



# Single-Cell RNA Sequencing Efficiently Predicts Transcription Factor Targets in Plants

**Yunjie Xie**<sup>1,2,3,4,5,6,7,8†</sup>, **Shenfei Jiang**<sup>1,2,3,4,5,6,7,8†</sup>, **Lele Li**<sup>2,3,4,5,6,7,8,9†</sup>, **Xiangzhen Yu**<sup>1,2,3,4,5,6,7,8</sup>, **Yupeng Wang**<sup>1,2,3,4,5,6,7,8</sup>, **Cuiqin Luo**<sup>1,2,3,4,5,6,7,8</sup>, **Qiuhua Cai**<sup>2,3,4,5,6,7,8</sup>, **Wei He**<sup>2,3,4,5,6,7,8</sup>, **Hongguang Xie**<sup>2,3,4,5,6,7,8</sup>, **Yanmei Zheng**<sup>1,2,3,4,5,6,7,8</sup>, **Huaan Xie**<sup>1,2,3,4,5,6,7,8,9</sup> and **Jianfu Zhang**<sup>1,2,3,4,5,6,7,8,9\*</sup>

<sup>1</sup> College of Plant Protection, Fujian Agriculture and Forestry University, Fuzhou, China, <sup>2</sup> Rice Research Institute, Fujian Academy of Agricultural Sciences, Fuzhou, China, <sup>3</sup> State Key Laboratory of Ecological Pest Control for Fujian and Taiwan Crops, Fuzhou, China, <sup>4</sup> Key Laboratory of Germplasm Innovation and Molecular Breeding of Hybrid Rice for South China, Ministry of Agriculture and Rural Affairs, Fuzhou, China, <sup>5</sup> Incubator of National Key Laboratory of Germplasm Innovation and Molecular Breeding Between Fujian and Ministry of Sciences and Technology, Fuzhou, China, <sup>6</sup> Fuzhou Branch, National Rice Improvement Center of China, Fuzhou, China, <sup>7</sup> Fujian Engineering Laboratory of Crop Molecular Breeding, Fuzhou, China, <sup>8</sup> Base of South China, State Key Laboratory of Hybrid Rice, Fuzhou, China, <sup>9</sup> College of Agronomy, Fujian Agriculture and Forestry University, Fuzhou, China

## OPEN ACCESS

### Edited by:

**Junhua Peng,**  
Huazhi Rice Bio-Tech Co., Ltd., China

### Reviewed by:

**Hong Luo,**  
Clemson University, United States  
**Chengdao Li,**  
Murdoch University, Australia

### \*Correspondence:

**Jianfu Zhang**  
jianfzhang@163.com

<sup>†</sup>These authors have contributed equally to this work

### Specialty section:

This article was submitted to  
Plant Biotechnology,  
a section of the journal  
Frontiers in Plant Science

**Received:** 06 September 2020

**Accepted:** 16 November 2020

**Published:** 08 December 2020

### Citation:

**Xie Y, Jiang S, Li L, Yu X, Wang Y, Luo C, Cai Q, He W, Xie H, Zheng Y, Xie H and Zhang J (2020) Single-Cell RNA Sequencing Efficiently Predicts Transcription Factor Targets in Plants.** *Front. Plant Sci.* 11:603302.  
doi: 10.3389/fpls.2020.603302

Discovering transcription factor (TF) targets is necessary for the study of regulatory pathways, but it is hampered in plants by the lack of highly efficient predictive technology. This study is the first to establish a simple system for predicting TF targets in rice (*Oryza sativa*) leaf cells based on 10 × Genomics' single-cell RNA sequencing method. We effectively utilized the transient expression system to create the differential expression of a TF (OsNAC78) in each cell and sequenced all single cell transcriptomes. In total, 35 candidate targets having strong correlations with OsNAC78 expression were captured using expression profiles. Likewise, 78 potential differentially expressed genes were identified between clusters having the lowest and highest expression levels of OsNAC78. A gene overlapping analysis identified 19 genes as final candidate targets, and various assays indicated that Os01g0934800 and Os01g0949900 were OsNAC78 targets. Additionally, the cell profiles showed extremely similar expression trajectories between OsNAC78 and the two targets. The data presented here provide a high-resolution insight into predicting TF targets and offer a new application for single-cell RNA sequencing in plants.

**Keywords:** transcription factor, TF targets, OsNAC78, expression trajectory, scRNA-seq

## INTRODUCTION

DNA-binding proteins mediate gene regulation and are involved in most cellular process. In particular, transcription factors (TFs) recognize DNA-binding sites in promoters to modulate gene expression (Nelson, 1995; Luscombe et al., 2000; Ptashne, 2005). Under both normal and stress conditions, thousands of targets are controlled by TFs to maintain intracellular stability.

In plants, numerous TFs still have unknown targets, which seriously affects the elucidation of transcriptional networks. At present, several approaches are being used to identify TF targets. One of the most effective approaches is chromatin immunoprecipitation sequencing (ChIP-seq), which has been widely used in animals and plants to identify DNA-binding proteins *in vivo* (Valouev et al., 2008; Barski and Zhao, 2009; Kaufmann et al., 2010). In principle, ChIP-seq relies on specific antigen-antibody recognition to precipitate DNA fragments bound by TFs. These DNA fragments are then isolated and identified to draw conclusions regarding TF transcriptional processes. Although the principle is simple, the technique is challenging, involving many experimental steps and requiring pre-optimization, and it is highly prone to operational errors and artifacts (Perna and Alberi, 2019; Yang et al., 2019). Another technique, microarray-based chromatin immunoprecipitation, requiring a specific antibody, is employed to identify TF targets, but often no significant DNA fragment enrichment is obtained (Mukherjee et al., 2004). Consequently, many researchers ultimately choose bulk RNA-sequencing (RNA-seq) as the simplest approach to predict TF targets through a transcriptome-wide analysis of differentially expressed genes (DEGs; Stark et al., 2019). For RNA-seq, millions of cells are processed in batches that may harbor potential differences, resulting in the generation of a large number of DEGs (Olsen and Baryawno, 2018). It is, however, difficult to subsequently verify the DNA-protein interactions. Thus, there are still many obstacles in using these techniques to discover TF targets.

Recently, single-cell RNA sequencing (scRNA-seq) has been applied to various fields. Gene expression profiles are refined to the single-cell level, and unique changes among cell types in a population are identified at a high resolution (Kolodziejczyk et al., 2015; Macosko et al., 2015; Olsen and Baryawno, 2018). ScRNA-seq has developed into a powerful tool for the analysis of cellular heterogeneity and the discovery of novel cell types in animals. The spatial distribution and pseudotime of numerous individual cells reveal dynamic cell differentiation and developmental trajectories (Han et al., 2018; Papalexi and Satija, 2018). However, its application in plants encountered challenges, in part because of the existence of the cell wall, which hinders the dissociation of tissues into single cells. Fortunately, a series of recent studies have successfully applied scRNA-seq to *Arabidopsis thaliana* roots. A continuous differentiation trajectory in roots and regulators of cell fate determination were revealed across pseudo-time. Simultaneously, transcriptional profiles in heterogeneous cell populations were observed from a high-resolution perspective, even under abiotic stress conditions (Denyer et al., 2019; Jean-Baptiste et al., 2019; Ryu et al., 2019; Zhang T.Q. et al., 2019).

Owing to the emergence of scRNA-seq and its successful application in plants, we hypothesized that this technology could be applied to predict TF targets. Here, we performed high-throughput scRNA-seq to determine expression profiles of rice (*Oryza sativa*) leaf cells. Based on differential expression levels of the NAC (NAM, ATAF, and CUC) TF OsNAC78 in cells, potential expression characteristics were used to

correlate the TF with its targets. Using DNA-binding assays for verification, we finally identified two targets of OsNAC78. Thus, this study provides a rapid and highly efficient approach to predict TF targets.

## MATERIALS AND METHODS

### Protoplast Isolation, Transfection, and Western Blotting

Rice protoplast isolations were performed as described previously (Zhang et al., 2011). Briefly, the indica cultivar *MH63* rice seedlings that had been grown for 3 weeks were selected as the source material. The leaves of the rice seedlings were cut and immediately transferred into the enzyme solution (1.5% cellulase R-10, 0.75% macerozyme R-10, 0.6 M mannitol, and 10 mM MES at pH 5.7). They were mixed with the enzyme solution for 3 h at 28°C and 70 rpm in the dark. After pouring off the enzyme solution, the W5 solution (154 mM NaCl, 125 mM CaCl<sub>2</sub>, 5 mM KCl and 2 mM MES at pH 5.7) was added and shaken vigorously to release the protoplasts, which were filtered through a 250-micron mesh and two layers of nylon mesh. The protoplasts were collected by centrifugation at 300 × g, washed with the W5 solution after which the supernatant was removed by centrifugation. The protoplasts were resuspended using MMg solution (0.6 M D-Mannitol, 15 mM MgCl<sub>2</sub> and 4 mM MES at pH 5.7). Each sample consisting of 5 µg green fluorescent protein (GFP)-containing plasmids and 100 µL rice protoplasts was transfected using the PEG-mediated method as described previously (Yoo et al., 2007). Samples were cultured for 4, 5, 8, and 12 h at 28°C in the dark, and then, the fluorescence of transfected cells was observed using a confocal microscope (Leica, Germany). The pRTVcOsNAC78-HA and pRTVcHA plasmids were independently transfected using the above methods and then cultured for 8 h at 28°C in the dark. The cell concentration and cell viability were determined using a hemocytometer and Trypan Blue staining, respectively, and then 10 × Genomics' scRNA-seq was performed. To confirm that the OsNAC78-HA plasmids were successfully transfected into the protoplasts, the total protein contents of transfected protoplasts were extracted, and western blotting assays were performed using anti-HA, as described previously (Zhang et al., 2011).

### ScRNA-Seq

The scRNA-seq libraries were prepared using Chromium Single Cell 3' Reagent v2 Kits in accordance with the manufacturer's protocol. Single cells were loaded onto the Chromium Single Cell Controller Instrument (10× Genomics) to generate single cell gel beads in emulsions (GEMs). After the generation of single cell GEMs, reverse-transcription reactions were performed to produce barcoded full-length cDNAs, followed by the disruption of GEMs using the recovery agent and cDNA clean up with DynaBeads Myone Silane Beads (Thermo Fisher Scientific). cDNAs were then amplified by PCR using 12 cycles. Subsequently, the amplified cDNAs were fragmented, end-repaired, A-tailed and index adaptor-ligated. The resulting libraries were amplified. Then, these libraries were sequenced on

the Illumina sequencing platform (NovaSeq 6000), and 150-bp paired-end reads were generated.

## Preprocessing and Analyzing ScRNA-Seq Data

Raw sequencing datasets were analyzed using the Cell Ranger software pipeline (version 2.2.0) provided by 10 $\times$  Genomics. The pipeline included demultiplexing cellular barcodes, generating map reads to the genome and transcriptome using STAR aligner, and creating down-sample reads as required to generate normalized aggregate data across samples, which produced a matrix of gene counts versus cells. The gene-cell matrices for samples were further processed using the R package Seurat (version 2.3.4). To remove low-quality cells and doublets, we filtered out cells with unique molecular identifiers (UMIs)/gene numbers outside of the mean value +/- two fold of the standard deviation limit, assuming a Gaussian distribution of each cells' UMI/gene number (Jean-Baptiste et al., 2019). Following visual inspection of the distribution of cells based on the fraction of mitochondrial genes expressed, we further discarded low-quality cells where >10% of the counts belonged to mitochondrial genes. After applying these QC criteria, 5,912 genes across 8,260 cells were used for further analyses in transiently over-expressed OsNAC78 samples (TOE). In total, 7,722 genes across 16,527 cells were used for further analyses in controls and TOEs.

Expression datasets were first normalized using the LogNormalize method, and 500 DEGs were detected with FindVariableFeatures function. After scaling using ScaleData, a principle component analysis (PCA) was performed using RunPCA with 100 potential PCs. Clusters were identified using FindClusters at a resolution of 1.4. The final data structure and cell clusters were visualized after the application of the RunTSNE function. Marker genes for each cluster were identified using FindMarkers, with the minimum cell percentage for a marker gene being more than 25% and the fold change of the average expression levels between two clusters being more than 1.5 (Jean-Baptiste et al., 2019). The expression levels of genes were plotted using the FeaturePlot function. To evaluate the correlations between OsNAC78 and candidate target genes, mean normalized expression levels for each cluster were calculated and Pearson's correlation coefficients were calculated using R v3.6.1.

To understand the basic roles of detected genes in our sample, gene ontology (GO) enrichment analysis was performed using topGO. A GO term with a *p*-value < 0.05 was regarded as significantly enriched.

## Construction of the Expression Vector and Rice Transformation

The coding region of OsNAC78 was amplified using specific primers containing the restriction sites *Pst I* and *BamH I*. The resulting OsNAC78 fragment was inserted into the *Pst I* and *BamH I* site of pCUBi1390, generating Ubipro::OsNAC78. The vector was introduced into *Agrobacterium tumefaciens* strain *EHA105*, and then transferred into embryogenic calli of *Nipponbare* via *Agrobacterium*-mediated transformation. T2 seeds were obtained for further research.

## Quantitative Real-Time PCR (qRT-PCR)

To investigate the expression levels of candidate genes in OsNAC78-overexpression lines, we extracted total RNAs from rice leaves using TRIzol reagent (TransGen Biotech, China). The total RNA was reverse transcribed into cDNA using a RNA reverse-transcription kit with gDNA Remover (Toyobo, Japan) in accordance with the manufacturer's instructions. qRT-PCR was performed using a FastStart Universal SYBR Green Master (ROX) Kit (Roche, United States) on an ABI Prism® 7500 Real-time PCR system. The reaction solution contained 10  $\mu$ L SYBR Green Mix, 8.25  $\mu$ L aseptic water, 0.375  $\mu$ L each primer and 1  $\mu$ L cDNA. The thermal cycling conditions were as follows: 95°C for 10 min, 40 cycles of 95°C for 10 s and 60°C for 30 s. The *ubiquitin* gene was used as an internal control, and the relative expression levels were obtained using the  $\Delta\Delta Ct$  method. All the experiments were repeated biologically three times. All the primers are listed in Supplementary Table 1.

## Yeast One-Hybrid Assay

The yeast one-hybrid assay was performed using the vector pLacZi or pJG4-5 and the yeast strain EGY48. The full-length CDS of OsNAC78 was amplified using specific primers and inserted into vector pJG4-5. The promoter fragments of candidate genes were amplified using the appropriate corresponding primers and inserted into the pLacZi vector. These constructed vectors and the pJG4-5-OsNAC78 vector were co-transformed into the EGY48 yeast strain using the PEG/LiAc method. The transformed yeast cells were grown on SD/-Trp/-Ura medium and then applied to the yeast plates containing 5-bromo-4-chloro-3-indolyl  $\beta$ -D-galactoside (X-Gal). Interactions were screened by the presence of a blue pigment. The empty vector pJG4-5 and recombinant pLacZi vectors were co-transformed as negative controls.

## Electrophoretic Mobility Shift Assay (EMSA)

The CDS of OsNAC78 was inserted into the pMAL-c5x vector. The maltose-binding protein (MBP) and MBP-OsNAC78 fusion proteins were express in *Escherichia coli* (Rosetta) and purified. The double-stranded Cy5.5-labeled probes used in this assay were synthesized (Biosun, China), and the sequences are listed in Supplementary Table 1. The EMSA assay was performed using an EMSA/Gel-Shift Kit (Beyotime, China) following the manufacturer's instructions. Briefly, 2 mg of purified MBP or MBP-OsNAC78 protein was added to the binding reaction and incubated for 20 min at 25°C in a thermal cycler (Bio-Rad, United States). The mixture was separated on a 4% polyacrylamide gel in 0.5 $\times$  Tris-Borate-EDTA buffer, and the gel images were taken using the Odyssey® Infrared Imaging System (LI-COR, United States).

## Dual-Luciferase (Dual-LUC) Transcriptional Activation Assay

To analyze transcriptional activity, we used a dual-LUC reporter assay system in rice protoplasts. First, the promoters of Os01g0934800 and Os01g0949900 were inserted into the LUC

reporter vector pGreen II 0800, which included a Renilla LUC (REN) gene driven by CaMV 35S as an internal control. Then, the pRTVcOsNAC78-HA vector acted as an effector. The reporter and effector plasmids were co-transformed into protoplasts using the above described method. The transformed protoplasts were cultured for 36 h at 28°C in the dark, and the activities of LUC and REN were measured in accordance with the Dual Luciferase Reporter Assay Kit's instructions (Vazyme, China). The min35s promoter was used as a negative control. The binding capacities of OsNAC78 to the candidate genes' promoters were expressed as LUC/REN ratios. All the experiments were repeated biologically three times.

## RESULTS

### Workflow of Predicting TF Targets Based on 10x Genomics' ScRNA-Seq

To predict TF-binding targets using 10x Genomics' scRNA-seq, leaves of 3-week-old rice seedlings were harvested and digested to obtain protoplasts. Using the PEG-mediated transfection method, the TF plasmids were introduced into protoplasts for expression. Then, cDNA fragments in each single cell were labeled using 10x Genomics' barcoded gel beads, and the gel beads containing barcode information bound to a mixture of cells and enzymes encapsulated by oil surfactant droplets located in the microfluidic "double cross" connection. Then, the droplets flowed into a pool and were collected. Using the barcode information, the sequencing libraries were constructed. Finally, based on the expression correlations between TF and targets, we utilized transient expression differences in TFs in each cell to capture the associated potential targets (Figure 1).

### Sample Preparation and 10x Genomics' ScRNA-Seq

Owing to the efficiency of the PEG-mediated transient expression, it is necessary to optimize the experimental conditions. We used the empty vector pRTVcGFP carrying the *ubiquitin* (Ubi) promoter to drive GFP as a reference, and the expression efficiency of GFP at different periods after transfection was observed. The confocal observations showed that GFP signals began to sporadically appear at 4 h and then gradually increased. The expression efficiency and level both peaked at 12 h (~65% of cells fluoresced). Notably, the expression efficiency of GFP was relatively complete, and cells had the most diverse GFP expression levels at 8 h (Figures 2A,B). Thus, we finally chose 8 h after transfection as the best time to collect samples.

We used the TF gene Os02g0822400 (OsNAC78) as our research object, and scRNA-seq as a tool to identify targets of OsNAC78. The TOE transfected with the pRTVcOsNAC78-HA vector, having the HA tag at the C-terminus, were used as the experimental group, and cells transfected with pRTVcHA were used as the control (Figure 2C). Samples were collected after transfection, and single-cell suspensions for each sample were examined, with more than 1,000 cells per μL present. Cell

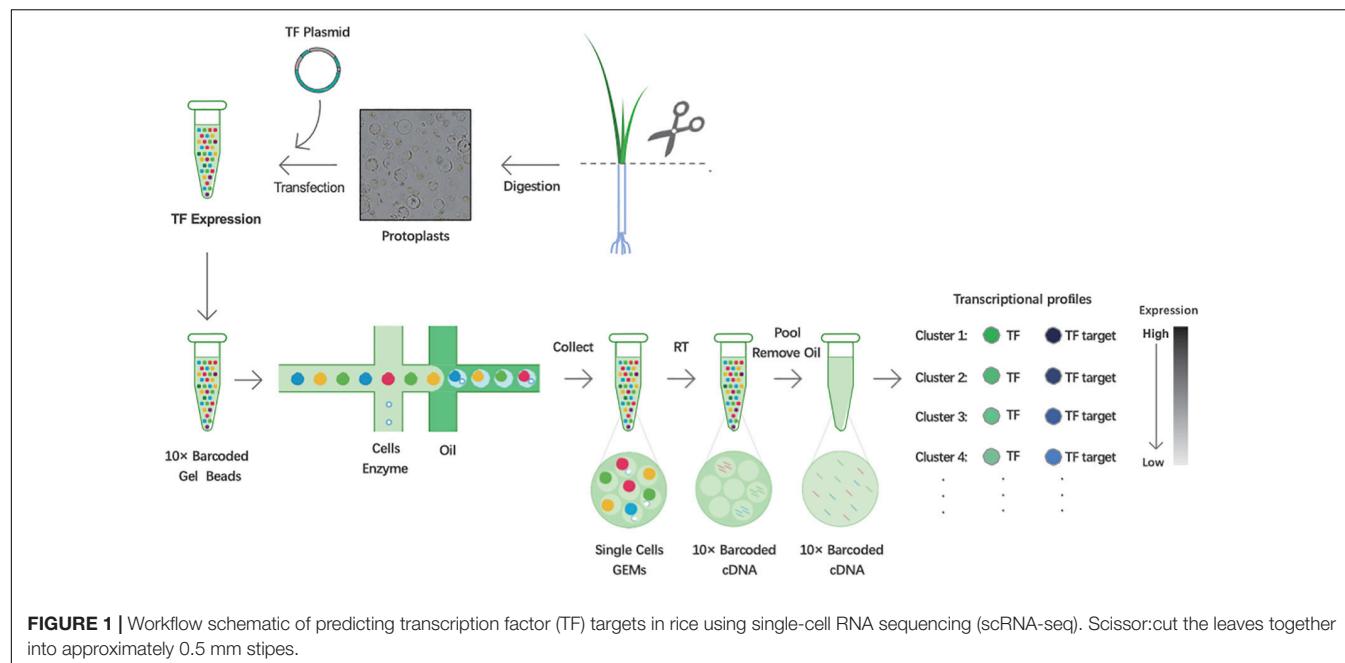
survival rates were more than 85%, which met the standard for sequencing (Supplementary Table 2). Additionally, the OsNAC78-HA protein in the TOE was detected by western blotting, indicating that OsNAC78-HA was overexpressed in cells, but it was not detected in the control (Figure 2D). Therefore, we successfully obtained qualified samples and then performed scRNA-seq.

### Predicting OsNAC78 Targets Using a ScRNA-Seq Analysis

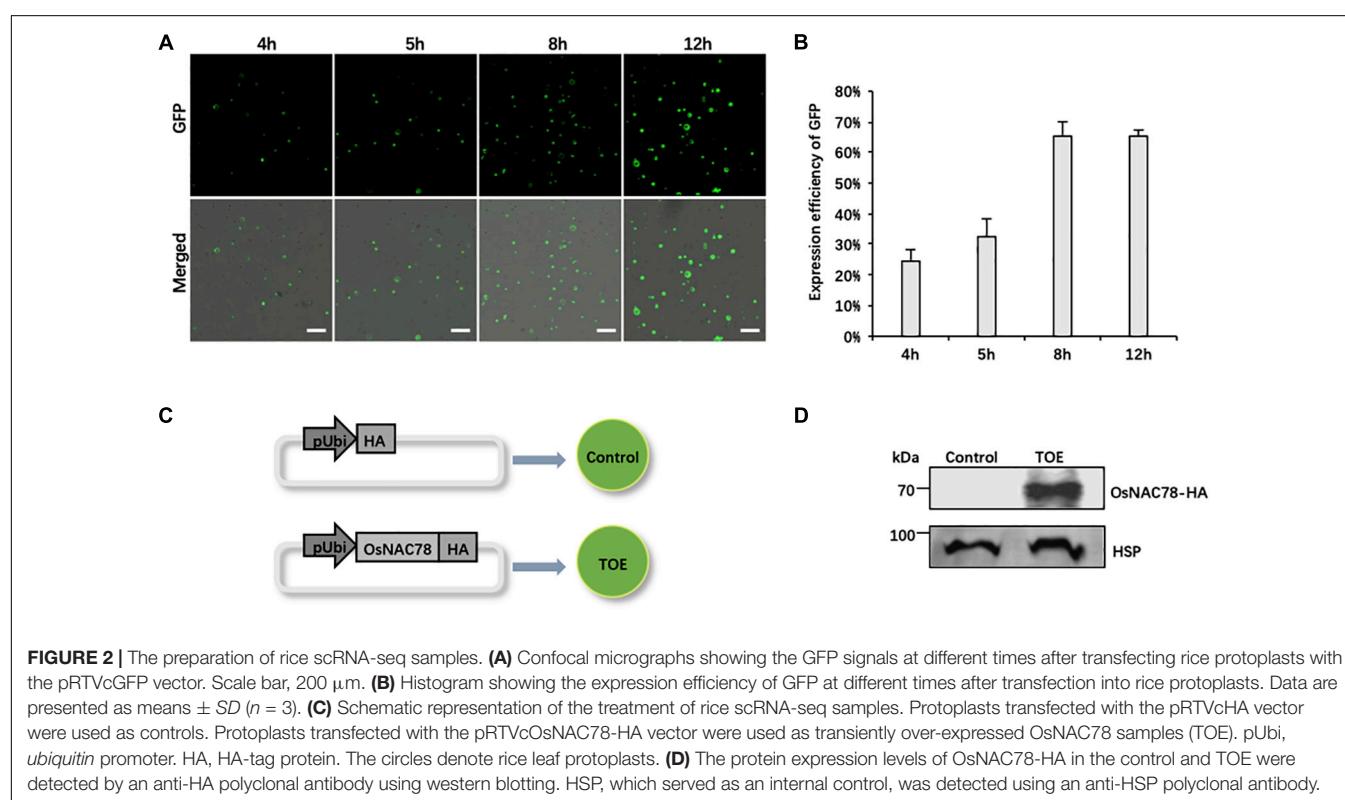
The transcriptome of rice leaf cells was profiled using the 10 × Genomics' platform. In total, 12,891 and 13,639 cells were captured in the control and TOE, respectively, containing means of 5,495 and 5,072 UMIs per cell, respectively. In total, 24,715 and 24,801 genes were detected in the control and TOE, respectively, with means of 1,913 and 1,765 genes, respectively, detected per cell (Supplementary Table 3). After quality filtering, some individual cells having gene/UMI numbers that exceeded the thresholds were discarded, and a mean of 2,050 genes was ultimately detected among 16,527 cells (Supplementary Figure 1). A total of 12,842 genes were detected for further analysis. To understand the basic information in our experiments, GO enrichment analysis was firstly performed (Supplementary Figure 2). Genes involved in basic ontology terms were enriched, such as cellular process, catabolic process, biosynthetic process and reproduction. Genes with binding function and structural molecule activity were detected in our analysis. In cellular component terms, genes located in nucleolus, cytosol, mitochondrion or membrane region were enriched in our dataset.

Following gene expression normalization for read depth, analysis PCA was performed using the top 500 highly variable expressed genes. An unsupervised clustering with a t-distributed stochastic neighbor embedding (t-SNE) analysis was performed on the transcriptomes using the Seurat software package to visualize and explore the datasets (van der Maaten and Hinton, 2008). The leaf cells in the control and TOE were grouped into 14 cell clusters on the basis of marker genes in the cell populations (Figure 3A). Interestingly, two distinct expression profiles were identified in the control and TOE, suggesting that OsNAC78 had effects on the TOE cells (Figure 3B). Compared with the control, the expression of OsNAC78 was significantly induced in the TOE (Figure 3C). The results indicate that OsNAC78 was successfully expressed and stimulated related gene expression levels in the TOE cells. In this t-SNE analysis, we obtained closely arranged clusters, derived from the low heterogeneity of rice leaf cells (Figures 3A–C). This indicated that single cell transcriptomes were almost identical, ensuring that the downstream transcriptional targets regulated by NAC78 were screened under the condition of relatively homogeneous transcriptomes of individual cell.

To investigate the effects of OsNAC78 expression on leaf cells, 13 clusters, each containing from 74 to 1,077 cells, were identified using t-SNE visualization based on the expression profile of the TOE cells (Figure 3D). A heat map illustrated the transcript accumulations of the top 10 marker genes in each cluster



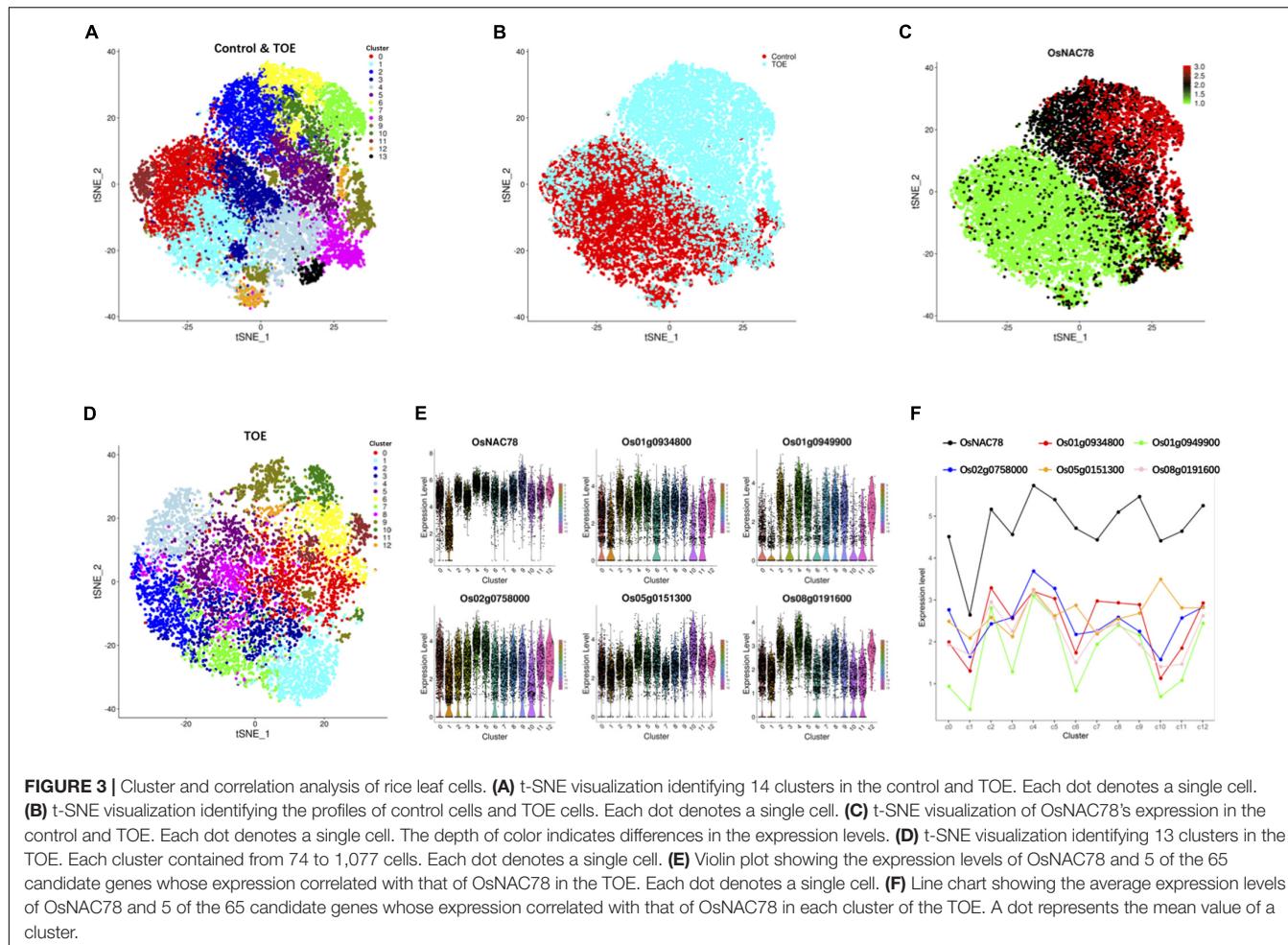
**FIGURE 1 |** Workflow schematic of predicting transcription factor (TF) targets in rice using single-cell RNA sequencing (scRNA-seq). Scissor:cut the leaves together into approximately 0.5 mm stipes.



**FIGURE 2 |** The preparation of rice scRNA-seq samples. **(A)** Confocal micrographs showing the GFP signals at different times after transfecting rice protoplasts with the pRTVcGFP vector. Scale bar, 200  $\mu$ m. **(B)** Histogram showing the expression efficiency of GFP at different times after transfection into rice protoplasts. Data are presented as means  $\pm$  SD ( $n = 3$ ). **(C)** Schematic representation of the treatment of rice scRNA-seq samples. Protoplasts transfected with the pRTVcHA vector were used as controls. Protoplasts transfected with the pRTVcOsNAC78-HA vector were used as transiently over-expressed OsNAC78 samples (TOE). pUbi, ubiquitin promoter. HA, HA-tag protein. The circles denote rice leaf protoplasts. **(D)** The protein expression levels of OsNAC78-HA in the control and TOE were detected by an anti-HA polyclonal antibody using western blotting. HSP, which served as an internal control, was detected using an anti-HSP polyclonal antibody.

(Supplementary Figure 3). Cluster-specific marker genes were identified on the basis of their expression levels in each cluster compared with all the other clusters (fold-change  $> 1.5$  and false discovery rate  $< 0.05$ ; Supplementary Table 4). OsNAC78 had diverse expression levels among the clusters, especially in clusters 1 and 4 (Figures 3E,F). On this basis, the expression

levels of 65 candidate genes were correlated with the expression of OsNAC78 (Figures 3E,F and Supplementary Figure 4). The average expression levels of the 65 candidate genes among the 13 clusters were calculated, and strong correlations for 35 candidate genes were observed (Pearson's correlation coefficient  $R^2 > 0.7$ ; Supplementary Table 5). Furthermore, DEGs were identified



between cluster 1, having the lowest OsNAC78 expression level, and cluster 4, having the highest OsNAC78 expression level (**Figures 3E,F**). In total, 265 genes changed their expression levels between cells in clusters 1 and 4 (fold-change > 1.5 and false discovery rate < 0.05; **Supplementary Table 6**), with 128 genes, including OsNAC78, significantly induced in cluster 4 and 78 DEGs expressing two-fold higher than in cluster 1 (**Figure 4A**).

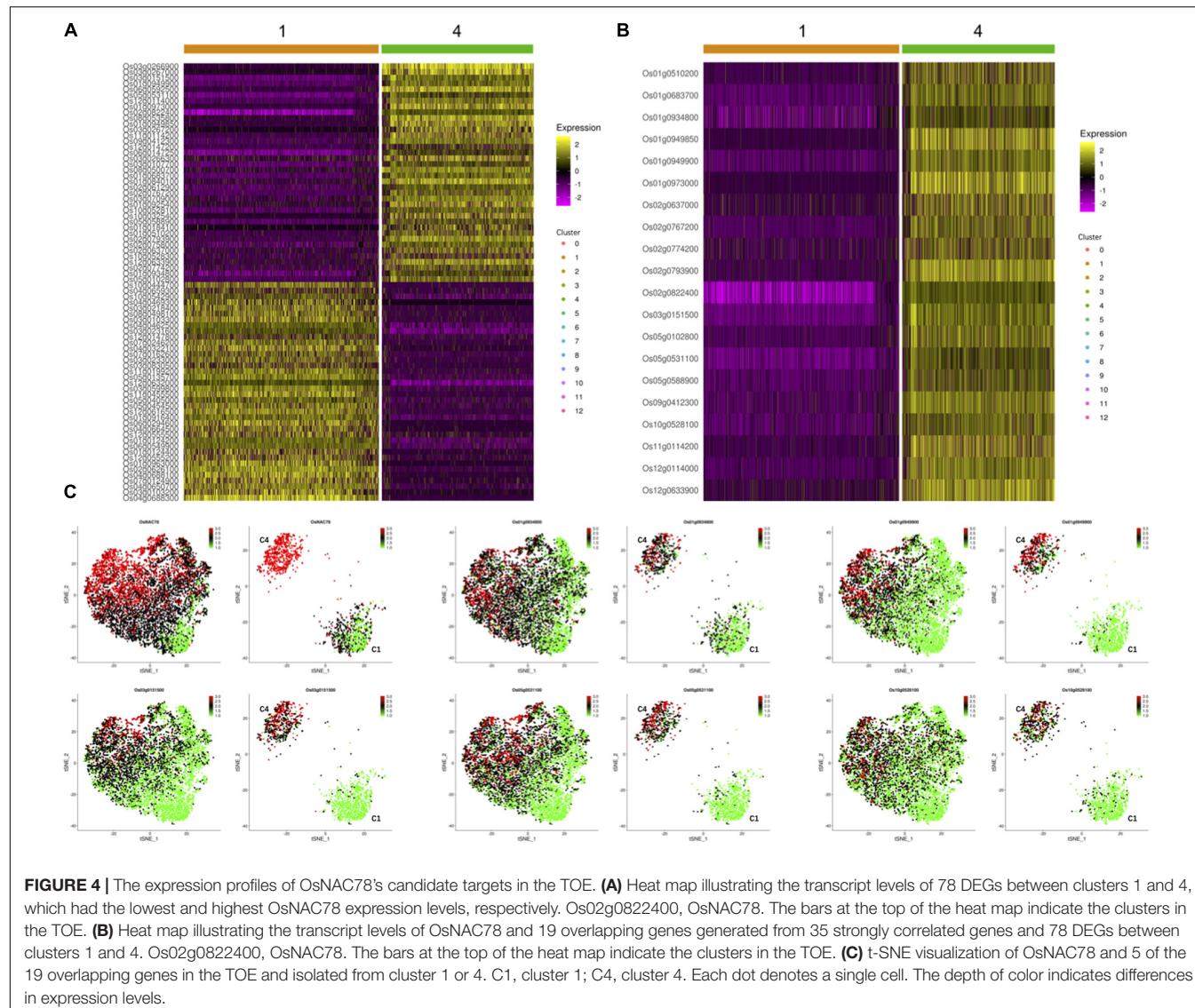
Using a gene overlapping analysis, 19 candidate genes were identified from the 35 strongly correlated genes and 78 DEGs (**Supplementary Table 7**). A heat map of all 19 overlapping genes in clusters 1 and 4 were constructed and their expression profiles in the TOE determined (**Figures 4B,C** and **Supplementary Figure 5**). We clearly observed that the expression patterns of these 19 genes were similar to that of OsNAC78. Thus, all 19 genes were selected as final candidate target genes of OsNAC78 and subjected to further experimental validation.

## OsNAC78 Directly Binds to the Promoters of Os01g0934800 and Os01g0949900

To conduct efficient validations of the 19 candidate genes, we generated transgenic lines that overexpressed OsNAC78,

and then, we assessed the expression levels of the candidate genes using qRT-PCR. The CDS of OsNAC78 was inserted into the pCUB1390 vector and transformed into the *Nipponbare* background. Os01g0934800 and Os01g0949900 were both significantly up-regulated when OsNAC78 was overexpressed (**Figure 5A**). Seven and eight CGT[GA] motifs in the promoter regions of Os01g0934800 and Os01g0949900, respectively, were identified by a sequence analysis. These elements are likely to be DNA-binding sites of OsNAC78, because NAC TFs recognize CGT[GA] core motifs (Tran et al., 2004; Jensen et al., 2010; Jensen and Skriver, 2014).

Thus, a yeast one-hybrid assay was used to determine the direct binding between OsNAC78 and the promoters of Os01g0934800 and Os01g0949900. OsNAC78 directly bound to the F2 (-1,395–1,116) promoter region of Os01g0934800 and to the F1 (-2,251–1,992) and F2 (-1,643–1,587) promoter regions of Os01g0949900 (**Figure 5B**). All three promoter fragments contained CGTG motifs. To further determine whether OsNAC78 specifically binds to these motifs, an EMSA was performed. The double-stranded Cy5.5-labeled probes used in this assay were synthesized based on these motif regions, and unlabeled probes were used as competitors. The OsNAC78 protein fused to an N-terminal MBP tag was



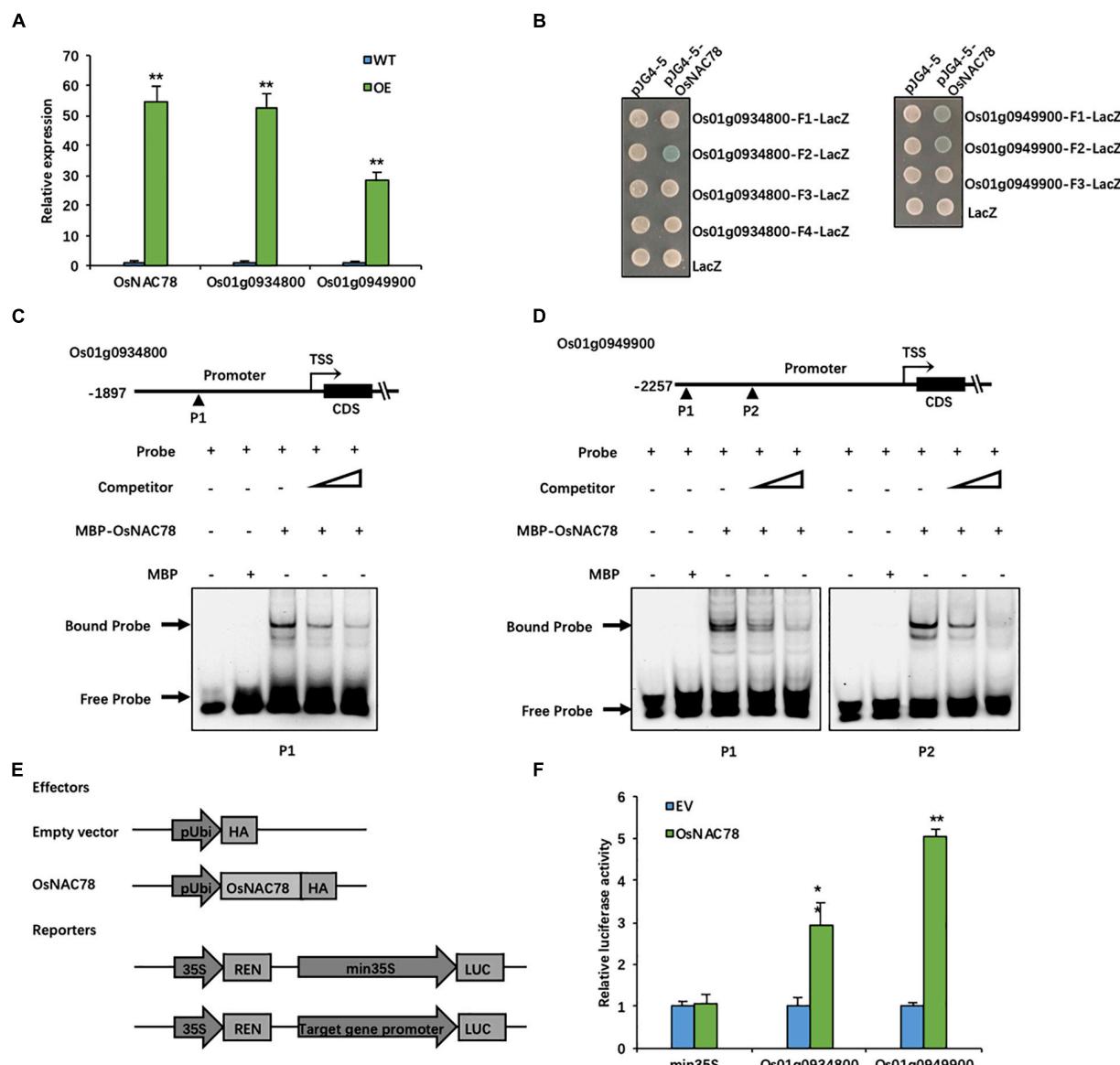
successfully expressed in *E. coli* (Rosetta). MBP-OsNAC78 bound labeled probes containing CGTG motifs to form DNA–protein complexes. Then, migration was attenuated by increasing doses of competitive probes (Figures 5C,D). Our results indicated that OsNAC78 directly bound the CGTG motifs in promoter fragments of Os01g0934800 and Os01g0949900.

The scRNA-seq results indicated that OsNAC78 is positively correlated with Os01g0934800 and Os01g0949900. However, this did not prove that it acts as an activator. We performed a dual-LUC assay to determine the correlations between OsNAC78 and both Os01g0934800 and Os01g0949900. The OsNAC78-HA fusion protein driven by the *ubiquitin* promoter and the Firefly LUC driven by the target gene promoter acted as an effector and a reporter, respectively (Figure 5E), when co-transfected into rice protoplasts. The LUC transcriptional activity driven by the target genes' promoters were

remarkably up-regulated in cells transiently over-expressed OsNAC78, demonstrating that OsNAC78 interacted with the two target genes' promoters, but not with the min35S promoter (Figure 5F). Additionally, OsNAC78 positively regulated the expression levels of Os01g0934800 and Os01g0949900. These observations showed that Os01g0934800 and Os01g0949900 are downstream targets of OsNAC78.

## Analyzing the Expression Profiles of Target Genes Os01g0934800 and Os01g0949900 in the Control and TOE

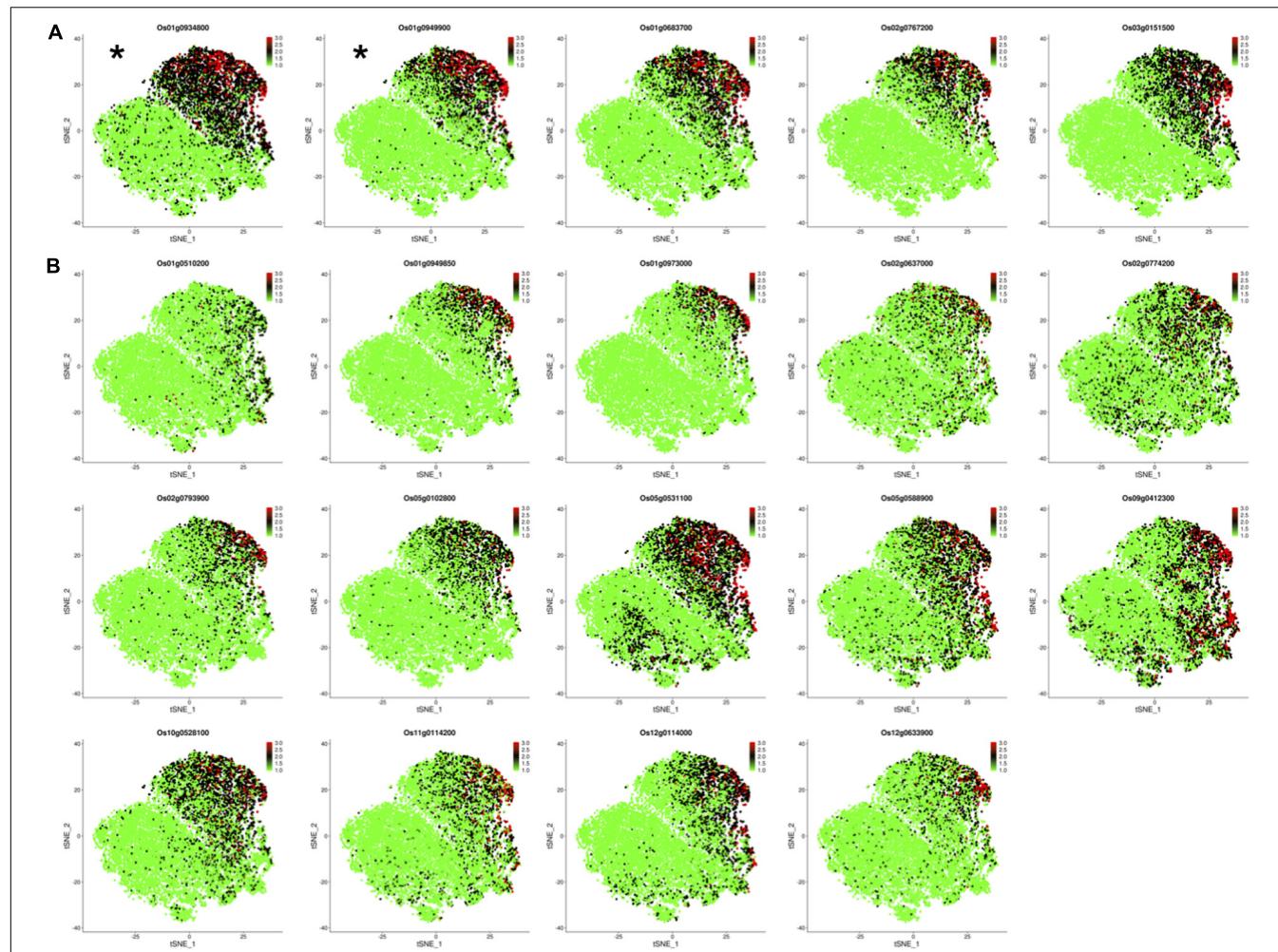
Os01g0934800 and Os01g0949900 have been identified as OsNAC78 targets from 19 candidate genes based on experimental results. The expression profiles of the two targets in the control and TOE need to be further studied for the efficient identification of targets. We determined the



**FIGURE 5 |** OsNAC78 directly binds to the promoters of Os01g0934800 and Os01g0949900. **(A)** The relative expression levels of OsNAC78, Os01g0934800, and Os01g0949900 were detected in wild-type (WT) and OsNAC78 overexpression (OE) lines as assessed by qRT-PCR. Data are presented as means  $\pm$  SDs ( $n = 3$ ) and were analyzed using Student's *t*-test; The asterisks indicate significant differences ( $**P < 0.01$ ). **(B)** The interactions between OsNAC78 and the promoter fragments of Os01g0934800 (represented as F1–4) and Os01g0949900 (represented as F1–3) were shown using a yeast one-hybrid assay. Yeast cells were grown on SD/-Ura/-Trp-X-gal medium. pJG4-5 was used as an effector vector. pLacZ (LacZ) was used as a reporter vector. **(C)** The interaction between OsNAC78 and a probe (P1) containing one CGTG motif in the Os01g0934800 promoter was shown using EMSA. Unlabeled probe was used as a competitor. The triangle denotes an increasing dose. MBP was used as a control. **(D)** The interaction between OsNAC78 and a probe (P1) containing two CGTG motifs or a probe (P2) containing one CGTG motif in the Os01g0949900 promoter was shown using EMSA. Unlabeled probe was used as a competitor. The triangle denotes an increasing dose. MBP was used as a control. **(E)** Schematic showing the vectors used for identifying interactions between OsNAC78 and the promoters of Os01g0934800 and Os01g0949900 using a dual-luciferase (LUC) assay. OsNAC78 was inserted into the pRTVcHA vector as an effector. The promoters of Os01g0934800 and Os01g0949900 were independently inserted into the pGreenII 0800-LUC vector as reporters, and Renilla luciferase (REN) was used as an internal control. **(F)** The relative LUC activity levels indicated the promoter activities of min35S, Os01g0934800 and Os01g0949900 in the absence and presence of OsNAC78 in protoplasts. The min35S promoter was used as a control. EV, empty vector. Data are presented as means  $\pm$  SDs ( $n = 3$ ) and were analyzed using Student's *t*-test; The asterisks indicate significant differences ( $**P < 0.01$ ).  $*$   $P < 0.05$ .

expression of 19 candidate genes using t-SNE visualization in the control and TOE (Figure 6). Notably, five genes, including targets Os01g0934800 and Os01g0949900, displayed

expression trajectories highly similar to that of OsNAC78 (Figures 3C, 6A). While the other 14 genes showed low correlations and different expression trajectories, compare



**FIGURE 6 |** The expression profiles of 19 candidate genes in the control and TOE. **(A)** t-SNE visualization of five candidate genes having expression trajectories that were highly similar to that of OsNAC78. \*OsNAC78 target. Each dot denotes a single cell. The depth of color indicates different expression levels. **(B)** t-SNE visualization of 14 candidate genes having expression trajectories that were somewhat similar to that of OsNAC78. Each dot denotes a single cell. The depth of color indicates different expression levels.

with that of OsNAC78 (**Figures 3C, 6B**). Although the interaction between OsNAC78 and the 19 candidate genes cannot be all clearly verified owing to experimental limitation, we found that genes having expression trajectories similar to the TF are more likely to be targets. Therefore, the efficiency of capturing targets is improved by screening candidate genes based on transcriptional similarity to the TF at a high resolution.

## DISCUSSION

With the development of droplet technology, scRNA-seq has been applied as a research tool in many fields (Klein et al., 2015; Zhang X.N. et al., 2019). However, it is unusual to identify TF targets using the high-throughput  $10 \times$  Genomics' platform. At present, directly testing the binding between TF and DNA fragments *in vivo* using Chip-seq is preferred.

However, its experimental operation is demanding, the steps are cumbersome and error-prone, and the pseudomorphisms are highly sensitive (Perna and Alberi, 2019; Yang et al., 2019). Therefore, bulk RNA-seq, which captures candidate targets using a single dimensional analysis of DEGs, is often used as a simple tool to predict TF targets instead of Chip-seq.

scRNA-seq is an upgrade of RNA-seq that is not limited to DEG analyses, but is a multidimensional analysis, which greatly improves the screening efficiency for identifying TF targets. For example, in the scRNA-seq data presented here, 78 DEGs were identified as candidate targets between clusters having the lowest and highest expression levels of OsNAC78 in the TOE (**Figure 4A**). The DEGs analysis of one method alone was not able to ensure accurate results. However, taking clusters as units, 35 genes strongly correlated with OsNAC78 were obtained (**Supplementary Table 5**). In total, 19 overlapping genes were acquired by comparing the two lists, and it was likely

that OsNAC78 targets were present. Compared with bulk RNA-seq, scRNA-seq has an additional dimension for scanning the transcriptome, which is an eligible predictor of TF targets.

Furthermore, the superiority of scRNA-seq is also reflected in the reduction of potential errors. In bulk RNA-seq, sampling from different individual plants makes it impossible to avoid identifying false positive genes, because the DEGs are derived from potential differences in growth conditions or external factors. However, scRNA-seq can avoid this error. A separate sample is sufficient for predicting TF targets, such as the TOE, in which each cell is treated almost the same during the collection process. Thus, external interference can be basically eliminated, resulting in a reduction in false positive results.

Although scRNA-seq has high-resolution visualization, there are also several factors that affect the data analysis. For example, the TOE was grouped into 13 clusters, taking the average expression of OsNAC78 for correlation analysis on the cluster unit. The greater the number of clusters, the fewer candidate targets with higher precision may be identified in a manner. Unfortunately, when we adjusted the resolution to cell unit for the correlation analysis, we were not able to acquire candidate targets associated with OsNAC78, and the data appeared irregular. There may be two reasons, as follows: (1) some noisy, sparse and high-dimensional cells cannot be eliminated and they interfere with the data analysis (Kulkarni et al., 2019); or (2) scRNA-seq easily captures highly expressed transcripts in cells, while genes with low and medium expression levels are unreliable quantified, including genes having zero reads (Kharchenko et al., 2014; Huang et al., 2018). When performing a correlation analysis on the cell unit, such genes may affect the accuracy. More data algorithms for single-cell expression recovery need to be developed to estimate the actual numbers of transcripts of all the genes. Thus, scRNA-seq still poses challenges when applied to plants.

Using t-SNE visualization, the cell clusters were closely arranged in our profiles (**Figures 3A,D**). This indicates that cell heterogeneity in rice leaves is low. Cell type-specific transcript abundance gradually decreases as chloroplasts mature, especially in bundle sheath cells and mesophyll cells (Sharpe et al., 2011). During the preparation of our samples, 3-week-old seedlings were chosen as the experimental material, and the rice leaves we investigated were mostly mature, which might explain the low level of cell heterogeneity. Thus, we hypothesize that juvenile leaves may be more suitable for scRNA-seq when studying cell type. Likewise, rice does not currently have as abundant cell-type marker genes as *Arabidopsis*; therefore, it is difficult to distinguish the cell type of each cluster in rice. But in this study, clustering was only used to correlate target genes effectively. The low cell heterogeneity reflects that single cell transcriptomes are almost identical, on which basis the accuracy of associated NAC78 target genes can be improved. In addition, we discovered that the arrangements of the two cell populations in the control and TOE were significantly different, because gene transcription in most TOE cells had altered (**Figure 3B**).

It is possible that the overexpression of OsNAC78 stimulated downstream related gene expression, resulting in two distinct expression profiles. This indicates the OsNAC78 may play an important role in rice.

OsNAC78 is a member of the NAC family. The NAC TFs have been identified as plant-specific proteins (Nuruzzaman et al., 2010; Welner et al., 2012). There are more than 150 NAC TFs in rice, and they contain a conserved N-terminal DNA-binding domain and a variable C-terminal transcriptional activation region (Ooka et al., 2003; Ernst et al., 2004). NAC TFs play essential roles in biotic and abiotic stress responses (Wu et al., 2009; Nuruzzaman et al., 2013; Sun et al., 2015). However, the function of OsNAC78 has not yet been identified in rice. Two targets, Os01g0934800 and Os01g0949900, were identified in this study. Os01g0934800 encodes esterase Pir7b, a defense-related rice protein, which may have a detoxification role in defense responses against *Pyricularia oryzae* (Luo et al., 2008). Os01g0949900 is a GST gene encoding a ROS-scavenging enzyme that may prevent oxidative bursts and inhibit oxidative damage under stress conditions (Wagner et al., 2002; Dean et al., 2005). The functions of these two targets provide clues to the roles of OsNAC78.

## DATA AVAILABILITY STATEMENT

The authors acknowledge that the data presented in this study must be deposited and made publicly available in an acceptable repository, prior to publication. Frontiers cannot accept a manuscript that does not adhere to our open data policies.

## AUTHOR CONTRIBUTIONS

JZ and HuX planned and designed the experiments. YX, XY, and JZ wrote the manuscript. YX, SJ, LL, and JZ performed the experiments and analyzed the data. YW, CL, QC, WH, HoX, and YZ were in charge of collection and interpretation some materials. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by the National Key R&D Program Foundation of China (Grant no. 2016YFD0300508) and the Special Foundation of Non-profit Research Institutes of Fujian Province (Grant no. 2018R1021-5).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2020.603302/full#supplementary-material>

**Supplementary Figure 1** | Each cells' gene/UMI numbers before or after quality filtering.

**Supplementary Figure 2** | Gene ontology analysis of detected genes in control and TOE sample.

**Supplementary Figure 3** | The transcript levels of top 10 marker genes in each cluster of the TOE.

**Supplementary Figure 4** | The expression levels of 60 of the 65 candidate genes in the TOE.

**Supplementary Figure 5** | The expression profiles of 14 of the 19 overlapping genes in the TOE and isolated from cluster 1 or 4.

**Supplementary Table 1** | Primers list.

## REFERENCES

- Barski, A., and Zhao, K. (2009). Genomic location analysis by ChIP-Seq. *J. Cell. Biochem.* 107, 11–18. doi: 10.1002/jcb.22077
- Dean, J. D., Goodwin, P. H., and Hsiang, T. (2005). Induction of glutathione S-transferase genes of *Nicotiana benthamiana* following infection by *Colletotrichum destructivum* and *C. orbiculare* and involvement of one in resistance. *J. Exp. Bot.* 56, 1525–1533. doi: 10.1093/jxb/eri145
- Denyer, T., Ma, X. L., Klesen, S., Scacchi, E., Nieselt, K., and Timmermans, M. C. P. (2019). Spatiotemporal developmental trajectories in the *Arabidopsis* root revealed using high-throughput single-cell RNA sequencing. *Dev. Cell* 48, 840–852. doi: 10.1016/j.devcel.2019.02.022
- Ernst, H. A., Olsen, A. N., Larsen, S., and Leggio, L. L. (2004). Structure of the conserved domain of ANAC, a member of the NAC family of transcription factors. *EMBO Rep.* 5, 297–303. doi: 10.1038/sj.emboj.7400093
- Han, X. P., Chen, H. D., Huang, D. S., Chen, H. D., Fei, L. J., Cheng, C., et al. (2018). Mapping human pluripotent stem cell differentiation pathways using high throughput single-cell RNA-sequencing. *Genome Biol.* 19:47. doi: 10.1186/s13059-018-1426-0
- Huang, M., Wang, J. S., Torre, E., Dueck, H., Shaffer, S., Bonasio, R., et al. (2018). SAVER: gene expression recovery for single-cell RNA sequencing. *Nat. Methods* 15, 539–542.
- Jean-Baptiste, K., McFaline-Figueroa, J. L., Alexandre, C. M., Dorrrity, M. W., Saunders, L., Bubb, K. L., et al. (2019). Dynamics of gene expression in single root cells of *Arabidopsis thaliana*. *Plant Cell* 31, 993–1011. doi: 10.1105/tpc.18.00785
- Jensen, M. K., Kjaersgaard, T., Nielsen, M. M., Galberg, P., Petersen, K., O’Shea, C., et al. (2010). The *Arabidopsis thaliana* NAC transcription factor family: structure-function relationships and determinants of ANAC019 stress signalling. *Biochem. J.* 426, 183–196. doi: 10.1042/BJ200991234
- Jensen, M. K., and Skriver, K. (2014). NAC transcription factor gene regulatory and protein-protein interaction networks in plant stress responses and senescence. *IUBMB Life* 66, 156–166. doi: 10.1002/iub.1256
- Kaufmann, K., Muiño, J. M., Østerås, M., Farinelli, L., Krajewski, P., and Angenent, G. C. (2010). Chromatin immunoprecipitation (ChIP) of plant transcription factors followed by sequencing (ChIP-SEQ) or hybridization to whole genome arrays (ChIP-CHIP). *Nat. Protoc.* 5, 457–472. doi: 10.1038/nprot.2009.244
- Kharchenko, P. V., Silberstein, L., and Scadden, D. T. (2014). Bayesian approach to single-cell differential expression analysis. *Nat. Methods* 11, 740–742. doi: 10.1038/nmeth.2967
- Klein, A. M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., et al. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* 161, 1187–1201. doi: 10.1016/j.cell.2015.04.044
- Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C., and Teichmann, S. A. (2015). The technology and biology of single-cell RNA sequencing. *Mol. Cell* 58, 610–620. doi: 10.1016/j.molcel.2015.04.005
- Kulkarni, A., Anderson, A. G., Merullo, D. P., and Konopka, G. (2019). Beyond bulk: a review of single cell transcriptomics methodologies and applications. *Curr. Opin. Biotechnol.* 58, 129–136. doi: 10.1016/j.copbio.2019.03.001
- Luo, Q., Han, W. W., Zhou, Y. H., Yao, Y., and Li, Z. S. (2008). The 3D structure of the defense-related rice protein Pir7b predicted by homology modeling and ligand binding studies. *J. Mol. Model.* 14, 559–569. doi: 10.1007/s00894-008-0310-3
- Luscombe, N. M., Austin, S. E., Berman, H. M., and Thornton, J. M. (2000). An overview of the structures of protein-DNA complexes. *Genome Biol.* 1, 1–37. doi: 10.1186/gb-2000-1-1-reviews001
- Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161, 1202–1214. doi: 10.1016/j.cell.2015.05.002
- Mukherjee, S., Berger, M. F., Jona, G., Wang, X. S., Muzzey, D., Snyder, M., et al. (2004). Rapid analysis of the DNA-binding specificities of transcription factors with DNA microarrays. *Nat. Genet.* 36, 1331–1339. doi: 10.1038/ng1473
- Nelson, H. C. (1995). Structure and function of DNA-binding proteins. *Curr. Opin. Genet. Dev.* 5, 180–189. doi: 10.1016/j.tca.2009.06.021
- Nuruzzaman, M., Manimekalai, R., Sharoni, A. M., Satoh, K., Kondoh, H., Ooka, H., et al. (2010). Genome-wide analysis of NAC transcription factor family in rice. *Gene* 465, 30–44. doi: 10.1016/j.gene.2010.06.008
- Nuruzzaman, M., Sharoni, A. M., and Kikuchi, S. (2013). Roles of NAC transcription factors in the regulation of biotic and abiotic stress responses in plants. *Front. Microbiol.* 4:248. doi: 10.3389/fmicb.2013.00248
- Olsen, T. K., and Baryawno, N. (2018). Introduction to single-cell RNA sequencing. *Curr. Protoc. Mol. Biol.* 122:e57. doi: 10.1002/cpmb.57
- Ooka, H., Satoh, K., Doi, K., Nagata, T., Otomo, Y., Murakami, K., et al. (2003). Comprehensive analysis of NAC family genes in *Oryza sativa* and *Arabidopsis thaliana*. *DNA Res.* 10, 239–247. doi: 10.1093/dnares/10.6.239
- Papalex, E., and Satija, R. (2018). Single-cell RNA sequencing to explore immune cell heterogeneity. *Nat. Rev. Immunol.* 18, 35–45. doi: 10.1038/nri.2017.76
- Perna, A., and Alberi, L. A. (2019). TF-ChIP method for tissue-specific gene targets. *Front. Cell. Neurosci.* 13:95. doi: 10.3389/fncel.2019.00095
- Ptashne, M. (2005). Regulation of transcription: from lambda to eukaryotes. *Trends Biochem. Sci.* 30, 275–279. doi: 10.1016/j.tibs.2005.04.003
- Ryu, K. H., Huang, L., Kang, H. M., and Schiefelbein, J. (2019). Single-cell RNA sequencing resolves molecular relationships among individual plant cells. *Plant Physiol.* 179, 1444–1456. doi: 10.1104/pp.18.01482
- Sharpe, R. M., Mahajan, A., Takacs, E. M., Stern, D. B., and Cahoon, A. B. (2011). Developmental and cell type characterization of bundle sheath and mesophyll chloroplast transcript abundance in maize. *Curr. Genet.* 57, 89–102. doi: 10.1007/s00294-010-0329-8
- Stark, R., Grzelak, M., and Hadfield, J. (2019). RNA sequencing: the teenage years. *Nat. Rev. Genet.* 20, 631–656. doi: 10.1038/s41576-019-0150-2
- Sun, L. J., Huang, L., Hong, Y. B., Zhang, H. J., Song, F. M., and Li, D. Y. (2015). Comprehensive analysis suggests overlapping expression of rice ONAC transcription factors in abiotic and biotic stress responses. *Int. J. Mol. Sci.* 16, 4306–4326. doi: 10.3390/ijms16024306
- Tran, L. S. P., Nakashima, K., Sakuma, Y., Simpson, S. D., Fujita, Y., Maruyama, K., et al. (2004). Isolation and functional analysis of *Arabidopsis* stress-inducible NAC transcription factors that bind to a drought-responsive cis-element in the *early responsive to dehydration stress 1* promoter. *Plant Cell* 16, 2481–2498. doi: 10.1105/tpc.104.022699
- Valouev, A., Johnson, D. S., Sundquist, A., Medina, C., Anton, E., Batzoglou, S., et al. (2008). Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. *Nat. Methods* 5, 829–834. doi: 10.1038/nmeth.1246
- van der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. *J. Mach. Learn. Res.* 9, 2579–2605. doi: 10.1007/s10846-008-9235-4
- Wagner, U., Edwards, R., Dixon, D. P., and Mauch, F. (2002). Probing the diversity of the *Arabidopsis* glutathione S-transferase gene family. *Plant Mol. Biol.* 49, 515–532. doi: 10.1023/a:1015557300450

- Welner, D. H., Lindemose, S., Grossmann, J. G., Møllegaard, N. E., Olsen, A. N., Helgstrand, C., et al. (2012). DNA binding by the plant-specific NAC transcription factors in crystal and solution: a firm link to WRKY and GCM transcription factors. *Biochem. J.* 444, 395–404. doi: 10.1042/BJ2011742
- Wu, Y. R., Deng, Z. Y., Lai, J. B., Zhang, Y. Y., Yang, C. P., Yin, B. J., et al. (2009). Dual function of *Arabidopsis* ATAF1 in abiotic and biotic stress responses. *Cell Res.* 19, 1279–1290. doi: 10.1038/cr.2009.108
- Yang, G. Z., Chao, D., Ming, Z. H., and Xia, J. X. (2019). A simple method to detect the inhibition of transcription factor-DNA binding due to protein-protein interactions in vivo. *Genes* 10:684. doi: 10.3390/genes10090684
- Yoo, S. D., Cho, Y. H., and Sheen, J. (2007). *Arabidopsis* mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat. Protoc.* 2, 1565–1572. doi: 10.1038/nprot.2007.199
- Zhang, T. Q., Xu, Z. G., Shang, G. D., and Wang, J. W. (2019). A single-cell RNA sequencing profiles the developmental landscape of *Arabidopsis* root. *Mol. Plant* 12, 648–660. doi: 10.1016/j.molp.2019.04.004
- Zhang, X. N., Li, T. Q., Liu, F., Chen, Y. Q., Yao, J. C., Li, Z. Y., et al. (2019). Comparative analysis of droplet-based ultra-high-throughput single-cell RNA-seq systems. *Mol. Cell* 73, 130–142. doi: 10.1016/j.molcel.2018.10.020
- Zhang, Y., Su, J. B., Duan, S., Ao, Y., Dai, J. R., Liu, J., et al. (2011). A highly efficient rice green tissue protoplast system for transient gene expression and studying light/chloroplast-related processes. *Plant Methods* 7:30.

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Xie, Jiang, Li, Yu, Wang, Luo, Cai, He, Xie, Zheng, Xie and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.