MF-PENS: An Efficient and Versatile Nuclear Extraction Method for Plant Single-Cell Sequencing

Lin Du[1,2,9], Xi Chen[2,9], Qing Liu[3,9], Mengyue Zhao[4,9], Fuyan Liu[2], Yali Song[2], Shuyuan Zhao[5], Xiaolin Sun[1,2], Cheng Tong[6], Mengjiao Zhang[7], Zhe Liang[6], Rouxi Chen[8,] *, Zhen Yue[8,] *, Jingmin Kang[2,8,] *

[1]College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China.
[2]BGI Research, Beijing 102601, China.
[3]State Key Laboratory of North China Crop Improvement and Regulation, College of Life Sciences, Hebei Agricultural University, Baoding 071000, China.
[4]State Key Laboratory of Molecular Developmental Biology, Institute of Genetics and Developmental Biology, Innovation Academy for Seed Design, Chinese Academy of Sciences, Beijing 100101, China
[5]Clin Lab, BGI Genomics, Tianjin 300308, China.
[6]Biotechnology Research Institute, Chinese Academy of Agricultural Sciences, Beijing, 100081, China
[7]College of Bioscience and Biotechnology, Yangzhou University, Yangzhou 225000，China
[8]BGI Research, Sanya 572025, China
[9]These authors contributed equally to this work.

*Correspondence authors: chenrouxi@genomics.cn (R.-X.C.), yuezhen@genomics.cn (Z.Y.), kangjingmin@genomics.cn (J.-M.K.)

Abstract

Single-cell sequencing is a crucial tool for exploring plant biology, yet the rigidity of plant cell walls and metabolite interference pose significant experimental challenges. To address this, we developed a novel nuclear preparation method, MF-PENS, which integrates methanol fixation with density gradient centrifugation. This method eliminates the need for flow cytometry, ensures nuclear integrity, prevents RNA leakage, and provides high-quality data for single-nucleus RNA sequencing (snRNA-seq) analysis. MF-PENS has been validated across various plant tissues, including fresh and frozen rice roots and leaves, as well as complex tissues such as soybean root nodules, soybean seeds, and tomato fruits. Experimental results demonstrate that MF-PENS effectively enhances the gene capture efficiency and cell clustering accuracy of snRNA-seq, providing reliable data support for downstream analysis. The method's broad compatibility and high efficiency provide a powerful impetus for plant single-cell research, advancing a deeper understanding of plant biology and developmental mechanisms.

Keywords: plant; single-cell sequencing; nuclear preparation; methanol fixation; density gradient centrifugation

Introduction

Cells are the basic units involved in organism growth and development. The development of complex tissues and organs results from the interactions of different cell types and their specific functions.[1] Numerous studies have found that cells of the same type in the same organ or tissue also present significant heterogeneity, and each cell has its own unique expression pattern.[2,3] Single-cell sequencing enables the sequencing of single-cell genomes, transcriptomes, and epigenomes. This technology can analyze the specificity of important rare cells, the heterogeneity of cells with the same phenotype, and the heterogeneity of the same genotype that cannot be solved by traditional sequencing technology.[4] Moreover, it can accurately identify the type and specific functions of cells as well as the changes in specific cells. It has been extensively applied in various fields such as tumor research, immunology research, stem cell research, neuroscience, infection research, and plant stress resistance studies. [5-9]

In the field of plant biology, single-cell sequencing technology initially focused primarily on foundational resource-oriented work, such as the construction of cellular atlases. For example, Liu *et al.* and Zhang *et al.* successfully mapped the single-cell atlas of rice root tips.[10,11] In recent years, this technology has gradually expanded into the study of plant stress response mechanisms: Liu *et al.* utilized single-cell technology to deeply dissect the molecular mechanisms of *Arabidopsis thaliana* in response to osmotic stress,[9] while Li *et al.* revealed the cellular response characteristics of cotton under salt stress conditions. Additionally, Swift *et al.* elucidated the key mechanisms

by which C4 plants evolved their highly efficient photosynthetic pathways using single-cell technology.[12] These studies mark a significant transition in plant single-cell sequencing technology from foundational atlas construction to functional mechanism research. However, the application of this technology in the plant field has significantly lagged behind that in the animal field.[13] One of the key factors contributing to this disparity lies in the unique biological characteristics of plant cells: they are encased within rigid cell walls and are rich in complex secondary metabolites such as polysaccharides and polyphenols. These structural and chemical properties collectively impose significant technical challenges for isolating high-quality protoplasts or nuclei from plant tissues that meet the requirements for single-cell sequencing.[14]

For protoplast preparation, it typically requires several hours of enzymatic digestion to release protoplasts,[15-17] and this process varies significantly depending on the species and tissue type. As a result, specific protoplast isolation protocols need to be developed for each plant species, making it nearly impossible to establish a universal standardized workflow. During the digestion process, cells remain active, and prolonged exposure to enzymatic conditions may not only induce stress-related transcriptional artifacts but also lead to the selective loss of rare or enzyme-sensitive cell populations, thereby compromising the accuracy of biological representation. Additionally, the reliance on fresh plant tissue for protoplast preparation imposes immediate processing constraints, limiting experimental flexibility and hindering retrospective studies on archived materials.

To overcome the limitations of protoplast-based methods, single-nucleus RNA sequencing (snRNA-seq) has emerged as a promising alternative. Farmer *et al.* demonstrated the reliability of snRNA-seq in plant research using Arabidopsis root tissues and revealed that snRNA-seq can identify additional cell subtypes compared to single-cell RNA sequencing (scRNA-Seq).[18] However, obtaining plant nuclei that meet the quality requirements for snRNA-seq remains a significant challenge. Existing solutions, such as those employed in studies on Arabidopsis and wheat roots,[18,19] involve crude nuclear extraction followed by purification using flow cytometry, ultimately enabling snRNA-seq. However, these methods have significant limitations: as a core instrument for ensuring purity, flow cytometers are not universally available in all laboratories, creating a high technical barrier for many researchers. Although some studies have attempted to conduct snRNA-Seq without relying on flow cytometers,[20] during the nuclear extraction process, the nuclear membrane is prone to damage, leading to RNA leakage. This manifests as a reduced number of detected cells (potentially due to the low number of genes detected per nucleus, which may be mistaken for cellular debris) or directly as a low average number of genes per nucleus. These technical constraints collectively hinder the progress of plant single-cell omics research.

In response to the critical challenges faced by current nuclear isolation techniques in plant snRNA-Seq research—including low gene capture efficiency due to insufficient nuclear integrity and strong reliance on high-end equipment such as flow cytometers—this study innovatively developed a novel plant nuclear preparation

method based on methanol fixation combined with density gradient centrifugation. This method enables the acquisition of high-integrity nuclei in a cost-effective and efficient manner, providing high-quality and stable data sources for snRNA analysis. This technological breakthrough not only significantly enhances the operability of experiments and the reliability of results but also lays a solid foundation for the widespread application of plant single-cell research. It will strongly empower global scientists to fully exploit the immense potential of single-cell sequencing technology in plant biology and agricultural applications.

Results

Establishing MF-PENS for plant single-cell sequencing

Our goal is to develop a nuclei extraction method suitable for frozen samples, featuring simplicity in operation, independence from flow cytometry, and the ability to ensure nuclear integrity and prevent RNA leakage for plant single-cell sequencing. To achieve this, we adopted the following strategies: for the suitability of frozen samples, we chose a nuclei extraction protocol; for operational simplicity and the elimination of flow cytometry dependence, we introduced Percoll Enhanced Nuclei Sorting (PENS) technology for impurity removal; and for nuclear integrity and RNA stability, we incorporated a tissue fixation protocol. To identify the optimal fixation condition, we tested 0.1% formaldehyde fixation (0.1%FF), 1% formaldehyde fixation (1%FF), 100% methanol fixation (MF), and a no-treatment control (NT), and combined these four fixation protocols with the PENS method, naming them 0.1%FF-PENS, 1%FF-PENS, MF-PENS, and NT-PENS, respectively. In the 0.1%FF-PENS, 1%FF-PENS, MF-PENS, and NT-PENS protocols, plant tissues were first cut into approximately 0.5 cm^2 pieces, the 0.1%FF-PENS, 1%FF-PENS, and MF-PENS groups were fixed with 0.1% formaldehyde, 1% formaldehyde, and 100% methanol, respectively, while the NT-PENS group proceeded directly to the next step. The fixed/non-fixed tissues underwent crude nuclei extraction, followed by purification of the crude nuclei using Percoll density gradient centrifugation. The purified nuclei were then labeled via microfluidic technology, and finally, library preparation and sequencing were completed **(Figure 1a)**.

MF-PENS enables high-quality single-nucleus RNA sequencing

To evaluate the impact of different fixation methods on the quality of snRNA-seq data, we systematically compared four treatment approaches in rice (*Oryza sativa*) root tissues **(Figure 1a)**: NT-PENS, 0.1%FF-PENS, 1%FF-PENS, and MF-PENS. Each method was performed with two biological replicates to ensure the reliability of experimental results. During data preprocessing, we implemented stringent quality control criteria, including the removal of doublets/multiplets, unbound or apoptotic cells, cells with high mitochondrial gene expression, and low-quality cells affected by enzymatic digestion, ultimately obtaining high-quality transcriptomic data.

Quality control analysis revealed that the MF-PENS method demonstrated significant advantages in several key metrics. Violin plots **(Figures 1b-c)** showed that

compared to the other three methods, MF-PENS produced significantly higher UMI counts and detected gene numbers, with high consistency between biological replicates. To explore the potential organelle residues that may occur due to tissue fixation, we quantified the percentage of chloroplast and mitochondrial genes across different methods **(Figure 1d)**. The results showed that all methods maintained low levels without significant differences, indicating that none of the three fixation methods led to the aggregation of organelles and nuclei, thereby resulting in organelle residues, confirming the reliability of the experimental data.

Pearson correlation analysis **(Figure 1e)** revealed high overall similarity among data generated by the four methods. Specifically, 0.1%FF-PENS showed the highest correlation with NT-PENS, while MF-PENS exhibited some divergence from other methods. This difference may stem from MF-PENS's unique advantage in capturing high-quality, high gene-expression cells, enabling it to acquire more biologically meaningful transcriptomic information.

MF-PENS provides higher precision for cell clustering in plant single-cell transcriptome data

In single-cell sequencing analysis, although UMI counts and gene detection numbers are fundamental metrics for evaluating data quality, cell clustering, and cluster annotation are the core steps for deciphering biological heterogeneity. To systematically evaluate the accuracy of cell clustering in the data generated by the four methods, we conducted an integrated analysis. Under the condition of uniformly setting nine cell clusters, both MF-PENS and 1%FF-PENS methods demonstrated excellent clustering performance, yielding well-defined and highly specific clusters. In contrast, NT-PENS and 0.1%FF-PENS methods exhibited noticeable cell aggregation issues **(Figure 2a)**.

To quantitatively assess clustering quality, we employed the graph-based Modularity metric,[21] which evaluates clustering performance by measuring the difference between intra-cluster and inter-cluster connection density, with values ranging from -1 to 1 (values closer to 1 indicate better clustering). Quantitative analysis revealed Modularity values of $-8.08e^{-5}$, $-1.09e^{-4}$, and $-4.94e^{-4}$ for 1%FF-PENS, NT-PENS, and 0.1%FF-PENS, respectively, while MF-PENS achieved the optimal Modularity value ($-5.07e^{-5}$), demonstrating its superior clustering performance. Furthermore, UMAP visualization showed no significant batch effects between biological replicates **(Figure 2b)**, confirming the stability and reproducibility of these methods.

In differential expression analysis, bubble plots of the top 10 upregulated differentially expressed genes (DEGs) identified by the four methods **(Figure 2c)** revealed that MF-PENS detected the most significant and specific DEGs. Further statistical analysis demonstrated that MF-PENS significantly outperformed the other three methods in the number of significantly upregulated genes (up-DEGs, $P_adj < 0.05$, $\log_2FC > 0.25$), significantly downregulated genes (down-DEGs, $P_adj < 0.05$, $\log_2FC < -0.25$), and total significant DEGs (All-DEGs, $P_adj < 0.05$, $|\log_2FC| > 0.25$) **(Figure 2d)**.

To validate the biological reliability of the identified DEGs, we constructed Venn diagrams of the top 100 upregulated DEGs from each method. The results showed that NT-PENS, 0.1%FF-PENS, 1%FF-PENS, and MF-PENS specifically identified 26, 54, 30, and 61 up-DEGs, respectively **(Figure 2e)**. Comparison with experimentally validated cell-type-specific marker genes from the PlantCellMarker database[22] revealed that 21.31% (13/61) of DEGs identified by MF-PENS matched known markers, significantly higher than NT-PENS (3.85%, 1/26), 0.1%FF-PENS (3.70%, 2/54), and 1%FF-PENS (13.33%, 4/30). These results demonstrate that MF-PENS not only identifies a greater number of DEGs but also achieves higher accuracy and biological relevance, highlighting its superior performance in downstream bioinformatics analysis.

MF-PENS is compatible with cryopreserved plant tissues

Through systematic comparison of four nuclear extraction methods, we have demonstrated that MF-PENS outperforms the others. Given that the applicability to frozen samples is a critical factor limiting experimental flexibility and an essential pathway for repurposing sample repository resources, we further evaluated the suitability of MF-PENS in processing frozen samples. We selected rice roots as representative of plant underground tissues and rice leaves as representative of aboveground tissues, storing them on dry ice for 15 days followed by transfer to a -80°C freezer for an additional 15 days to simulate the actual scenarios of sample transportation and storage. Subsequently, we performed nuclear extraction using the MF-PENS technique on these two frozen samples and their corresponding fresh samples, followed by single-cell sequencing analysis **(Figures 5a and 5f)**. Each sample type was processed with two biological replicates and subjected to stringent quality control, yielding high-quality snRNA-seq data. In the root tissue analysis, we observed remarkable consistency between fresh and frozen samples in key quality metrics, including UMI counts, detected gene numbers, and mitochondrial gene percentages **(Figure 5b)**. These results demonstrate that the MF-PENS method can effectively extract high-quality transcriptomic information from root samples under different storage conditions.

Cell clustering analysis further confirmed the reliability of the data. UMAP visualization revealed distinct and independent cell cluster distributions across samples **(Figure 5c)**, while batch effect analysis indicated the absence of significant technical biases between samples **(Figures 5d-e)**. These characteristics ensure the suitability of the data for various downstream bioinformatics analyses, including cell type annotation, differential expression gene identification, and cell-cell interaction analysis.

In parallel experiments with leaf tissues, we observed similar results to those in roots. Fresh and frozen leaf samples showed high consistency in key metrics such as UMI counts, gene numbers, and mitochondrial gene percentages **(Figure 5g)**. Cell clustering results displayed well-separated distributions **(Figure 5h)**, with no significant batch effects detected between samples **(Figures 5i-j)**. These findings further confirm the stability and reliability of the MF-PENS method across different plant tissue types.

In summary, the MF-PENS method demonstrates exceptional sample adaptability, enabling efficient nuclear extraction from both fresh and frozen samples of conventional plant tissues (root/leaf) and yielding high-quality snRNA-seq data for reliable downstream analysis. This characteristic establishes MF-PENS as a valuable experimental approach in plant single-cell transcriptomics research.

MF-PENS possesses outstanding compatibility with various plant species and tissues

In plants, tissues with high economic and research value often contain abundant metabolites, the presence of which makes it particularly challenging to extract high-quality nuclei, thereby limiting the feasibility of conducting snRNA-Seq studies. To comprehensively evaluate the broad applicability of MF-PENS for challenging sample types, we selected four representative complex plant tissues for single-nucleus RNA sequencing analysis **(Figure 4a)**: infected tissue (soybean root nodules), oil-rich tissue (soybean seeds), polysaccharide-rich tissue (tomato fruit), and less differentiated tissue (more fragile cells, maize callus). Two biological replicates were performed for each tissue type, and after stringent quality control, we found that MF-PENS successfully obtained 17,565, 17,329, 17,318, and 12,725 high-quality cells, respectively **(Figure 4b)**, from these complex tissues.

Sequencing data analysis revealed that the MF-PENS method demonstrated excellent performance across tissues with varying levels of complexity. Specifically, soybean seed samples showed an average of 3,815 genes and 7,446 UMI counts per nucleus. Soybean root nodule samples achieved an average of 2,136 genes and 3,128 UMI counts per nucleus, representing a significant improvement over previously reported soybean nodule snRNA-seq data (average genes per nucleus: 1,026 at 3-dpi, 1,607 at 5-dpi, 1,487 at 12-dpi, and 1,240 at 20-dpi; average UMI counts per nucleus: 1,873 at 3-dpi, 2,165 at 5-dpi, 2,082 at 12-dpi, and 1,598 at 20-dpi).[23,24] Tomato fruit samples exhibited an average of 2,349 genes and 5,137 UMI counts per nucleus, while maize endosperm samples demonstrated the highest sequencing efficiency, with an average of 4,535 genes and 8,724 UMI counts per nucleus **(Figure 4c)**.

Quality control analysis further validated the reliability of the data. Mitochondrial gene percentages remained consistently low across all eight samples **(Figure 4d)**, and biological replicates of the same tissue type showed high reproducibility **(Figure 4e)**. These results highlight the capability of MF-PENS to overcome the limitations of traditional single-cell sequencing technologies when processing complex plant tissues, providing robust and high-quality data for advanced biological studies.

In summary, MF-PENS proves to be a versatile and powerful tool for snRNA-Seq, enabling researchers to obtain high-quality data from a wide range of challenging plant tissues, thereby facilitating deeper insights into plant biology and development.

Discussion

In the field of plant single-cell sequencing, research efforts lag significantly behind

those in animal studies, primarily due to the challenges associated with plant cell and nucleus preparation, which have limited the widespread application of related technologies. To address this challenge, we propose an innovative solution: using methanol to fix tissues for stabilizing nuclear structures and employing Percoll density gradient centrifugation to replace traditional flow cytometry for nucleus purification. We have named this method MF-PNES **(Figure 1a)**.

To validate the reliability and applicability of MF-PENS, we applied this method to generate snRNA-seq data from both fresh and frozen rice root and leaf samples. Through systematic comparison, the results demonstrated that MF-PENS is capable of producing high-quality snRNA data for conventional plant samples **(Figures 3b and 3g)**, and the cell clustering based on these data was significantly improved **(Figures 3c and 3h)**. Furthermore, another notable advantage of MF-PENS lies in its compatibility with frozen plant tissues **(Figures 3b, 3d, 3g, and 3i)**, which not only makes snRNA-Seq studies feasible for samples collected from remote areas that cannot be processed immediately but also provides researchers with technical support for retrospective analysis of valuable samples stored in biobanks.

We further conducted an in-depth evaluation of the applicability and efficacy of the MF-PENS method in processing samples with complex compositions, including those rich in polysaccharides, polyphenols, lipids, and symbiotic microorganisms. By performing snRNA-seq on samples such as tomato fruits, soybean seeds, soybean root nodules, and maize callus **(Figure 4a)**, and systematically assessing the quality of the generated datasets **(Figures 4b-e)**, the results demonstrated that even when confronted with these highly challenging sample types, MF-PENS was able to achieve data quality comparable to that of conventional samples reported in the literature. This finding not only confirms the exceptional performance of MF-PENS in handling complex biological samples but also highlights its broad applicability.

In summary, MF-PENS is a low-barrier, highly versatile technique suitable for various plant sample types and conditions. We predict that the widespread application of this technology will further advance plant single-cell research, facilitating a shift from tissue-level understanding to the elucidation of cell type-specific regulatory mechanisms. This will open new avenues for future plant biology research.

Methods

Nucleus isolation

Plant samples were cut into ~0.5 square centimeter fragments using a gillette blade, followed by fixation in prechilled 100% methanol on ice for 30 minutes. The methanol was then discarded, and the tissue was washed twice with NIBTA buffer (10 mM Tris-HCl (pH 8.0), 10 mM EDTA (pH 8.0), 80 mM KCl, 1 mM Spermine-4HCl, 1 mM Spermidine-3HCl, 2% PVP, 0.3% Triton X-100, 1 mM DTT, 1x ProtectRNA™ RNase Inhibitor, 1x Protease Inhibitor Cocktail). The tissue was transferred to a petri dish and chopped with a Gillette razor blade on ice. The chopped tissue was resuspended in

NIBTA buffer and homogenized on ice for 5 minutes. The homogenate was then filtered through a 40 μm strainer, and the filtrate was collected into a new 50 mL centrifuge tube. The tube was centrifuged at 1,260 g for 10 minutes at 4℃. The supernatant was discarded, and the tissue pellet was resuspended in 4 mL of NIBTA buffer. In a new 15 mL centrifuge tube, 70% Percoll solution (3.5 mL Percoll and 1.5 mL NIBTA buffer) was added, and the 4 mL resuspended tissue homogenate was carefully overlaid onto the 70% Percoll layer. The tube was centrifuged at 650 g for 30 minutes at 4℃, during which most nuclei banded at the interface between the NIBTA buffer and the Percoll layer. The nuclei band was gently collected into a new 15 mL centrifuge tube, 10 mL of NIBTA buffer was added, and the tube was centrifuged again at 1,260 g for 5 minutes. The nuclei pellet was washed twice with NIBTA buffer, then resuspended in PBS containing 0.04% BSA. After counting the nuclei, the suspension was kept on ice for later use (it is recommended not to exceed 30 minutes).

Single-nucleus RNA library construction and Sequencing

The single-nucleus RNA-seq libraries were prepared as previously described with DNBelab C Series High-throughput Single-Cell RNA Library Preparation Set V2.0 (MGI, 940-000508-00) and Sequencing was performed using DNBSEQ T7.

Preprocessing of snRNA-seq data

The snRNA-Seq data were processed and analyzed using DNBelab_C_Series_HT_single cell-analysis-software (version 2.0),[25] an open-source and flexible pipeline. Sequencing reads were aligned to respective reference genomes: rice reads to MSU v7.0,[26] soybean reads to GCF_000004515.6,[27] tomato reads to GCF_036512215.1,[28] and maize reads to GCF_902167145.1.[29] Stringent quality control measures were implemented across all samples. Potential doublets or multiplets were identified and filtered using the DoubletFinder package,[30] followed by the removal of unbound or apoptotic cells. We established strict inclusion criteria for cell quality: (1) cells expressing fewer than 200 or more than 6,000 genes were excluded; (2) cells with UMI counts below 500 or exceeding 40,000 were removed; and (3) cells demonstrating mitochondrial gene expression exceeding 5% of total expression were discarded. These rigorous quality control steps ensured the retention of high-quality cells for subsequent bioinformatics analysis.

Cell clustering and visualization

We processed the snRNA-seq data using the Seurat v4.3.2 package.[31] To normalize gene expression across cells, we applied the “NormalizeData” function, which utilizes a global-scaling normalization method. This approach involves dividing the gene expression values by the total expression per cell, scaling the results by a factor of 10,000 (default), and applying a natural log transformation to stabilize the variance. To capture the most biologically relevant transcriptional variations, we identified the top 2,000 highly variable genes using the “FindVariableFeatures” function. To ensure robust integration of datasets and minimize batch effects, we employed Canonical Correlation Analysis (CCA) through the “FindIntegrationAnchors” and “IntegrateData”

functions, which align shared biological features across samples. For dimensionality reduction, we performed principal component analysis (PCA) using the “RunPCA” function (npcs = 40) to extract the most informative components of variation. Subsequently, we applied graph-based clustering with the Louvain algorithm via the “FindClusters” function (resolution = 0.3) to group cells based on their transcriptional profiles. Finally, to visualize the high-dimensional data in an interpretable two-dimensional space, we utilized Uniform Manifold Approximation and Projection (UMAP) with the “RunUMAP” function (dims = 1:40, min.dist = 0.3). This approach preserves the local and global structure of the data, enabling clear visualization of distinct cell clusters and their relationships.

Acknowledgments

We thank Dr. Xuelei Lin (Institute of Genetics and Developmental Biology, Chinese Academy of Sciences) for providing the B73, Williams 82, and Zhonghuang 13 materials; Assoc. Res. Fang Zhang (Institute of Genetics and Developmental Biology, Chinese Academy of Sciences) for providing the Zhonghua 11 materials; and Prof. Tao Lin (College of Horticulture, China Agricultural University) for providing the Alisa Craig materials. Additionally, we sincerely appreciate the support provided by the China National GeneBank (CNGB).

Figures

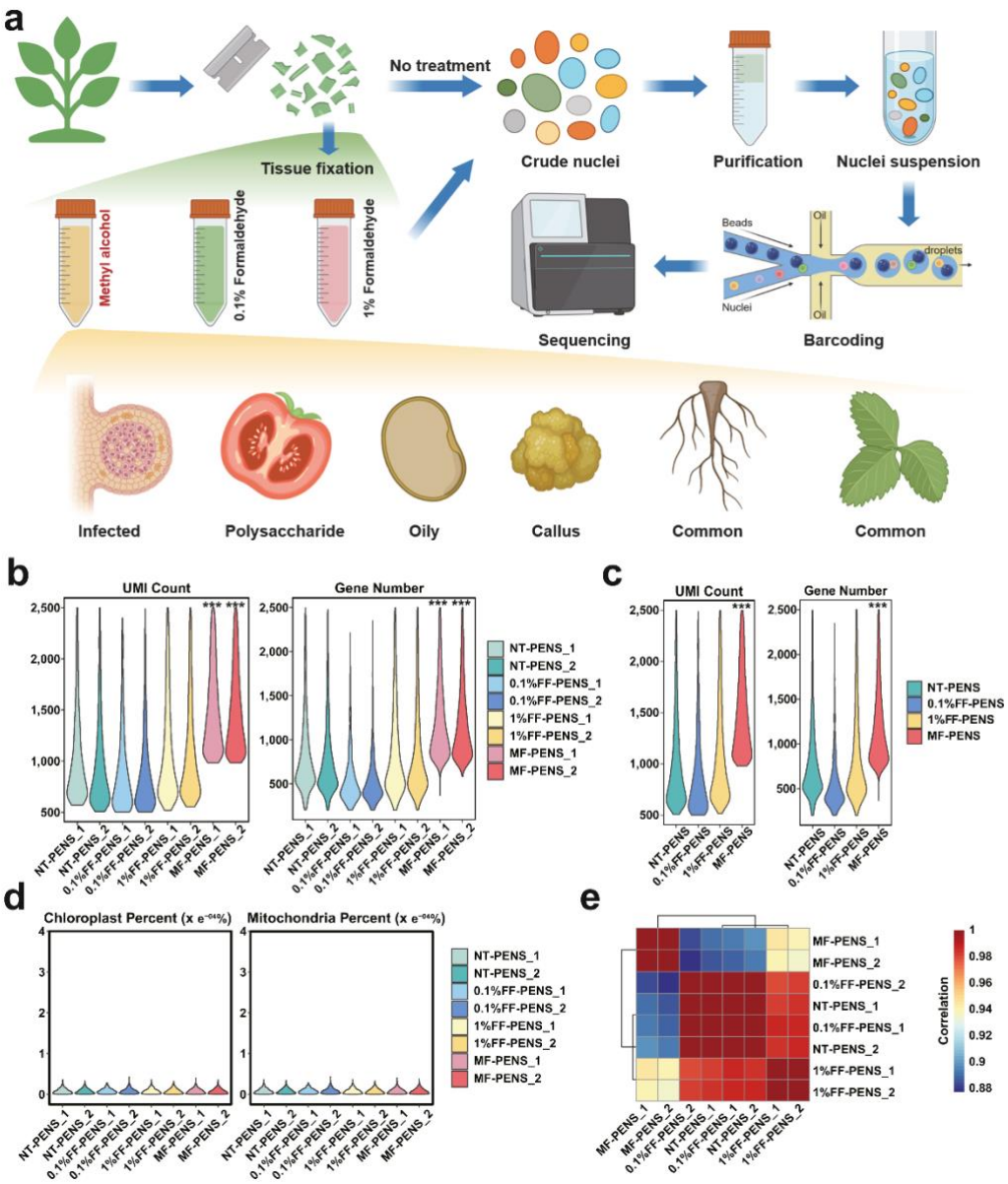

Figure1. Design principle and verification of optimal performance of MF-PENS on rice roots.

a. Pipeline and design principle of MF-PENS, which is suitable for various plant species. **b.** Violin plots showing UMI counts and gene numbers for eight biological samples across four methods: NT-PENS, 0.1%FF-PENS, 1%FF-PENS, and MF-PENS. Each method includes two replicates. **c.** Violin plots displaying UMI counts and gene numbers for four distinct methods. *** in **b-c** indicates that the sample data from the MF-PENS method are significantly higher ($P < 0.05$) than those from the other methods. **d.** Mitochondrial and chloroplast percentages (y-axis values × 10^{-4}%) for the eight biological samples corresponding to the four methods. **e.** Pearson correlation heatmap of the eight samples, with colors ranging from blue to red indicating correlation coefficients increasing from 0.88 to 1.0. Hierarchical clustering was applied to the

results.

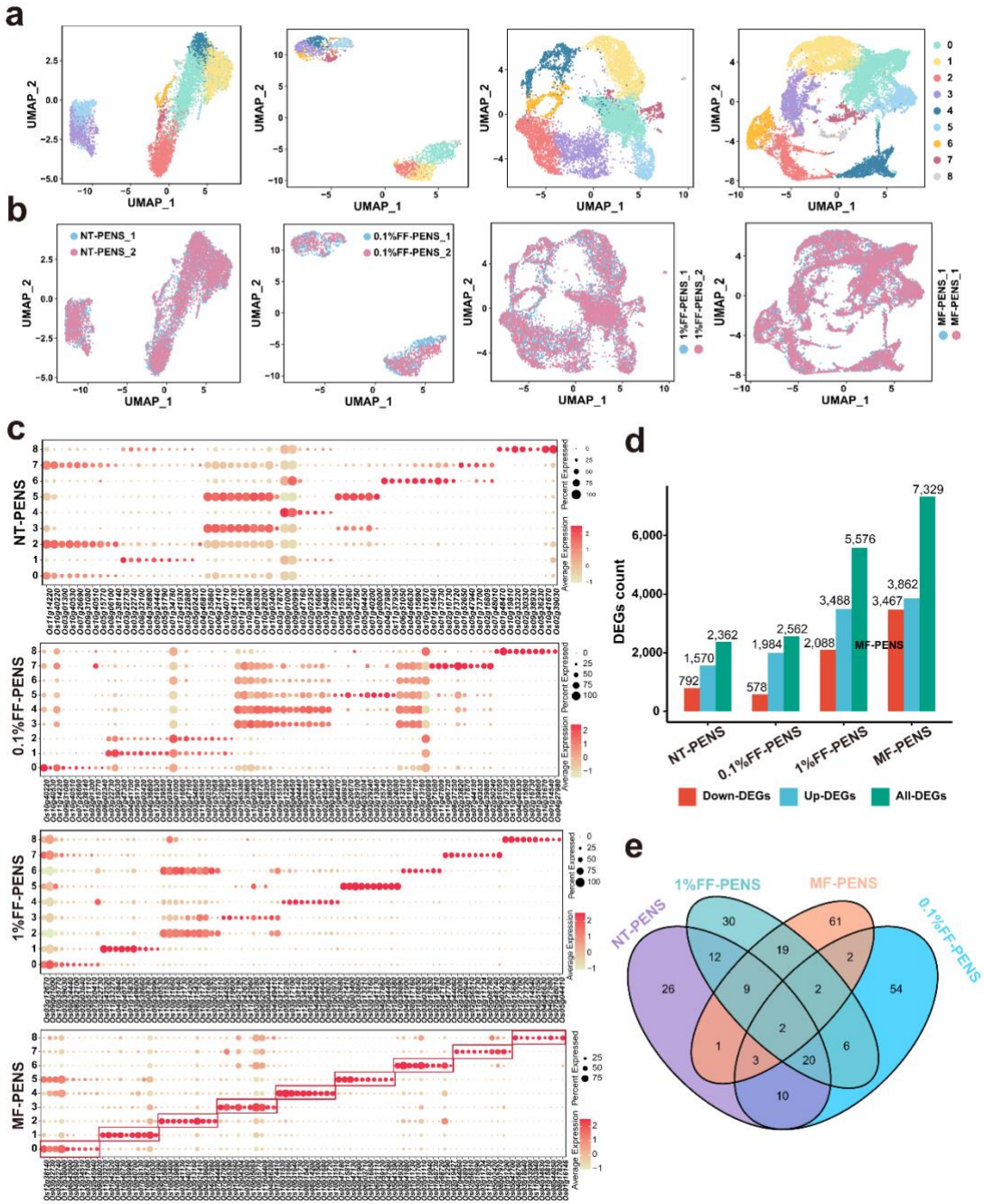

Figure 2. MF-PENS-generated snRNA-seq data enables superior cell clustering performance.

a. UMAP plots showing Seurat clustering results for nine clusters derived from snRNA-seq data generated by four methods (from left to right): NT-PENS, 0.1%FF-PENS, 1%FF-PENS, and MF-PENS. **b.** UMAP plots of two biological replicates for each of the four methods, based on the clustering shown in **(a)**. **c.** Bubble plot displaying the top 10 upregulated differentially expressed genes (DEGs) for each cluster in the cell clustering results of the four methods. **d.** Bar plot illustrating the number of upregulated DEGs (up-DEGs, $P_adj < 0.05$, $\log_2FC > 0.25$), downregulated DEGs (down-DEGs,

P_adj < 0.05, $\log_2$FC < -0.25), and significant DEGs (All-DEGs, *P_adj* < 0.05, |$\log_2$FC| > 0.25) identified by the four methods. **e.** Venn diagram showing the overlap of the top 100 up-DEGs identified by the four methods.

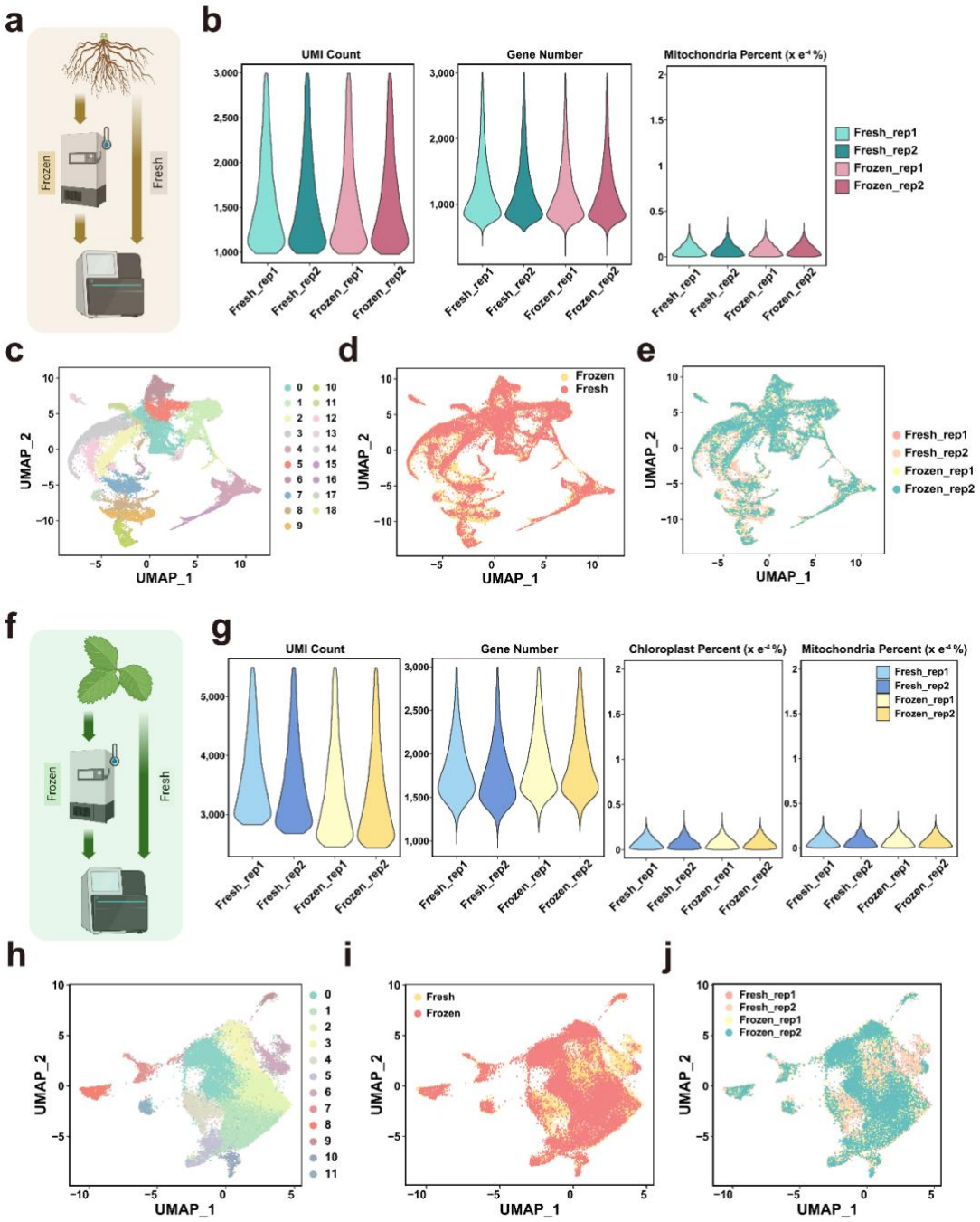

Figure 3. MF-PENS is suitable for both fresh and cryopreserved plant samples. **a.** Schematic workflow of snRNA-seq using the MF-PENS method for fresh and cryopreserved rice root tissue samples. **b.** Violin plots showing UMI counts, gene numbers, and mitochondrial percentages (y-axis × 10^{-4}) for four samples, including two replicates each of fresh and frozen methods. **c.** UMAP plot of integrated cell clustering

results for the root samples. **d-e.** Distribution of cell types from fresh and frozen samples and cells from the four individual samples on **(c)**. **f.** Schematic workflow of snRNA-seq using the MF-PENS method for fresh and cryopreserved rice leaf tissue samples. **g.** Violin plots showing UMI counts, gene numbers, mitochondrial percentages (y-axis × 10^{-4}%), and chloroplast percentages (y-axis × 10^{-4}%) for four samples, including two replicates of fresh and frozen methods. **h.** UMAP plot of integrated cell clustering results for the leaf samples. **i-j.** Distribution of cell types from fresh and frozen samples and cells from the four individual samples on **(f)**.

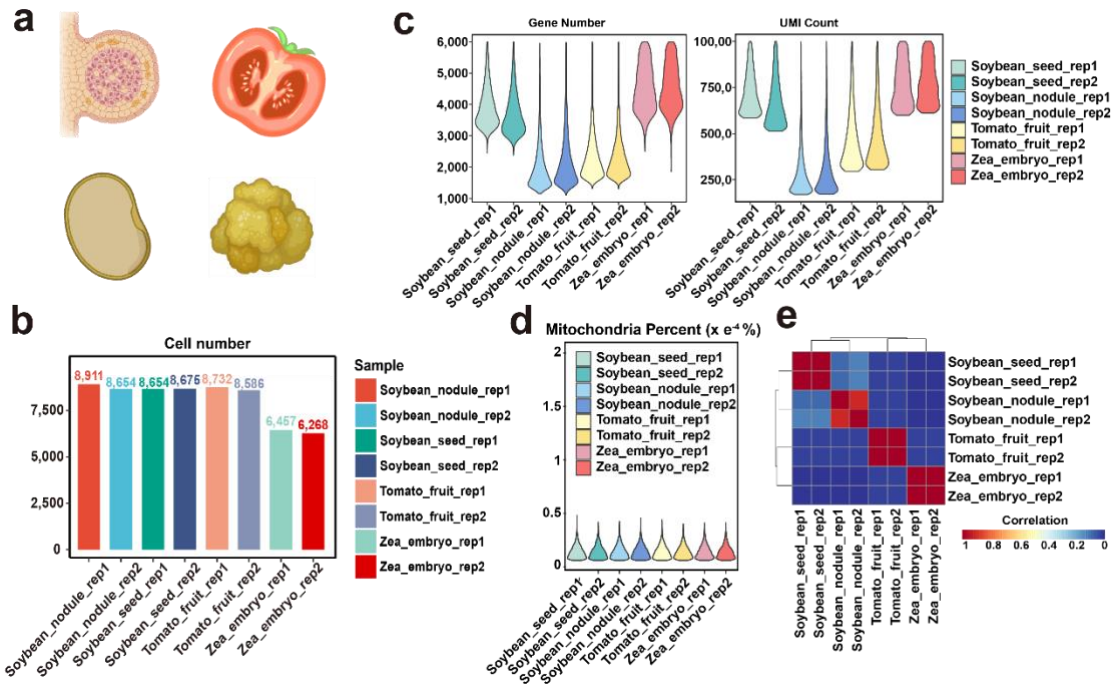

Figure 4. Universal applicability across plant species and tissues for MF-PENS. **a.** Application of MF-PENS for snRNA-seq in diverse tissue types, including infected tissue (soybean root nodules), polysaccharide-rich tissue (tomato fruit), lipid-rich tissue (soybean seeds), and callus tissue (maize callus). **b.** Cell number of eight samples **c.** Violin plots showing gene numbers and UMI counts for eight biological samples, including two replicates for each of the four tissue types. **d.** Mitochondrial percentages (y-axis × 10^{-4}%) for the eight biological samples. **e.** Pearson correlation heatmap of the eight biological samples.

REFFERENCE

1 Birnbaum, K. *et al.* A gene expression map of the Arabidopsis root. *Science* **302**, 1956-1960 (2003). https://doi.org/10.1126/science.1090022

2 Zhao, J. *et al.* Single-cell RNA sequencing reveals the heterogeneity of liver-resident immune cells in human. *Cell Discov* **6**, 22 (2020). https://doi.org/10.1038/s41421-020-0157-z

3 Tsagiopoulou, M., Rashmi, S., Aguilar-Fernandez, S., Nieto, J. & Gut, I. G. Multi-organ single-cell transcriptomics of immune cells uncovered organ-specific gene expression and functions. *Sci Data* **11**, 316 (2024). https://doi.org/10.1038/s41597-024-03152-z

4 Jovic, D. *et al.* Single-cell RNA sequencing technologies and applications: A brief overview. *Clin Transl Med* **12**, e694 (2022). https://doi.org/10.1002/ctm2.694

5 Bian, S. *et al.* Single-cell multiomics sequencing and analyses of human colorectal cancer. *Science* **362**, 1060-1063 (2018). https://doi.org/10.1126/science.aao3791

6 Terekhova, M. *et al.* Single-cell atlas of healthy human blood unveils age-related loss of NKG2C(+)GZMB(-)CD8(+) memory T cells and accumulation of type 2 memory T cells. *Immunity* **56**, 2836-2854.e2839 (2023). https://doi.org/10.1016/j.immuni.2023.10.013

7 Mazid, M. A. *et al.* Rolling back human pluripotent stem cells to an eight-cell embryo-like stage. *Nature* **605**, 315-324 (2022). https://doi.org/10.1038/s41586-022-04625-0

8 Ochocka, N. *et al.* Single-cell RNA sequencing reveals functional heterogeneity of glioma-associated brain macrophages. *Nat Commun* **12**, 1151 (2021). https://doi.org/10.1038/s41467-021-21407-w

9 Liu, Q. *et al.* Multiome in the Same Cell Reveals the Impact of Osmotic Stress on Arabidopsis Root Tip Development at Single-Cell Level. *Adv Sci (Weinh)* **11**, e2308384 (2024). https://doi.org/10.1002/advs.202308384

10 Liu, Q. *et al.* Transcriptional landscape of rice roots at the single-cell resolution. *Mol Plant* **14**, 384-394 (2021). https://doi.org/10.1016/j.molp.2020.12.014

11 Zhang, T. Q., Chen, Y., Liu, Y., Lin, W. H. & Wang, J. W. Single-cell transcriptome atlas and chromatin accessibility landscape reveal differentiation trajectories in the rice root. *Nat Commun* **12**, 2053 (2021). https://doi.org/10.1038/s41467-021-22352-4

12 Swift, J. *et al.* Exaptation of ancestral cell-identity networks enables C(4) photosynthesis. *Nature* **636**, 143-150 (2024). https://doi.org/10.1038/s41586-024-08204-3

13 Singh, S., Praveen, A., Dudha, N., Sharma, V. K. & Bhadrecha, P. Single-cell transcriptomics: a new frontier in plant biotechnology research. *Plant Cell Rep* **43**, 294 (2024). https://doi.org/10.1007/s00299-024-03383-9

14 Sun, Y. *et al.* Single-Cell Transcriptomics Applied in Plants. *Cells* **13** (2024). https://doi.org/10.3390/cells13181561

15 Yang, P., Sun, Y., Sun, X., Li, Y. & Wang, L. Optimization of preparation and transformation of protoplasts from Populus simonii × P. nigra leaves and subcellular localization of the major latex protein 328 (MLP328). *Plant Methods* **20**, 3 (2024). https://doi.org/10.1186/s13007-023-01128-5

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

17 Li, C., Zhang, S., Yan, X., Cheng, P. & Yu, H. Single-nucleus sequencing deciphers

developmental trajectories in rice pistils. *Dev Cell* **58**, 694-708.e694 (2023). https://doi.org/10.1016/j.devcel.2023.03.004

18 Farmer, A., Thibivilliers, S., Ryu, K. H., Schiefelbein, J. & Libault, M. Single-nucleus RNA and ATAC sequencing reveals the impact of chromatin accessibility on gene expression in Arabidopsis roots at the single-cell level. *Mol Plant* **14**, 372-383 (2021). https://doi.org/10.1016/j.molp.2021.01.001

19 Zhang, L. *et al.* Asymmetric gene expression and cell-type-specific regulatory networks in the root of bread wheat revealed by single-cell multiomics analysis. *Genome Biol* **24**, 65 (2023). https://doi.org/10.1186/s13059-023-02908-x

20 Wang, K. *et al.* An optimized FACS-free single-nucleus RNA sequencing (snRNA-seq) method for plant science research. *Plant Sci* **326**, 111535 (2023). https://doi.org/10.1016/j.plantsci.2022.111535

21 Alcalá-Corona, S. A., Sandoval-Motta, S., Espinal-Enríquez, J. & Hernández-Lemus, E. Modularity in Biological Networks. *Front Genet* **12**, 701331 (2021). https://doi.org/10.3389/fgene.2021.701331

22 Jin, J. *et al.* PCMDB: a curated and comprehensive resource of plant cell markers. *Nucleic Acids Res* **50**, D1448-d1455 (2022). https://doi.org/10.1093/nar/gkab949

23 Liu, Z. *et al.* Integrated single-nucleus and spatial transcriptomics captures transitional states in soybean nodule maturation. *Nat Plants* **9**, 515-524 (2023). https://doi.org/10.1038/s41477-023-01387-z

24 Fan, J. *et al.* A large-scale integrated transcriptomic atlas for soybean organ development. *Molecular Plant* (2025). https://doi.org/https://doi.org/10.1016/j.molp.2025.02.003

25 Huang, Y. *et al.* Protocol for optimized nasal mucosa sample processing to obtain high-quality scRNA-seq and scATAC-seq data. *STAR Protoc* **5**, 103298 (2024). https://doi.org/10.1016/j.xpro.2024.103298

26 Kawahara, Y. *et al.* Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map data. *Rice (N Y)* **6**, 4 (2013). https://doi.org/10.1186/1939-8433-6-4

27 Schmutz, J. *et al.* Genome sequence of the palaeopolyploid soybean. *Nature* **463**, 178-183 (2010). https://doi.org/10.1038/nature08670

28 The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* **485**, 635-641 (2012). https://doi.org/10.1038/nature11119

29 Hayford, R. K. *et al.* Functional annotation and meta-analysis of maize transcriptomes reveal genes involved in biotic and abiotic stress. *BMC Genomics* **25**, 533 (2024). https://doi.org/10.1186/s12864-024-10443-7

30 McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst* **8**, 329-337.e324 (2019). https://doi.org/10.1016/j.cels.2019.03.003

31 Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411-420 (2018). https://doi.org/10.1038/nbt.4096