



# Single-cell transcriptome analyses reveal cellular and molecular responses to low nitrogen in burley tobacco leaves

Yuqing Feng<sup>1</sup> | Yuanyuan Zhao<sup>1</sup> | Yanjun Ma<sup>2</sup> | Deshui Liu<sup>2</sup> | Hongzhi Shi<sup>1</sup>

<sup>1</sup>College of Tobacco, Henan Agricultural University, Zhengzhou, Henan, P.R.China

<sup>2</sup>Technology Center, Shanghai Tobacco Group Beijing Cigarette Factory Co., Ltd, Beijing, China

**Correspondence**  
 Hongzhi Shi,  
 Email: [shihongzhi88@163.com](mailto:shihongzhi88@163.com)

**Funding information**  
 the Shanghai Tobacco Corporation Research Program [grant number TP2019-C4]

**Edited by** R.M. Rivero

## Abstract

Tobacco (*Nicotiana tabacum*) is cultivated and consumed worldwide. It requires great amounts of nitrogen (N) to achieve the best yield and quality. With a view to sustainable and environmentally friendly agriculture, developing new genotypes with high productivity under low N conditions is an important approach. It is unclear how genes in tobacco are expressed at the cellular level and the precise mechanisms by which cells respond to environmental stress, especially in the case of low N. Here, we characterized the transcriptomes in tobacco leaves grown in normal and low-N conditions by performing scRNA-seq. We identified 10 cell types with 17 transcriptionally distinct cell clusters with the assistance of marker genes and constructed the first single-cell atlas of tobacco leaves. Distinct gene expression patterns of cell clusters were observed under low-N conditions, and the mesophyll cells were the most important responsive cell type and displayed heterogeneous responses among its three subtypes. Pseudo-time trajectory analysis revealed low-N stress decelerates the differentiation towards mesophyll cells. In combination with scRNA-seq, WGCNA, and bulk RNA-seq results, we found that genes involved in porphyrin metabolism, nitrogen metabolism, carbon fixation, photosynthesis, and photosynthesis-antenna pathway play an essential role in response to low N. Moreover, we identified COL16, GATA24, MYB73, and GLK1 as key TFs in the regulation of N-responsive genes. Collectively, our findings are the first observation of the cellular and molecular responses of tobacco leaves under low N stress and lay the cornerstone for future tobacco scRNA-seq investigations.

## 1 | INTRODUCTION

Tobacco (*Nicotiana tabacum*) is an economically important crop cultivated worldwide. It usually requires great amounts of nitrogen (N) to achieve the best yield and quality and is usually over-fertilized (Kaiser et al., 2015; Sifola et al., 2018). Excessive use of nitrogen fertilizers adversely impacts nitrogen use efficiency (NUE), which not only leads to yield loss but surpasses the tolerable limits of environmental sustainability (Sutton et al., 2021; Wang et al., 2022). Kaiser et al. 2015 found that nitrate both in soil solution from tobacco fields or in underground water below tobacco crop was greater than that in natural conditions. Improving NUE will not only avert the environmental hazards

caused by the application of excessive N but also critically reduce the estimated billions of dollars in annual expense on N-based fertilizers worldwide. In addition, as tobacco is a leaf-harvested crop, the leaf traits are directly responsible for yield and marketability. Thus, leaf traits are often the primary targets for breeders. However, lower NUE of tobacco leaves will exert adverse effects on the leaf quality, such as producing higher levels of nitrate and tobacco-specific nitrosamines (TSNAs). Considering the negative effect that an excess N usually determines on cured product quality, developing new genotypes with high NUE and high productivity under low-N conditions becomes indispensable (Bailey, 2014; Drake et al., 2015; Sifola, 2003). Therefore, understanding how plants respond to low N conditions and the mechanism of

low-N tolerance is very important for the sustainable development of modern crop production.

Plants quickly perceive and respond to low-N stress via a large number of physiological and metabolic events, such as the degradation of proteins, decrease of enzyme activities related to nitrogen metabolism and photosynthesis, accumulation in carbohydrates, and so on (Liu et al., 2020a). Among those events, the reduction of photosynthetic capacity is one of the major low-N-induced damages that inhibit plant growth and development (Boussadia et al., 2010). While many studies have provided great insights into the mechanism of leaf response to low-N conditions, they have failed to capture the heterogeneity of individual cells (Huang et al., 2022; Kiba et al., 2018; Sun et al., 2021; Shi et al., 2022). Elucidating these behaviors may expand our understanding of leaf functionality.

In recent years, single-cell RNA sequencing (scRNA-Seq) techniques have emerged for analyzing transcriptomes expressed in single cells. scRNA-seq has been mostly applied to animal tissues, albeit a fast-growing number of applications to plants have been implemented since 2019. Its high-resolution molecular information contributes to elucidating cellular heterogeneity and molecular processes specific to cell types in complex tissues, such as flowers (Kang et al., 2022), leaves (Sun et al., 2022), shoot apical meristems (Zhang et al., 2021), roots (Dorrity et al., 2021). Many single-cell transcriptomic atlases have been established successfully for various crops, including *Arabidopsis thaliana* (Kim et al., 2021), allotetraploid peanut (*Arachis hypogaea* L.; Liu et al., 2021), Chinese cabbage (Sun et al., 2022), rice (Wang et al., 2021) and maize (Marand et al., 2021). These atlases lay the foundation for the functional characterization and manipulation of candidate genes that are needed to develop elite cultivars. In addition to revealing the gene expression atlas, single-cell transcriptomics is also used to investigate plant responses to biotic and abiotic stresses (Bai et al., 2022; Liang et al., 2023; Sun et al., 2022). This technology has provided a previously unattainable view of gene expression dynamics in plant tissues and the alteration of these dynamics under various physiological conditions (Rich-Griffin et al., 2020). However, the application of scRNA-seq remains limited to only a few plant species, presumably due to the technical obstacles caused by the difficulty in protoplast isolation and the limited prior knowledge of cell identity and developmental biology.

Burley tobacco, a chloroplast and N-deficient phenotype, is low in photosynthesis and accumulates high levels of nitrate in its leaves (Liu et al., 2016). Besides their photosynthetic role, leaves serve as an interface between plants and the environment and respond to environmental stimulation. Therefore, understanding the complexity of leaves at a single-cell level will be an active area of interest in plant biology. In this study, we constructed the first single-cell transcriptomic atlas for the burley tobacco leaf and identified cluster-specific marker genes for each type of leaf cell. Implementing scRNA-seq to burley tobacco leaf, we further explored the transcriptional and developmental activities in response to low-N stress and identified major response cell clusters. Based on the expression patterns for cell lineage responses to low-N conditions, we characterized pathways and genes associated with low-N resistance. We present here the first scRNA-seq expression atlas of the tobacco leaf at the single-cell level

and transcriptional responses to low-N conditions. Our findings will enable improvements in tobacco cultivation and help to shape future research toward the genetic improvement of tobacco for different cell types in plant response to stress.

## 2 | MATERIALS AND METHODS

### 2.1 | Preparation of leaf Samples for scRNA-seq

The burley tobacco (*Nicotiana tabacum*) cultivar TN90 was used. The seeds were sterilized twice with 2% (v/v) sodium hypochlorite for 5 min and sown on a sponge at  $25 \pm 2^\circ\text{C}$  in long-day conditions (14 h light/10 h dark) with a light intensity of  $400 \mu\text{mol m}^{-2} \text{s}^{-1}$  in a greenhouse located at the National Tobacco Cultivation & Physiology & Biochemistry Research Center of Henan Agricultural University (China). Seven days after sowing, seedlings were transferred to a hydroponic solution containing 1.25 mM NH<sub>4</sub>NO<sub>3</sub> for one week, followed by 2.5 mM NH<sub>4</sub>NO<sub>3</sub> for two weeks. Then half of the seedlings were transferred to a low nitrogen solution (0.25 mM NH<sub>4</sub>NO<sub>3</sub>; TL0) for five days, while the other half remained in a high nitrogen condition (2.5 mM NH<sub>4</sub>NO<sub>3</sub>; TL5) for five days. Leaves (the third leaf from the top) were cut into pieces and frozen immediately in liquid nitrogen for scRNA-seq analyses, bulk RNA-seq, and physiological index measurement.

### 2.2 | Tobacco leaf nuclei isolation and scRNA-seq library construction

The leaf nuclei isolation, library synthesis, and RNA sequencing were completed by Gene Denovo (Guangzhou, China). Nuclei were isolated following the method described in (Long et al. 2021). In brief, leaves were dissected and transferred immediately into a 1.5-ml RNase-free Eppendorf tube kept in liquid nitrogen and ground into fine powder by a 1000-μl pipette tip in the tube. The powder was placed on a glass plate with 300 μL ice-cold Extraction Buffer (EB)—0.4 M sucrose, 10 mM Tris-HCl pH 8.0, 10 mM MgCl<sub>2</sub>, 0.2% (w/v) Triton X-100, 1 mM di-thiothreitol (DTT), 1× protease inhibitor (Roche), 0.4 U/μl RNase inhibitor (RNase-OUT, Thermo Fisher Scientific). Then, extracted nuclei were sorted using 4,6-diamidino-2-phenylindole (DAPI) and loaded into a flow cytometer with a 70-μm nozzle for further analysis. The cells of each group were mixed into one sample and adjusted to 1000 cells/μL. The indexed sequencing libraries were prepared using Chromium Single Cell 3' Reagent Kits (v3) according to the manufacturer's instructions. The barcoded sequencing libraries were checked by Agilent Bioanalyzer 2100. Finally, library sequencing was performed by Illumina NovaSeq 6000.

### 2.3 | Pre-processing of Raw scRNA-Seq Data

Sequencing reads of the two scRNA-seq samples (TL0 and TL5) were mapped to the TN90 reference genome (<https://www.ncbi.nlm.nih.gov>.

[gov/genome/?term=GCF\\_000715135.1\\_Ntab-TN90](https://www.ncbi.nlm.nih.gov/genome/?term=GCF_000715135.1_Ntab-TN90)) by cellranger (version 2.0.0); (Hao et al., 2020). Cells with the criteria of gene counts between 390 and 13000 per cell, unique molecular identifier (UMI) counts less than 47 000 per cell, and percentage of mitochondrial genes less than 10% were filtered using Seurat (v4.0.4) R package (Hao et al., 2020).

## 2.4 | Clustering, visualization, and marker gene selection

Harmony was used to normalize and filter the gene-barcode matrix to reduce feature dimensions (Korsunsky et al., 2019). And t-Distributed Stochastic Neighbor Embedding (tSNE) (Van der Maaten & Hinton, 2008) and uniform manifold approximation and projection (UMAP) were performed for cell cluster visualization. During these analyses, 0.5 resolution was used for two groups of sample combinations. The cell type of each cluster was defined by known marker genes. Based on the marker gene results, heatmap visualization was achieved using Seurat (v4.0.4) R package (Hao et al., 2020). Then, the number of sorted cells from leaf samples was counted.

After cell type annotations, the novel marker genes for each cell type were identified according to a  $\log_2$  (fold change [FC])  $> 1$  and  $P \leq 0.01$ , and the genes needed to be expressed in 25% of cells of the specific type. In addition, we further selected the top 5 genes according to the result of differentially expressed genes. Then, the expression distribution of each marker gene and percentages of cells expressing the marker were displayed by color and size in dot plots or violin plots, heat maps, and bubble diagrams.

## 2.5 | Identification of differentially expressed genes (DEGs) after low-N stress and functional enrichment analysis

DEGs for each cell type were detected using parameters of “ $\log_2$  [FC]  $> 0.36$ ”, “min.pct = 0.10” and “ $P \leq 0.05$ ” in the Mann–Whitney U test between a nitrogen-treated sample and a control sample. DEGs identified in this study were subjected to gene ontology (GO) and kyoto encyclopedia of genes and genomes (KEGG) enrichment analyses (Ashburner et al., 2000; Kanehisa & Goto, 2000). The top 20 enrichment pathways with high significance were annotated and used to construct an enrichment heatmap for the different cell types. In addition, the UpSet plot and heatmap of DEGs were performed using the OmicShare tools, a free online platform for data analysis (<https://www.omicshare.com/tools>).

## 2.6 | Pseudotime trajectory from primordium cells to mesophyll cells analysis

The Monocle2 (Version2.10.1); (Trapnell et al., 2014) were used to construct single-cell trajectory for selected clusters. We identified key

genes related to the development and differentiation process with false discovery rate (FDR)  $<1e-5$ , and grouped genes with similar trends in expression, reasoning that such groups might share common biological functions and regulators. We also performed a KEGG enrichment analysis to reveal the regulatory mechanism. Heatmap of transcription factors (TFs) analysis was performed using the OmicShare tools, a free online platform for data analysis (<https://www.omicshare.com/tools>).

## 2.7 | Weighted gene co-expression network analysis (WGCNA)

To further identify the module associated with the key cell clusters, gene co-expression networks were constructed using WGCNA (v1.47) package in (R Langfelder & Horvath, 2008). The parameters: “power = 6, minModuleSize = 50, mergeCutHeight = 0.3” were used to identify modules. GO (Ashburner et al., 2000) and KEGG (Kanehisa & Goto, 2000) pathway enrichment analyses were conducted to analyze the biological functions of selected modules. Finally, the network file was visualized using Cytoscape (v3.7.1); (Shannon et al., 2003) software to present the gene biological interaction.

## 2.8 | Sequencing and analysis of bulk RNA-seq

Total RNA was isolated via a Trizol reagent kit (Invitrogen) according to the manufacturer's protocol. The resulting cDNA library was sequenced using Illumina Novaseq6000 by Gene Denovo Biotechnology Co. (Guangzhou, China). Quality control of the dataset was performed using fastp (version 0.18.0); (Chen et al., 2018). Next, the sequencing clean reads were mapped to the TN90 tobacco reference genome ([https://www.ncbi.nlm.nih.gov/genome/?term=GCF\\_000715135.1\\_Ntab-TN90](https://www.ncbi.nlm.nih.gov/genome/?term=GCF_000715135.1_Ntab-TN90)) using HISAT2. 2.4 (Kim et al., 2015).

The gene expression level of samples was calculated using fragments per kilobase of transcript per million mapped reads (FPKM) and RNA differential expression analysis was performed by DESeq2 software (Love et al., 2014). A statistical test was performed using the negative binomial Wald test followed by a Benjamini-Hochberg correction to obtain the FDR Benjamini & (Hochberg, 1995; Love et al., 2014). DEGs were defined using the following criteria:  $|\log_2(\text{FC})| \geq 1$  and  $\text{FDR} < 0.05$ . Venn diagrams and heatmaps were generated using OmicShare tools, a free online platform for data analysis (<https://www.omicshare.com/tools>).

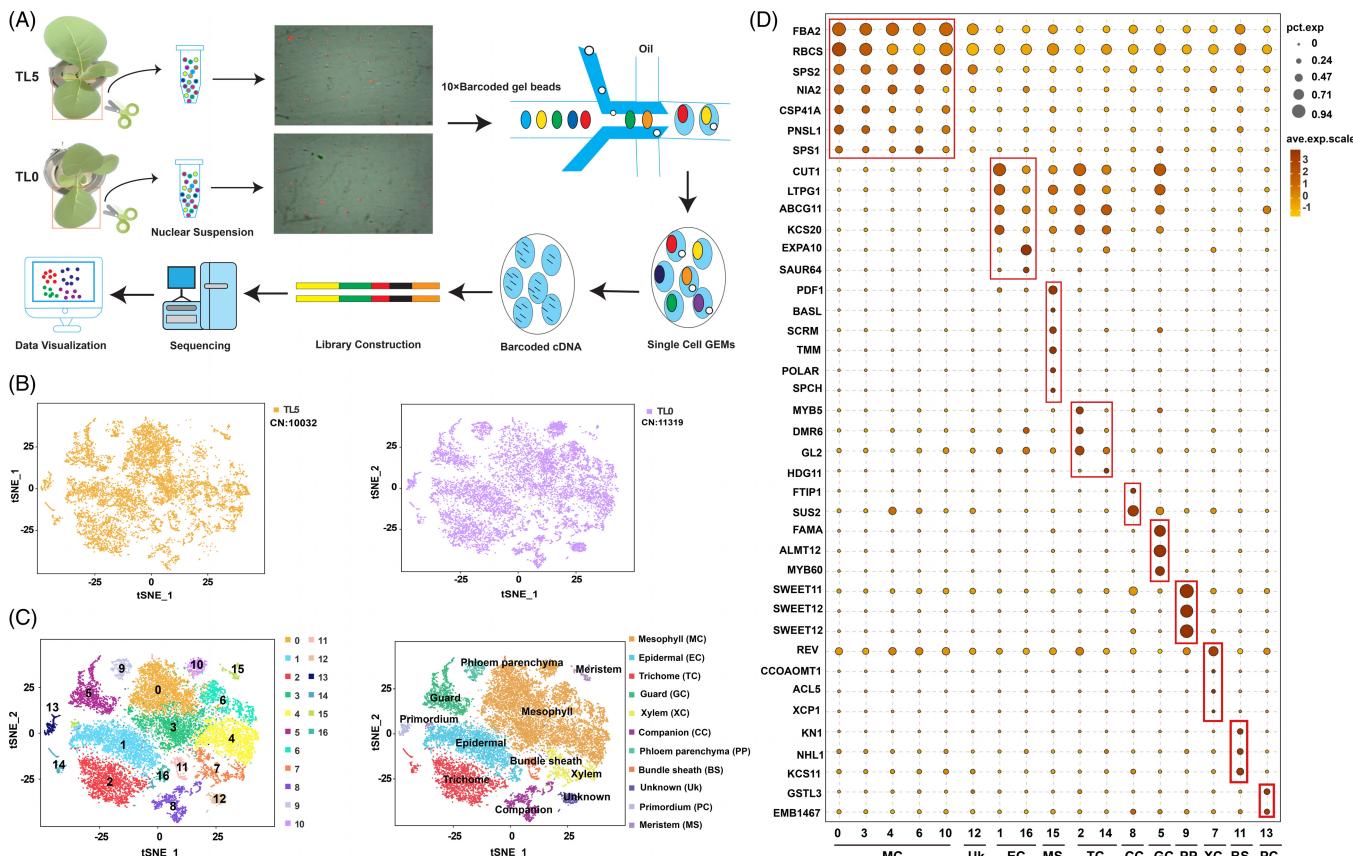
## 2.9 | Real-time PCR

Reverse transcription PCR was performed using reverse transcriptase (Vazyme). qRT-PCR was performed with SYBR qPCR Master Mix (Vazyme). Reactions were performed in an ABI StepOnePlus Real-Time

PCR System (Massachusetts). The PCR mixtures were first incubated in a 384-well optical plate (Roche) at 95°C for 90 s, followed by 40 cycles at 95°C for 5 s, 60°C for 15 s, and 72°C for 20 s. The expression levels of mRNAs were normalized and calculated using the  $2^{-\Delta\Delta Ct}$  method (Livak & Schmittgen, 2001). The primers are given in Table S1. And gene expression was normalized using the EIF5A1 gene.

## 2.10 | Physiological index measurements

The pigment content was determined using the methods described by (Zou 2000). Sucrose synthetase activity (SSA), sucrose phosphate synthase (SPS), rubisco enzyme activity, nitrate reductase activity (NRA), glutamine synthetase activity (GSA), and glutamate synthase (GOGAT) activity were determined using micro determination kits (Suzhou Comin Biotechnology Co., Ltd). The NO<sub>3</sub>-N content was determined using the method described by (Cataldo et al. 1975). A student t-test between samples was performed using the OmicShare tools, a free online platform for data analysis (<https://www.omicshare.com/tools>). All data correspond to the mean of three biological replicates (n = 3). And GraphPad Prism (v. 8.0.1, GraphPad Software Inc.) was used to obtain the figures.



**FIGURE 1** Single-cell transcriptome atlas of the tobacco seedling leaf. (A) Brief workflow of single-cell library preparation. (B) t-SNE plot showing the distribution and numbers of cells for leaf samples. (C) t-SNE visualization for the identification of 17 cell clusters in leaves. (D) Expression pattern of representative cell-specific marker genes in 17 cell clusters. The average expression level (color) and the proportion of cells expressing the gene (dot size) are shown.

## 3 | RESULTS

### 3.1 | Single-cell RNA sequencing of tobacco leaf cells and leaf cell atlas generation

To systematically determine gene expression patterns during tobacco leaf development, we isolated nuclear suspension from burley tobacco developing leaves cultivated at 5 mM and 0.5 mM N levels and profiled them using droplet-based scRNA-seq to generate a single-cell transcriptomic atlas (Figure 1A). A total of 10,032 individual leaf cells in the TL5 sample and 11,319 individual leaf cells in the TLO sample were labeled (Figure 1B). A total of 1.77 billion clean reads were obtained based on the tobacco transcriptomes of 28,343 cells, achieving an average read of 32,592 per cell. The ratio of high-quality reads to quality scores at Q30 was more than 93.00%. The TLO sample contained 393,656,011 reads, 84.50% of which were mapped to the reference genome. The median number of UMIs per cell was 4985, and an average of 3362 genes were expressed per cell. The TL5 sample contained 377,928,797 reads, 81.20% of which were mapped to the reference genome. The median number of UMIs per cell was 4159, and an average of 2803 genes were expressed per cell (Table S2).



To generate a cell atlas of tobacco leaf development, we merged two leaf samples (TL5 and TL0) for cell clustering and annotation. In total, 28,343 single-cell transcriptomes were used to identify distinct cell populations, and they were grouped into 17 distinct clusters (Figure 1C). We assigned cells to ten types according to the cluster-specific expression of reported marker genes – primordium, meristem, epidermal, mesophyll, trichome, guard, xylem, companion, phloem parenchyma, and bundle sheath cells (Figure 1D; Table S3). The code genes for NADH DEHYDROGENASE [UBIQUINONE] IRON SULFUR protein 1 (EMB1467; Satterlee et al., 2020) and GLUTATHIONE S-TRANSFERANSE L3 (GSTL3; Liu et al., 2021) are highly expressed and specific to primordium cells. In mesophyll cells, we identified the code genes for FRUCTOSE-BISPHOSPHATE 1 (FBA2), PHOTOSYNTHETIC NDH SUBUNIT OF LUMENAL LOCATION 1 (PNL1), SOLANESYL-DIPHOSPHATE SYNTHASE 3 (SPS1), RIBULOSE BISPHOSPHATE CARBOXYLASE SMALL CHAIN S41 (RBCS), SOLANESYL-DIPHOSPHATE 3 (SPS2), NITRATE REDUCTASE [NADH] 2 (NIA2), and CHLOROPLAST STEM-LOOP BINDING protein of 41 kDa a (CSP41A) as marker genes (Bai et al., 2022; Guo et al., 2022a; Kim et al., 2021; Liu et al., 2021; Wang et al., 2021). The code genes for BIDIRECTIONAL SUGAR TRANSPORTER SWEET12 (SWEET12) and BIDIRECTIONAL SUGAR TRANSPORTER N3 (SWEET11) served as markers for phloem parenchyma cells (Kim et al., 2021). We identified the code genes for EXPANSIN-A10 (EXPA10), 3-KETOACYL-COA SYNTHASE 5 (CUT1), 3-KETOACYL-COA SYNTHASE 20 (KCS20), NON-SPECIFIC LIPID TRANSFER protein GPI-ANCHORD 1 (LTPG1), AUXIN-RESPONSIVE protein SAUR68 (SAUR64), and ABC transporter G family member 11 (ABCG11) as markers for epidermal cells (Endo et al., 2014; Kim et al., 2021; Liu et al., 2022; Liang et al., 2023; Tenorio et al., 2022). The cell cluster enriched in genes encoding for SUCROSE TRANSPORT protein (SUS2) and QUIRKY protein (FTP1) was identified as the companion cells (Kim et al., 2021; Zhang et al., 2021). And the cluster enriched in the genes encoding for MYB-related protein 306 (MYB60), ALUMINUM-ACTIVATED MALATE TRANSPORTER 12 (ALMT12), and transcription factor FAMA (FAMA) was identified as the guard (Xu et al., 2021; Zhang et al., 2021). HOMEOTIC PROTEIN KNOTTED-1 (KN1), 3-KETOACYL-COA SYNTHASE 11 (KCS11), and NDR1/HIN1-Like protein 3 (NHL1) are known marker genes for bundle sheath cells (Bai et al., 2022; Liu et al., 2021; Jackson et al., 1994). The genes encoding for DOWNTIL MILDEW RESISTANCE protein 6 (DMR6), HOMEOBOX-LEUCINE ZIPPER protein GLABRA 2 (GL2), MYB-RELATED protein 308 (MYB5), and HOMEOBOX-LEUCINE ZIPPER protein HDG11 (HDG11) served as markers for trichome cells (Wei et al., 2019; Wang et al., 2021). The genes encoding for HOMEOBOX-LEUCINE ZIPPER protein (REV), XYLEM CYSTEINE PROTEINASE 1 (XCP1), CAFFEOYL-COA O-METHYLTRANSFERASE 1 (CCOAMT1), and THERMOSPERMINE SYNTHASE ACAULIS 5 (ACL5) are highly expressed in xylem and were used as xylem cell marker genes (Bai et al., 2022; Liu et al., 2022; Liang et al., 2023). The genes encoding for PROTODERMAL FACTOR 1 (PDF1), BREAKING OF ASYMMETRY IN THE STOMATAL LINEAGE (BASL), TOO MANY MOUTHS protein

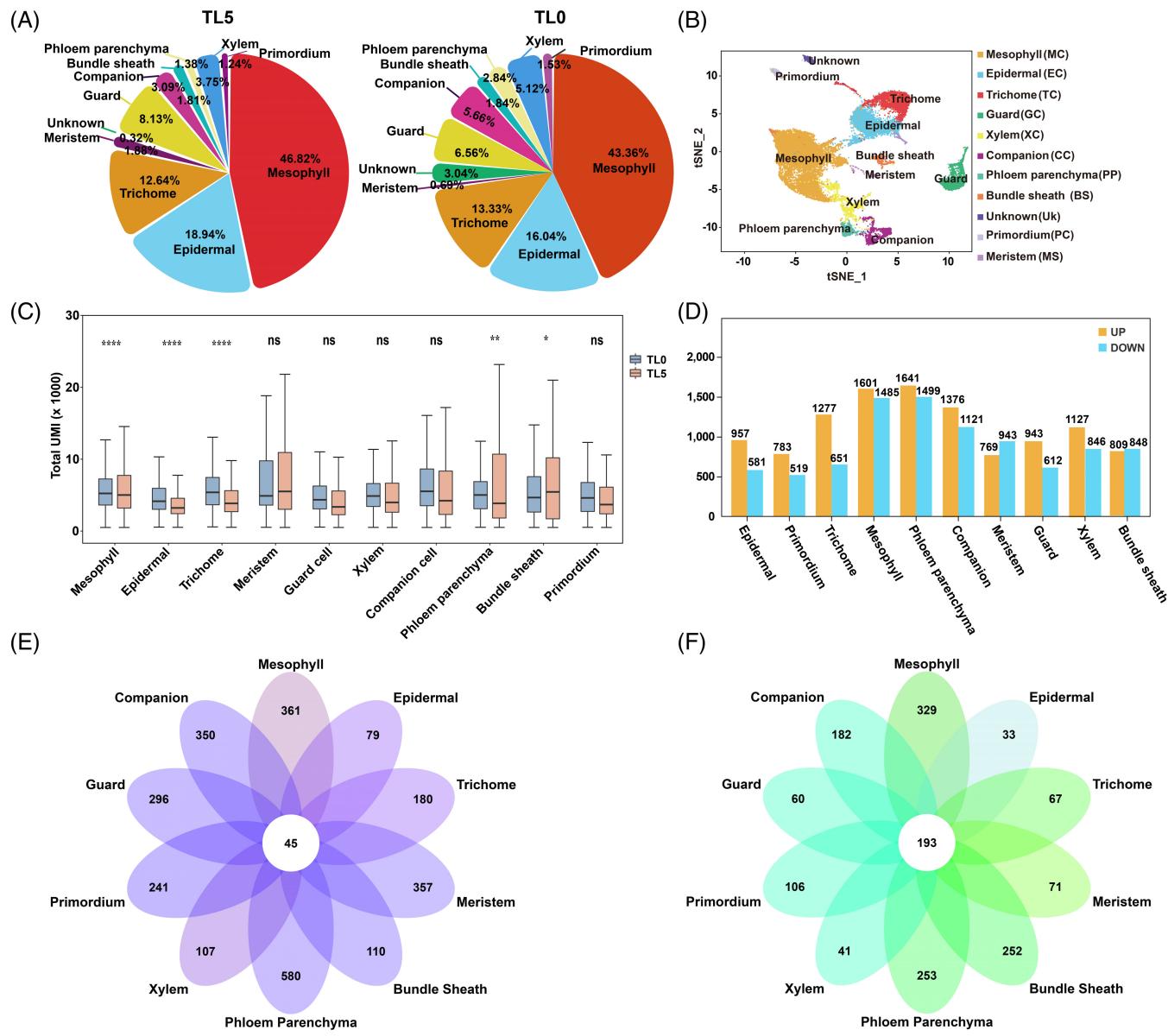
(TMM), transcription factor ICE1 (SCRM), transcription factor SPEECHLESS (SPCH), and POLAR LOCALIZATION DURING ASYMMETRIC DIVISION AND REDISTRIBUTION protein (POLAR) were used as representative meristem marker genes (Liang et al., 2023; Pillitteri et al., 2011; Yang & Sack, 1995; Zhang et al., 2021) (Figure 1D; Table S3).

### 3.2 | Identification of new cell-type marker genes

To explore the potential marker genes for different cell types, we analyzed gene expression profiles in the mesophyll, epidermal, trichome, meristem, primordium, guard, companion, bundle sheath, xylem, and phloem parenchyma. We confirmed the genes with high (average expression value in the target cluster > that of the others) and cell-type-specific expression (genes must be expressed in 25% of cells within the target cell type and < 25% of cells in all the other cell types;  $p \text{ value} \leq 0.01$ ;  $\log_{2}\text{FC} \geq 1$ ). In total, 700 genes in the mesophyll, 452 genes in epidermal, 561 genes in the trichome, 701 genes in the meristem, 522 genes in primordium, 699 genes in guard, 648 genes in companion, 326 genes in bundle sheath, 246 genes in xylem as well as 676 genes in phloem parenchyma were identified (Figure S1). Some genes with known gene functions were included (Figure S1; Table S4): genes encoding for CATION/H (+) ANTIPORTER gene (CHX18) and GUARD CELL S-TYPE ANION CHANNEL gene (SLAC1) in guard cells; gene encoding for WRKY transcription factor (WRKY9) in primordium cells; gene encoding for BORON transporter (BOR1) in xylem cells; genes encoding for HIGH MOBILITY GROUP B protein (HMB9) and ABC transporter C family member (ABCC2) in mesophyll cells; gene encoding for ACYL-COENZYME A OXIDASE gene (ACX3) in epidermal cells; gene encoding for BIDIRECTIONAL SUGAR TRANSPORTER (SWEET11) in phloem parenchyma cells; gene encoding for DIACYLGLYCEROL KINASE (DGK5) in bundle sheath cells; gene encoding for TUBULIN ALPHA CHAIN (TUBA) in meristem cells; and gene encoding for PHENYLALANINE AMMONIA-LYASE (TPA1) in trichome cells. Moreover, many genes whose functions were unknown were identified from the transcriptomic datasets. In addition, the top three marker genes with the highest expression in each cell type were selected for display in the t-SNE plots to show the cell type specificity (Figure S1).

### 3.3 | Single-cell transcriptomic changes vary among cell types' responses to low N stress

Not unexpectedly, the seedling samples of control and low-N contained all ten cell types (Figure 2B). The analysis of cell proportions between the control (TL5) and treatment (TL0) showed that the tobacco leaves responded markedly to the low-N stress (Figure 2A; Table S5). In particular, mesophyll cells have greater cell percentages than the other cell types (Figure 2A). There was a significant decrease in the proportion of mesophyll

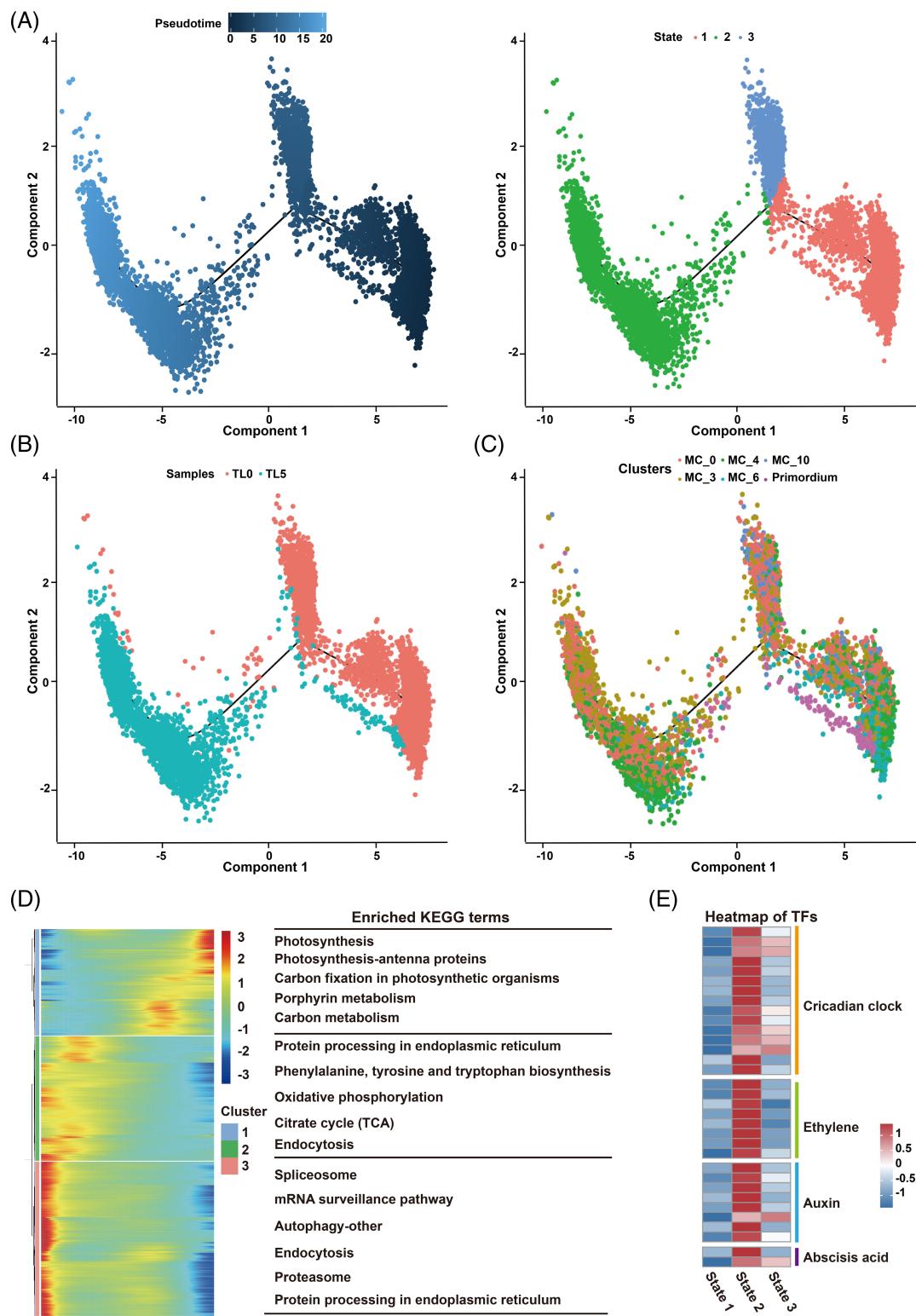


**FIGURE 2** Single-cell transcriptome analysis of control and low N samples. **(A)** The proportion of each cell type in the TL5 samples and the TL0 samples. **(B)** Visualization of the cell clusters using UMAP plot based on marker gene analysis. **(C)** UMI values of each cell type (*P*-values are given by the Student's *t*-tests, significance level: \*\*\*,  $P < 0.001$ ; \*\*,  $P < 0.01$ ; \*,  $P < 0.05$ ; ns., not significant). **(D)** The number of down-DEGs and up-DEGs in different cell types. **(E)** Venn diagram of up-DEGs. **(F)** Venn diagram of down-DEGs. DEG, differentially expressed gene; UMAP, uniform manifold approximation projection; UMI, unique molecular identifier.

cells (Figure 2A). The UMI analysis indicated that different cell clusters exhibited distinct transcriptional characteristics during low-N stress (Figure 2C). To detect if the expression of some genes is affected by low-N in a cell-type-specific manner, we identified DEGs for each cell type (Figure 2D; Table S6). Except for the phloem parenchyma, mesophyll cells have the highest DEGs in the different cell types (Figure 2D). Venn diagram analysis showed that there was some overlap of DEGs across different cell types, but the DEGs mainly appeared in a single cell type. Notably, mesophyll cells have greater specific up and down-regulated DEGs than the other cell types (Figure 3E-F). These observations suggest mesophyll cells are one major low-N responsive cell type.

### 3.4 | Pseudo-time trajectory analysis of mesophyll cells during the different stages of low N stress

The mesophyll layer serves as the primary location of photosynthesis and it has been well-known to originate from the primordium (Runions et al., 2017). We found the composition of cell populations changed upon low-N treatment (Figure 2A). In particular, the proportion of mesophyll cells decreased, and that of primordium cells increased in response to low N. To identify the implicit developmental mechanisms, a pseudo-time trajectory from primordium to mesophyll cells was performed (Figure 3A). The two samples were projected to two ends of the pseudo-time trajectory, including three trajectory



**FIGURE 3** The developmental trajectory from primordium to mesophyll cells in TL0 and TL5 samples. **(A)** Pseudo-time trajectory from primordium to mesophyll cells. Each dot represents a single cell. Color represents the pseudo time score (left). Color represents different states (right). **(B–C)** Sample and cell cluster distributions along the pseudo-time trajectory of mesophyll development. **(D)** Pseudo-time heatmap and KEGG analysis of DEGs ( $FDR \leq .05$ ). Color bar indicates the relative gene expression level. **(E)** Heatmap visualization of the expression of representative identified TFs involved in the circadian clock and plant hormone pathways from the first cluster. The color bar indicates the relative gene expression level.

states (Figure 3B-C). We separated the cell trajectories of the three stages and found that the TL5 sample was at stage 2. Besides, the expression of most DEGs was upregulated with increasing proposed time, and the majority of DEGs were highly expressed in state 2 (Figures 3D and S2), which revealed low-N stress decelerates the differentiation towards mesophyll cells.

To further explore the regulatory mechanisms that slow down the differentiation towards mesophyll cells upon low-N stress, we identified 15,915 DEGs in the pseudo-time trajectory (Figure 3D; Table S7). Genes expressed in cluster 1 were associated with the transition from primordium to mesophyll cells. They were enriched in the GO terms related to photosynthesis and carbon metabolism, and their expression was suppressed in mesophyll cells of the TL0 sample upon low-N treatment (Figure 3D), echoing the delayed development towards mesophyll cells. Additionally, 138 TFs existed in cluster 1, of which 33 TFs were identified as being related to the circadian clock and plant hormone pathways (Figure 3E; Table S8), which play an essential role in the transition from primordium to mesophyll cells. These TFs can be classified into four types: circadian regulation-related proteins, auxin, ethylene, and abscisic acid transcription factors (Figure 3E).

### 3.5 | Tobacco leaf responses of mesophyll cells under low N stress

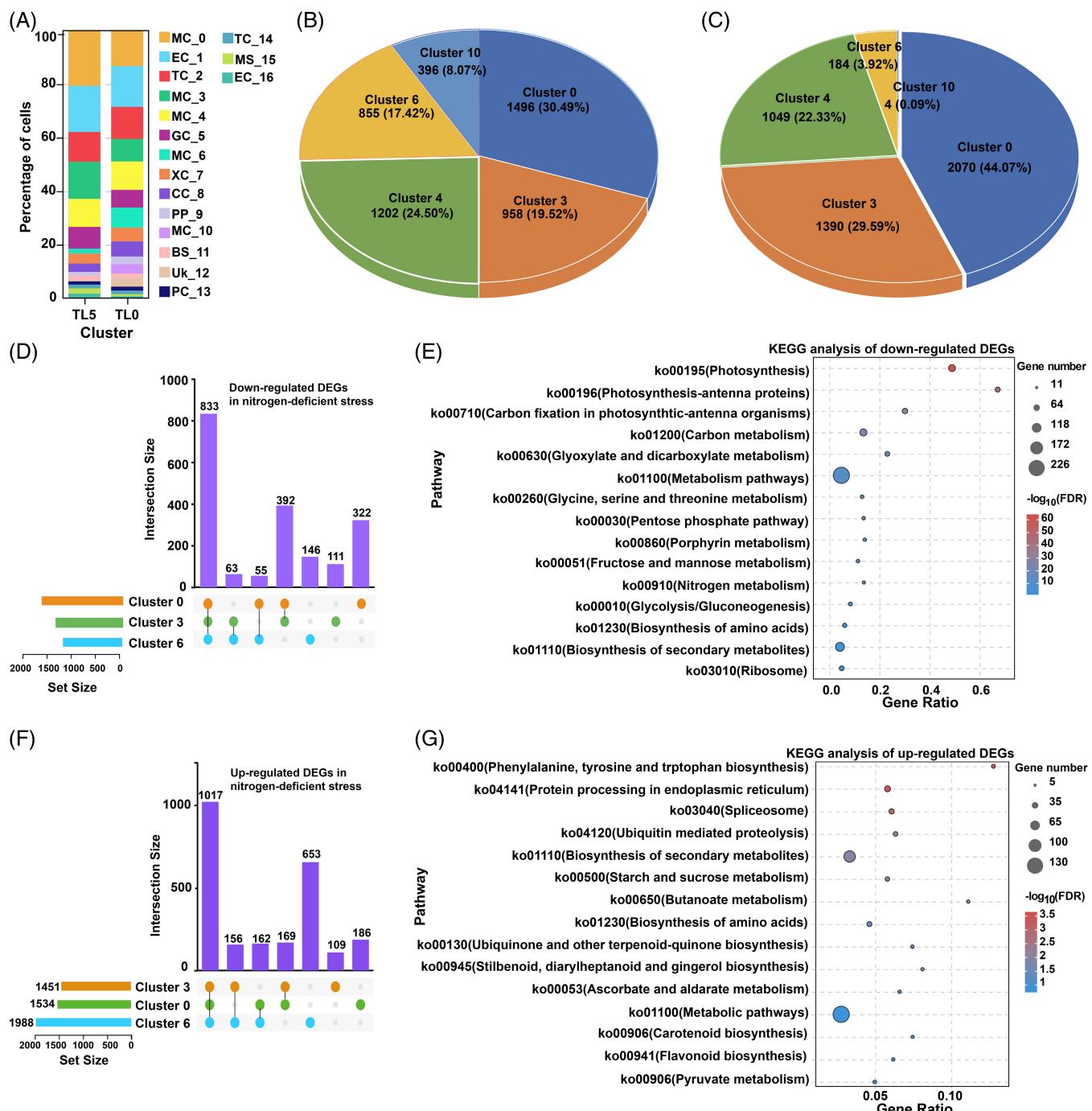
The scRNA-seq analysis confirmed that the mesophyll cell was the major low N-responsive cell type (Figure 4A; Table S9). To further identify and verify major cell subpopulations of mesophyll cells, we analyzed the proportion of subpopulations of mesophyll cells in samples. The results indicated that the largest changes were observed in clusters 0, 3, and 6, revealing their relatively sensitive response (Figure 4B-C; Table S10). To observe their responses, we analyzed their DEGs. Using the upset plot, we found that there was some overlap of DEGs across different cell types, but each cell subpopulation had a unique response (Figure 4D-F). In the following, we mainly focused on their common responsive mechanisms. The common upregulated DEGs were enriched in the GO terms related to the starch biosynthesis process and starch metabolism process, whereas the common downregulated DEGs were enriched in the GO terms related to photosynthesis and light reaction (Figure S3). Subsequently, we observed their common upregulated DEGs were mainly involved in phenylalanine, tyrosine, and tryptophan biosynthesis, protein processing in the endoplasmic reticulum, spliceosome, and ubiquitin-mediated proteolysis (Figure 4G), while their common downregulated DEGs were mainly involved in photosynthesis, photosynthesis-antenna proteins, carbon fixation in photosynthetic organisms, and nitrogen metabolism (Figure 4E). Subsequently, we focussed on describing the downregulated DEGs involved in these key pathways. For instance, the *ncbi\_107764264* (PSAO) and *ncbi\_107774864* (PSBO) genes involved in photosynthesis, the *ncbi\_107774398* (RBCS) and *ncbi\_107771489* (FBA2) genes involved in carbon fixation, the *ncbi\_107774251* (LHC4.1) and *ncbi\_107772871* (CAB16) genes

involved in photosynthesis-antenna proteins, the *ncbi\_107795891* (POR1) and *ncbi\_107791852* (CHLH) genes involved in porphyrin metabolism, and the *ncbi\_107823732* (NIA1) and *ncbi\_107766022* (GLN2) genes involved in nitrogen metabolism were all suppressed under lowN stress (Figure S4). The main mechanism was that low N decelerated the porphyrin biosynthesis, photosynthesis, and nitrogen metabolism.

### 3.6 | Identification of genes in key pathways using WGCNA and RNA-seq analysis

To further analyze the correlation between the key clusters and gene expression patterns in two samples, WGCNA was used to construct a gene co-expression network. The soft threshold power value was set to 6 for subsequent analyses (Figure S5). Besides, a dendrogram was shown to quantify module similarity based on the correlation, and a total of twenty-one gene modules were obtained (Figure S5). We further analyzed the gene expression pattern in the key clusters (clusters 0, 3, 6) of each sample. We found that the genes in the brown module were highly expressed in TL5 key clusters; in contrast, the genes in TL0 were lowly expressed (Figure 5A). KEGG analysis showed that pathways of photosynthesis-antenna, carbon fixation in photosynthetic organisms, porphyrin metabolism, photosynthesis, and nitrogen metabolism were significantly regulated in the brown module (Figure 5C). At the same time, we performed RNA-seq to screen key genes. A total of 6541 DEGs were identified, of which 4752 genes were upregulated and 1789 were downregulated (Figure 5B). Notably, the downregulated DEGs were also enriched in photosynthesis and nitrogen metabolism according to KEGG analysis (Figure 5D), demonstrating genes related to these pathways are most likely to be influenced by low-N stress. We thus focused on the DEGs involved in these pathways. Through overlap analysis of the DEGs using WGCNA, scRNA, and RNA-seq method, we identified 10 DEGs in porphyrin metabolism, five DEGs in nitrogen metabolism, 18 DEGs in carbon fixation in photosynthetic organisms, 19 DEGs in photosynthesis, and 18 DEGs in photosynthesis-antenna proteins as key genes (Figure 5E; Table S11).

Remarkably, great changes were found in the physiological traits of tobacco leaves under low-N conditions (Figure S6). The enzyme activities associated with photosynthesis and nitrogen, including the rubisco enzyme activity, nitrate reductase activity, glutamine synthetase activity, and glutamate synthase activity, declined under low-N stress (Figure S6). Furthermore, a significant decrease in chlorophyll *a*, chlorophyll *b*, leaf biomass, and NO<sub>3</sub>-N content was observed in response to low-N conditions (Figure S6). Subsequently, Pearson's correlation was conducted to explore the relationships between the key gene expression and changes in physiological traits. The results revealed that the expression levels of most key genes were correlated with these physiological traits (Figure S6). Collectively, these results implied low-N stress induced key gene expression changes, thus resulting in a change in physiological traits.

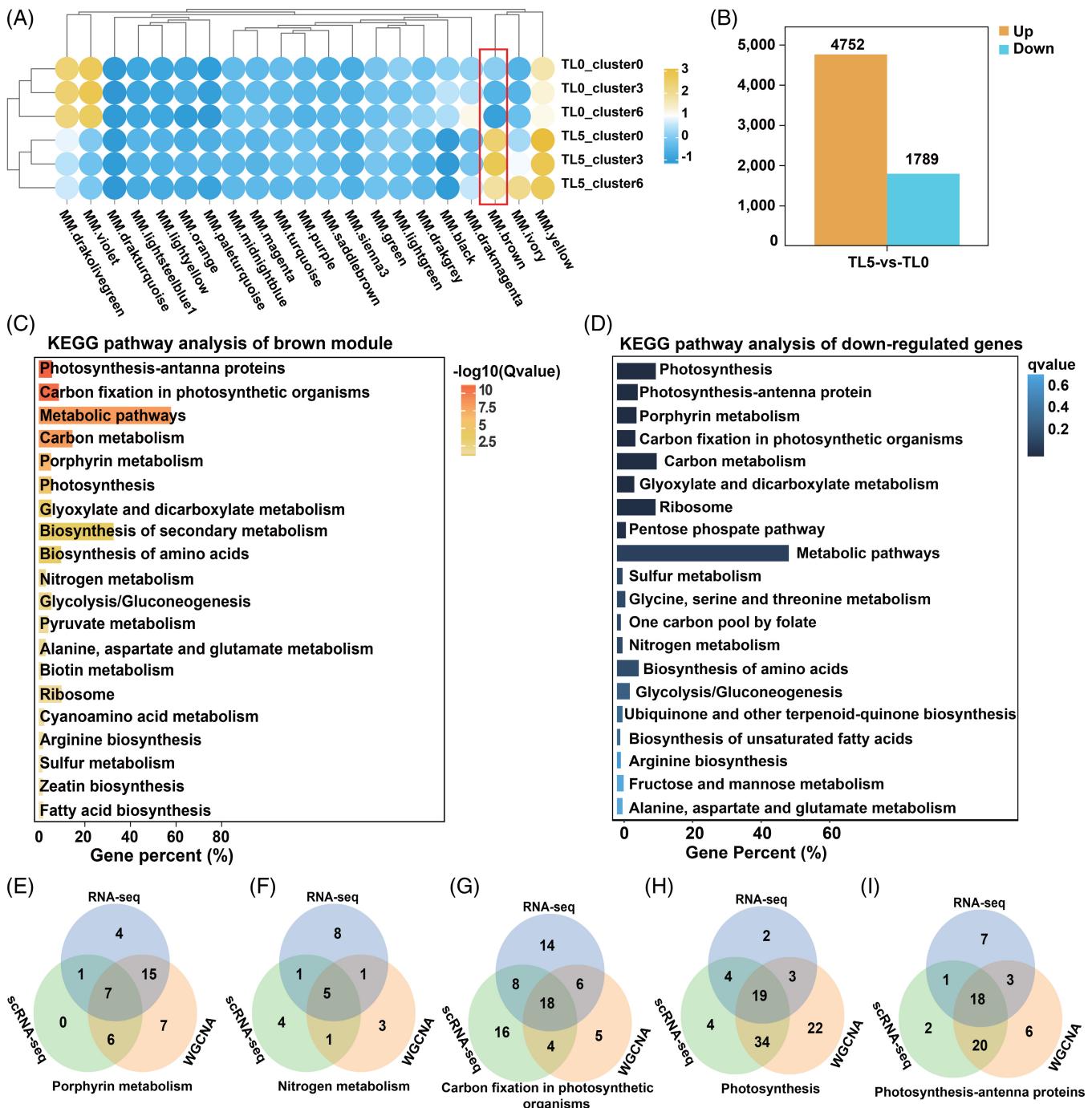


**FIGURE 4** Analysis of responses of key subpopulations of mesophyll cells under low N stress. **(A)** The proportion of each cell cluster in the TLO and TL5 samples. **(B)** The proportion of subpopulations of mesophyll cells in the TLO sample. **(C)** The proportion of subpopulations of mesophyll cells in the TL5 sample. **(D, F)** UpSet plot of down and upregulated DEGs in clusters 0, 3, and 6. The height of the bar under each column reflects the number of DEGs shared by the cell types filled in the matrix. **(E, G)** KEGG pathway analysis of common down and upregulated DEGs in clusters 0, 3, and 6.

### 3.7 | Identification of the TFs affecting the key genes and co-expression network construction

TFs are key regulators of all biological processes in plants, including response to low-N conditions. In the above sections, we have illustrated that genes involved in photosynthesis, nitrogen, and porphyrin

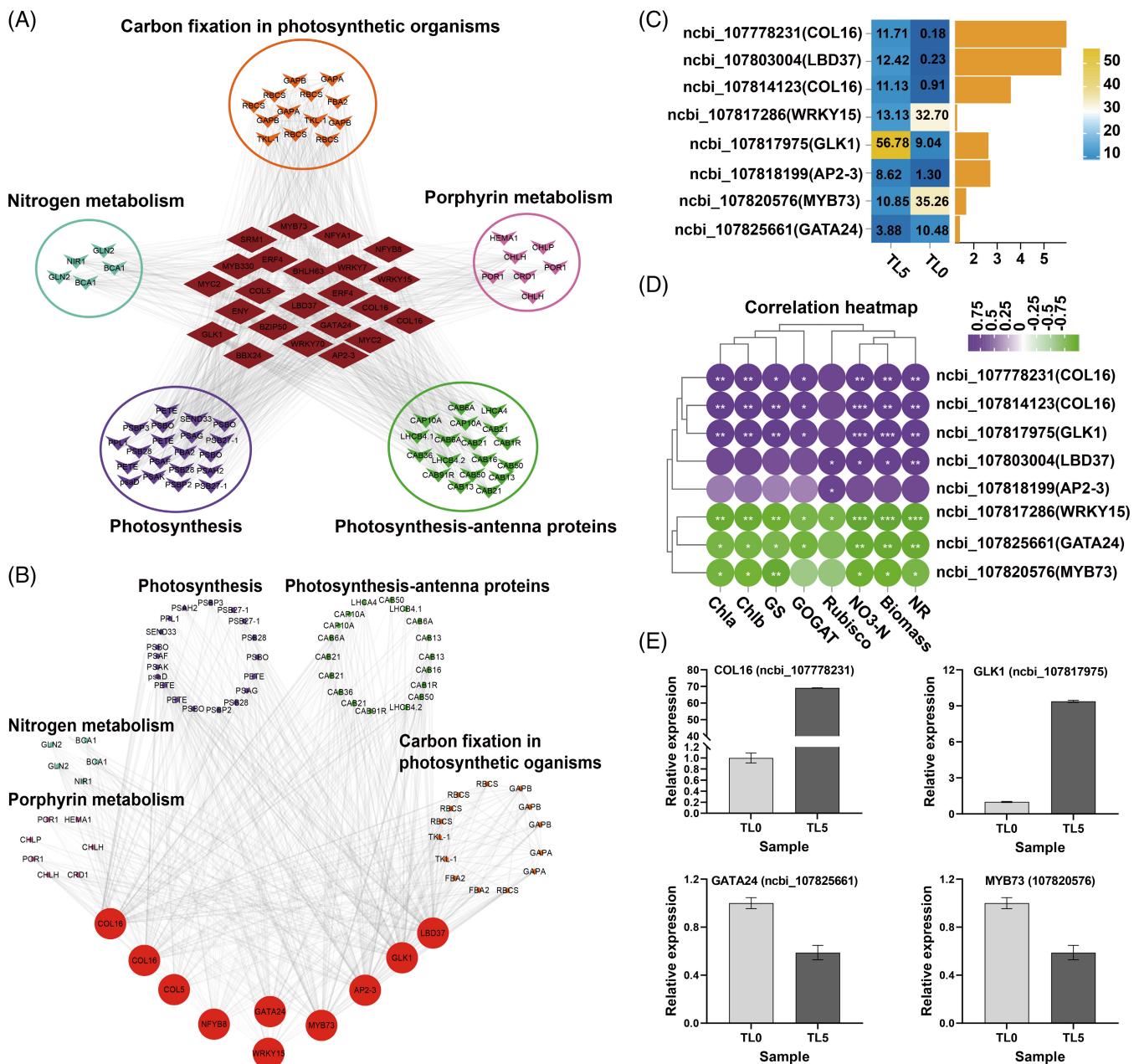
metabolism were important for resistance to low-N stress and leaf development. As a result, these genes were selected as the guide to screen the co-expressed TFs for constructing the protein–protein interaction (PPI) network. In the network, a total of 23 TFs were found to be correlated with these key genes, which were divided into 14 TF families, including MYB, NF-YA, NF-YB, ERF, bHLH, WRKY,



**FIGURE 5** Identification of key genes using WGCNA and RNA-seq analysis. **(A)** The gene expression patterns in the key clusters of two samples. **(B)** DEG numbers based on RNA-seq analysis. **(C)** KEGG enrichment analysis of brown module. **(D)** KEGG pathway analysis of downregulated DEGs. **(E-I)** Venn diagram of DEGs involved in porphyrin metabolism, nitrogen metabolism, carbon fixation in photosynthetic organisms, photosynthesis, and photosynthesis-antenna using WGCNA, RNA-seq, and scRNA-seq methods.

CO-like, LBD, C2H2, bZIP, GATA, AP2, G2-like, DBB (Figure 6A). To narrow down the potential TFs in the regulatory network, the top 10 hub TFs, COL16 (*ncbi\_107778231* and *ncbi\_107814123*), COL5 (*ncbi\_107826614*), NFYB8 (*ncbi\_107828035*), WRKY15 (*ncbi\_107817286*), GATA24 (*ncbi\_107825661*), MYB73 (*ncbi\_107820576*), AP2-3 (*ncbi\_107818199*), GLK1 (*ncbi\_107817975*), and LBD37 (*ncbi\_107803004*), were identified

by CytoHubba plug-in using MCC algorithm in Cytoscape software (Figure 6B). Furthermore, we verified the identified TFs in RNA-seq results. The expression levels of screened genes in the TL5 and TLO samples and the correlation between physiological traits and the expression level of these TFs were performed (Figure S6A). Interestingly, the expression of CO-like, NF-YB, AP2, G2-like, and LBD TFs was



**FIGURE 6** Co-expression network construction and identification of the TFs. **(A)** Identification of the TFs correlated with the key genes in the brown module. **(B)** Co-expression network showed the top 10 TFs in the brown module. **(C)** The expression levels of screened TFs in TL5 and TLO samples. Heatmap illustrated the expression level of TFs. Bar plots showed the value of  $|\log_2\text{FC}|$ . **(D)** The correlation between physiological traits and the expression level of screened TFs. **(E)** RT-PCR validation of screened TFs. The \* symbol represents  $0.01 < P < 0.05$ . The \*\*\* symbol represents  $0.001 < P < 0.01$ . The \*\*\* symbol represents  $0.01 < P < 0.001$ .

downregulated in TLO sample, while the expression of WRKY, MYB, and GATA TFs was upregulated (Figure 6C). In addition, four of these TFs, including COL16, WRKY15, MYB73, and GLK1 were significantly correlated with physiological traits (Figure 6D). Based on the above results, COL16 ([ncbi\\_107778231](#)), GATA24 ([ncbi\\_107825661](#)), MYB73 ([ncbi\\_107820576](#)), and GLK1 ([ncbi\\_107817975](#)) may be considered as the key regulating TFs in response to low-N conditions. And RT-PCR result were consistent with the transcriptome expression trend.

## 4 | DISCUSSION

In worldwide agriculture, low NUE is a major concern because of its negative impact on the environment and crop yield and quality of plants, such as tobacco. On the premise of ensuring crop yield, reduced N application and enhanced NUE demand a pressing investigation. In this context, there is an urgent need to have a detailed understanding of the response mechanism in tobacco leaves under low-N stress. Currently, diverse omics techniques have been

increasingly used to unravel molecular mechanisms of tobacco adaptation to low N stress (Guo et al., 2022b; Wei et al., 2018; Yang et al., 2022). However, conventional omics mainly provide insights into tobacco plant response to low N at the whole-tissue level, failing to capture the characteristics of different cell populations. The advent of single-cell transcriptomics makes it possible to dissect the plant resistance processes at the level of individual cell types. For this reason, single-cell transcriptome technology may help study how tobacco responds to low-N stress. In the present study, we characterized the distinct gene expression patterns of tobacco leaf cells under low-N conditions and successfully identified several TFs for resistance to low-N conditions in tobacco leaves.

In this study, we first captured all major cell types of tobacco leaves and constructed a high-resolution transcriptome atlas (Figure 1B). Moreover, we recapitulated the cellular, transcriptional, and developmental responses to low N stress. We annotated the tobacco leaf cell type using several known marker genes. Based on the gene expression patterns in tobacco, we separated tobacco leaf cells into 10 cell types: epidermal, mesophyll, trichome, primordium, bundle sheath, xylem, meristem, guard, companion, and phloem (Figure 1D; Table S3). In addition, the cluster-specific marker genes characterized could be used in categorizing diverse tissues of tobacco into specific cell types (Table S4). The single-cell transcriptome of tobacco leaf was previously unreported and the single-cell atlas will profoundly benefit the functional characterization of leaf cells in tobacco and other plant species.

Based on a single-cell expression atlas from two samples of low-N and normal -N treatment, we showed that low-N treatment does not alter cell type identity but does lead to changes in the relative proportions of cell numbers. We found that mesophyll cells have greater cell percentages than the other cell types and there was a significant decrease in the proportion of mesophyll cells under low-N conditions (Figure 2A). In addition, heterogeneity in the gene expression response during low-N stress was observed among cell types; although some genes had similar gene expression patterns, the majority of genes responded to stress in a cell-type-specific manner. More DEGs were identified in mesophyll cells (Figure 2D), which indicates mesophyll cells may have the greatest initial response to stress. These results are consistent with previous studies reported in other abiotic stress, such as those involving *Botrytis Cinerea*-infected woodland strawberry (*Fragaria vesca*) leaves and low-N-treated rice seedlings (Bai et al., 2022; Wang et al., 2020a). Further, we observed that the largest changes were identified in clusters 0, 3, and 6 (Figure 4B-C; Table S10), revealing their relative sensitivity for low-N exposure. The common responses analysis of three cell clusters of mesophyll was conducted, and we found photosynthesis, photosynthesis-antenna proteins, carbon fixation in photosynthetic organisms, and nitrogen metabolism were enriched in all three cell clusters (Figure 4E). Photosynthesis is the process of energy synthesis and biomass accumulation and occurs in mesophyll. Low-N stress could inhibit photosynthetic pigment biosynthesis and photosynthesis product formation, which influences nitrogen metabolism, carbon balance, and growth in plants (Sun et al., 2021).

The cultivation of high N-deficiency tolerant genotypes is a primary approach for the maintenance of crop production and the reduction of N fertilizer (Chen et al., 2003; Govindasamy et al., 2023). Thus, it is imperative for breeders to identify the genes responsible for enhancing the N-deficiency tolerance in tobacco. Previous studies have shown that the enhancement of genes responsible for carbon and nitrogen metabolism was beneficial for NUE improvement (Zhang et al., 2023). The genes identified by comparative genome study for wheat nitrate reductase and glutamine synthetase are good candidates to improve nitrogen metabolism and N deficiency resistance (Shi et al., 2022). Some genes of the photosynthesis pathway were also identified (Zhang et al., 2023). Moreover, a few of the genes belonging to the *Pet* and *Psb* gene families were found to be associated with photosynthesis and photosystem II, and these genes were significantly down-regulated in response to N deficiency (Liu et al., 2020b). Previous results showed that the changes in chlorophyll content have a close relationship with the N-deficient tolerant responses in wheat seedlings (Liu et al., 2020a). In view of the fact that these pathways are activated in response to low-N conditions, it is crucial to identify the genes associated with these activated signaling pathways. In the present work, we jointly analyzed scRNA-seq data, WGCNA results, and the gene expression pattern of bulk RNA-seq data. The results enabled us to identify correlated key genes for the above key pathways under low-N conditions (Figure 5E-I; Table S11). Meanwhile, we found that the key genes were correlated with the activities of Rubisco, GS, NR, GOGAT, biomass, NO<sub>3</sub>-N, Chl *a*, and Chl *b* content (Figure S6). Interestingly, the many correlated key genes worked together to affect a specific physiological trait. Therefore, it may be assumed these genes are co-regulated by the key regulators and influence many physiological traits simultaneously.

TFs act as key regulators by regulating gene expression (Gao et al., 2018). Many TFs have been proven to influence the expression of N-response genes. Effah et al. (2022) detected that WRKY, NAC, AP2/ERF, and bZIP TFs were mostly expressed higher under N-limited conditions. In the present study, a total of 23 TFs were identified to be associated with porphyrin, carbon, and nitrogen metabolism in response to N starvation, including MYB73, WRKY15, GLK1, GATA24, AP2-3, and other TFs (Figure 6A). A recent study by Wang et al. (2020b) revealed that OsMYB305 is a potential target for enhancing N assimilation in rice. Jiang et al. (2018) reported GLK5 and bZIP108 genes play an important role in regulating genes in response to N deficiency. Similarly, ERF and WRKY TFs exerted preferential regulatory effects on N utilization and assimilation genes (Li et al., 2023). In soybean, previous studies have shown that GATA44 and GATA58 have crucial roles in modulating nitrogen metabolism and chlorophyll biosynthesis (Zhang et al., 2015). In this work, the co-expression of the regulatory network between carbon and nitrogen-associated genes and TFs was constructed by PPI analysis. The identified top 10 hub TFs (CO-like: COL16 and COL5, NF-YB: NF-YB8, WRKY: WRKY15, AP2: AP2-3, MYB: MYB73, GATA: GATA24, LBD: LBD37, and G2-like: GLK1) were verified in RNA-seq results and eight of them were screened, and the correlation between physiological trait and the expression level of these TFs were performed (Figure 6B-D). Interestingly, the

expression of CO-like, NF-YB, AP2, G2-like, and LBD TFs was downregulated in TLO sample, while the expression of WRKY, MYB, and GATA TFs was upregulated (Figure 6C). Additionally, four of these TFs, including COL16, WRKY15, MYB73, and GLK1 were significantly correlated with physiological traits (Figure 6D), which may be considered as the key regulating TFs in response to low-N stress. To the best of our knowledge, the key regulating TFs from this study were confirmed in the qRT-PCR. The findings in the current study could provide further hints about the genetic regulation of the process under N-deficiency conditions and the identified transcription factors can serve as a foundation to breed tobacco varieties with high tolerance to low-N growing conditions.

Except for describing the gene expression atlas in diverse cell types, a pseudo-time trajectory was conducted to unravel the response in mesophyll cells (Figure 5). Currently, it is considered that the mesophyll originates from the primordium (Runions et al., 2017). Thus, a pseudo-time trajectory from primordium to mesophyll cells was performed. Pseudotime analysis in our study revealed that the TLO samples were at the beginning of the pseudo-time trajectory (Figure 3E), which indicated low-N stress decelerates the differentiation towards mesophyll cells. Furthermore, the genes associated with the transition from primordium to mesophyll cells were enriched in the GO terms related to photosynthesis and carbon metabolism, and their expression was suppressed in mesophyll cells of the TLO sample (Figure 3E), explaining the lower leaf biomass and rubisco enzyme activity in TLO sample (Figure 3E). Additionally, several circadian regulation-related proteins, auxin, ethylene, and abscisic acid transcription factors were simultaneously involved in the transition from primordium to mesophyll cells (Figure 3F; Table S8), which implied the leaf development is regulated by hormone-responsive TFs. The results are in line with a previous study by Liu et al. (2021). These results could expand our understanding of the developmental mechanisms of the mesophyll in the leaf blade.

## 5 | CONCLUSION

Overall, we constructed a single-cell transcriptome atlas and analyzed the leaf cell responses to low-N stress. This is the first report of a single-cell transcriptome in tobacco leaves, and we developed cell-type marker genes. Our scRNA-seq also revealed mesophyll cells were the major responsive cell populations under N-deficiency conditions, and pseudo-time trajectories analysis revealed low-N stress decelerates the differentiation towards mesophyll cells. In combination with scRNA-seq, WGCNA, and bulk RNA-seq results, we found that genes involved in porphyrin metabolism, nitrogen metabolism, carbon fixation in photosynthetic organisms, photosynthesis, and photosynthesis-antenna pathway play an essential role in response to low N condition. Moreover, we identified COL16, WRKY15, MYB73, and GLK1 to be the key TFs in the regulation of N-responsive genes. These findings improve our understanding of leaf development at single-cell resolution and facilitate the investigation of the effects of low-N stress at the single-cell level.

## AUTHOR CONTRIBUTIONS

Shi Hongzhi, Ma Yanjun, and Feng Yuqing conceived and designed the experiments. Feng Yuqing and Zhao Yuanyuan conducted the experiments. Feng Yuqing, Liu Deshui, and Zhao Yuanyuan performed the transcriptome sequencing and data analysis. Feng Yuqing wrote the paper. Shi Hongzhi reviewed and revised the manuscript. The author(s) read and approved the final manuscript.

## FUNDING INFORMATION

The work was supported by the Shanghai Tobacco Corporation Research Program [grant number TP2019-C4].

## DATA AVAILABILITY STATEMENT

The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive (Chen et al., 2021) in the National Genomics Data Center (Xue et al., 2022), China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (GSA: CRA012100) that are publicly accessible at <https://ngdc.cncb.ac.cn/gsa/s/db4248Qd>.

## ORCID

Yuqing Feng  <https://orcid.org/0000-0001-6346-875X>  
Hongzhi Shi  <https://orcid.org/0000-0002-2400-0721>

## REFERENCES

- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M. et al. (2000) Gene ontology: tool for the unification of biology. *Nature genetics*, 25(1), 25–29.
- Benjamini, Y. & Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1), 289–300.
- Boussadia, O., Steppe, K., Zgallai, H., El Hadj, S. B., Braham, M., Lemeur, R. et al. (2010) Effects of nitrogen deficiency on leaf photosynthesis, carbohydrate status and biomass production in two olive cultivars ‘Meski’ and ‘Koroneiki’. *Scientia Horticulturae*, 123(3), 336–342.
- Bailey, W. A. (2014) Effect of nitrogen rate on growth, yield, quality, and leaf chemistry of dark tobacco. *Tobacco Science*, (51), 13–22.
- Bai, Y., Liu, H., Lyu, H., Su, L., Xiong, J. & Cheng, Z.M. (2022) Development of a single-cell atlas for woodland strawberry (*Fragaria vesca*) leaves during early *Botrytis cinerea* infection using single-cell RNA-seq. *Horticulture Research*, 9, uhab055.
- Cataldo, D.A., Haroon, M., Schrader, L.E., Youngs, V.L. Cataldo, D. A., Maroon, M., Schrader, L. E., & Youngs, V. L. (1975) Rapid colorimetric determination of nitrate in plant tissue by nitration of salicylic acid. *Communications in soil science and plant analysis*, 6(1), 71–80.
- Chen, Y., Murchie, E. H., Hubbard, S., Horton, P. & Peng, S. (2003) Effects of season-dependent irradiance levels and nitrogen-deficiency on photosynthesis and photoinhibition in field-grown rice (*Oryza sativa*). *Physiologia plantarum*, 117(3), 343–351.
- Chen, S., Zhou, Y., Chen, Y. & Gu, J. (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34(17), i884–i890.
- Chen, T.T., Chen, X., Zhang, S.S., Zhu, J.W., Tang, B.X., Wang, A.K. et al. (2021) The genome sequence archive family: toward explosive data growth and diverse data types. *Genomics Proteomics Bioinformatics*, 19(4):578–83.
- Drake, M. P., Vann, M. C. & Fisher, L. R. (2015) Nitrogen application rate influence on yield, quality, and chemical constituents of flue-cured tobacco, Part I: Application timing. *Tobacco Science*, (52), 11–17.

- Dorrity, M. W., Alexandre, C. M., Hamm, M. O., Vigil, A. L., Fields, S., Queitsch, C. et al. (2021) The regulatory landscape of *Arabidopsis thaliana* roots at single-cell resolution. *Nature communications*, 12(1), 3334.
- Endo, M., Shimizu, H., Nohales, M. A., Araki, T. & Kay, S. A. (2014) Tissue-specific clocks in *Arabidopsis* show asymmetric coupling. *Nature*, 515(7527), 419–422.
- Effah, Z., Li, L.L., Xie, J.H., Karikari, B., Liu, C., Xu, A.X. & Zeng, M. (2022) Transcriptome profiling reveals major structural genes, transcription factors and biosynthetic pathways involved in leaf senescence and nitrogen remobilization in rainfed spring wheat under different nitrogen fertilization rates. *Genomics*, 114(2), 110271.
- Gao, J., Bi, W.S., Li, H.P., Wu, J.J., Yu, X.M., Liu, D.Q. & Wang, X.D. (2018) WRKY transcription factors associated with NPR1-mediated acquired resistance in barley are potential resources to improve wheat resistance to *Puccinia triticina*. *Frontiers in plant science*, 14, 1486.
- Guo, X.L., Liang, J.L., Lin, R.M., Zhang, L.P., Zhang, Z.C., Wu, J. & Wang, X.W. (2022a) Single-cell transcriptome reveals differentiation between adaxial and abaxial mesophyll cells in *Brassica rapa*. *Plant Biotechnology Journal*, 20(12), 2233.
- Guo, H., He, X.Y., Zhang, H., Tan, R.L., Yang, J.P., Xu, F.S. et al. (2022b) Physiological responses of cigar tobacco crop to nitrogen deficiency and genome-wide characterization of the *NtNPF* family genes. *Plants*, 11(22), 3064.
- Govindasamy, P., Muthusamy, S. K., Bagavathianan, M., Mowrer, J., Jagannadham, P. T. K., Maity, A. et al. (2023) Nitrogen use efficiency—a key to enhance crop productivity under a changing climate. *Frontiers in Plant Science*, 14, 1121073.
- Hao, Y.H., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S.W., Butler, A. et al. (2020) Integrated analysis of multimodal single-cell data. *Cell*, 184(13), 3573–3587.
- Huang, W. T., Zheng, Z. C., Hua, D., Chen, X. F., Zhang, J., Chen, H. H. et al. (2022) Adaptive responses of carbon and nitrogen metabolisms to nitrogen-deficiency in *Citrus sinensis* seedlings. *BMC Plant Biology*, 22(1), 370.
- Jackson, D., Veit, B. & Hake, S. (1994) Expression of maize *KNOTTED1* related homeobox genes in the shoot apical meristem predicts patterns of morphogenesis in the vegetative shoot. *Development*, 120(2), 405–413.
- Jiang, L., Ball, G., Hodgman, C., Coules, A., Zhao, H., & Lu, C.G. (2018) Analysis of gene regulatory networks of maize in response to nitrogen. *Genes*, 9(3), 151.
- Kanehisa, M., & Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, 28(1), 27–30.
- Kaiser, D. R., Sequinatto, L., Reinert, D. J., Reichert, J. M., Rheinheimer, D. S., & Dal Bianco, L. (2015) High nitrogen fertilization of tobacco crop in headwater watershed contaminates subsurface and well waters with nitrate. *Journal of Chemistry*, 2015.
- Kim, D., Langmead, B., & Salzberg, S. L. (2015) HISAT: a fast spliced aligner with low memory requirements. *Nature methods*, 12(4), 357–360.
- Kiba, T., Inaba, J., Kudo, T., Ueda, N., Konishi, M., Mitsuda, N. et al. (2018) Repression of nitrogen starvation responses by members of the *Arabidopsis* GARP-type transcription factor *NIGT1/HRS1* subfamily. *The Plant Cell*, 30(4), 925–945.
- Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K. et al. (2019) Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature methods*, 16(12), 1289–1296.
- Kim, J. Y., Symeonidi, E., Pang, T. Y., Denyer, T., Weidauer, D., Bezrutczyk, M. et al. (2021). Distinct identities of leaf phloem cells revealed by single cell transcriptomics. *The Plant Cell*, 33(3), 511–530.
- Kang, M., Choi, Y., Kim, H. & Kim, S. G. (2022) Single-cell RNA-sequencing of *Nicotiana attenuata* corolla cells reveals the biosynthetic pathway of a floral scent. *New Phytologist*, 234(2), 527–544.
- Langfelder, P. & Horvath, S. (2008) WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics*, 9(1), 1–13.
- Livak, K. J., & Schmittgen, T. D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta CT}$  method. *methods*, 25(4), 402–408.
- Love, M. I., Huber, W. & Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, 15(12), 1–21.
- Lu, J.L., Zhang, L.C., Lewis, R. S., Bovet, L., Goepfert, S., Jack, A. M. et al. (2016) Expression of a constitutively active nitrate reductase variant in tobacco reduces tobacco-specific nitrosamine accumulation in cured leaves and cigarette smoke. *Plant biotechnology journal*, 14(7), 1500–1510.
- Liu, X.X., Wang, S.W., Deng, X.P., Zhang, Z.Y., & Yin, L.N. (2020a) Comprehensive evaluation of physiological traits under nitrogen stress and participation of linolenic acid in nitrogen-deficiency response in wheat seedlings. *BMC Plant Biology*, 20, 1–16.
- Liu, X., Yin, C.M., Xiang, L., Jiang, W.T., Xu, S.Z., & Mao, Z.Q. (2020b) Transcription strategies related to photosynthesis and nitrogen metabolism of wheat in response to nitrogen deficiency. *BMC plant biology*, 20(1), 1–13.
- Liu, H., Hu, D.X., Du, P.X., Wang, L.P., Liang, X.Q., Li, H.F. et al. (2021) Single-cell RNA-seq describes the transcriptome landscape and identifies critical transcription factors in the leaf blade of the allotetraploid peanut (*Arachis hypogaea* L.). *Plant Biotechnology Journal*, 19(11), 2261–2276.
- Long, Y.P., Liu, Z.J., Jia, J.B., Mo, W.P., Fang, L., Lu, D.D. et al. (2021) FlsnRNA-seq: protoplasting-free full-length single-nucleus RNA profiling in plants. *Genome biology*, 22, 1–14.
- Liu, Z.X., Wang, J.J., Zhou, Y.P., Zhang, Y.X., Qin, A.Z., Yu, X.L. et al. (2022) Identification of novel regulators required for early development of vein pattern in the cotyledons by single-cell RNA-sequencing. *The Plant Journal*, 110(1), 7–22.
- Li, P., Du, R., Li, Z., Chen, Z., Li, J., & Du, H. (2023) An integrated nitrogen utilization gene network and transcriptome analysis reveal candidate genes in response to nitrogen deficiency in *Brassica napus*. *Frontiers in Plant Science*, 14, 1187552.
- Liang, X.Y., Ma, Z., Ke, Y.H., Wang, J.L., Wang, L.F., Qin, B. et al. (2023) Single-cell transcriptomic analyses reveal cellular and molecular patterns of rubber tree response to early powdery mildew infection. *Plant, Cell & Environment*, 46(7).
- Marand, A. P., Chen, Z., Gallavotti, A. & Schmitz, R. J. (2021) A cis-regulatory atlas in maize at single-cell resolution. *Cell*, 184(11), 3041–3055.
- Pillitteri, L. J., Peterson, K. M., Horst, R. J. & Torii, K. U. (2011) Molecular profiling of stomatal meristems reveals new component of asymmetric cell division and commonalities among stem cell populations in *Arabidopsis*. *The Plant Cell*, 23(9), 3260–3275.
- Runions, A., Tsiantis, M. & Prusinkiewicz, P. (2017) A common developmental program can produce diverse leaf shapes. *New Phytologist*, 216(2), 401–418.
- Rich-Griffin, C., Stechemesser, A., Finch, J., Lucas, E., Ott, S. & Schäfer, P. (2020) Single-cell transcriptomics: a high-resolution avenue for plant functional genomics. *Trends in Plant Science*, 25, 186–197.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D. et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, 13(11), 2498–2504.
- Sifola, M. I. & Postiglione, L. (2003) The effect of nitrogen fertilization on nitrogen use efficiency of irrigated and non-irrigated tobacco (*Nicotiana tabacum* L.). *Plant and Soil*, 252(2), 313–323.
- Sifola, M. I., Carrino, L., Cozzolino, E., Ianuario, S., Lucibelli, A. & Coppola, A. (2018) A survey of fertility program responses of kentucky dark fire-cured Tobacco (L.) yield and quality for cigars manufacture in the Benevento province (Southern Italy). *Contributions to Tobacco & Nicotine Research*, 28(1), 14–29.

- Satterlee, J. W., Strable, J. & Scanlon, M. J. (2020) Plant stem-cell organization and differentiation at single-cell resolution. *Proceedings of the National Academy of Sciences*, 117(52), 33689–33699.
- Sun, T.T., Zhang, J.K., Zhang, Q., Li, X.L., Li, M.J., Yang, Y.Z. et al. (2021) Integrative physiological, transcriptome, and metabolome analysis reveals the effects of nitrogen sufficiency and deficiency conditions in apple leaves and roots. *Environmental and Experimental Botany*, 192, 104633.
- Sutton, M.A., Howard, C.M., Kanter, D.R., Lassaletta, L., Möring, A., Raghuram, N. et al. (2021) The nitrogen decade: mobilizing global action on nitrogen to 2030 and beyond. *One Earth*, 4(1), 10–14.
- Shi, X.L., Cui, F., Han, X.Y., He, Y.L., Zhao, L., Zhang, N. et al. (2022) Comparative genomic and transcriptomic analyses uncover the molecular basis of high nitrogen-use efficiency in the wheat cultivar Kenong 9204. *Molecular Plant*, 15(9), 1440–1456.
- Sun, X.X., Feng, D.L., Liu, M.Y., Qin, R.X., Li, Y., Lu, Y. et al. (2022) Single-cell transcriptome reveals dominant subgenome expression and transcriptional response to heat stress in Chinese cabbage. *Genome Biology*, 23(1), 1–19.
- Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S.Q., Morse, M. et al. (2014) The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature biotechnology*, 32(4), 381–386.
- Tenorio Berrio, R., Verstaen, K., Vandamme, N., Pevernagie, J., Achon, I., Van Duyse, J. et al. (2022) Single-cell transcriptomics sheds light on the identity and metabolism of developing leaf cells. *Plant physiology*, 188(2), 898–918.
- Van der Maaten, L. & Hinton, G. (2008) Visualizing data using t-SNE. *Journal of machine learning research*, 9(11), 2579–2605.
- Wei, Y.H., Shi, A.B., Jia, X.T., Zhang, Z.Y., Ma, X.M., Gu, M.X. et al. (2018). Nitrogen supply and leaf age affect the expression of *TaGS1* or *TaGS2* driven by a constitutive promoter in transgenic tobacco. *Genes*, 9(8), 406.
- Wei, Z.L., Cheng, Y.L., Zhou, C.C., Li, D., Gao, X., Zhang, S.X. et al. (2019) Genome-wide identification of direct targets of the TTG1-bHLH-MYB complex in regulating trichome formation and flavonoid accumulation in *Arabidopsis thaliana*. *International Journal of Molecular Sciences*, 20(20), 5014.
- Wang, Y., Huan, Q., Chu, X., Li, K., & Qian, W. (2020a). Single-cell transcriptome analyses recapitulate the cellular and developmental responses to abiotic stresses in rice. *BioRxiv*, 2020–01.
- Wang, D.J., Xu, T.Q., Yin, Z.K., Wu, W.J., Geng, H.T., Li, L. et al. (2020b) Overexpression of *OsMYB305* in rice enhances the nitrogen uptake under low-nitrogen condition. *Frontiers in plant science*, 11, 369.
- Wang, Y., Huan, Q., Chu, X., Li, K. & Qian, W.F. (2021) Single-cell transcriptome atlas of the leaf and root of rice seedlings. *Journal of Genetics and Genomics*, 48(10), 881–898.
- Wang, J.J., Hussain, S., Sun, X., Zhang, P., Javed, T., Dessoky, E.S. et al. (2022) Effects of nitrogen application rate under straw incorporation on photosynthesis, productivity and nitrogen use efficiency in winter wheat. *Frontiers in Plant Science*, 13, 862088.
- Xu, M., Du, Q., Tian, C., Wang, Y. & Jiao, Y. (2021) Stochastic gene expression drives mesophyll protoplast regeneration. *Science Advances*, 7(33), eabg8466.
- Xue, Y.B., Bao, Y.M., Zhang, Z., Zhao, W.M., Xiao, J.F., He, S.M. (2022) Database resources of the national genomics data center, China National Center for Bioinformation in 2022. *Nucleic Acids Research*, 50(D1):D27–38.
- Yang, M. & Sack, F. D. (1995) The too many mouths and four lips mutations affect stomatal production in *Arabidopsis*. *The Plant Cell*, 7(12), 2227–2239.
- Yang, R.Z., Yang, J., Yu, J., Wang, S.L., Yang, C.L. & Xu, F.S. (2022) Effects of different nitrogen application rates on the quality and metabolomics of cigar tobacco. *Agronomy Journal*, 114(2), 1155–1167.
- Zou, Q. (2000) *Guide of plant physiological experiments*. Beijing: China Agricultural Press; 56–59.
- Zhang, C.J., Hou, Y.Q., Hao, Q.N., Chen, H.F., Chen, L.M., Yuan, S.L. et al. (2015) Genome-wide survey of the soybean GATA transcription factor gene family and expression analysis under low nitrogen stress. *PLoS one*, 10(4), e0125174.
- Zhang, T.Q., Chen, Y. & Wang, J.W. (2021). A single-cell analysis of the *Arabidopsis* vegetative shoot apex. *Developmental Cell*, 56(7), 1056–1074.
- Zhang, X.B., Ding, Y.G., Ma, Q., Li, F.J., Tao, R.R., Li, T. et al. (2023) Comparative transcriptomic and metabolomic analysis revealed molecular mechanism of two wheat near-isogenic lines response to nitrogen application. *Plant Physiology and Biochemistry*, 195, 47–57.

## SUPPORTING INFORMATION

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

**How to cite this article:** Feng, Y., Zhao, Y., Ma, Y., Liu, D. & Shi, H. (2023) Single-cell transcriptome analyses reveal cellular and molecular responses to low nitrogen in burley tobacco leaves. *Physiologia Plantarum*, 175(6), e14118. Available from: <https://doi.org/10.1111/ppl.14118>