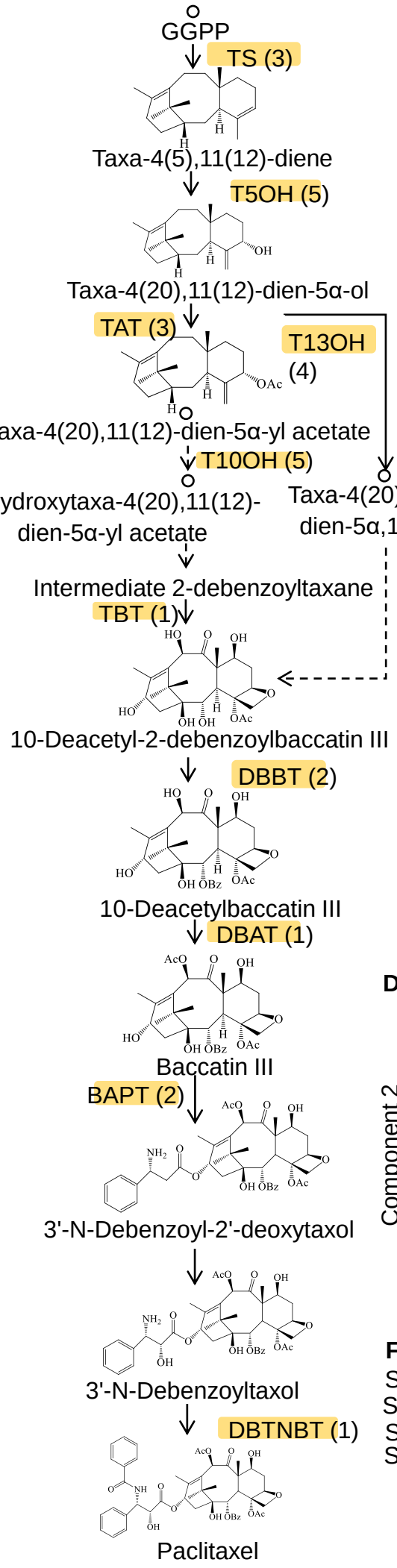


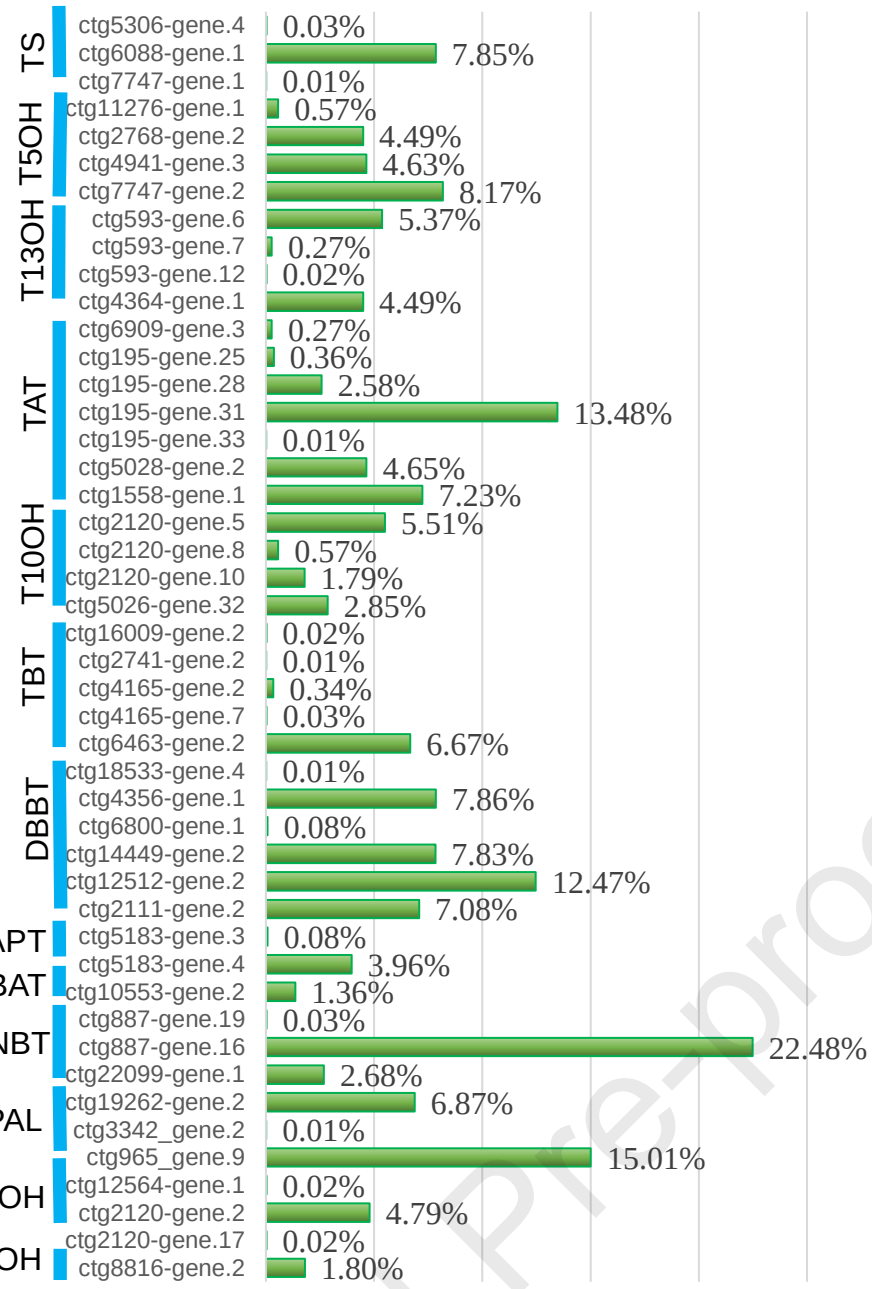


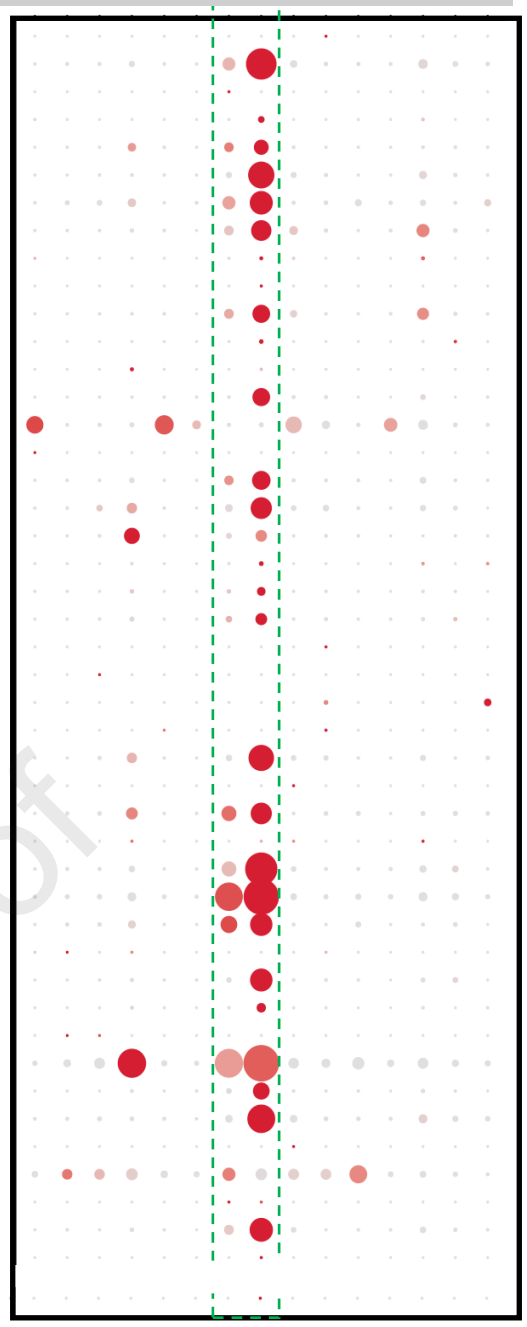
Leaf mesophyll cell (6, 7)
Leaf stomatal complex (8,12)
Vascular cell (9)
Procambium cell (9)
Phloem (14)

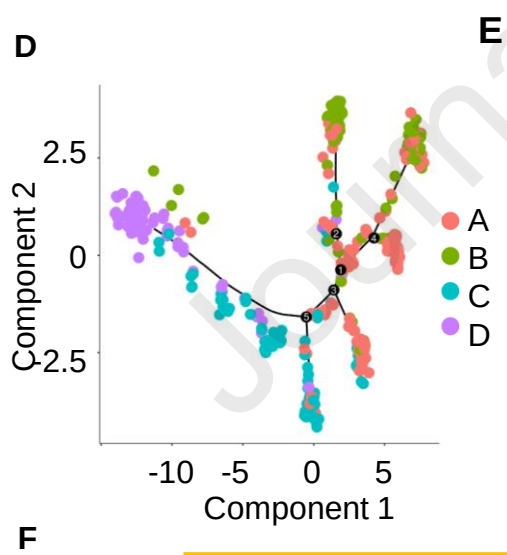
Leaf pavement cell (13)
Bundle sheath cell (3)
Vein cell (3)
Leaf guard cell (8,12)
Leaf epidermis (12)

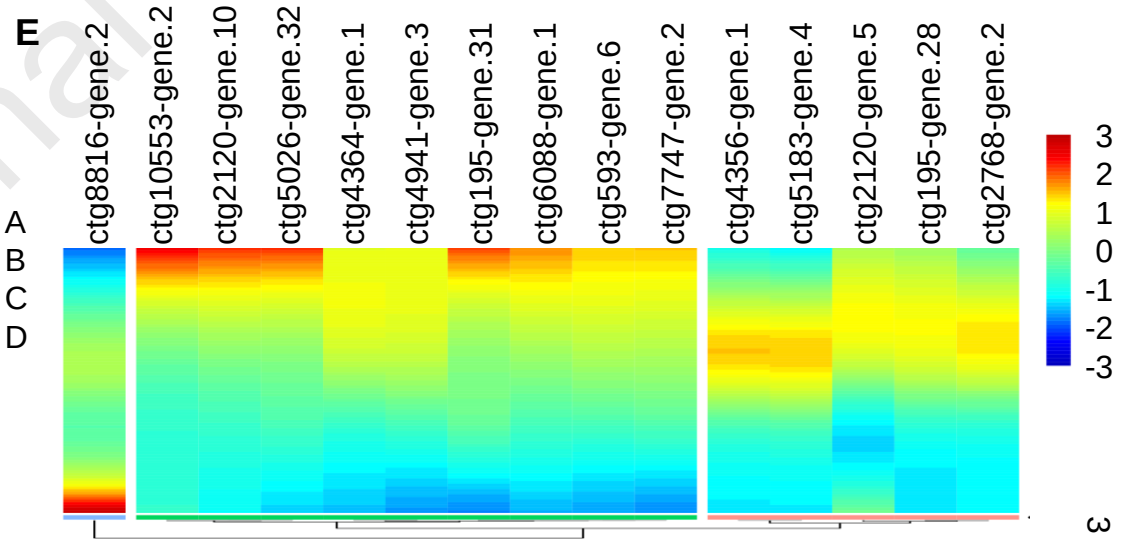
A

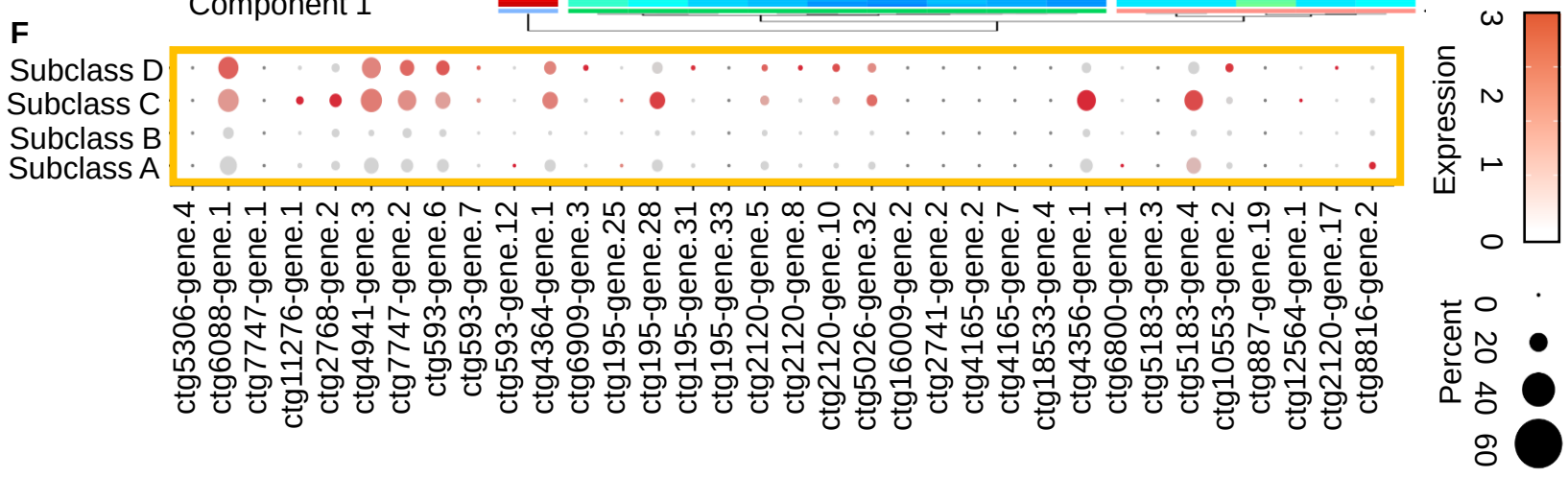


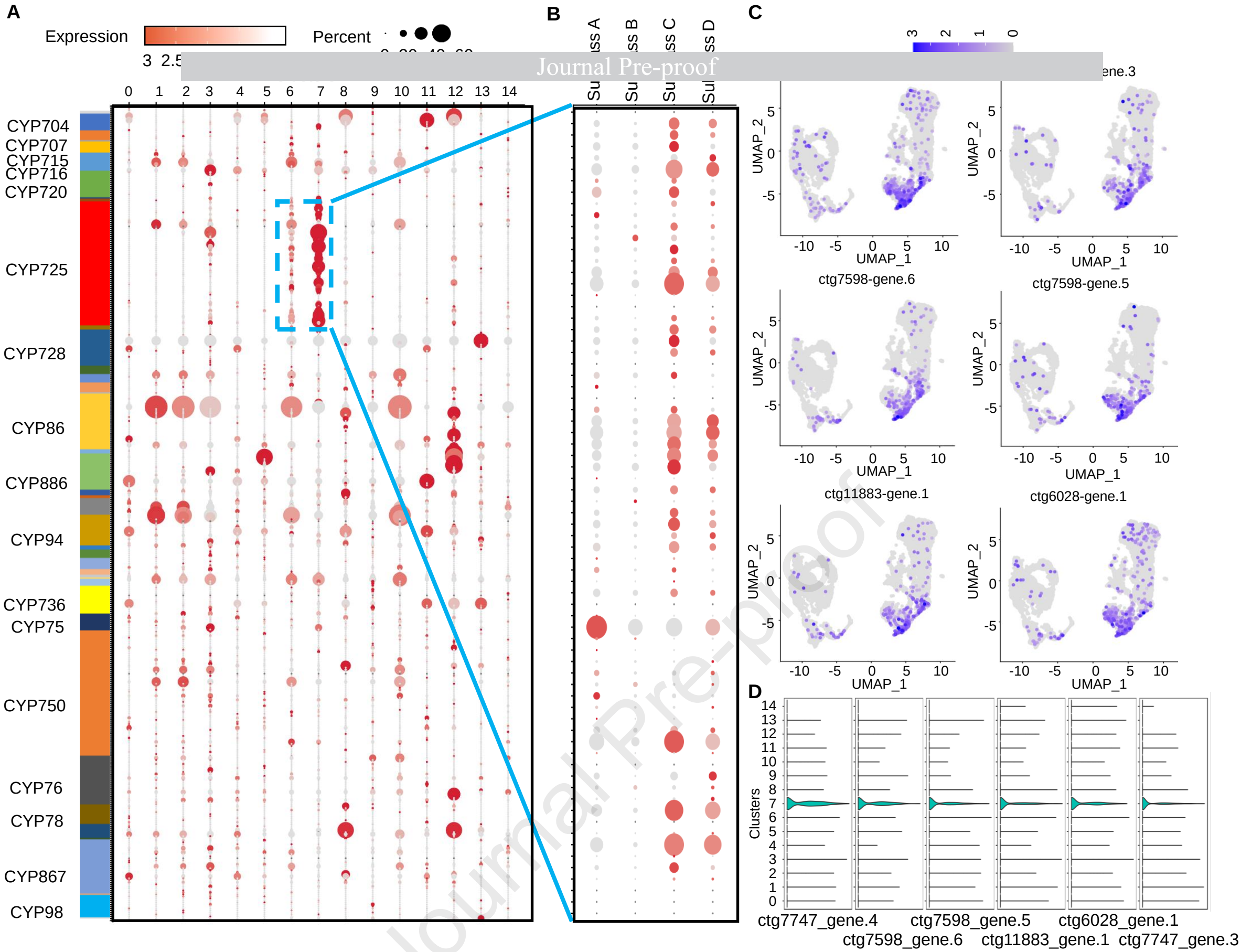




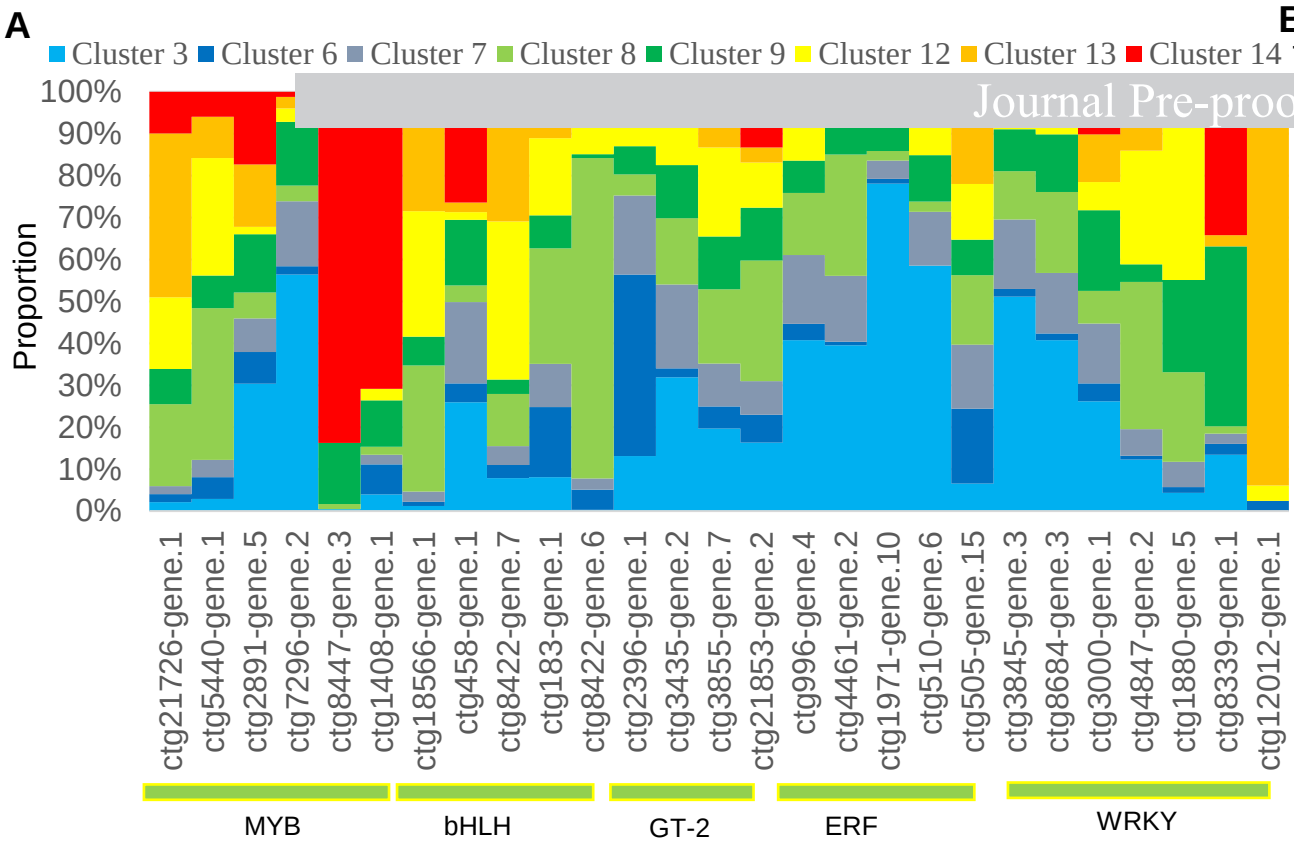




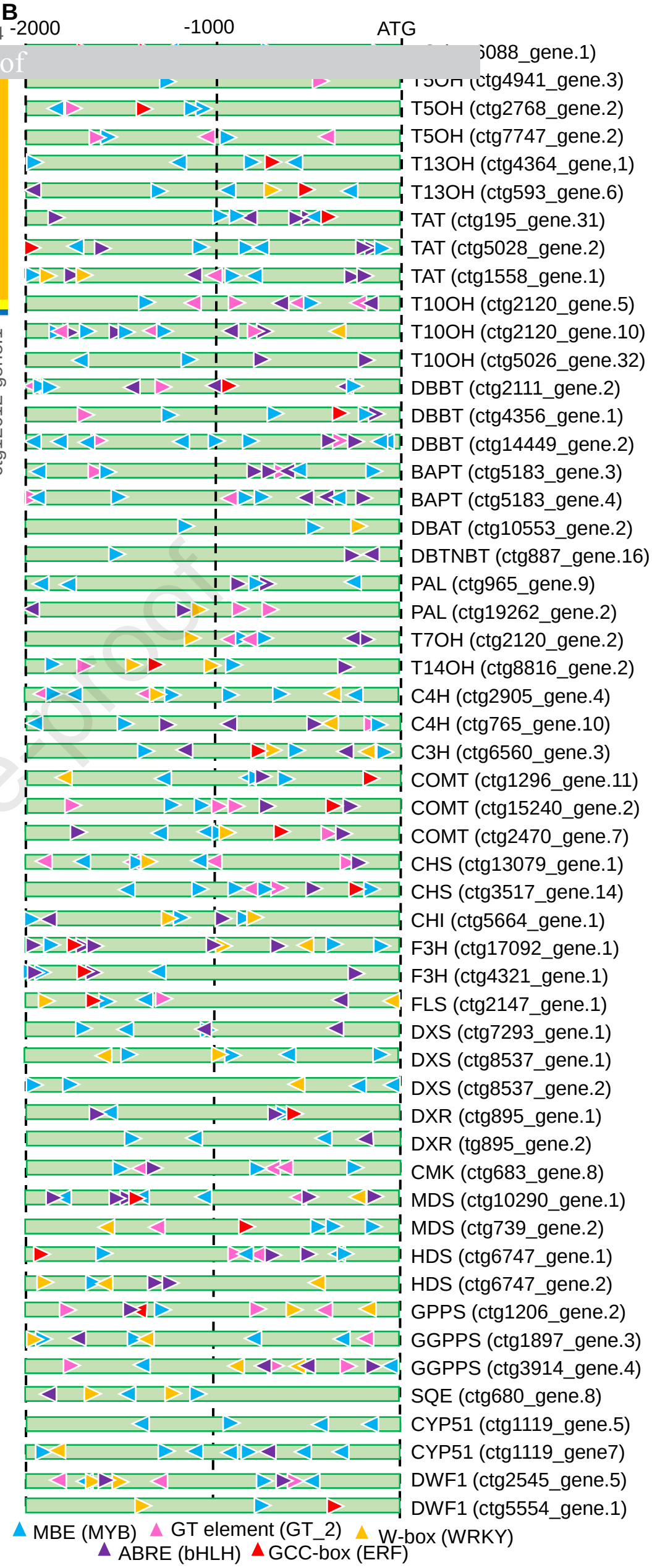




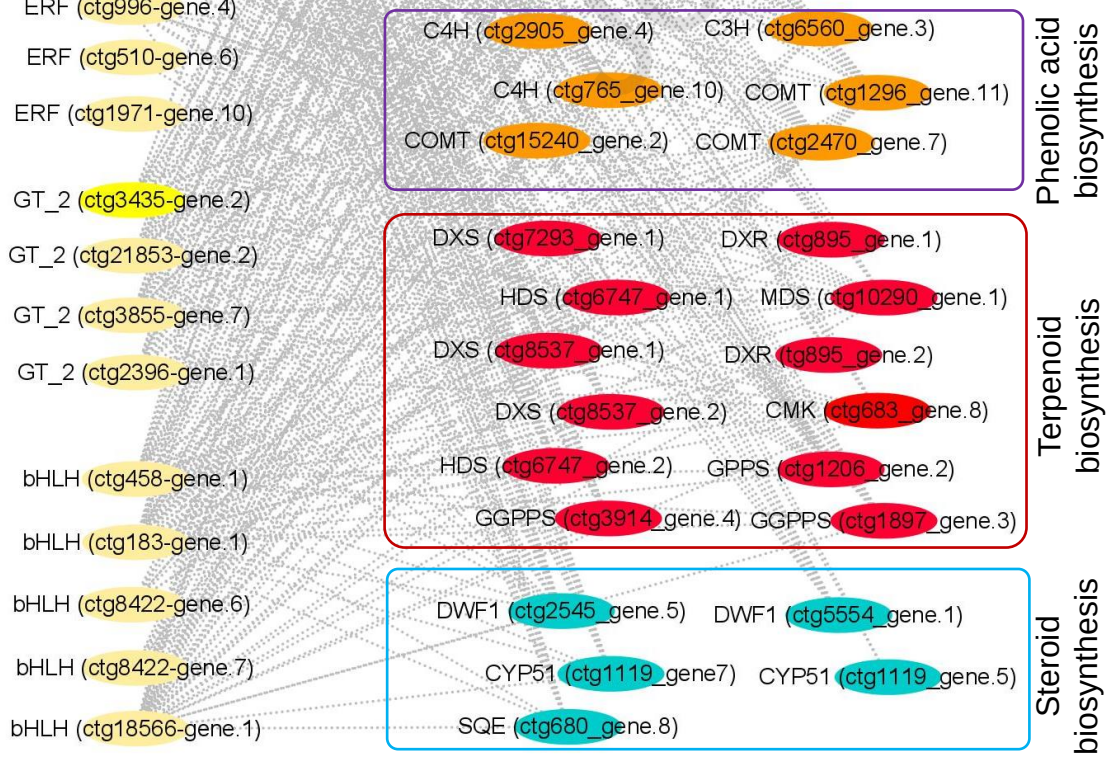
A

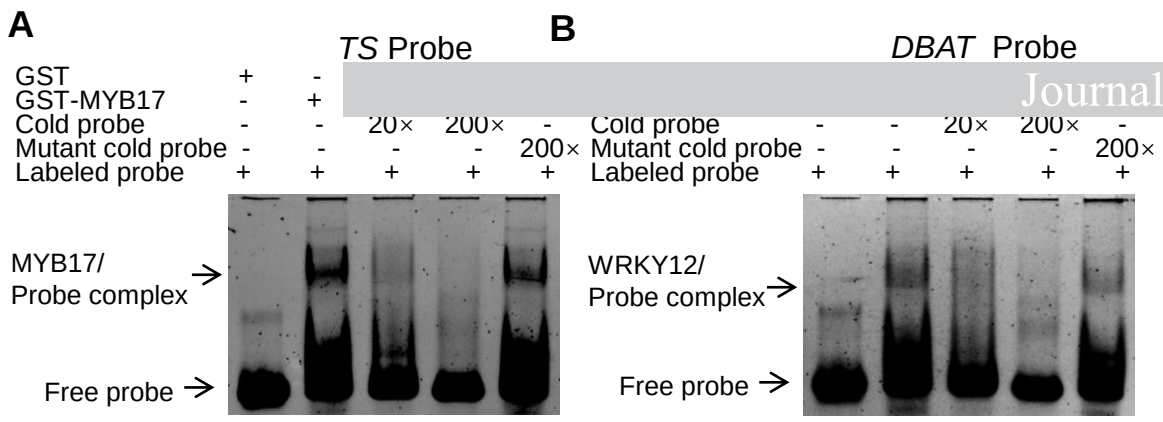


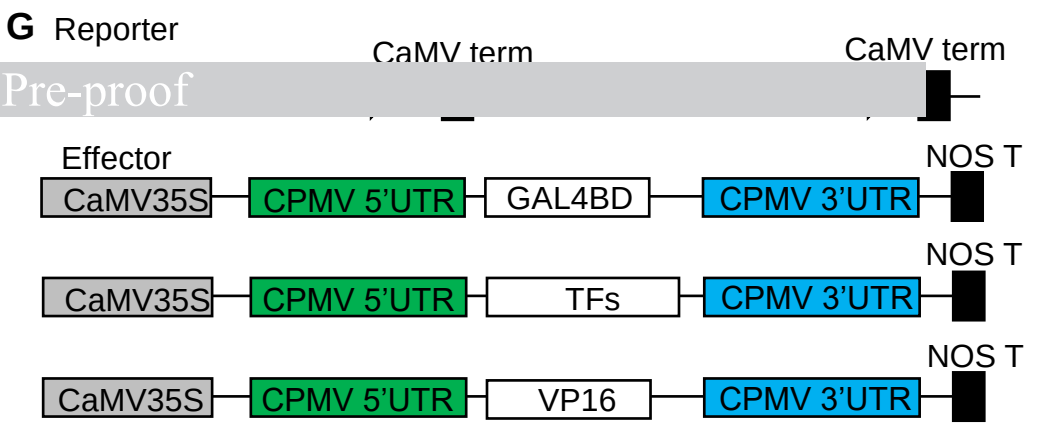
B

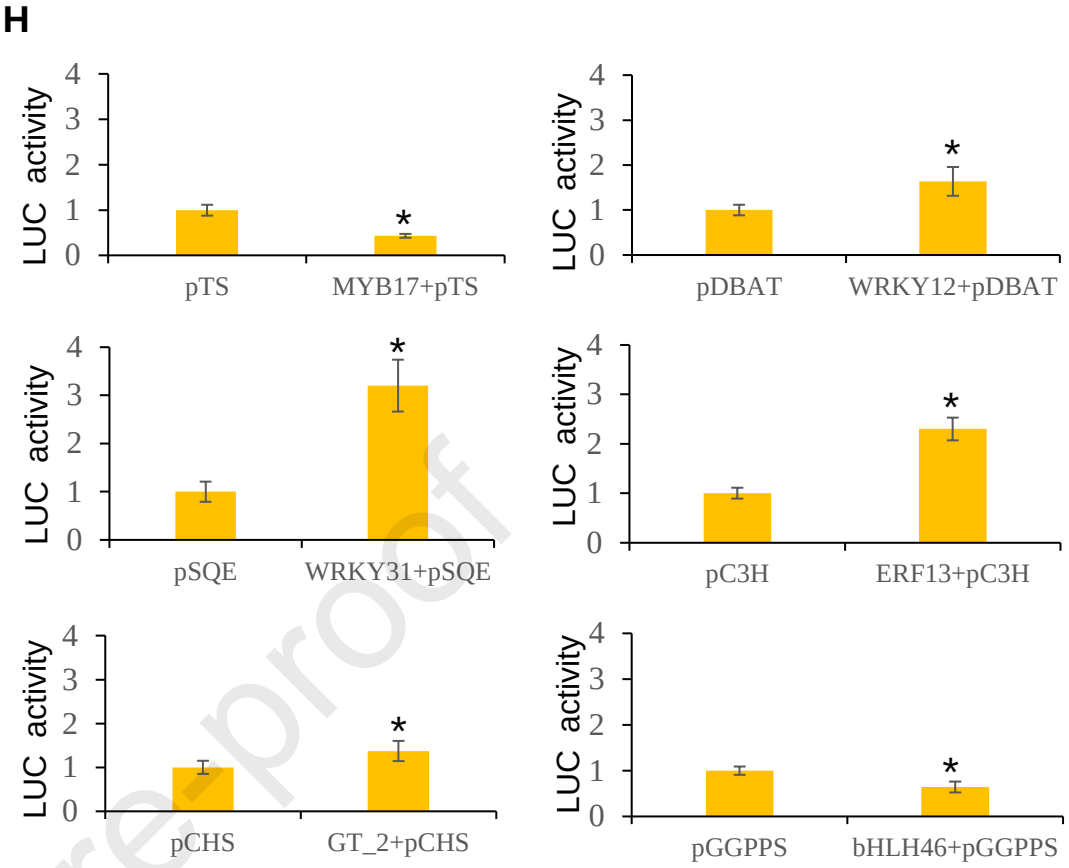


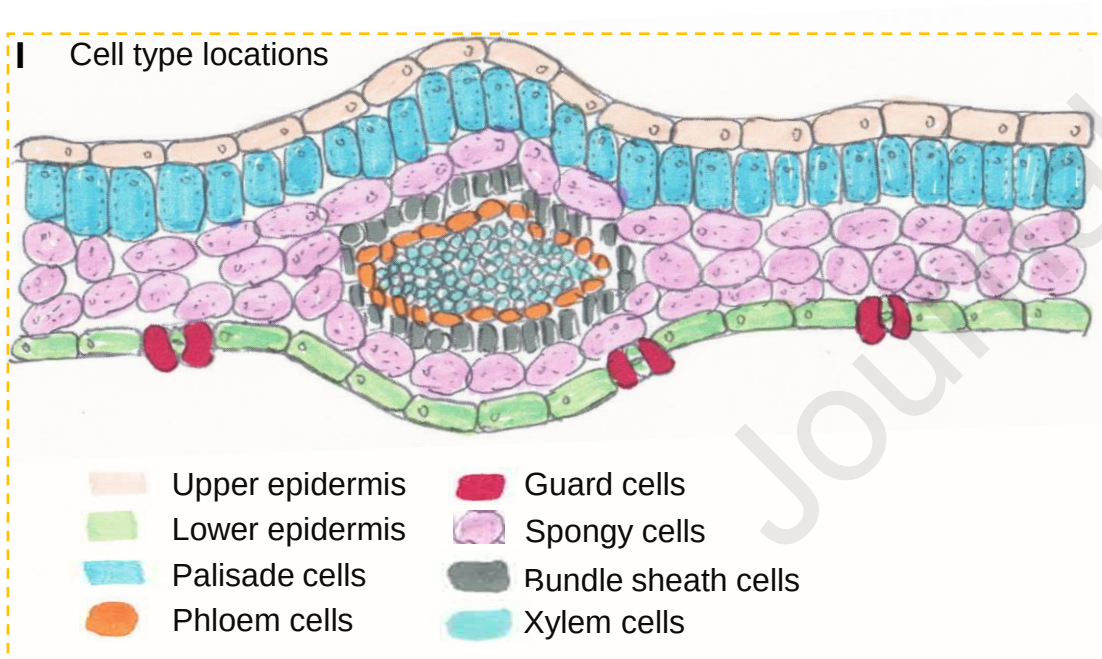
C

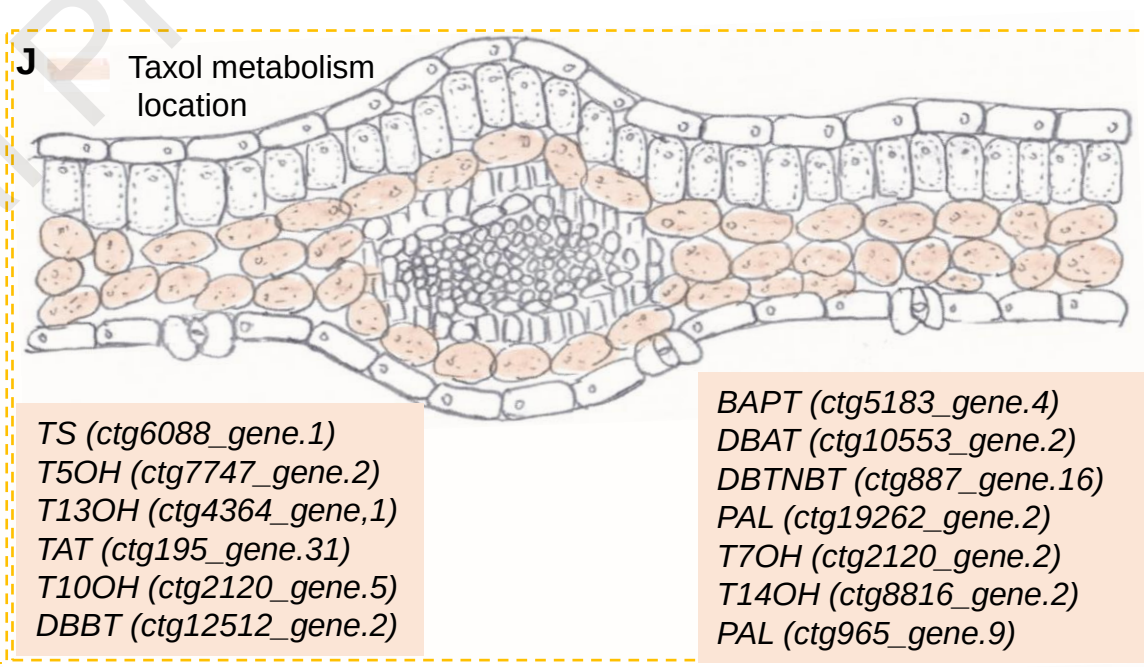






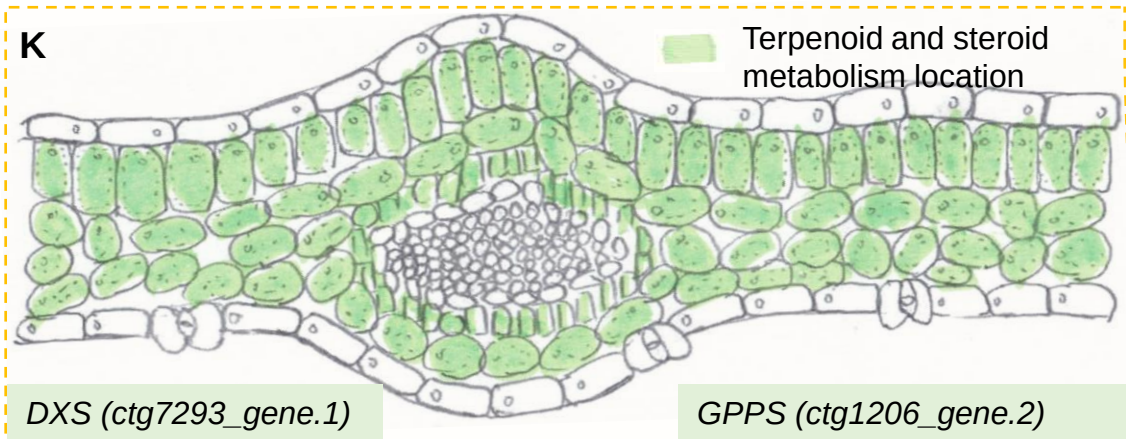


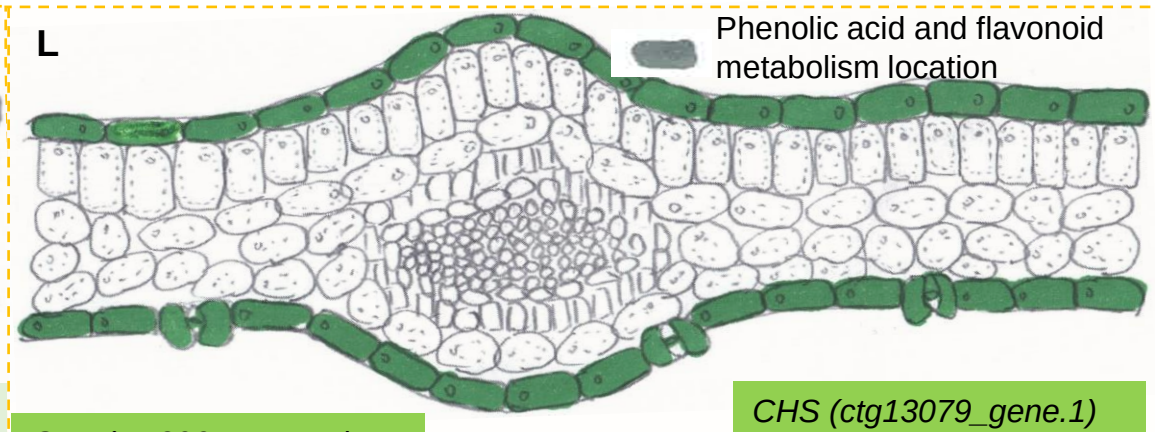




TS (ctg6088_gene.1)
T5OH (ctg7747_gene.2)
T13OH (ctg4364_gene,1)
TAT (ctg195_gene.31)
T10OH (ctg2120_gene.5)
DBBT (ctg12512_gene.2)

BAPT (ctg5183_gene.4)
DBAT (ctg10553_gene.2)
DBTNBT (ctg887_gene.16)
PAL (ctg19262_gene.2)
T7OH (ctg2120_gene.2)
T14OH (ctg8816_gene.2)
PAL (ctg965_gene.9)





DXS (ctg7293_gene.1)
DXR (ctg895_gene.1)
CMK (ctg683_gene.8)
MDS (ctg10290_gene.1)
MDS (ctg739_gene.2)
HDS (ctg6747_gene.2)

GPPS (ctg1206_gene.2)
GGPPS (ctg1897_gene.3)
GGPPS (ctg3914_gene.4)
SQE (ctg680_gene.8)
CYP51 (ctg1119_gene.5)
DWF1 (ctg5554_gene.1)

C4H (ctg2905_gene.4)
C4H (ctg765_gene.10)
C3H (ctg6560_gene.3)
COMT (ctg1296_gene.11)
COMT (ctg15240_gene.2)

CHS (ctg13079_gene.1)
CHS (ctg3517_gene.14)
CHI (ctg5664_gene.1)
F3H (ctg17092_gene.1)
F3H (ctg4321_gene.1)
FLS (ctg2147_gene.1)