

# Single-cell ATAC sequencing illuminates the *cis*-regulatory differentiation of taxol biosynthesis between leaf mesophyll and leaf epidermal cells in *Taxus mairei*

Xiaori Zhan<sup>a,b</sup>, Hongshan Zhang<sup>b,c</sup>, Xueshuang Liang<sup>a,b</sup>, Hou Kailin<sup>a,b</sup>, Wanting Lin<sup>a,b</sup>, Ruoyun Ma<sup>a,b</sup>, Tian Qiu<sup>a,b</sup>, Cheng Chen<sup>d</sup>, Zhijing Wang<sup>d</sup>, Qicong Wu<sup>a,b</sup>, Dan Mao<sup>a,b</sup>, Yipin Ji<sup>a,b</sup>, Xiao-lin Li<sup>e</sup>, Chunna Yu<sup>a,b</sup>, Mingshuang Wang<sup>a,b</sup>, Shangguo Feng<sup>a,b</sup>, Qicai Ying<sup>a,b</sup>, Huizhong Wang<sup>a,b,\*</sup>, Chenjia Shen<sup>a,b,c,\*</sup>

<sup>a</sup> College of Life and Environmental Sciences, Hangzhou Normal University, Hangzhou 311121, China

<sup>b</sup> Zhejiang Provincial Key Laboratory for Genetic Improvement and Quality Control of Medicinal Plants, Hangzhou Normal University, Hangzhou 311121, China

<sup>c</sup> Kharkiv Institute, Hangzhou Normal University, Hangzhou 311121, China

<sup>d</sup> College of Pharmacy, Hangzhou Normal University, Hangzhou 311121, China

<sup>e</sup> 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

## ARTICLE INFO

### Keywords:

Chromatin accessibility  
*cis*-regulatory  
 ScATAC-seq  
*Taxus mairei*  
 Taxol biosynthesis  
 Transcription regulation

## ABSTRACT

Taxol is a widely used anticancer drug; however, its accumulation pattern can vary largely among leaf tissues. Taxol biosynthesis is driven by a series of transcription factors (TFs), but the extent of chromatin changes that functionally regulate gene expression associated with taxol biosynthesis remain poorly understood. To assess the effects of tissue heterogeneity on taxol biosynthesis, we applied single-cell transposase-accessible chromatin sequencing to *Taxus mairei* leaves and identified differentially accessible regions. We captured 9488 independent cells with 42,192 fragment reads per cell. In total, 66.64% were confidently mapped onto the *T. mairei* genome. Projection analysis divided all the cells into nine clusters, which suggests a high degree of cell heterogeneity in *T. mairei* leaves. Based on the cell type markers, Clusters 2 and 7 were annotated as leaf mesophyll cells, and Clusters 5 and 6 were annotated as leaf epidermal cells. A differential accessibility analysis identified 9600 and 8538 leaf mesophyll and epidermal, respectively, cell-specific open chromatin regions. We predicted the genomic features of the genes involved in different stages of taxol biosynthesis, and there was a close correlation between chromatin open peaks and the expression levels of several taxol biosynthesis-related genes. We further identified a number of cell type-specific TF motifs and reference TFs that appeared to be involved in taxol biosynthesis. The study of *T. mairei* leaves at a single-cell resolution provides a valuable resource for determining the basic principles of the cell type-specific regulation of taxol biosynthesis.

## 1. Introduction

*Taxus* is the solitary natural source of paclitaxel, also named Taxol in trade, a well applied chemotherapy agent used to treat different cancers (Marupudi et al., 2007). Owing to its extremely low content in *Taxus* tissues, plant-derived taxol is in short supply, which greatly limits its clinical application (Feng et al., 2023; Zhu and Chen, 2019). As incidences of cancer increase, so do the market and medicinal values of taxol. To address supply issues, several chemical and semichemical

synthesis strategies have been developed to produce taxol from *Taxus* cell lines or through heterologous systems (Ajikumar et al., 2010; Liang et al., 2014). Understanding taxol biosynthesis is essential for exploring alternative production methods (Yu et al., 2020).

As a tetracyclic diterpene, taxol, like many other secondary metabolites, is naturally synthesized using geranylgeranyl diphosphate (GGPP) as a starting material (Horiguchi et al., 2008). First, one dimethylallyl pyrophosphate molecule and three isopentenyl diphosphate molecules are catalyzed into GGPP molecule by GGPP synthase (GGPPS)

\* Corresponding authors at: College of Life and Environmental Sciences, Hangzhou Normal University, Hangzhou 311121, China.

E-mail addresses: [whz62@163.com](mailto:whz62@163.com) (H. Wang), [shencj@hznu.edu.cn](mailto:shencj@hznu.edu.cn) (C. Shen).

<https://doi.org/10.1016/j.indcrop.2023.117411>

Received 2 April 2023; Received in revised form 22 August 2023; Accepted 30 August 2023

Available online 14 September 2023

0926-6690/© 2023 Elsevier B.V. All rights reserved.

(Howat et al., 2014). Then, a specific sub-pathway begins with the catalyzation of GGPP into the taxadiene core, which is modified by a series of oxidases, hydroxylases, and acetylases (Han et al., 2022a). Several key acyltransferases, such as taxadienol 5 $\alpha$ -O-acetyl transferase (TAT), taxane-2 $\alpha$ -O-benzoyltransferase (TBT), and baccatin III-13-O-phenylpropanoyl transferase (BAPT), were predicted to play a role in taxol biosynthesis pathway (Kuang et al., 2019). The final taxol is formed by 3'-N-debenzoyl-2'-deoxytaxol-N-benzoyl transferase (DBTNBT) (Long et al., 2008). Many transcription factors (TFs) have been functionally identified, including MYBs (TmMYB3 and TmMYB39), MYCs (TcMYC1, TcMYC2, and TcMYC4), WRKYs (TcWRKY1, TcWRKY7, and TcWRKY47), and ERFs (TcERF12 and TcERF15) (Li et al., 2013; Yu et al., 2022; Yu et al., 2020; Zhang et al., 2018; Zhang et al., 2015).

Chromatin is organized by the complex of DNA and different histone proteins (Armeev et al., 2019). Epigenetic modifications of DNA and histone proteins control chromatin packing and accessibility, which can restrict TF access to *cis*-regulatory elements (Sijacic et al., 2018). The efficiency with which TFs trigger the expression of downstream genes is positively correlated with chromatin accessibility (Zhang et al., 2021). The chromatin-dependent regulation of secondary metabolite biosynthesis has been well-studied in fungi (Collemare and Seidl, 2019). However, a lack of information about chromatin's spatial organization in plants has limited our understanding of the regulatory genes involved specifically in secondary metabolism.

High-throughput sequencing integrated with several newly developed methods has been used in plants to investigate chromatin accessibility at a genome-wide level (Zhong et al., 2016). Transposase-accessible chromatin sequencing (ATAC-seq) is a useful method for identifying open chromatin regions (OCRs) (Buenrostro et al., 2013). Owing to its high sensitivity, ATAC-seq requires a very small amount of plant material, and library construction is simple (Adey et al., 2010). In plants, ATAC-seq has uncovered the roles of OCRs in sex determination, anthocyanin biosynthesis, floral scent production, and environmental stimuli and developmental signal responses (Cheng et al., 2022; Han et al., 2022b; Maher et al., 2018; Zhou et al., 2022). Plant cells, from different tissues and developmental processes, employ various mechanisms to modify the accessibility of regulatory chromatin regions (Burton and Torres-Padilla, 2014). To date, only a few tissue-specific *cis*-regulatory elements have been reported in plants.

Leaves possess various cell types, including mesophyll, vein, stomatal complex, procambium, and guard, and they are the primary sites of light acquisition and nutrition utilization (Kalve et al., 2014). Some metabolites have shown cell type-specific accumulation patterns in fresh sage, ginkgo, and tea leaves (Wu et al., 2022a). For examples, 176 alkaloids have been detected in the vein cells of mature *Gelsemium elegans* leaves, and a number of alcohols significantly accumulate in the surface cells of adaxial *Populus trichocarpa* leaves (Kulkarni et al., 2018; Wu et al., 2022b). A large number of active compounds, such as flavonoids, taxoids, and polysaccharides, have been detected in *Taxus mairei* leaves (Hao et al., 2017; Monacelli et al., 2002; Wang et al., 2019; Yang et al., 2016). Several previous studies have confirmed that taxol biosynthesis occurs in the leaves of *T. mairei* (Zhan et al., 2023). The contents of four important taxoids, such as taxol, 10-deacetyl baccatin III, baccatin III, cephalomannine, 10-deacetyl taxol, were determined by UPLC in the leaves of *T. mairei* (Yang et al., 2016). Among different tissues, the leaf of *T. mairei* has the highest taxol content ( $5.18 \times 10^{-5}$ , w/w) (Dong et al., 2010). Leaves of *T. mairei* are used for industrial extraction of taxol, because they have sufficient biomass and are renewable. Understand taxoid localization is important for the comprehensive utilization of *Taxus* resources.

Recent single-cell (scRNA) RNA sequencing method has been applied to investigate chromatin accessibility at a higher resolution (Marand et al., 2021b). The single-cell regulatory landscapes was first studied in model plants *Arabidopsis* and maize (Dorrity et al., 2021; Marand et al., 2021a). Cell-specific chromatin-accessible landscapes also aid in

understanding the regulatory mechanisms underlying cell type-specific metabolite biosynthesis. However, the effect of cell-specific chromatin-accessible on secondary metabolism in *Taxus* trees are largely unknown. In our study, scATAC-seq technique has been used extensively to reveal the changes in chromatin accessibility between leaf epidermal and mesophyll cells. A number of cell type-specific TF motifs involved in taxol biosynthesis have been predicted. Our study provides a useful information for predicting the genome-based regulatory motifs that underlie secondary metabolite biosynthesis in medicinal plants.

## 2. Experimental procedures

### 2.1. Nuclei isolation and scATAC-seq

Leaves from new twigs of three-year-old *T. mairei* were harvested for nuclei isolation. Nuclei were isolated according to a published protocol (Dorrity et al., 2021). In detail, 1 g of leaves was cut into small pieces with a razor blade in 1 mL of extraction solution containing 0.8 M sucrose, 25 mM Tris-HCl, 10 mM MgCl<sub>2</sub> and 1  $\times$  Protease Inhibitor. The extract was kept on ice for 5 min

The nuclear suspension was adjusted to 500–1500 cells/ $\mu$ L. The cell nucleus suspension was incubated with the transposase mixing solution. The transposase enters the cell nucleus, preferentially cuts the open chromatin region to fragment DNA, and adds sequencing primer to the end of the DNA fragment. The gel beads with barcode and single nuclei were wrapped in an oil droplet to form 'Gel Beads in Emulsion' (GEM). After the gel beads were dissolved, the DNA fragments were released from the nucleus splits. The single DNAs with P5 connector and 10  $\times$  barcode were linear amplified by thermal cycle produce. The labeled DNAs were recovered from the break GEMs, and the P7 connector was introduced to DNA fragments to construct the library by PCR method. After quality checking, high-throughput sequencing of the library was performed on a 10  $\times$  Genomics Chromium platform.

### 2.2. Quality control of the raw data

To ensure high-quality sequencing data, raw sequence data was subjected to quality control, including the amount of sequencing data, the available barcode carrier rate, and the sequencing quality Q30. The 16 bp barcode obtained by sequencing was compared with the correct barcode (whitelist) and the frequency of the correct barcode was counted. Extract error barcodes that differ from the whitelist in two bases (Hamming Distance $\leq$ 2) were extracted and sorted according to the abundance (read number) corresponding to the barcode and the quality value of the error base. Error barcodes with a correction probability that the observed barcode is the real barcode is  $> 90\%$  were corrected.

### 2.3. Processing of raw *T. mairei* leaf scATAC data

The 3' connector sequence of raw data was removed using the cutadapt tool and the sequence data without connector were compared to the *T. mairei* genome using the BWA-MEM software with the default parameter. The sequences less than 25 bp in length were stored in un-mapped BAM output file. The reads with same R1 and R2 alignment positions were considered to be from the same DNA fragment. Among all the reads with the same barcode, one read is selected and defined as "original", and the other reads are considered as duplicates. After filtering with the following parameters: R1 and R2 MAPQ  $> 30$ , no mitochondria, and non-chimeras, clean original reads were used for downstream comparison. The 5' end of original reads were shifted 4 bp to the right and the 3' end of original reads were shifted 5 bp to the left. The adjusted position is the Tn5 enzyme cutting center. The sequence between the enzyme cutting sites after the position adjustment is defined as fragment, and the information is stored in the fragments.tsv.gz file.

## 2.4. Peak annotation and cell calling

The number of transposition events of each site in genome was calculated using the sites determined at the end of fragment in the fragments.tsv.gz file. The smooth contour of transposition events is generated by counting the 400 bp moving window, and a mixed model similar to ZINBA is used to fit the data. Merge the events within 500 bp into one peak and store it in the BED file. Chromatin accessibility analysis of known, annotated and epigenetic related regions in the *T. mairei* genome. Several sites, including transcription start site (TSS), DNase highly sensitive region sites, enhancer binding region sites, promoter binding region sites, binding sites defined by ENCODE, and peaks binding sites, were used for peak annotation. GO/KEGG enrichment analysis of the genes that can be scanned to peak was performed to understand the possible biological process or biological function.

For cell calling, the proportion of fragments in peak of each barcode and the proportion of fragments in 2000 bp around the peak were compared. Then subtract a fixed value, is the estimated number of fragments of each barcode from different GEMs, related to sequencing depth from all barcode counts. Two mixed models of negative binomial distribution are fitted to obtain signals and noises. The odds ratio is set to 100000 to distinguish the corresponding barcodes of cells and non-cells.

## 2.5. Cell visualization

To visualize the data, dimensionalities of most resulting cells were reduced into 2D-space using the cluster mode of Cell Ranger ATAC software (ver. 1.2.0) with Latent Semantic Analysis method. 'CellRanger Aggr' program was applied to normalize the number of fully mapped reads per cells.

## 2.6. Cell clustering and marker identification

The aggregated peak-cell matrix was uploaded into R (ver. 4.0.4) using the single cell analysis packages Seurat (ver. 4.0.0) and Signac (ver. 1.2.0). The gene expression matrix was generated and Chrom\_Var deviation score matrix were uploaded onto the Seurat as objects. In order to reducing batch effects, Harmony algorithm was used iteratively corrects LSI embeddings. Components were used to produce a Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction. A Shared-Nearest-Neighbor graph was produced from the second 30 Latent Semantic Indexing components by the elbow plot method. The Shared-Nearest-Neighbor graph was applied to calculate Seurat's LSM algorithm, which was used to cluster cells. Find\_Transfer\_Anchors and Transfer\_Data programs were used to identify the cell-type of scATAC data. After label transfer, cell type-specific TF motifs were identified using Seurat's Find\_All\_Markers function with the logistic regression method.

## 2.7. Leaf tissue type annotation

A series of known marker genes were selected to screen cell type specific markers of *T. mairei*. To screen cell type specific markers, the orthologous genes of the reported leaf-related markers in model plant *Arabidopsis* were aligned. The detailed sequences of *Arabidopsis* marker genes were downloaded from the Plant Cell Marker DataBase (Jin et al., 2022). Using the *Arabidopsis* genes as the queries, the *T. mairei* orthologous genes were searched by BLAST function in TBtools software. The hits with high score were selected to annotate the cell types in *T. mairei*. To identify and annotate cell types, the average feature values from the nearest known marker genes were used.

## 2.8. Functional enrichment analysis of peak targets

For target functional analysis, Compare\_Cluster function in the cluster\_Profiler program was applied. GO and KEGG enrichment terms

were got using adjust *P* value < 0.05 with the Benjamini-Hochberg method. Enrichment analysis is implemented with the Fisher's exact test.

## 2.9. Motif enrichment analysis

*T. mairei* TF motifs were Predicted by PlantTFDB ver. 5.0. Then, Motif\_Scan program were used to predict all *T. mairei* TF binding sites in the chromatin accessibility region with cut-off *P* value < 1E-4. Two programs, 'genome\_compile' and 'motif\_compile' were used to generate reference index and background score. At last, the 'add\_Peak\_Annotations' function was used to add annotation to our TF motif-binding site and the 'add\_Deviations\_Matrix' function was used to get a motif deviation score matrix. All peaks were matched with the given transcription factor basing on the TF motif. After filtering, the fragments of each peak that can be matched to peak-barcode matrix were integrated, and the number of fragment in barcode that can be matched to the same TF motif was calculated.

## 2.10. cis-Element scanning of OCR sequences

The OCR sequences were scanned using the PlantCARE program. Several classic binding elements, including ERF binding sites (GCCGCGC and CCGAC), bHLH binding sites (CANNTG, CACG(A/C)G and CACGTG), WRKY binding site (TTGACC), and MYB binding sites (CAACTG and CAACCA), were screened.

## 2.11. Gene cloning, prokaryotic expression and EMSA

Total RNAs were isolated using a RNeasy MiniKit (Qiagen, Hilden, Germany) and were used to generate cDNAs using a cDNA library preparation kit (Illumina, San Diego, USA). The full-length ORF sequences of bHLH18, MYB108, WRKY22, and ERF39 were cloned into pGEX-4 T vector and transformed into *Escherichia coli* cells to produce recombinant His-tag proteins. The recombinant proteins were purified using His60 Ni Superflow Resin (Clontech) and isolated by 12% SDS-PAGE electrophoresis. EMSA was performed according to our previous works (Yu et al., 2022; Yu et al., 2020). All probe sequences are listed in Table S1.

## 2.12. Dual-luciferase reporter assay

The full-length sequences of bHLH18, MYB108, WRKY22, and ERF39 were cloned into the pGreenII62-SK vector as effectors, and the promoter sequences of TS, T5OH, and DBTNBT genes were cloned into the pGreenII 0800-LUC vector as reporters. After co-transformed into tobacco leaf cells, LUC activities were calculated using a Promega dual-luciferase assay kit. The primer sequences are listed in Table S2.

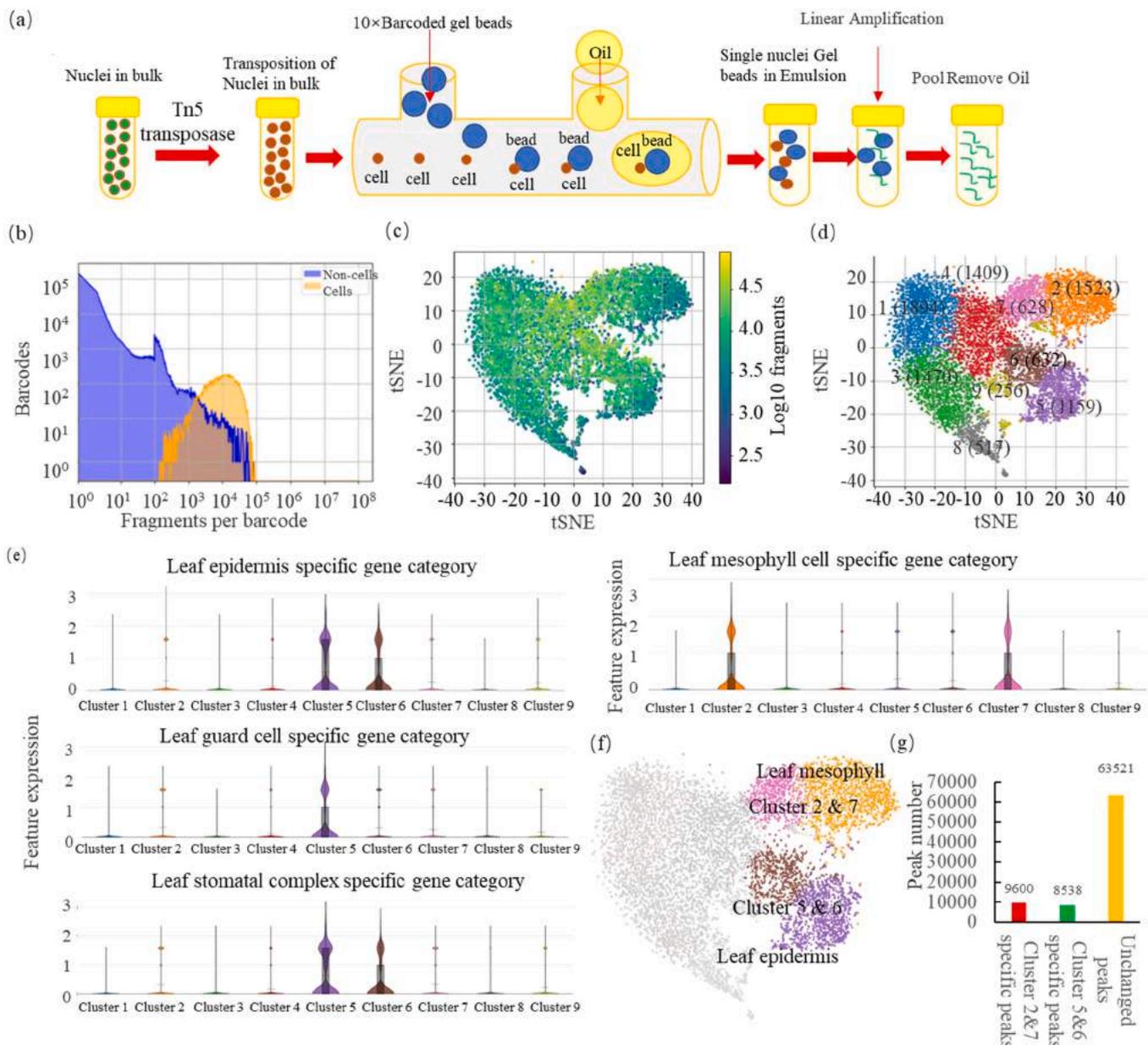
## 2.13. Data availability

The *T. mairei* reference genome was downloaded from NCBI database (ID: PRJNA730337). The ScATAC datasets of *T. mairei* leaves have been uploaded to the NCBI database under accession number PRJNA909435.

## 3. Results

### 3.1. Overview of the scATAC data from *T. mairei* leaves

We performed scATAC-seq on *T. mairei* leaves using a 10 × Genomics microfluidic device and mixed the sequences with single-cell barcodes (Fig. 1a). The scATAC-seq generated 400 million reads, with 93.63% containing a valid barcode. Approximately, 66.6% of the raw reads were mapped onto the *T. mairei* reference genome. Based on the fragment coverage that corresponded to each barcode, all the captured cells were classified into two different types: cells and non-cells (Fig. 1b). An



**Fig. 1.** scATAC-seq and cell clustering of *T. mairei* leaves. (a) Workflow of the scATAC-seq of *T. mairei* leaves. Nucleus isolated the leaves were loaded into a 10x Genomics Chromium Controller. (b) Fragment distribution in cells (indicated by different barcodes). Blue color indicated non-cells and yellow color indicated cells. (c) t-SNE projections of cells colored by fragment number. (d) Cell clustering using SNE method. The Number of cells in each cluster. Dots indicated individual cells and colour indicated different cell clusters. (e) The Violin charts of cell type specific expression marker genes. (f) Visualization of leaf mesophyll (cluster 2 & 7) cells and leaf epidermis (Cluster 5 & 6) cells. (g) The number of peaks in the leaf mesophyll (cluster 2 & 7) cells and leaf epidermis (Cluster 5 & 6) cells.

estimation using the valid barcodes predicted 9488 cells with 42,192 raw reads per cell. The fraction of high-quality fragments in the cells was 89.3%, and the fraction of transposition events that fell into the peak regions was 4.7% (Table S3).

### 3.2. Peak annotation

To reveal the genomic locations of putative *cis*-regulatory elements, OCRs were analyzed to produce fragment peaks. The transcription start site (TSS) enrichment score was 2.44, with 7.69% of the high-quality fragments overlapping a TSS (Table S4). Moreover, 66.64% of the fragment reads were confidently mapped onto the *T. mairei* genome, whereas 2.37% of the fragment reads were unmapped. The distribution map of fragments overlapping peaks is shown in Fig. S1a, and the relative position of each peak from a TSS is shown in Fig. S1b. The

positional relationship between peaks and genes provides important reference information for peak function predictions (Table S5). All the peaks were grouped into three types, promoter region (29,410 peaks), distal region (62,701 peaks), and intergenic region (3171 peaks) (Fig. S1c). A Gene Ontology (GO) enrichment analysis of peak targeting genes showed that ‘transcription’ (GO:0006351), ‘nucleosome’ (GO:0000786), and ‘DNA binding transcription factor activity’ (GO:0003700) were the most significantly enriched GO terms (Fig. S1d). A Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis of peak targeting genes showed that ‘RNA polymerase’ (ko03020), ‘Pyrimidine metabolism’ (ko00240), and ‘Purine metabolism’ (ko00230) were the most significantly enriched KEGG terms (Fig. S2).

### 3.3. Cell clustering

Single-cell studies require the assistance of visualization tools to compare the data among different cell groups. The *t*-SNE projection of cells is displayed as colored fragment counts (Fig. 1c). The detected fragments were imported into Seurat loupe software, and nine cell clusters with uneven cell numbers were identified (Table S6). Cluster 1 contained the largest number of cells (1894) and Cluster 9 contained the smallest number of cells (256) (Fig. 1d). Across all the detected cells, 27,721 peaks (33.9%) contained cell-type information. The number of differentially expressed peaks in each cluster ranged from 800 (Cluster 9) to 26,624 (Cluster 2).

### 3.4. Cell-type annotation

Different cell types were defined according to the differentially expressed genes near the peaks. In total, 18 leaf epidermal, 13 guard, 15

leaf phloem, 6 bundle sheath and vein, 12 mesophyll, 18 leaf pavement, 5 leaf procambium, and 23 stomatal complex cell-specific expressed genes were applied as markers to annotate the scATAC-seq cell types (Table S7). The peaks near the leaf epidermis specifically expressed genes were enriched in Clusters 5 and 6, and the peaks near the leaf mesophyll cell markers were enriched in Clusters 2 and 7. The peaks near the leaf guard cell markers were enriched in Cluster 5, and the peaks near the stomatal complex markers were enriched in Clusters 5 and 6 (Fig. 1e).

Although thousands of differential OCRs have been detected, they do not precisely predict the expression level of a specific nearby gene (Dorrity et al., 2021). Based on the current data, Clusters 1, 3, 4, 8, and 9 remained un-annotated. We considered Clusters 2 and 7 as a group (leaf mesophyll cells) and Clusters 5 and 6 as another group (leaf epidermal cells) (Fig. 1f). A differential accessibility analysis identified 9600 leaf mesophyll cell-specific OCRs and 8538 leaf epidermal cell-specific OCRs (Fig. 1g).

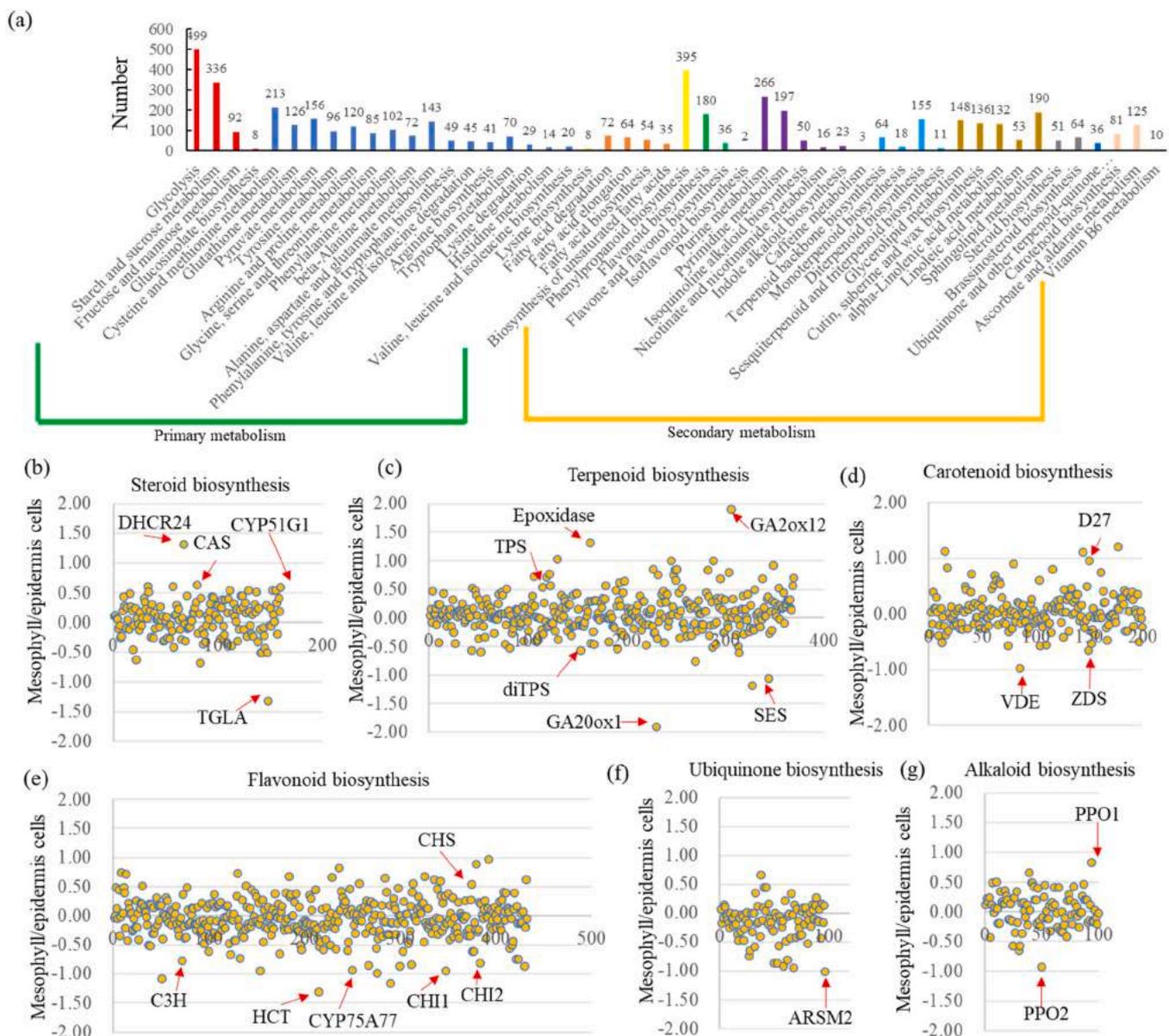


Fig. 2. Cell type specific chromatin accessibility of primary and secondary metabolism. (a) KEGG analysis of the primary and secondary metabolism-related genes. The chromatin accessibility differences of peak targeting genes, including steroid biosynthesis-related genes (b), terpenoid biosynthesis-related genes (c), carotenoid biosynthesis-related genes (d), flavonoid biosynthesis-related genes (e), ubiquinone biosynthesis-related genes (f), and alkaloid biosynthesis-related genes (g), between leaf mesophyll cell and leaf epidermis cell. Red arrows indicated the differential genes with  $|\log(\text{mesophyll/epidermis cells})| > 1$ .

### 3.5. Cell type-specific chromatin accessibility of metabolism-related genes

To investigate the relationship between metabolic regulation and chromatin accessibility, a KEGG enrichment analysis of all the peak targeting genes was performed (Table S8). For primary metabolism, four saccharide-related pathways, consisting of 935 peak targeting genes, 17 amino acid metabolism-related pathways, consisting of 1389 peak targeting genes, and 4 fatty acid metabolism-related pathways, consisting of 225 peak targeting genes, were identified. For secondary metabolism, 395 phenylpropanoid-, 218 flavonoid-, 555 alkaloid-, 248 terpenoid-, 659 lipid-, 115 steroid-, 36 quinone-, and 216 oxidation-related peak targeting genes were identified (Fig. 2a).

A tissue-specific analysis of the peaks near the secondary metabolism-related genes was performed (Fig. S3). For the steroid biosynthesis pathway, peaks corresponding to *DHCR24*, *CAS*, and *CYP51G1* were significantly enriched in the mesophyll cells, and the peak corresponding to *TGLA* was significantly enriched in the epidermal cells; for the terpenoid biosynthesis pathway, peaks corresponding to *Epoxidase*, *GA2ox12*, and *TPS* were significantly enriched in the mesophyll cells, and peaks corresponding to *ditPS*, *GA2Oox1*, and *SES* were significantly enriched in the epidermal cells; for the carotenoid biosynthesis pathway, the peak corresponding to *D27* was significantly enriched in the mesophyll cells, and the peaks corresponding to *VDE* and *ZDS* were significantly enriched in the epidermal cells; for the flavonoid biosynthesis pathway, the peak corresponding to *CHS* was significantly enriched in the mesophyll cells, and peaks corresponding to *C3H*, *HCT*, *CYP75A77*, *CH11*, and *CH12* were significantly enriched in the epidermal cells; and for the ubiquinone and alkaloid biosynthesis pathways, the peak corresponding to *PPO1* was significantly enriched in the mesophyll cells, and peaks corresponding to *ARSM2* and *PPO2* were significantly enriched in the epidermal cells (Fig. 2b).

### 3.6. Chromatin accessibility of the genes involved in the MEP pathway in *T. mairei* leaves

After searching the *T. mairei* genome, three 1-deoxy-D-xylulose 5-phosphate synthase (*DXS*)-encoding genes, two 1-deoxy-D-xylulose 5-phosphate reductoisomerase (*DXR*)-encoding genes, one 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase (*CMK*)-encoding gene, three 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (*MDS*)-encoding genes, two 4-hydroxy-3-methylbut-2-enyl-diphosphate synthase (*HDS*)-encoding genes, four *GGPPS*-encoding genes, and six *GGPPS*-encoding genes were identified (Fig. 3a).

Using the scATAC and *T. mairei* genome datasets, we predicted the features of the MEP pathway genes. Most of the peaks were located in the promoter and distal regions (Fig. 3b). For the *DXS* genes, one peak was detected in each of the promoter regions of *ctg7293\_gene.1* and *ctg953\_gene.7*, and no peak was detected near the *ctg8537\_gene.2*; for the *DXR* gene, one peak was detected in the distal region of *ctg895\_gene.1*; for the *CMK* gene, one peak was detected in the distal region of *ctg683\_gene.8*; for the *MDS* genes, one peak was detected in each of the distal regions of *ctg10290\_gene.1* and *ctg6748\_gene.1*, and two peaks were detected in the promoter region of *ctg739\_gene.2*; for the *HDS* genes, one peak was detected in the distal region of *ctg6747\_gene.2*, and no peak was detected near *ctg6747\_gene.1*; for the *GGPPS* genes, at least one peak was identified in the distal regions of *ctg1206\_gene.2*, *ctg12725\_gene.8*, *ctg12725\_gene.14*, and *ctg35859\_gene.2*; and for the *GGPPS* genes, three peaks each were detected near *ctg12725\_gene.16*, *ctg1897\_gene.3*, and *ctg3914\_gene.4*.

A published transcriptome dataset was used to investigate whether there is a close correlation between the peak number and gene expression level (Xiong et al., 2021). The *GGPPS*-encoding genes (*ctg12725\_gene.8* and *ctg35859\_gene.2*) with more than two peaks displayed higher expression levels than their homologous genes with only single peaks. For the *GGPPS*-encoding genes, *ctg12725\_gene.16* (three peaks), *ctg1897\_gene.3* (three peaks), and *ctg3914\_gene.4* (two

peaks) showed the highest expression levels, whereas *ctg3914\_gene.1* (one peak) and *ctg4334\_gene.2* (one peak) were not expressed (Fig. 3c).

### 3.7. Chromatin accessibility of taxol biosynthesis-related genes in *T. mairei* leaves

Our scATAC-seq detected a number of genes involved in taxol biosynthesis, including three *TS* genes, five *T5OH* genes, three *TAT* genes, one *TBT* gene, five *T10OH* genes, four *T13OH* genes, two *DBBT* genes, one *DBAT* gene, two *BAPT* genes, and one *DBTNBT* gene (Fig. 4a). For the *TS* genes, two peaks were identified near *ctg6088\_gene.1*, and no peak was detected near *ctg5306\_gene.4* or *ctg7747\_gene.1*; for the *T5OH* genes, four peaks were detected near *ctg7747\_gene.2*, whereas one peak each was detected near *ctg2768\_gene.2* and *ctg11276\_gene.1*; for the *T13OH* genes, three peaks were detected near *ctg4364\_gene.1* and one peak was detected near *ctg593\_gene.6*; for the *TAT* genes, *ctg195\_gene.25*, *ctg195\_gene.31*, and *ctg5028\_gene.2* contained at least two peaks; for the *BAPT* gene, one peak was identified in the distal region of *ctg10553\_gene.2*; for the *DBAT* gene, two peaks were identified in the promoter region of *ctg5183\_gene.4*; for the *T10OH* genes, one peak each was identified near *ctg2120\_gene.10* and *ctg5026\_gene.32*; for the *TBT* genes, at least one peak each was identified near *ctg7327\_gene.1*, *ctg1505\_gene.1*, and *ctg4165\_gene.7*; for the *DBBT* genes, at least one peak each was identified near *ctg18533\_gene.4*, *ctg4356\_gene.1*, *ctg6800\_gene.1*, *ctg12512\_gene.2*, and *ctg2111\_gene.2*; and for the *DBTNBT* genes, two peaks were identified near *ctg887\_gene.16*, and one peak each was identified near *ctg887\_gene.19*, *ctg887\_gene.21*, *ctg4088\_gene.8*, and *ctg6800\_gene.1* (Fig. 4b).

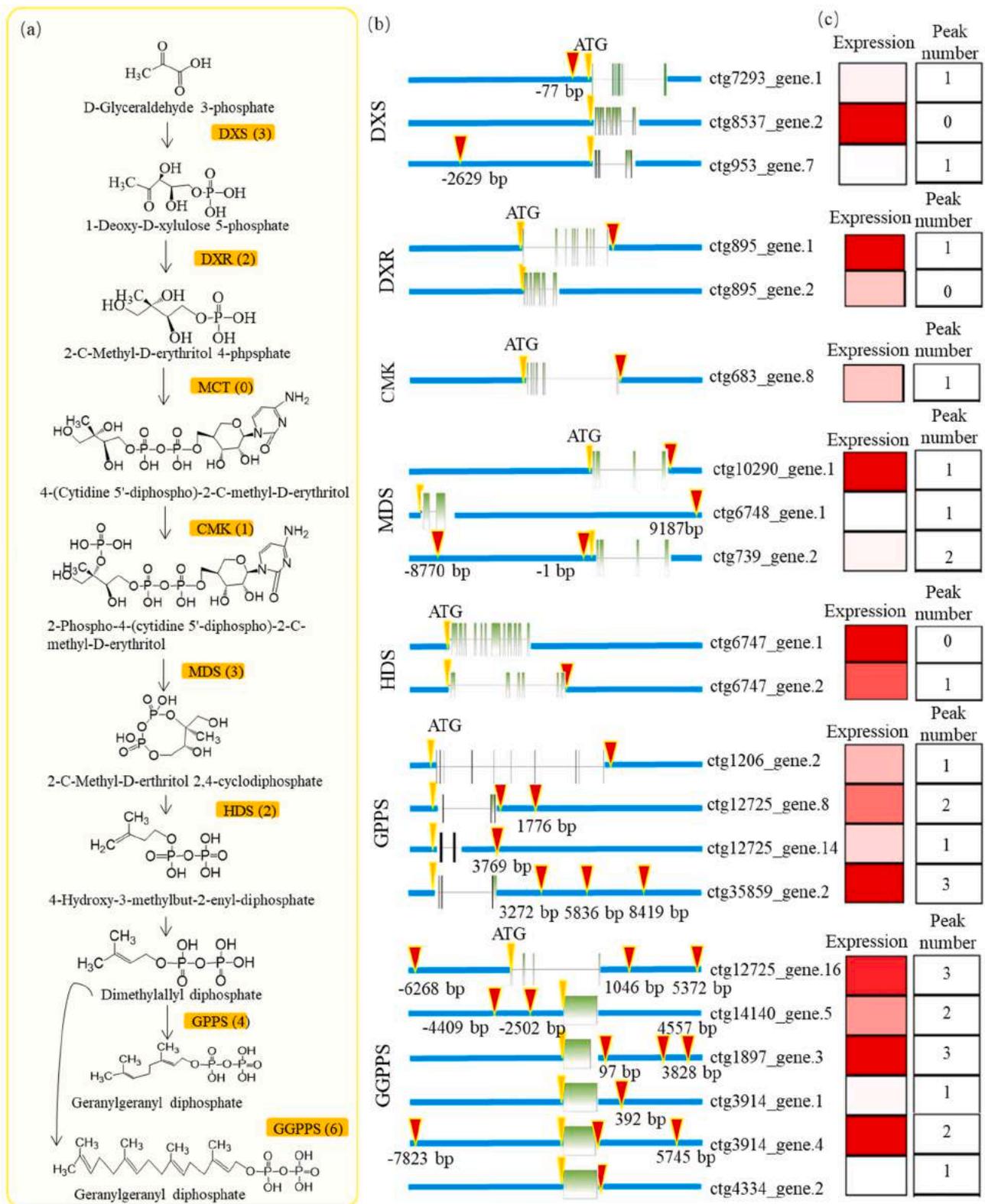
The expression levels of several taxol biosynthesis-related genes showed close correlations with peak number. Among homologous genes, those with the maximum peak numbers, such as *ctg6088\_gene.1* (*TS*, two peaks), *ctg7747\_gene.2* (*T5OH*, four peaks), *ctg4364\_gene.1* (*T13OH*, three peaks), *ctg195\_gene.25* (*TAT*, three peaks), and *ctg18533\_gene.4* (*DBBT*, two peaks), had the highest expression levels (Fig. 4b).

### 3.8. Cell type-specific chromatin accessibility of taxol biosynthesis-related genes

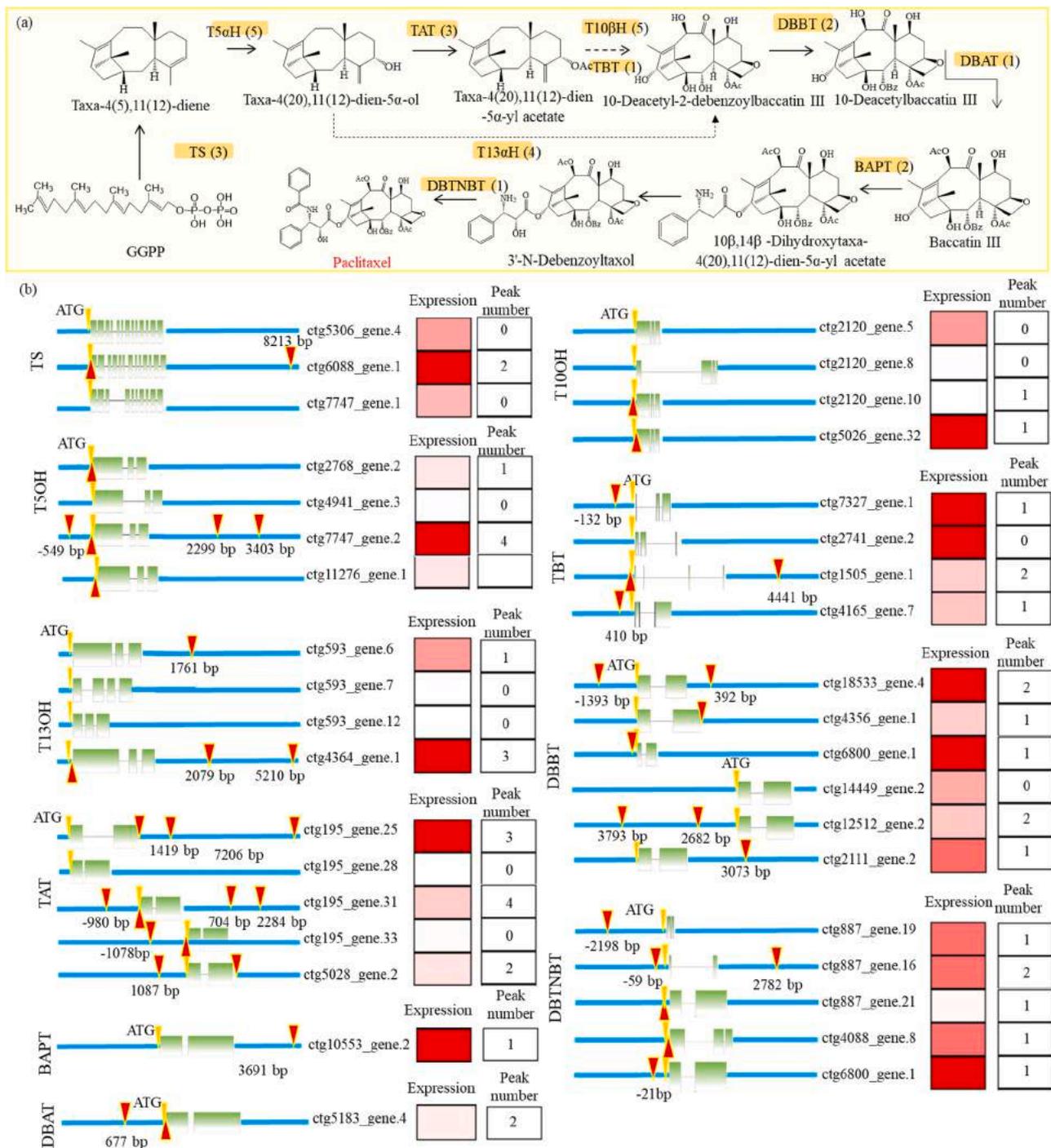
scATAC data revealed cell type-specific chromatin accessibility. Although most peaks were shared by leaf mesophyll and epidermal cells, some of chromatin regions were significantly more accessible in one cell type than the other. In our study, we evaluated differences in peaks surrounding taxol biosynthesis-related genes between leaf mesophyll and epidermal cells. In total, 15 taxol biosynthesis-related peaks with significant variabilities were found between Clusters 2 and 7 (leaf mesophyll cells) and Clusters 5 and 6 (leaf epidermal cells). Four taxol biosynthesis-related genes, *TS* (*ctg6088\_gene.1*), *TAT* (*ctg195\_gene.31*), and *DBTNBTs* (*ctg887\_gene.16* and *ctg6800\_gene.1*), were comparatively more accessible in the leaf epidermal cells. However, nine taxol biosynthesis-related genes, *TAT* (*ctg195\_gene.33*), *T5OHs* (*ctg2768\_gene.2*, *ctg7747\_gene.2*, and *ctg11276\_gene.1*), *TBTs* (*ctg4165\_gene.7* and *ctg7327\_gene.1*), *DBTNBTs* (*ctg887\_gene.21* and *ctg4088\_gene.8*), and *T13OH* (*ctg4364\_gene.1*), were comparatively more accessible in the leaf mesophyll cells (Fig. 5).

### 3.9. Screening of cell type-specific TF motifs

In most cases, TF binding sites, which can enroll potential TFs to the TSSs of their targets, were likely enriched in accessible chromatin regions. In total, 656 TF motifs were identified using the scATAC-seq data (Fig. 6a and Table S9). Among these TF motifs, 122 were significantly enriched in the leaf mesophyll cells, and 153 were significantly enriched in the leaf epidermal cells (Fig. 6b). In the leaf mesophyll cells, the most significantly enriched TF motifs were TGA1A (MA0129.1), ERF1B (MA0567.1), CRF2 (MA0975.1), ERF13 (MA1004.1), PHYPADRAFT



**Fig. 3. Chromatin accessibility of the MEP pathway-related genes in *T. mairei* leaves.** (a) Overview of the MEP pathway. The numbers in brackets represented the number of coding genes. (b) The genomic features of the MEP pathway genes. Yellow triangles indicated the sites of ATG and red triangles indicated the sites of OCR. (c) The heatmaps showed the expression levels of different encoding genes of each MEP pathway-related enzyme. Red indicated high expression level. The numbers in black boxes indicated the number of OCRs nearby each gene.

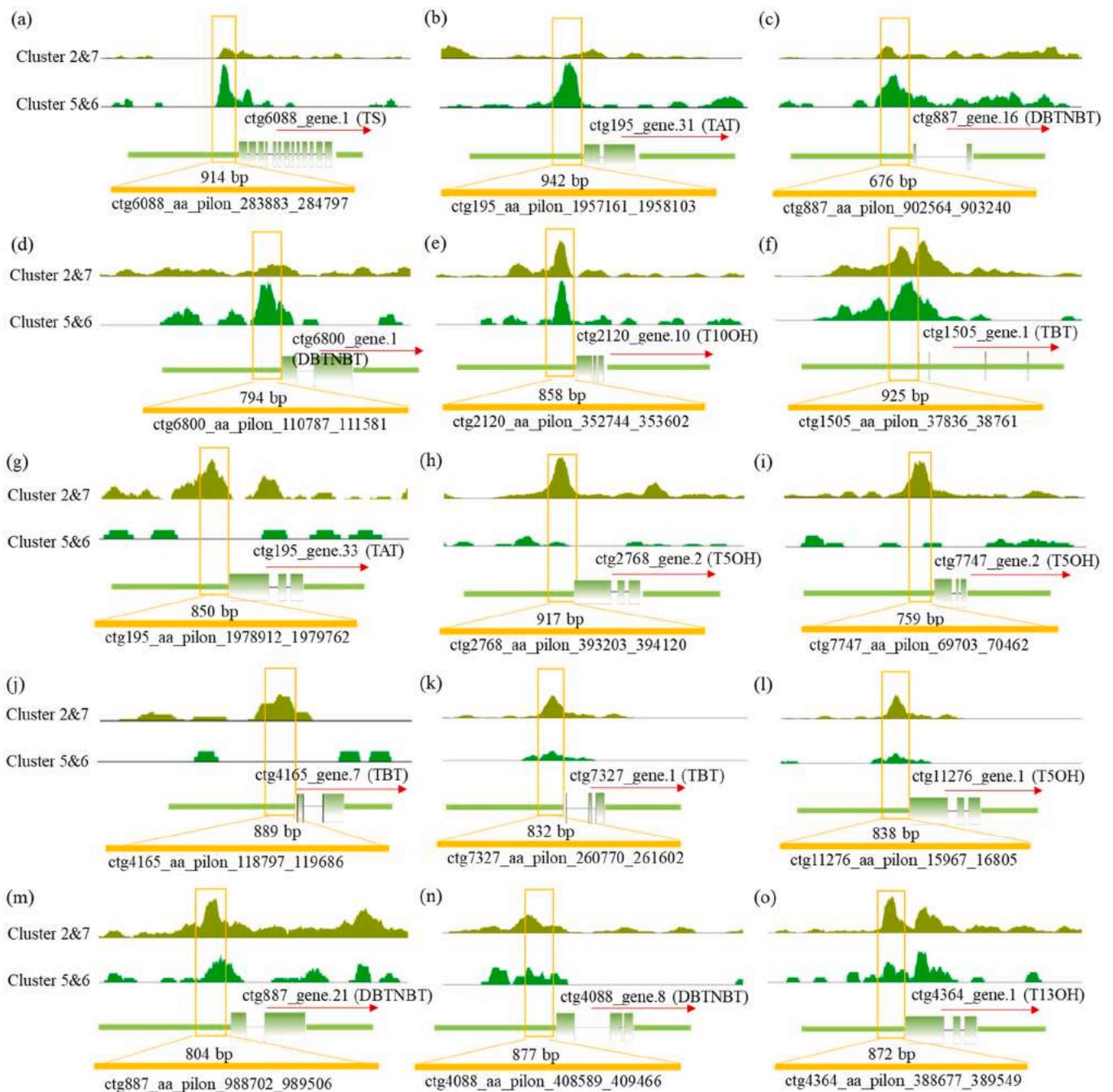


**Fig. 4. Chromatin accessibility of taxol biosynthesis-related genes in *T. mairei* leaves.** (a) Overview of the taxol biosynthesis pathway. The numbers in brackets represented the number of coding genes. (b) The genomic features of the taxol biosynthesis pathway genes. Yellow triangles indicated the sites of ATG and red triangles indicated the sites of OCR. The heatmaps showed the expression levels of different encoding genes of each taxol biosynthesis pathway-related enzyme. Red indicated high expression level. The numbers in black boxes indicated the number of OCRs nearby each gene.

(MA1023.1), KAN4 (MA1028.1), TCP4 (MA1035.1), RAP2-3 (MA1051.1), ERF109 (MA1053.1), TCP5 (MA1067.1), and NID1 (MA1793.1). In the leaf epidermal cells, the most significantly enriched TF motifs were TGA1A (MA0129.1), CCA1 (MA0972.1), TCP4 (MA1035.1), MYB46 (MA1040.1), MYB55 (MA1041.1), TCP19 (MA1063.1), TCP5 (MA1067.1), ARALYDRAFT (MA1096.1), and ATHB-23 (MA1327.2).

The ERF, WRKY, MYB, and bHLH families play important roles in the regulation of taxol metabolism (Zheng et al., 2023). For the ERF-type motifs, ERF01 (MA1424.1), ERF03 (MA0995.2), and ERF025

(MA1752.1) were significantly enriched in the mesophyll cells; for the WRKY-type motifs, WRKY21 (MA1079.2) and WRKY55 (MA1305.1) were significantly enriched in the mesophyll cells, whereas WRKY22 (MA1303.1) was significantly enriched in the epidermal cells; for the MYB-type motifs, MYB65 (MA1177.1), MYB3R1 (MA1178.2), MYB56 (MA1174.1), and MYB105 (MA1169.1) were significantly enriched in the mesophyll cells, whereas MYB108 (MA2026.1) was significantly enriched in the epidermal cells; and for the bHLH-type motifs, bHLH130 (MA1358.1) was significantly enriched in the mesophyll cells, whereas bHLH18 (MA1361.1) was significantly enriched in the epidermal cells



**Fig. 5.** Cluster-aggregated chromatin accessibility surrounding the taxol biosynthesis-related genes. The detail information of the OCRs nearby ctg6088\_gene.1 (a), ctg195\_gene.31 (b), ctg887\_gene.16 (c), ctg6800\_gene.1 (d), ctg2120\_gene.10 (e), ctg1505\_gene.1 (f), ctg195\_gene.33 (g), ctg2768\_gene.2 (h), ctg7747\_gene.2 (i), ctg4165\_gene.7 (j), ctg7327\_gene.1 (k), ctg11276\_gene.1 (l), ctg887\_gene.21 (m), ctg4088\_gene.8 (n), and ctg4364\_gene.1 (o), was shown. Yellow lines indicated OCRs.

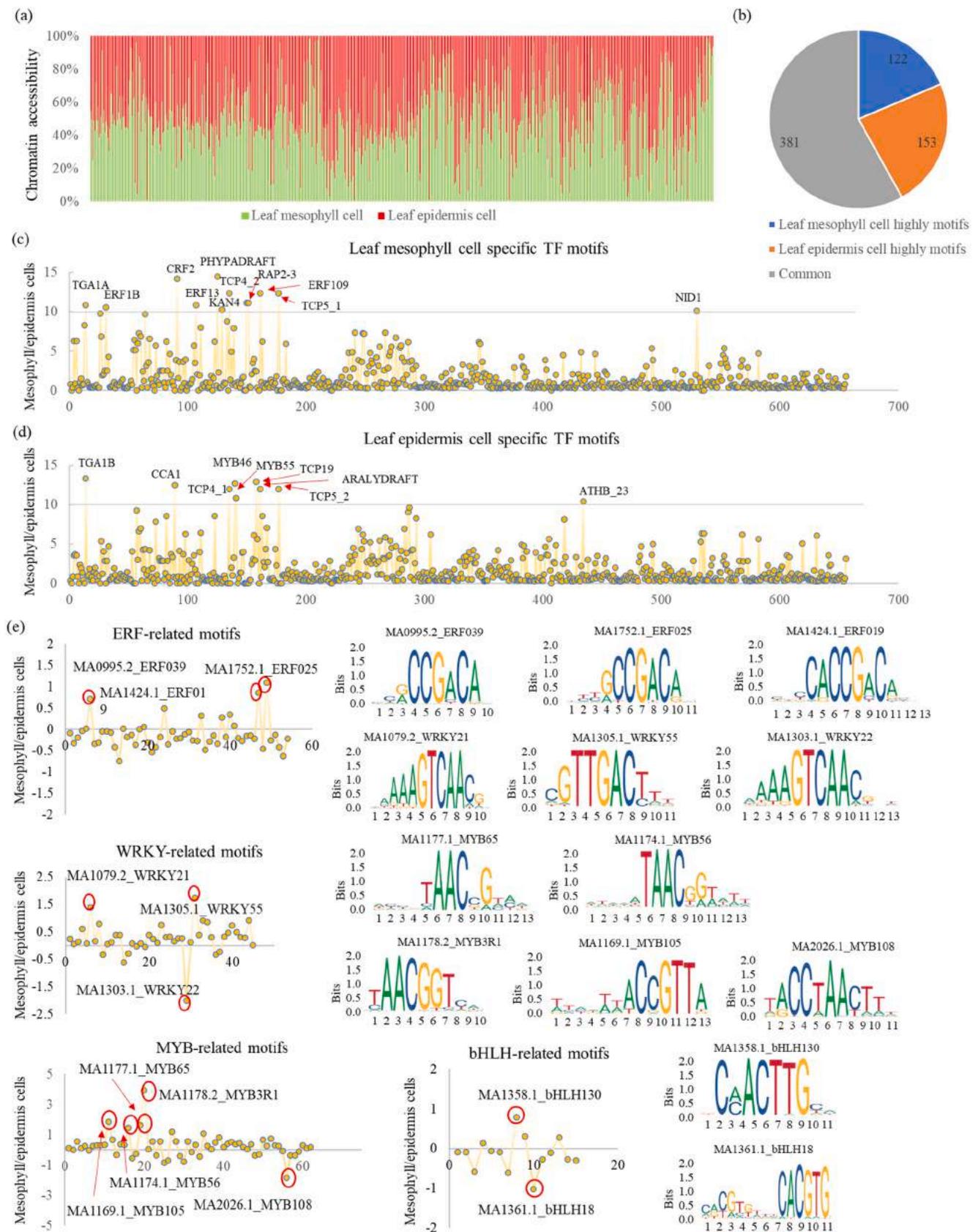
(Fig. 6d).

### 3.10. Prediction of TFs involved in the regulation of taxol biosynthesis

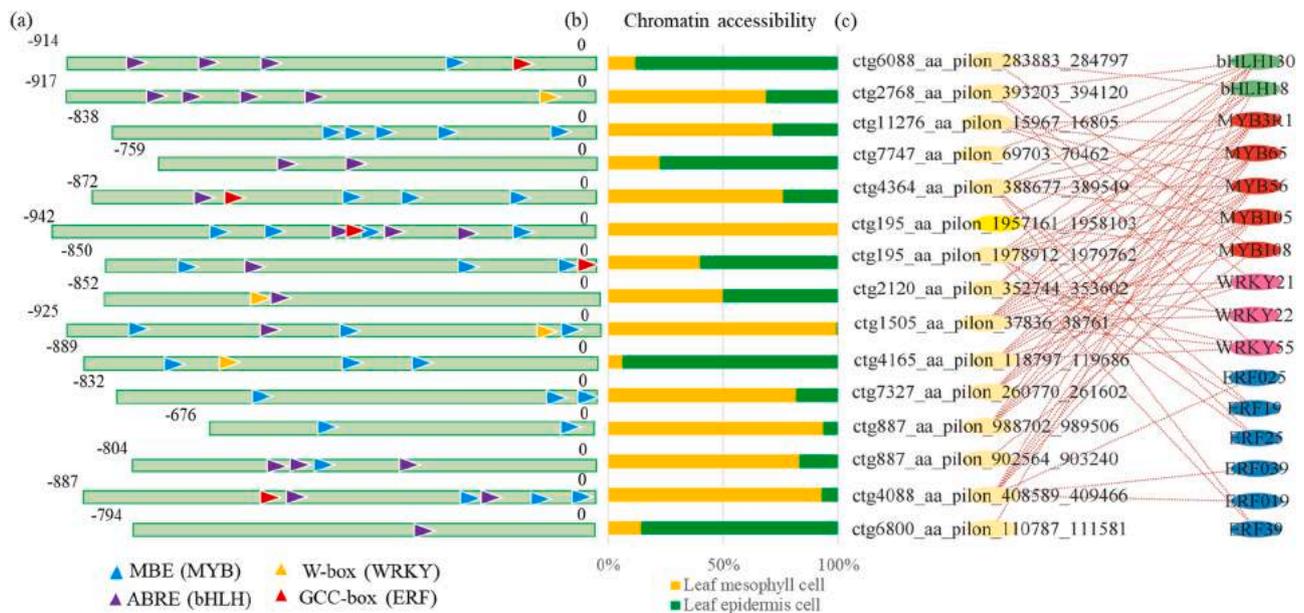
Although several taxol biosynthesis-related TFs were uncovered in different *Taxus* species, information on tissue-specific transcriptional regulation of taxol biosynthesis in leaf is limited. Although a specific motif may be recognized by a type of TF, we could not distinguish among members with similar sequence preferences. To identify the potential TFs involved in taxol biosynthesis, the sequences of taxol biosynthesis-related OCRs were extracted from the public *T. mairei* genome. The distribution of various *cis*-elements, such as MBE, W-box,

W-box, ABRE, and GCC-box, in all the selected OCRs is shown in Fig. 7a. The cell type-specific chromatin accessibility of the selected 15 taxol biosynthesis-related genes is shown in Fig. 7b.

Two criteria, similar chromatin accessibility pattern and matching binding elements, were applied to screen the cell type-specific TFs. Two bHLHs (bHLH18 and bHLH130), five MYBs (MYB3R1, MYB65, MYB56, MYB105, and MYB108), three WRKYs (WRKY21, WRKY22, and WRKY55), and three ERFs (ERF025, ERF019, and ERF039) were considered to be potential regulators of the taxol biosynthesis pathway. Furthermore, the regulatory network for taxol biosynthesis is shown in Fig. 7c.



**Fig. 6. Analysis of cell type specific TF motifs.** (a) The biased chromatin accessibility between leaf mesophyll cells and epidermis cells. The read counts of each TF motif in cluster 2&7 (leaf mesophyll cells) and cluster 5&6 (leaf epidermis cells) was used to calculate the percentage of chromatin accessibility. (b) The number of tissue-specific TF motifs. (c) Leaf mesophyll cell specific TF motifs. (d) Leaf epidermis cell specific TF motifs.  $|\log_2 \text{fold change}| > 10$  was set as threshold. (e) Analysis of tissue-specific TF family-related motifs, including ERF-, WRKY-, MYB-, and bHLH-related motifs between leaf mesophyll cells and epidermis cells. Sequence Logo of the recovered TF motifs were shown.



**Fig. 7. Prediction of TFs involved in the regulation of taxol biosynthesis.** (a) The distribution of various *cis*-elements in OCRs nearby the taxol biosynthesis-related genes. Blue triangles indicated MBE elements, yellow triangles indicated W-boxes, purple triangles indicated ABRE elements, and red triangles indicated GCC-boxes. (b) The tissue-specific chromatin accessibility of taxol biosynthesis-related genes. The read counts of each OCR in cluster 2&7 (leaf mesophyll cells) and cluster 5&6 (leaf epidermis cells) was used to calculate the percentage of chromatin accessibility.

### 3.11. Verification of TFs binding to their potential taxol biosynthesis-related target genes

Four TFs, bHLH18 (ctg16366\_gene.1), MYB108 (ctg7156\_gene.2), WRKY22 (ctg2531\_gene.1), and ERF39 (ctg11028\_gene.1), were randomly selected to verify their binding activities to potential targets. The full-length sequences of bHLH18, MYB198, WRKY22, and ERF39 were independently used to produce recombinant proteins. Several probes surrounding the binding sites in different OCRs were prepared. Electrophoretic Mobility Shift Assay (EMSA) results showed that bHLH18 bound directly to the ABREs in the OCRs near *TS* (ctg6088\_gene.1) and *T5OH* (ctg7747\_gene.2); MYB108 bound directly to the MBEs from the OCRs near *TS* (ctg6088\_gene.1) and *T5OT* (ctg11276\_gene.1); WRKY22 bound physically to the W-boxes from the OCRs near *T5OH* (ctg2768\_gene.2); and ERF39 bound physically to the GCC-boxes from the OCRs near *DBTNBT* (ctg4088\_gene.8). No significant signals were detected between WRKY22 and *TAT* (ctg195\_gene.33) or between ERF39 and *TBT* (ctg4165\_gene.7) (Fig. 8a-h).

To analyze the transcriptional activities of bHLH18, MYB108, WRKY22, and ERF39, a dual-LUC reporter system was carried out. The full-length sequences of bHLH18, MYB108, WRKY22, and ERF39 and the partial promoter sequences of the *TS*, *T5OH*, and *T10OH* were used in this study (Fig. S4). The results showed that bHLH18 significantly inhibited the expression of its downstream target genes (*TS* and *T5OH*), respectively (Fig. 8i-j). MYB108 significantly up-regulated the expression of its downstream targets (*TS* and *T5OH*), respectively (Fig. 8k-l). WRKY22 significantly induced the expression of its downstream target (*T5OH*, and Fig. 8n). ERF39 significantly increased the expression of its downstream target (*DBTNBT*, and Fig. 8n).

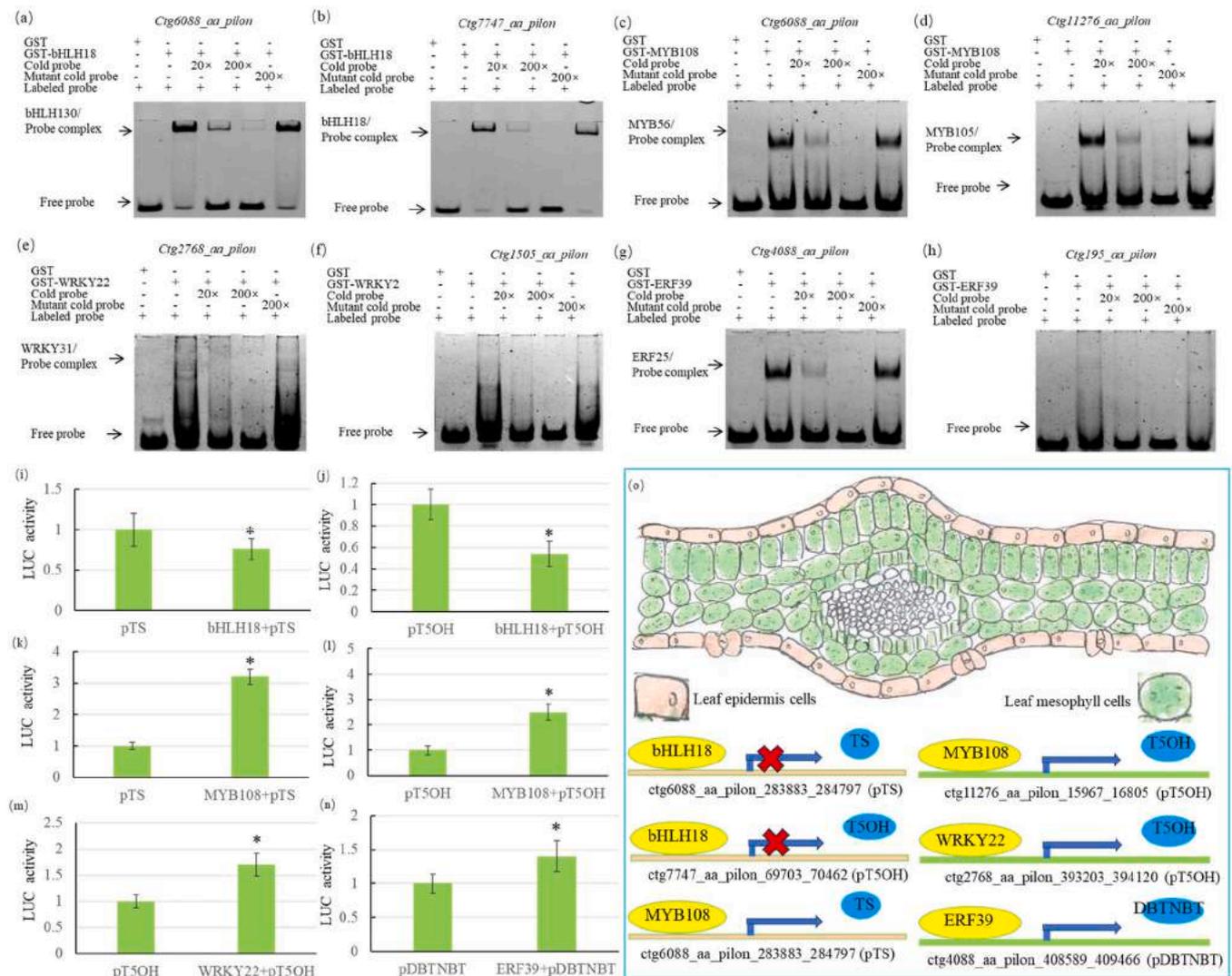
## 4. Discussion

Complex secondary metabolism in plants is programmed by both genetic and epigenetic information (Zhan et al., 2022). A large number of active ingredients, including flavonoids, taxoids, and polysaccharides, have been functionally identified in *T. mairei* leaves (Hao et al., 2017; Monacelli et al., 2002; Wang et al., 2019; Yang et al., 2016). Traditional studies on the regulation of secondary metabolism were carried out at

tissue and organ level. Our study focused on the cell-specific regulation of taxol biosynthesis in *T. mairei* leaves, providing a higher resolution.

Epigenetic marks include different types of regulatory events (Springer and Schmitz, 2017). Recent studies have focused on the effects of cytosine methylation on taxol biosynthesis. A higher level of DNA methylation has been observed in long-term cultured *Taxus* cells, suggesting that taxol biosynthesis can be inhibited by DNA methylation (Fu et al., 2012). In *Taxus × media*, the core promoter regions of *GGPPS*, *TS*, and *DBTNBT* are protected from cytosine methylation accumulation (Escrich et al., 2022). The methylation of *BAPT* promoter alters a potentially flux-limiting step of the taxane biosynthesis (Sanchez-Munoz et al., 2018). In our study, chromatin accessibility was evaluated using a newly applied scATAC-seq method. The ground-breaking application of 10 × Genomics technology in plant ATAC-seq promotes the analysis of heterogeneity in chromatin structure remodeling at the single-cell level (Marand et al., 2021b). To date, the single-cell regulatory landscapes of only a few plants, such as Arabidopsis, maize, and rice, have been reported. In Arabidopsis, 5283 root nuclei with 22,749 open chromatin peaks have been identified. In rice, 25,312 root nuclei with 14,602 unique Tn5 integration sites have been identified. In maize, 56,575 nuclei with an average of 31,660 unique high-quality sites have been identified (Dorrity et al., 2021; Feng et al., 2022; Marand et al., 2021a). In our study, 9488 cells with an average of 10,333 high-quality integration sites per cell were detected in *T. mairei* leaves. Although the number of captured cells varied in different experiments, the median number of Tn5 integration sites per cell was more than 10,000, suggesting sufficient sequencing depth to reveal epigenetic information for *T. mairei*.

In fully-extended leaves, the main tissue types are epidermis (including stomata and guard cells), mesophyll (including palisade and spongy mesophyll cells), and vascular system (Tsukaya, 2002). Because the specific distributions of the active ingredients, their biosynthesis would be expected to vary; however, leaf tissue-specific *cis*-regulatory information is currently unavailable. In model plants, a number of cell-type marker genes have been reported. Due to the lack of marker genes, cell-type annotations of scATAC experiments are difficult in woody plants. PCMDB provided a series of potential expression marker genes that can be selected to annotate cell clusters (Jin et al., 2022). All the



**Fig. 8.** Verification of the binding of TFs to their potential targets. (a-h) EMSA for the binding of cell type specific TFs to their targets. The GST only or TF-GST fusion protein was incubated with the probes containing the binding elements derived from the promoters. ‘-’ and ‘+’ represent absence and presence, respectively, and ‘20 ×’ or ‘200 ×’ show increasing amounts of probes for competition. (i-n) The dual luciferase assays in tobacco leaves showed that co-transformation of TFs activates both secondary metabolism-related promoters. (o) A model of cell specific regulation and transport of taxol. Green color indicated leaf mesophyll cells and yellow color indicated epidermis cells.

*T. mairei* leaf cells were classified into nine clusters, five of which could not be annotated by known marker genes. Post-transcriptional regulation, such as RNA methylation and RNA transport, might diminish the correlation between chromatin accessibility and expression level (Feng et al., 2022). Thus, further experimental validation should be done to establish a species-specific marker gene library.

The taxol biosynthesis pathway consists of a number of rate-limiting enzymes, most of which are encoded by multiple genes (Yu et al., 2023; Zhou et al., 2019). In our study, the differences in the chromatin accessibility of different enzyme-encoding genes were investigated. The TS catalyzes the cyclization of GGPP into the key intermediate taxadiene (Hao da et al., 2009). Among the three TS-encoding genes, ctg6088\_gene.1 showed a greater chromatin accessibility than ctg5306\_gene.4 and ctg7747\_gene.1. Taxane 5a-hydroxylase (T5OH) represents the initial oxygenation of the core taxadiene precursor (Jennewein et al., 2005). Among the T5OH-encoding genes, ctg7747\_gene.2 showed the greatest chromatin accessibility. Taxoid 13 $\alpha$ -hydroxylase (T13OH) favors taxadien-5 $\alpha$ -ol as a substrate to produce 10-deacetyl-2-debenzoylbaccatin III (Jennewein et al., 2001). In the four T13OH-encoding genes, ctg4364\_gene.1 showed a higher chromatin accessibility than the other genes. The elevation of the

chromatin accessibility near the limiting enzyme-encoding genes is a promising way to promote taxol biosynthesis.

Moreover, the *trans*-TFs control gene expression by binding to *cis*-regulatory elements embedded within OCRs (Lu et al., 2018); therefore, a comprehensive screening of *cis*-elements and *trans*-TFs across leaf mesophyll and epidermal cells will increase our understanding of taxol biosynthesis regulation at the spatial level. In *T. mairei*, a number of differentially enriched TF motifs between mesophyll and epidermal cells were identified. For example, TCP4, CRF2, and KAN4 were significantly enriched in the mesophyll cells, whereas CCA1, TCP4\_1, and TCP19 were significantly enriched in the epidermal cells. The distinct TF motifs found between mesophyll and epidermal cells may function in the differential accumulation of secondary metabolites. In *Taxus chinensis*, ERF12 and ERF15, two jasmonate responsive factors, function as repressor and activator of *TS* expression, respectively (Zhang et al., 2015). Interestingly, three identified ERF family motifs were significantly enriched in the epidermal cells, suggesting that their downstream target genes might be preferentially expressed in the epidermal cells. In *T. media*, MYB3 is a phloem-specific transcriptional regulator of taxol biosynthesis, and MYB39 is involved in the sexually dimorphic biosynthesis of taxol (Yu et al., 2022; Yu et al., 2020). Our data revealed that

four mesophyll cell-specific MYB motifs, MYB65, MYB3R1, MYB56, and MYB105 types, were significantly enriched in the mesophyll cells, suggesting that MYB family TFs also have cell-type specificity.

Taxol biosynthesis is a complex process that is controlled by a series of TFs (Chen et al., 2020; Yu et al., 2021). After searching the taxol biosynthesis-related OCRs, several classic *cis*-elements, including MBE, W-box, ABRE, and GCC-box, were found. To narrow the TF list, we propose that tissue-specific *cis*-regulation is important. Using this criterion, five MYBs, two bHLHs, three WRKYs, and four ERFs, were considered to be involved in the taxol biosynthesis regulatory network of *T. mairei* leaves. An EMSA assay indicated two targets of bHLH18, two targets of MYB108, one target of WRKY22, and one target of ERF39, improving our understanding of the taxol biosynthesis regulatory network. Our study identified several cell type-specific TFs and provided novel resources for genetic improvement of cultivated *Taxus* trees.

## 5. Conclusion

A scATAC-seq analysis was performed to identify differentially accessible regions of the *T. mairei* genome. In total, 9488 captured cells were divided into nine clusters, suggesting a high degree of cell heterogeneity. A differential accessibility analysis identified 9600 and 8538 leaf mesophyll and epidermal cell-specific OCRs, respectively. We predicted a close correlation between chromatin open peaks and expression levels of various taxol biosynthesis genes. Furthermore, several cell type-specific TFs were identified, suggesting their functions in regulation of taxol biosynthesis. Our data provide valuable resource for studying the basic principles of the cell type-specific regulation of taxol biosynthesis.

## CRedit authorship contribution statement

C.S. and H.W. conceived the original screening and research plans; C.S. and X.Z. supervised the experiments; T.Q. and H.Z. took care the trees; H.K. and X.L. performed the experiments for nuclei isolation; C.C. and Z.W. performed the experiments based on EMSA and Dual-luciferase methods; Q.W., X.W., X.L., W.L., R.M., and M.W. provided technical assistance to X.Z., T.Q., H.Z., and H.K. S.F. and H.Z. designed the experiments and analyzed the data; H.W. and C.S. conceived the project and wrote the article with contributions of all the authors; C.S. supervised and completed the writing. H.W. and C.S. agree to serve as the author responsible for contact and ensures communication.

## Declaration of Competing Interest

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

## Data Availability

Data will be made available on request.

## Acknowledgements

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). We are also grateful to LC Sciences company (Hangzhou, China) for metabolomic and transcriptomic analyses.

## Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.indcrop.2023.117411](https://doi.org/10.1016/j.indcrop.2023.117411).

## References

- Adey, A., Morrison, H.G., Asan, Xun, X., Kitzman, J.O., Turner, E.H., Stackhouse, B., MacKenzie, A.P., Caruccio, N.C., Zhang, X., Shendure, J., 2010. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biol.* 11, R119.
- Ajikumar, P.K., Xiao, W.H., Tyo, K.E., Wang, Y., Simeon, F., Leonard, E., Mucha, O., Phon, T.H., Pfeifer, B., Stephanopoulos, G., 2010. Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*. *Science* 330, 70–74.
- Armeev, G.A., Gribkova, A.K., Pospelova, I., Komarova, G.A., Shaytan, A.K., 2019. Linking chromatin composition and structural dynamics at the nucleosome level. *Curr. Opin. Struct. Biol.* 56, 46–55.
- Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., Greenleaf, W.J., 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* 10, 1213–1218.
- Burton, A., Torres-Padilla, M.E., 2014. Chromatin dynamics in the regulation of cell fate allocation during early embryogenesis. *Nat. Rev. Mol. Cell Biol.* 15, 723–734.
- Chen, Y., Zhang, M., Jin, X., Tao, H., Wang, Y., Peng, B., Fu, C., Yu, L., 2020. Transcriptional reprogramming strategies and miRNA-mediated regulation networks of *Taxus media* induced into callus cells from tissues. *BMC Genom.* 21, 168.
- Cheng, C., Guo, Z., Li, H., Mu, X., Wang, P., Zhang, S., Yang, T., Cai, H., Wang, Q., Lu, P., Zhang, J., 2022. Integrated metabolic, transcriptomic and chromatin accessibility analyses provide novel insights into the competition for anthocyanins and flavonols biosynthesis during fruit ripening in red apple. *Front. Plant Sci.* 13, 975356.
- Collemare, J., Seidl, M.F., 2019. Chromatin-dependent regulation of secondary metabolite biosynthesis in fungi: is the picture complete? *FEMS Microbiol. Rev.* 43, 591–607.
- Dong, Q.F., Liu, J.J., Yu, R.M., 2010. Taxol content comparison in different parts of *Taxus madia* and *Taxus chinensis* var. *mairei* by HPLC. *Zhong Yao Cai* 33, 1048–1051.
- Dorriy, M.W., Alexandre, C.M., Hamm, M.O., Vigil, A.L., Fields, S., Queitsch, C., Cuperus, J.T., 2021. The regulatory landscape of *Arabidopsis thaliana* roots at single-cell resolution. *Nat. Commun.* 12, 3334.
- Escrich, A., Cusido, R.M., Bonfill, M., Palazon, J., Sanchez-Munoz, R., Moyano, E., 2022. The epigenetic regulation in plant specialized metabolism: DNA methylation limits paclitaxel in vitro biotechnological production. *Front. Plant Sci.* 13, 899444.
- Feng, D., Liang, Z., Wang, Y., Yao, J., Yuan, Z., Hu, G., Qu, R., Xie, S., Li, D., Yang, L., Zhao, X., Ma, Y., Lohmann, J.U., Gu, X., 2022. Chromatin accessibility illuminates single-cell regulatory dynamics of rice root tips. *BMC Biol.* 20, 274.
- Feng, S., Kailin, H., Zhang, H., Chen, C., Huang, J., Wu, Q., Zhang, Z., Gao, Y., Wu, X., Wang, H., Shen, C., 2023. Investigation of the role of TmMYB16/123 and their targets (TmMTP1/11) in the tolerance of *Taxus media* to cadmium. *Tree Physiol.* 43, 1009–1022.
- Fu, C., Li, L., Wu, W., Li, M., Yu, X., Yu, L., 2012. Assessment of genetic and epigenetic variation during long-term *Taxus* cell culture. *Plant Cell Rep.* 31, 1321–1331.
- Han, N., Geng, W.J., Li, J., Liu, S.T., Zhang, J., Wen, Y.J., Xu, H.H., Li, M.Y., Li, Y.R., Han, P.P., 2022a. Transcription level differences in *Taxus wallichiana* var. *mairei* elicited by Ce3+, Ce4+ and methyl jasmonate. *Front. Plant Sci.* 13 (1040596).
- Han, Y., Lu, M., Yue, S., Li, K., Dong, M., Liu, L., Wang, H., Shang, F., 2022b. Comparative methylomics and chromatin accessibility analysis in *Osmanthus fragrans* uncovers regulation of genic transcription and mechanisms of key floral scent production. *Hortic. Res.* 96.
- Hao, J., Guo, H., Shi, X., Wang, Y., Wan, Q., Song, Y.B., Zhang, L., Dong, M., 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 Physiol.* 58, 1878–1890.
- Hao da, C., Yang, L., Huang, B., 2009. Molecular evolution of paclitaxel biosynthetic genes *TS* and *DBAT* of *Taxus* species. *Genetica* 135, 123–135.
- Horiguchi, T., Rithner, C.D., Croteau, R., Williams, R.M., 2008. Studies on Taxol(R) biosynthesis. Preparation and tritium labeling of biosynthetic intermediates by deoxygenation of a taxadiene tetra-acetate obtained from Japanese Yew. *J. Label. Comp. Radio.* 51, 325–328.
- Howat, S., Park, B., Oh, I.S., Jin, Y.W., Lee, E.K., Loake, G.J., 2014. Paclitaxel: biosynthesis, production and future prospects. *New Biotechnol.* 31, 242–245.
- Jenneweine, S., Rithner, C.D., Williams, R.M., Croteau, R.B., 2001. Taxol biosynthesis: taxane 13 alpha-hydroxylase is a cytochrome P450-dependent monooxygenase. *Proc. Natl. Acad. Sci. USA* 98, 13595–13600.
- Jenneweine, S., Park, H., DeJong, J.M., Long, R.M., Bollon, A.P., Croteau, R.B., 2005. Coexpression in yeast of *Taxus* cytochrome P450 reductase with cytochrome P450 oxygenases involved in Taxol biosynthesis. *Biotechnol. Bioeng.* 89, 588–598.
- Jin, J., Lu, P., Xu, Y., Tao, J., Li, Z., Wang, S., Yu, S., Wang, C., Xie, X., Gao, J., Chen, Q., Wang, L., Pu, W., Cao, P., 2022. PCMDB: a curated and comprehensive resource of plant cell markers. *Nucleic Acids Res.* 50, D1448–D1455.
- Kalve, S., De Vos, D., Beemster, G.T., 2014. Leaf development: a cellular perspective. *Front. Plant Sci.* 5, 362.
- Kuang, X., Sun, S., Wei, J., Li, Y., Sun, C., 2019. Iso-Seq analysis of the *Taxus cuspidata* transcriptome reveals the complexity of Taxol biosynthesis. *BMC Plant Biol.* 19, 210.
- Kulkarni, P., Dost, M., Bulut, O.D., Welle, A., Bocker, S., Boland, W., Svatos, A., 2018. Secondary ion mass spectrometry imaging and multivariate data analysis reveal co-aggregation patterns of *Populus trichocarpa* leaf surface compounds on a micrometer scale. *Plant J.* 93, 193–206.
- Li, S., Zhang, P., Zhang, M., Fu, C., Yu, L., 2013. Functional analysis of a WRKY transcription factor involved in transcriptional activation of the *DBAT* gene in *Taxus chinensis*. *Plant Biol.* 15, 19–26.
- Liang, Z., Xie, Z., Lam, S., Xu, X., 2014. Optimization of the fractional precipitation of paclitaxel from a *Taxus chinensis* cell culture using response surface methodology and

- its isolation by consecutive high-speed countercurrent chromatography. *J. Sep. Sci.* 37, 2322–2330.
- Long, R.M., Lagisetti, C., Coates, R.M., Croteau, R.B., 2008. Specificity of the *N*-benzoyl transferase responsible for the last step of Taxol biosynthesis. *Arch. Biochem. Biophys.* 477, 384–389.
- Lu, Z., Ricci, W.A., Schmitz, R.J., Zhang, X., 2018. Identification of *cis*-regulatory elements by chromatin structure. *Curr. Opin. Plant Biol.* 42, 90–94.
- Maher, K.A., Bajic, M., Kajala, K., Reynoso, M., Pauluzzi, G., West, D.A., Zumstein, K., Woodhouse, M., Bubb, K., Dorrity, M.W., Queitsch, C., Bailey-Serres, J., Sinha, N., Brady, S.M., Deal, R.B., 2018. Profiling of accessible chromatin regions across multiple Plant species and cell types reveals common gene regulatory principles and new Control modules. *Plant Cell* 30, 15–36.
- Marand, A.P., Chen, Z., Gallavotti, A., Schmitz, R.J., 2021a. A *cis*-regulatory atlas in maize at single-cell resolution. *Cell* 184 (3041–3055), e3021.
- Marand, A.P., Zhang, X., Nelson, J., Braga Dos Reis, P.A., Schmitz, R.J., 2021b. Profiling single-cell chromatin accessibility in plants. *STAR Protoc.* 2, 100737.
- Marupudi, N.L., Han, J.E., Li, K.W., Renard, V.M., Tyler, B.M., Brem, H., 2007. Paclitaxel: a review of adverse toxicities and novel delivery strategies. *Expert Opin. Drug Saf.* 6, 609–621.
- Monacelli, B., Pasqua, G., Botta, B., Vinciguerra, V., Gacs-Baitz, E., Monache, G.D., 2002. Abietane diterpenoids from callus cultures of *Taxus baccata*. *Planta Med.* 68, 764–766.
- Sanchez-Munoz, R., Bonfill, M., Cusido, R.M., Palazon, J., Moyano, E., 2018. Advances in the regulation of *in vitro* paclitaxel production: methylation of a Y-patch promoter region alters *BAPT* gene expression in *Taxus* cell cultures. *Plant Cell Physiol.* 59, 2255–2267.
- Sijacic, P., Bajic, M., McKinney, E.C., Meagher, R.B., Deal, R.B., 2018. Changes in chromatin accessibility between *Arabidopsis* stem cells and mesophyll cells illuminate cell type-specific transcription factor networks. *Plant J.: Cell Mol. Biol.* 94, 215–231.
- Springer, N.M., Schmitz, R.J., 2017. Exploiting induced and natural epigenetic variation for crop improvement. *Nat. Rev. Genet.* 18, 563–575.
- Tsukaya, H., 2002. Leaf development. *Arab. Book 1*, e0072.
- Wang, T., Zhang, F., Zhuang, W., Shu, X., Wang, Z., 2019. Metabolic variations of flavonoids in leaves of *T. media* and *T. mairei* obtained by UPLC-ESI-MS/MS. *Molecules* 24, 3323.
- Wu, L., Qi, K., Liu, C., Hu, Y., Xu, M., Pan, Y., 2022a. Enhanced coverage and sensitivity of imprint DESI Mass Spectrometry Imaging for plant leaf metabolites by post-photoionization. *Anal. Chem.* 94, 15108–15116.
- Wu, Z.H., Wang, R.Z., Sun, Z.L., Su, Y., Xiao, L.T., 2022b. A mass spectrometry imaging approach on spatiotemporal distribution of multiple alkaloids in *Gelsemium elegans*. *Front. Plant Sci.* 13, 1051756.
- Xiong, X., Gou, J., Liao, Q., Li, Y., Zhou, Q., Bi, G., Li, C., Du, R., Wang, X., Sun, T., Guo, L., Liang, H., Lu, P., Wu, Y., Zhang, Z., Ro, D.K., Shang, Y., Huang, S., Yan, J., 2021. The *Taxus* genome provides insights into paclitaxel biosynthesis. *Nat. Plants* 7, 1026–1036.
- Yang, L., Zheng, Z.S., Cheng, F., Ruan, X., Jiang, D.A., Pan, C.D., Wang, Q., 2016. Seasonal dynamics of metabolites in needles of *Taxus wallichiana* var. *mairei*. *Molecules* 21, 1403.
- Yu, C., Luo, X., Zhang, C., Xu, X., Huang, J., Chen, Y., Feng, S., Zhan, X., Zhang, L., Yuan, H., Zheng, B., Wang, H., Shen, C., 2020. Tissue-specific study across the stem of *Taxus media* identifies a phloem-specific TmMYB3 involved in the transcriptional regulation of paclitaxel biosynthesis. *Plant J.* 103, 95–110.
- Yu, C., Zhang, C., Xu, X., Huang, J., Chen, Y., Luo, X., Wang, H., 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 Biol.* 21, 104.
- Yu, C., Huang, J., Wu, Q., Zhang, C., Li, X.L., Xu, X., Feng, S., Zhan, X., Chen, Z., Wang, H., Shen, C., 2022. Role of female-predominant MYB39-bHLH13 complex in sexually dimorphic accumulation of taxol in *Taxus media*. *Hortic. Res.* 62..
- Yu, C., Hou, K., Zhang, H., Liang, X., Chen, C., Wang, Z., Wu, Q., Chen, G., He, J., Bai, E., Li, X., Du, T., Wang, Y., Wang, M., Feng, S., Wang, H., Shen, C., 2023. Integrated mass spectrometry imaging and single-cell transcriptome atlas strategies provide novel insights into taxoid biosynthesis and transport in *Taxus mairei* stems. *Plant J.* <https://doi.org/10.1111/tj.16315>.
- Zhan, X., Chen, Z., Chen, R., Shen, C., 2022. Environmental and genetic factors involved in plant protection-associated secondary metabolite biosynthesis pathways. *Front. Plant Sci.* 13, 877304.
- 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., 2023. Mass spectrometry imaging and single-cell transcriptional profiling reveal the tissue-specific regulation of bioactive ingredient biosynthesis in *Taxus* leaves. *Plant Commun.*, 100630
- Zhang, M., Li, S., Nie, L., Chen, Q., Xu, X., Yu, L., 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.
- Zhang, M., Jin, X., Chen, Y., Wei, M., Liao, W., Zhao, S., Fu, C., 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.
- Zhang, X., Pandey, M.K., Wang, J., Zhao, K., Ma, X., Li, Z., Zhao, K., Gong, F., Guo, B., Varshney, R.K., Yin, D., 2021. Chromatin spatial organization of wild type and mutant peanuts reveals high-resolution genomic architecture and interaction alterations. *Genome Biol.* 22, 315.
- Zheng, H., Fu, X., Shao, J., Tang, Y., Yu, M., Li, L., Huang, L., Tang, K., 2023. Transcriptional regulatory network of high-value active ingredients in medicinal plants. *Trends Plant Sci.* 28, 429–446.
- Zhong, J., Luo, K., Winter, P.S., Crawford, G.E., Iversen, E.S., Hartemink, A.J., 2016. Mapping nucleosome positions using DNase-seq. *Genome Res.* 26, 351–364.
- Zhou, G., Yin, H., Chen, F., Wang, Y., Gao, Q., Yang, F., He, C., Zhang, L., Wan, Y., 2022. The genome of *Areca catechu* provides insights into sex determination of monoecious plants. *New Phytol.* 236, 2327–2343.
- Zhou, T., Luo, X., Yu, C., Zhang, C., Zhang, L., Song, Y.B., Dong, M., 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 Biol.* 19, 33.
- Zhu, L., Chen, L., 2019. Progress in research on paclitaxel and tumor immunotherapy. *Cell. Mol. Biol. Lett.* 24, 40.