Plant Communications
Resource article

CellPress
Partner Journal

scPlant: A versatile framework for single-cell transcriptomic data analysis in plants

Shanni Cao[1], Zhaohui He[1], Ruidong Chen[1], Yuting Luo[1], Liang-Yu Fu[1], Xinkai Zhou[1], Chao He[2], Wenhao Yan[2], Chen-Yu Zhang[1,*] and Dijun Chen[1,*]

[1]State Key Laboratory of Pharmaceutical Biotechnology, School of Life Sciences, Nanjing University, Nanjing 210023, China

[2]National Key Laboratory of Crop Genetic Improvement, Hubei Hongshan Laboratory, Huazhong Agricultural University, Wuhan, China

***Correspondence: Chen-Yu Zhang (cyzhang@nju.edu.cn), Dijun Chen (dijunchen@nju.edu.cn)**

https://doi.org/10.1016/j.xplc.2023.100631

ABSTRACT

Single-cell transcriptomics has been fully embraced in plant biological research and is revolutionizing our understanding of plant growth, development, and responses to external stimuli. However, single-cell transcriptomic data analysis in plants is not trivial, given that there is currently no end-to-end solution and that integration of various bioinformatics tools involves a large number of required dependencies. Here, we present scPlant, a versatile framework for exploring plant single-cell atlases with minimum input data provided by users. The scPlant pipeline is implemented with numerous functions for diverse analytical tasks, ranging from basic data processing to advanced demands such as cell-type annotation and deconvolution, trajectory inference, cross-species data integration, and cell-type-specific gene regulatory network construction. In addition, a variety of visualization tools are bundled in a built-in Shiny application, enabling exploration of single-cell transcriptomic data on the fly.

Cao S., He Z., Chen R., Luo Y., Fu L.-Y., Zhou X., He C., Yan W., Zhang C.-Y., and Chen D. (2023). scPlant: A versatile framework for single-cell transcriptomic data analysis in plants. Plant Comm. **4**, 100631.

INTRODUCTION

The rapid development of single-cell technologies, in particular single-cell transcriptomics, offers a unique opportunity to analyze molecular and cellular heterogeneity within a tissue or organism. With breakthrough advances in single-cell RNA sequencing (scRNA-seq) based on protoplast isolation or single-nucleus RNA-seq (snRNA-seq) (Seyfferth et al., 2021), recent years have witnessed an explosion of single-cell transcriptomics-based studies and applications in various plant species, including Arabidopsis (Zhang et al., 2019a, 2021b; Denyer et al., 2019; Jean-Baptiste et al., 2019; Ryu et al., 2019; Shulse et al., 2019; Wendrich et al., 2020; Farmer et al., 2021; Kim et al., 2021; Neumann et al., 2022), rice (Liu et al., 2021a; Zhang et al., 2021a; Wang et al., 2021; Zong et al., 2022), maize (Satterlee et al., 2020; Bezrutczyk et al., 2021; Marand et al., 2021), and others. These efforts have not only made it possible to dissect the cellular architecture of complex plant tissues, such as the root system (Zhang et al., 2019a; Denyer et al., 2019; Jean-Baptiste et al., 2019; Ryu et al., 2019; Liu et al., 2021a; Dorrity et al., 2021; Farmer et al., 2021) and the floral meristem (Xu et al., 2021; Neumann et al., 2022; Zong et al., 2022), in a thorough and unbiased manner but have also enabled the characterization of gene expression dynamics and heterogeneity over time (Denyer et al., 2019; Liu et al., 2020; Shahan et al., 2022) or in response to external stimuli (Jean-Baptiste et al., 2019; Wendrich et al., 2020; Wang et al., 2021; Cervantes-Pérez et al., 2022; Ye et al., 2022) at single-cell resolution.

Because the technical challenges of applying single-cell transcriptomics to plants have been overcome, many plant organs and tissues can now routinely be subjected to single-cell analyses for specific purposes (Ryu et al., 2021). The transcriptomes of hundreds to thousands of individual cells can be established in a single experiment. Analyzing such single-cell data to make new biological discoveries requires both hands-on expertise in data analysis and an integrated analytical framework with versatile functions (Pisco et al., 2021). However, existing tools for analysis of single-cell transcriptomics data are designed for specific analysis tasks. It is still challenging, especially for biologists without bioinformatics expertise, to navigate different analysis tools to establish a computational pipeline for analysis of their new data (Luecken and Theis, 2019). Although guidelines and best practices for mapping single-cell transcriptomic atlases have been proposed and are valuable to the scientific community (Luecken and Theis, 2019; Clarke et al., 2021), they are not immediately applicable to the plant community owing to lack of a plant knowledgebase,

Published by the Plant Communications Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and CEMPS, CAS.

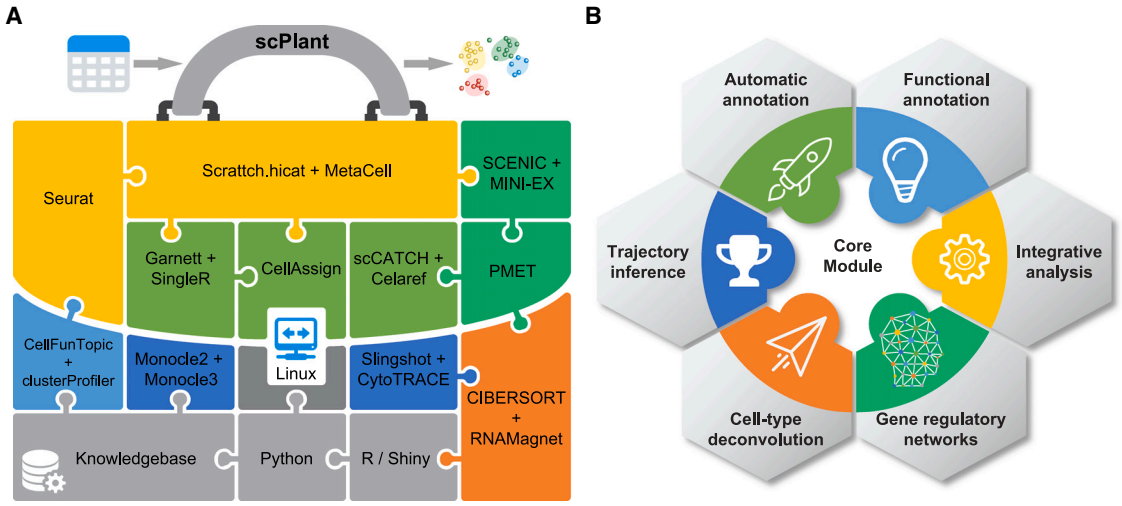

Figure 1. The scPlant workflow.

(A) The method schema of scPlant. Major software packages or environments used in scPlant are highlighted. Blocks are colored according to the main category of the corresponding software packages (Supplemental Table 1).

(B) scPlant consists of one core module and several advanced modules.

which is the key to data analysis and pipeline development (Pisco et al., 2021). Therefore, implementing an up-to-date workflow is not trivial and would be indispensable for building a navigable plant cell atlas (Rhee et al., 2019; Jha et al., 2021).

Here, we describe an end-to-end computational framework, termed scPlant, for exploration of single-cell transcriptomic data in plants. The scPlant pipeline is implemented by bundling popular analysis tools and an associated knowledgebase and thus allows for diverse analytical tasks. scPlant provides a core module for basic single-cell transcriptomic data processing and several add-on modules for advanced topics, including automatic cell-type annotation (wherever possible) and deconvolution (for bulk data), trajectory inference, cross-species data integration, and cell-type-specific gene regulatory network construction. In addition, scPlant offers a built-in Shiny application with multiple visualization toolboxes, enabling data exploration on the fly. scPlant is freely available at https://github.com/compbioNJU/scPlant for non-commercial purposes.

RESULTS AND DISCUSSION

Overview of scPlant

scPlant has been developed for single-cell transcriptomic data analysis in plants by incorporating popular analysis tools such as Seurat (Butler et al., 2018), SCENIC (Aibar et al., 2017), and others (Supplemental Table 1). The only required input for the scPlant framework is one or several matrixes of single-cell transcriptomic data, which can routinely be obtained from scRNA-seq or snRNA-seq experiments. We provide the knowledgebase for the corresponding plant species in order to fully run the pipeline. The method schematic of scPlant is shown in Figure 1A.

The scPlant framework consists of one core module and several advanced add-on modules (Figure 1B). The core module is designed for preprocessing single-cell transcriptomic data, a multistep analytical process that includes quality control (e.g., removal of potential noise), correction of unwanted variation (e.g., batch effects), normalization, dimensionality reduction, cell clustering, analysis of differential expression, and integration (if needed; see methods). After running the core module with a few commands (see the last section practice guidelines for using scPlant), cell clusters can be determined using outstanding unsupervised clustering methods such as the Louvain clustering algorithm (Blondel et al., 2008), and the high-dimensional expression matrix can be embedded into a low-dimensional space using dedicated dimensionality reduction tools such as t-distributed stochastic neighbor embedding (t-SNE) (Maaten and Hinton, 2008) or uniform manifold approximation and projection (UMAP) (McInnes et al., 2018) (Figure 1B). To annotate the identified cell clusters, differential expression analysis is used to find marker gene sets for each cluster by comparing cells in the cluster of interest with all other cells in the dataset. Therefore, mapping cell clusters and defining marker gene sets are typically the major tasks for any single-cell analysis in the core module. In principle, the core module is a prerequisite for all advanced modules. In the following sections, we demonstrate how to use the advanced modules of scPlant to extract biological insights using published scRNA-seq datasets.

Case study 1: Cell-type annotation and deconvolution

From the above core module, single-cell data are analyzed by identifying cell clusters and their associated marker genes. The cell identity of each cluster can be determined by overlaying the expression of known marker genes. Thanks to recent and ongoing efforts such as reference marker gene databases (Chen et al., 2021; Jin et al., 2022) and single cell atlases (Seyfferth et al., 2021), these resources are very valuable for cell identity annotation. In this regard, we have developed automated cluster annotation methods in scPlant to facilitate cell-type annotation, which can be based on an existing marker gene database, functional terms such as Gene Ontology (GO) terms, or a well-established reference cell map (Figure 2A). We have implemented several different solutions for automatic cell-type annotation that are based on popular tools from the mammalian field, including SingleR

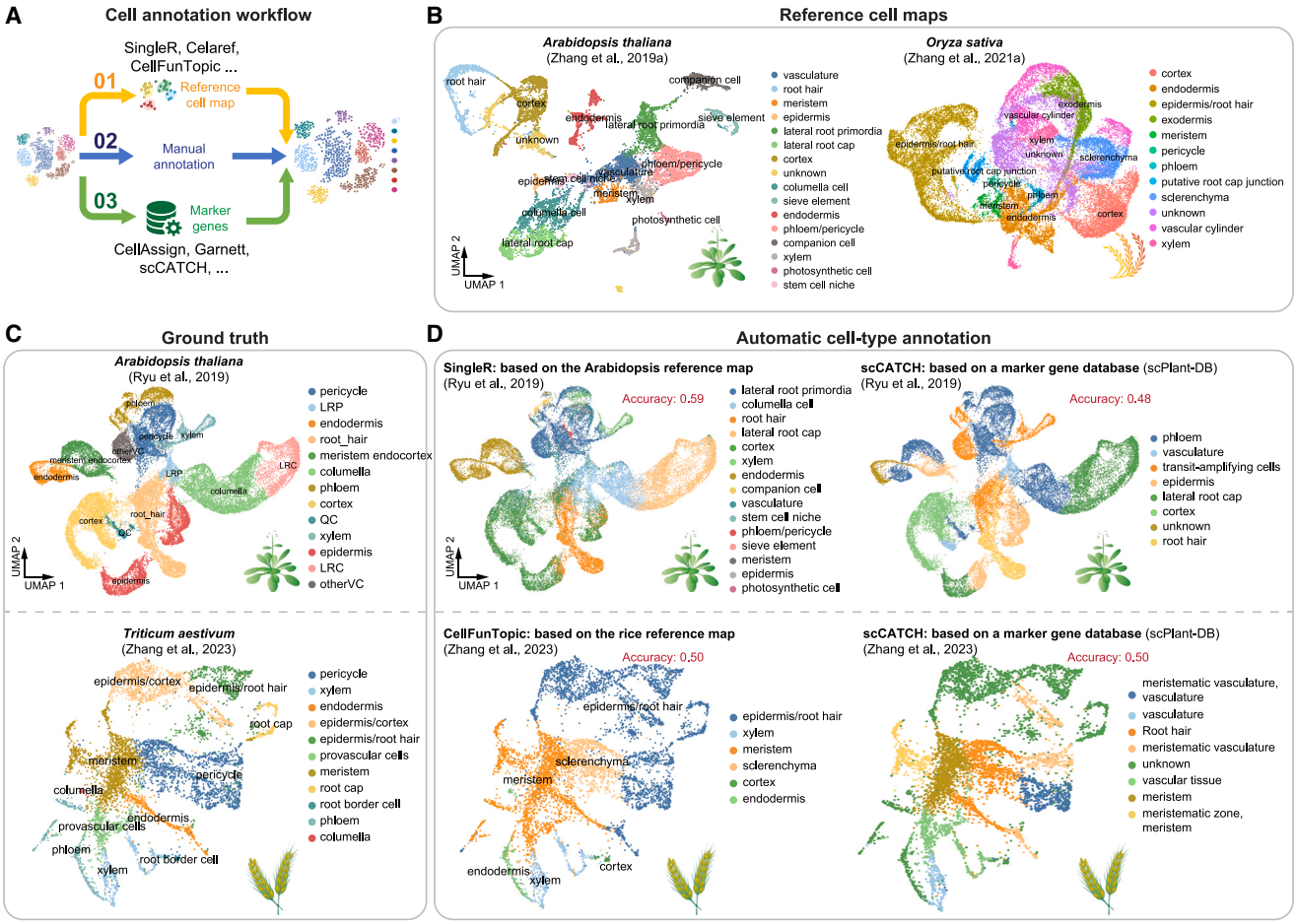

Figure 2. scPlant enables automatic cell-type annotation and deconvolution.

(A) Conceptual schema of cell-type annotation workflow.

(B) Manual annotation of the reference cell map in Arabidopsis (Zhang et al., 2019a) or rice (Zhang et al., 2021a).

(C) Test datasets (Ryu et al., 2019; Zhang et al., 2023) used for automatic cell-type annotation. The manual cell-type annotation is considered the ground truth.

(D) Automatic cell-type annotation using different strategies. Cell-type prediction of an Arabidopsis single-cell dataset (Ryu et al., 2019) based on a reference cell map (Zhang et al., 2019a) using SingleR (top left) or based on known marker genes using scCATCH (top right). Prediction of cell types for a wheat dataset (Zhang et al., 2023) based on a reference cell map (Zhang et al., 2021a) using CellFunTopic (bottom left) or based on known marker genes using scCATCH (bottom right). The prediction accuracy is provided.

(Aran et al., 2019), CellAssign (Zhang et al., 2019b), Garnett (Pliner et al., 2019), scCATCH (Shao et al., 2020), and so on. As a proof of concept, we manually annotated two scRNA-seq datasets (PRJNA517021 and PRJNA706435) from Arabidopsis (Zhang et al., 2019a) and rice roots (Zhang et al., 2021a) as the reference cell map (Figure 2B). We used the scPlant automatic annotation tool to predict cell types in new scRNA-seq/snRNA-seq datasets from distinct species (PRJNA507252 and CRA008788) (Ryu et al., 2019; Zhang et al., 2023) and compared the outcomes with manual annotations (ground truth; Figure 2C). Our evaluation of diverse prediction strategies based on scPlant marker database (scPlant-DB) or reference cell maps resulted in promising predictions (Figure 2D and Supplemental Figure 1). For instance, scPlant achieved highly accurate automatic cell-type annotation for a complex genome (hexaploid wheat) using reference annotations or marker gene databases from a different species. Note that manual annotation is still necessary, given that the current plant marker gene database is far from complete. However, because reference databases are becoming increasingly available in various plants, our automatic cell-type annotation solution offers important alternatives to manual annotation for the plant community.

The biological role and functional specificity of a given cell type is tightly linked to the tissue context in which the cells exert their functions. As an alternative to cell-type annotation, the cell-type composition and its contributions to global gene expression changes in tissues can be inferred from bulk RNA-seq data using regression on a reference expression matrix constructed from a set of cell-cluster-specific marker genes. For this purpose, we have implemented a cell-type deconvolution function in our scPlant pipeline. The deconvolution function is especially useful for comparing changes in cell-type composition in bulk data generated under different conditions (e.g., normal conditions versus stress responses) and thus identifying critical cell types or cell clusters for further functional investigation. To illustrate the concept, we used single-cell data obtained from the Arabidopsis root

(Zhang et al., 2019a) to deconvolve bulk RNA-seq data generated under different stress conditions. The composition of vasculature and companion cells displayed substantial alterations in response to various stressors (Supplemental Figure 2), highlighting the significance of these cell types in stress responses.

Case study 2: Functional annotation of cell types

One of the ultimate goals of cell-type annotation is to characterize each cell cluster with meaningful biological labels based on identified gene signatures for the cluster. For this purpose, we have provided a new module for functional annotation of cell types or clusters using complicated functions from our recently developed *CellFunTopic* package. In brief, using Arabidopsis root data (Zhang et al., 2019a) for a case study, we first used gene set enrichment analysis (GSEA) (Subramanian et al., 2005) to score the enrichment of well-defined biological processes or pathways (called functional terms) in each cluster (Figures 3A and 3B). Reference annotations of functional terms can be taken from the GO or KEGG database. We then used topic modeling to map functional terms with reliable biological meanings to relevant cell clusters on the basis of their topic contributions. In this manner, we were able to annotate each cell cluster with specific functional topics (Figures 3C–3E). For instance, the photosynthetic cell shows enriched functions related to "chloroplast," "photosystem," and "response to radiation," and the endodermis corresponds to "response to water" and "water channel activity." Furthermore, scPlant enables network visualization of the functional similarity or specificity of different cell clusters (Figure 3F). It reveals that an unannotated cluster (cluster 18 with "unknown" marker genes) is tightly connected to the cortex clusters, consistent with the observation that this unknown cluster is located close to the cortex clusters on the cell map (Figure 2B).

Case study 3: Inference of developmental trajectories

Computational modeling can facilitate the inference of lineage relationships between different cell types. These models can be used to identify key regulatory genes and signaling pathways that control development. In plant systems, the inference of developmental trajectories in single cells has been used to study the formation of specific tissues and organs, such as leaves (Lopez-Anido et al., 2021) and roots (Zhang et al., 2019a; Denyer et al., 2019; Liu et al., 2021b; Shahan et al., 2022). To perform this analysis, scPlant incorporates various widely used pseudotime inference tools such as Monocle2 (Qiu et al., 2017), Monocle3 (Cao et al., 2019), CytoTRACE (Gulati et al., 2020), and SlingShot (Street et al., 2018). We applied scPlant to infer developmental trajectories using single-cell transcriptomics data from the Arabidopsis root (Zhang et al., 2019a) with manually annotated cell types as described above (Figures 4A–4D). Slingshot connects the clusters by constructing a minimum spanning tree and then fits principle curves for each identified branch (Figure 4A), and CytoTRACE predicts the relative differentiation state of cells from scRNA-seq data (Figure 4B). Neither SlingShot nor CytoTRACE requires prior information for inferring cellular trajectories. However, CytoTRACE predicted the wrong root (xylem cell) of trajectories in this dataset. Therefore, prior information can be supplied as a starting cell from which the trajectory may originate. Monocle2 and Monocle3 allow specification of root cells (meristem cells in this example), which may help the method to find the correct trajectory (Figures 4C and 4D). The order of cells along developmental trajectories varied among CytoTRACE, Monocle2, and Monocle3 in the test dataset (Figure 4E). Nevertheless, the pseudotimes predicted by these methods show an overall positive correlation (Figure 4F). Therefore, the choice of method should be based primarily on the expected topology of the trajectory and the dimensions of the dataset (Saelens et al., 2019).

Case study 4: Comparative analysis of single-cell atlases across plant species

With a growing number of single-cell atlases generated in diverse plant species, it is appealing to compare single-cell gene expression across species, which provides an opportunity to study conserved gene expression programs in the same cell type, generate cell-type phylogenies, and deduce the evolutionary origin of cell identities (Ryu et al., 2021). For such analyses, we have provided tools in the scPlant framework for cross-species integration of single-cell data in matched organs/tissues using one-to-one orthologous genes as anchors. Note that scPlant mainly uses the strategy of canonical correlation analysis (CCA) and reciprocal principal-component analysis (RPCA) for data integration (Stuart et al., 2019). Although these methods have been widely used in different applications, it would be appealing to test the performance of other state-of-the-art integration methods (Luecken et al., 2021) for analysis of plant single-cell data.

As a case study, we endeavored to generate an integrated cell map across species in roots by analyzing published scRNA-seq data from Arabidopsis (Zhang et al., 2019a), rice (Zhang et al., 2021a), and maize (Ortiz-Ramírez et al., 2021) (Figures 5A–5D). We manually annotated the single-cell atlases for each species and the integrated cell atlas using known marker genes (Figures 5B, 5E, and 5F). We showed that most cell identities were consistent before and after integration and that species-specific cell types were still preserved after integration (Figure 5C), confirming the robustness of the integration analysis. For example, meristem cells have been independently annotated in each species (Zhang et al., 2019a, 2021a; Ortiz-Ramírez et al., 2021) (Figure 5B), and these cells were clustered in the integrated cell map and expressed marker genes with high conservation and specificity (Figures 5F and 5G). However, it will be interesting to further investigate conserved and/or unique genes, as well as gene regulatory networks, that are expressed in the same cell type across species.

Case study 5: Construction of cell-type-specific gene regulatory networks

Gene regulatory networks (GRNs), which consist of interactions between transcription factors (TFs) and their target genes (TGs), orchestrate cell-specific gene expression patterns and, in turn, determine cell function. To construct cell-type-specific GRNs using single-cell data, we have incorporated the widely used single-cell GRN analysis tool SCENIC (Aibar et al., 2017) into the scPlant pipeline and prepared the databases required to support SCENIC for data analysis in plants. Using the annotated cell atlas from the above analysis (Zhang et al., 2019a), we can use SCENIC to identify cell-type-specific GRNs and regulons based on co-expression analysis and TF motif enrichment (Figure 6). We have provided various visualization tools alongside the scPlant

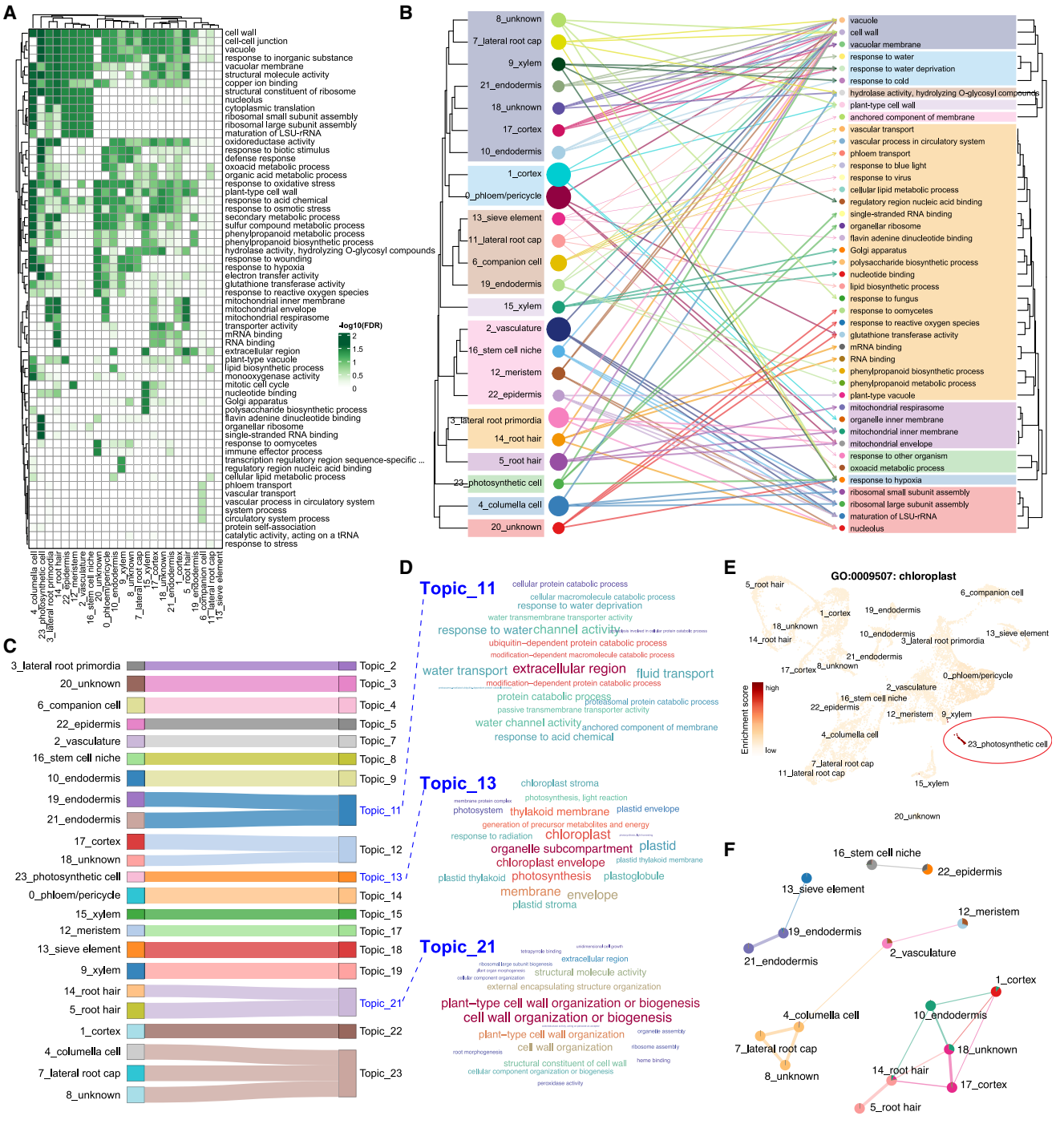

Figure 3. scPlant enables functional annotation of cell types.

(A) Gene set enrichment analysis (GSEA) of different cell clusters. Heatmap shows the enrichment of biological pathways (rows) in different cell clusters (columns).

(B) The relationship between cell clusters (left) and representative enriched pathways (right).

(C) Cell clusters characterized by specific "functional topics" based on topic modeling of gene set enrichment analysis results.

(D) Word clouds showing examples of functional topics.

(E) UMAP plot showing the enrichment of a specific biological pathway over cell clusters.

(F) Networks displaying functional similarity of different cell clusters. The case study is based on Arabidopsis root data (Zhang et al., 2019a).

pipeline to display single-cell GRN results in different ways. For example, the pattern of regulon activity across cell clusters can be shown in heatmaps (Gu, 2022) (Figures 6A and 6B). Highly specific regulons can be displayed in a scatterplot (Figure 6C) and cell-type-specific GRNs in a network view (Figure 6D). We found that the phytochrome-interacting TF PIF5

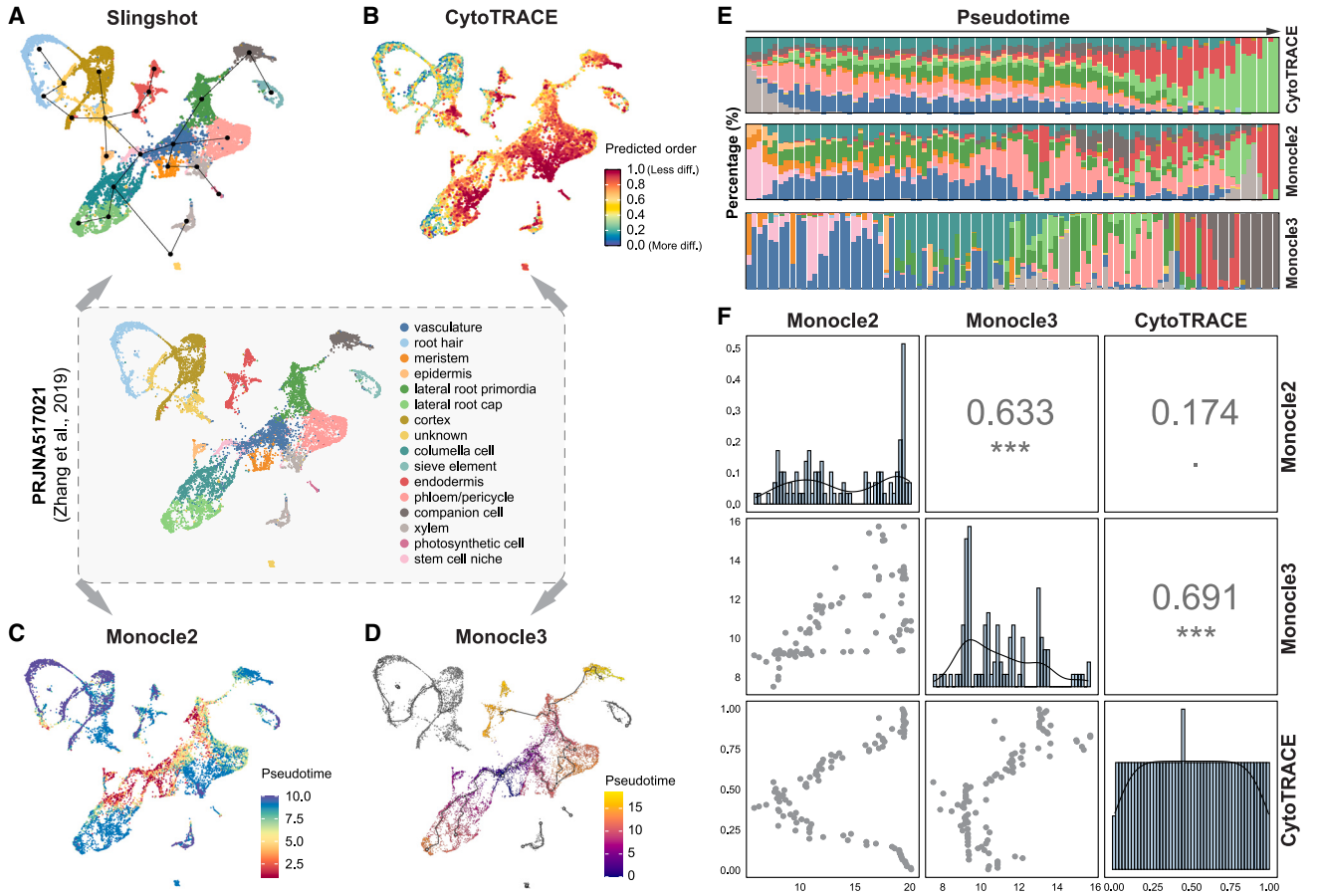

Figure 4. scPlant enables inference of developmental trajectories.

(A–D) UMAP showing the predicted pseudotime or trajectories over an example of a cell map. Developmental trajectories were inferred by CytoTRACE **(A)**, Slingshot **(B)**, Monocle2 **(C),** and Monocle3 **(D)**. The test data were taken from Zhang et al. (2019a).

(E) Bar plots showing the percentage of cell types over the pseudotime (n = 100 bins) along the trajectory.

(F) Comparison of developmental trajectories inferred by CytoTRACE, Monocle2, and Monocle3 based on estimated pseudotime (n = 100 bins).

is the top regulon TF for photosynthetic cells, in line with its regulatory role in photosynthesis and photoprotection (Sakuraba et al., 2014; Toledo-Ortiz et al., 2014).

Recently, the plant-specific GRN inference tool MINI-EX was developed (Ferrari et al., 2022). We also incorporated this tool into our scPlant pipeline. To compare results from MINI-EX and SCENIC, we visualize their outputs in a similar way (Supplemental Figure 3). Overall, the regulons predicted by MINI-EX and SCENIC were consistent among different cell types (Figures 6E and 6F). For instance, both methods identified bZIP44 as a specific regulon TF for companion cells, and ANAC030 was identified as a regulon TF for xylem cells (Pawittra et al., 2020). However, we noticed that MINI-EX produces a higher signal-to-noise ratio than SCENIC in terms of regulon enrichment score (Figure 6E). This discrepancy probably reflects the different strategies used by each tool for regulon identification. MINI-EX uses TF and TG expression levels to filter the regulons when assigning them to a particular cell type. By contrast, SCENIC calculates the regulon activity score based on the overall expression of the regulon (TFs + TGs), resulting in the inclusion of regulons with low TF expression values in specific cell types.

We have also adapted the paired motif enrichment tool (PMET) (Rich-Griffin et al., 2020) to predict pairs of TF-binding motifs within the promoter regions of cell-type-specific marker genes. To detect co-associations of regulons in a cell-type-specific manner, only regulon TFs from the above SCENIC analysis are used in the PMET analysis. As an example, the enrichment of motif pairs within promoters of identity genes of a specific cell cluster (the photosynthetic cell cluster 23, in this case) is shown in Figure 6G. A highly significant pairing of motifs of TCP family was identified as specific for the photosynthetic cell cluster, including TCP14 and TCP15, which are known to be involved in light responses (Viola et al., 2023). In addition, pairing of ABI5 and ABF motifs was also enriched, and these have been shown to inhibit photosynthesis and promote chlorophyll catabolism and leaf senescence (Collin et al., 2021).

Practice guidelines for using scPlant

To demonstrate the use of scPlant for single-cell transcriptomic data analysis, we have provided the real code used to run the case studies above (Box 1). The present protocol would require 2–3 days to set up the workflow and perform a thorough analysis of several published single-cell datasets

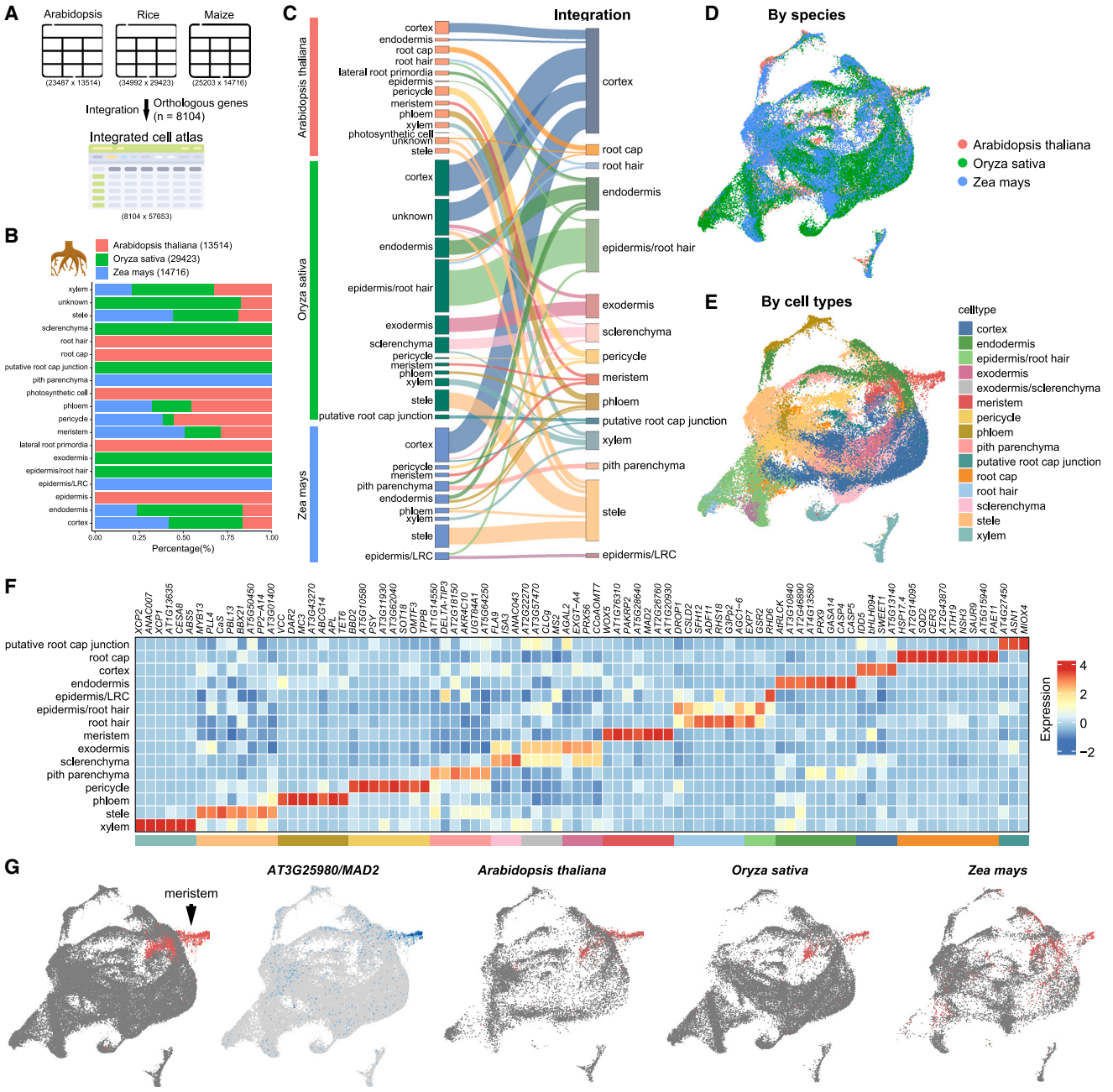

Figure 5. scPlant enables cross-species analysis of single-cell transcriptomic data.

(A) Schema of single-cell transcriptomic data integration in Arabidopsis (Zhang et al., 2019a), rice (Zhang et al., 2021a), and maize (Ortiz-Ramírez et al., 2021).

(B) Manual annotation of root cells for each species according to the corresponding publication.

(C) Sankey diagram showing cell-type annotation before and after integration.

(D) Integrated cell map, colored according to the species of origin.

(E) Cell-type annotation of the integrated cell map.

(F) Representative marker genes for each cell type.

(G) Distribution of meristem cells annotated based on the integrated cell map and the expression pattern of the marker gene *MAD2* (left two plots); the distribution of meristem cells annotated from the original datasets for each species is shown in the right three plots.

(Zhang et al., 2019a, 2021a; Ryu et al., 2019; Ortiz-Ramírez et al., 2021). The pipeline can easily be adapted to analyze new datasets in various plant species. The workflow is targeted to plant biologists who have basic familiarity with computer software and the Linux command-line environment. Guiding principles and a practice notebook are provided online (https://compbionju.github.io/scPlant/) to help users run the recommended workflow.

Finally, we have designed a built-in Shiny application (scPlant-App), which combines the visualization power of the R programming language with user-friendly web interfaces (Chen et al.,

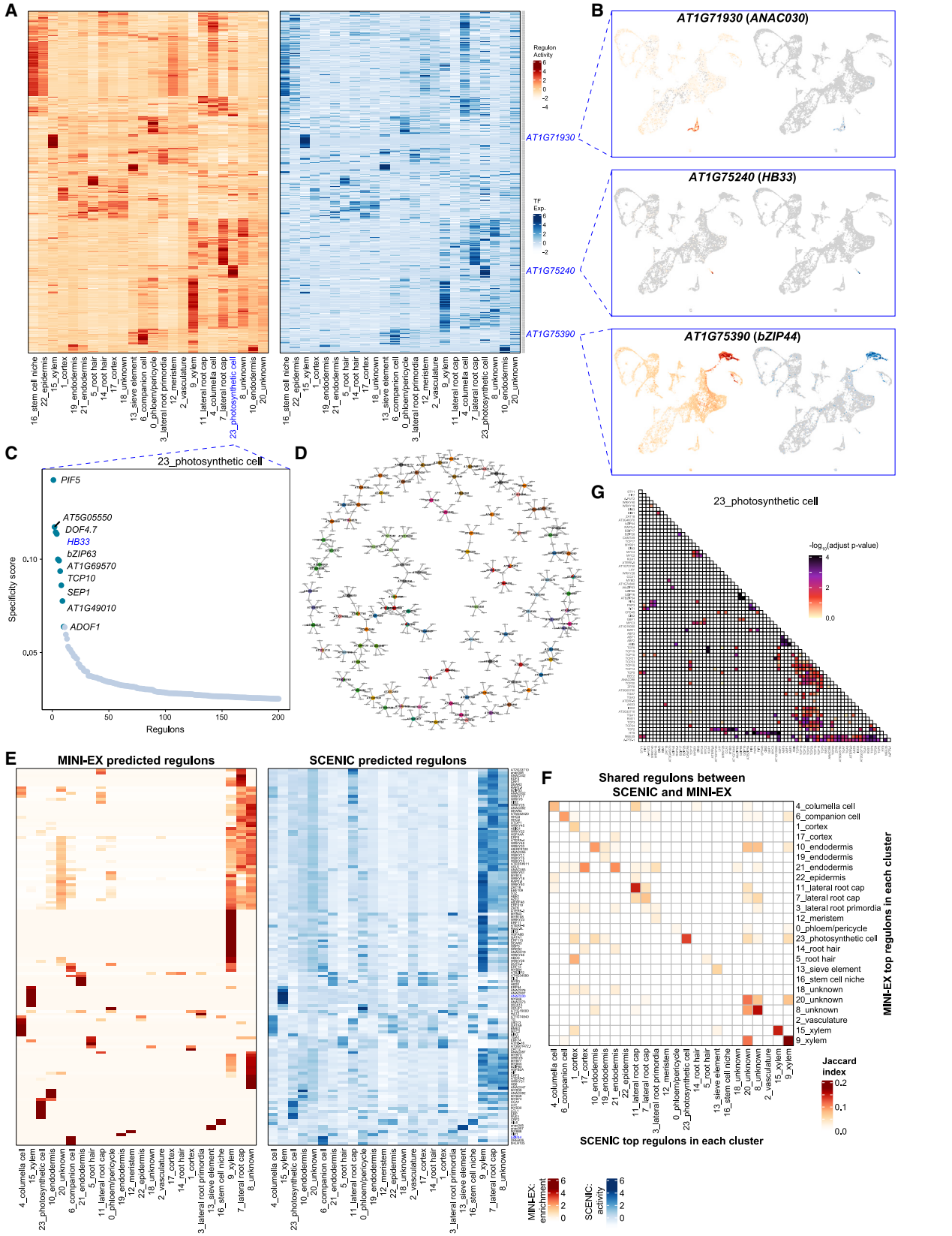

Figure 6. Construction of single-cell gene regulatory networks by scPlant.

(A–D) Outcome of the SCENIC analysis.

(A) Heatmaps showing the regulon activity per cell type (left) and the expression specificity of the corresponding regulon TFs (right).

(legend continued on next page)

2018), for interactive exploration of the analysis results from the workflow described above (Figure 7). This function is particularly beneficial for deep mining of specific cell clusters, genes, and GRNs of interest.

scPlant features

scPlant is a platform specifically designed for analysis of scRNA-seq data from plants. Some of the features that make it useful for plant scRNA-seq analysis include the following:

(1) Plant-specific genome annotations: the scPlant platform uses marker genes, regulatory motif databases, and functional annotations that are specific to plants, which can help in accurately identifying and functionally annotating cell types in plant scRNA-seq datasets. Moreover, the incorporated GRN inference tool MINI-EX (Ferrari et al., 2022) is specifically designed for plant single-cell data analysis.

(2) Plant-specific parameters: the scPlant platform may use specific parameters that are tailored to analysis of plant scRNA-seq data. For example, the number of detected genes per cell is lower in plants than in human and animals. As a result, when using the scPlant platform, less stringent cutoff values may be used for data quality control to avoid filtering out cells with low gene expression. In addition, the presence of plant-specific organelles, such as chloroplasts and mitochondria, which have unique gene expression profiles, can be leveraged in quality control.

(3) Visualization and exploration: the scPlant platform includes interactive visualization tools that enable users to explore and analyze their scRNA-seq data in a user-friendly way. This includes tools for visualizing gene expression patterns, identifying co-expressed gene modules, and exploring gene networks and pathways.

Overall, the scPlant platform provides a powerful set of tools and resources for analyzing plant scRNA-seq data, and it can help to advance our understanding of the complex regulatory mechanisms that control plant growth and development.

METHODS

The implementation of scPlant

The scPlant pipeline is mainly implemented in R by taking the advantage of its statistical and graphic power. We chose the Seurat object (Hao et al., 2021) to store the underlying data structures of single-cell data. We have incorporated several important and popular R packages that are designed specifically for single-cell transcriptome data analysis in mammalian species but have not yet been tested in plants, such as scrattch.hicat (Tasic et al., 2018) and MetaCell (Baran et al., 2019). A few analytical tasks need to run outside of the R environment because the required software packages are designed in Python (see online tutorial).

Preprocessing and quality control of single-cell data

Input files are loaded into individual Seurat objects using the Seurat R package. The resulting feature-by-barcode count matrix is then normalized and scaled for each sample. Cells retained for further analysis must meet a set of user-defined criteria: a minimum number of expressed genes and/or UMI counts and a maximum percentage of mitochondrial gene expression.

Batch correction and integration

When datasets include many samples, data integration is essential for overcoming the complex, non-linear, nested batch effects in the data, which can arise from unwanted technical variation (e.g., sequencing depth, sequencing lanes, read length, plates or flow cells, protocol, experimental laboratories, sample acquisition and handling, sample composition, reagents or media, and/or sampling time) or biological factors (e.g., tissues, spatial locations, species, time points, or variation among individuals). In the current version of scPlant, the canonical correlation analysis and robust principal-component analysis methods in Seurat v.3 (Stuart et al., 2019) have been used for batch correction to account for sample-specific effects. We will continue to test the performance of various single-cell data integration tools (Luecken et al., 2021) for comprehensive analysis of plant data and improve the scPlant pipeline accordingly.

Dimensionality reduction

Principal-component analysis (PCA) is performed on the filtered or integrated feature-by-barcode matrix. Based on the top principal components, the low-dimensional embeddings can be obtained using t-distributed stochastic neighbor embedding (t-SNE) (Maaten and Hinton, 2008) or uniform manifold approximation and projection (UMAP) (McInnes et al., 2018).

Cell clustering and annotation

Cell clusters can be detected using the Louvain community detection–based method with a given resolution. Differentially expressed genes in each cluster are identified using the Seurat "FindAllMarkers" with non-parametric Wilcoxon rank-sum test method. To annotate cell clusters, we have compiled a plant cell-type-specific marker gene database by collecting marker gene information from the literature and databases (Chen et al., 2021; Jin et al., 2022). This curated marker database also enables automatic cell-type annotation using prediction tools such as scCATCH (Shao et al., 2020) and Garnett (Pliner et al., 2019).

Cell-type deconvolution

Using a cell-type expression matrix of top marker genes from single-cell data, the CIBERSORT (Newman et al., 2015) algorithm as implemented in the RNAMagnet package (Baccin et al., 2019) is used to deconvolute cell-type abundance from bulk RNA-seq datasets.

Pseudotime trajectory inference

Several popular pseudotime inference tools, Monocle2 (Qiu et al., 2017), Monocle3 (Cao et al., 2019), CytoTRACE (Gulati et al., 2020), and

(B) UMAP depicting the activity and expression specificity of three selected regulons.
(C) Scatterplot showing the rank of regulons for a specific cell cluster.
(D) A network view of the top five target genes per representative regulon. Small nodes denote target genes, and large nodes denote regulons.
(E) Heatmaps showing the regulon activity per cell type predicted by either MINI-EX (left) or SCENIC (right). Only common regulons predicted by both programs are shown.
(F) Heatmap showing the similarity (measured by the Jaccard index) of cell-type-specific regulons. Top 30 regulons for each cell type were used for comparisons.
(G) Heatmap representing the enrichment analysis of regulon pairs in the photosynthetic cell cluster (cluster 23).

Box 1. Example code for the case studies in this paper.

```
## Quality control and pre-processing

library(CellFunTopic)

SeuratObj <- readData(data = expMatrix, type = 'expMatrix',
species = "Arabidopsis thaliana")

SeuratObj <- QCfun(SeuratObj)

SeuratObj <- RunSeurat(SeuratObj, resolution = 1)

# Visualization of cell clusters

constellationPlot(SeuratObj)

RunMetacell(SeuratObj)

## Auto-identification of cell types (for Figure 2)

# Six methods were tested for automatic cell annotation.

SeuratObj <- AutoAnnotate_SingleR(SeuratObj, ref)

SeuratObj <- AutoAnnotate_CellAssign(SeuratObj, marker_list)

SeuratObj <- AutoAnnotate_Garnett(SeuratObj, marker_file_-
path = "markers.txt")

result <- AutoAnnotate_Celaref(SeuratObj, counts_ref,
cellinfo_ref)

result <- AutoAnnotate_scCATCH(SeuratObj, marker_custom)

result <- CellFunTopicpredictFun(SeuratObj,
reference_SeuratObj)

## Estimating the cell-type composition of a bulk sample using a
single-cell reference

CIBER <- decomposeBulk(SeuratObj, exprs, meta)

fractionHmp(CIBER, meta)

## Functional annotation (for Figure 3)

# Gene Set Enrichment Analysis (GSEA)

SeuratObj <- RunGSEA_plant(SeuratObj, by = 'GO')

# Topic modeling (for Figure 3)

SeuratObj <- runLDA(SeuratObj, k = 20)

# Refer to https://compbioNJU.github.io/CellFunTopic for more
visualization tools.

## Pseudotime trajectory inference (Figure 4)

result <- RunMonocle2(SeuratObj)

result <- RunMonocle3(SeuratObj)

result <- RunSlingshot(SeuratObj)

result <- RunCytoTRACE(SeuratObj)

## Cross-species integration (for Figure 5)

crossSpecies_integrate(matrices, species = c('Ath',
'Osa', 'Zma'))

## Gene regulatory network construction (for Figure 6)

## using SCENIC

bash RunGRN_SCENIC.sh -f SeuratObj.rds -s At -t 2 -d cisTar-
get_databases -o output.

# visualization of regulatory network

ras_exp_hmp(SeuratObj, rasMat)

ras_exp_scatter(SeuratObj, rasMat, gene = colnames(rasMat)[1])

topRegulons(rssMat, topn = 5)

toptargets(tf_target, topn = 5)

SpecificityRank(rssMat, cluster = 23, topn = 10)

## using MINI-EX

bash RunGRN_MINIEX.sh -f SeuratObj.rds.

# visualization of regulatory network.

MINIEXouputPath <- "/your_path/regulons_output"

enrich_exp_hmp(SeuratObj, MINIEXouputPath)

enrich_exp_scatter(SeuratObj, MINIEXouputPath, gene =
'AT1G75390')

topRegulons_MINIEX(MINIEXouputPath, topn = 5)

targets_MINIEX(MINIEXouputPath, cluster = 23, regulons =
"AT3G16770")

BordaRank_MINIEX(MINIEXouputPath,
cluster = 23, topn = 10)

## Run the Paired Motif Enrichment Tool (PMET)

# Note: the command below is run in **Shell**

bash RunPMET.sh -o output -m DEG.txt -s At.

# Visualization of PMET result

processPMET(PMETresult, species = 'Arabidopsis thaliana')

PMEThmp(PMETresult)

triHmp(PMETresult, clus)

topPairsNet(PMETresult, clus)

## Explore interactively using the built-in
Shiny web application (Figure 7)

load_shinyApp(SeuratObj)
```

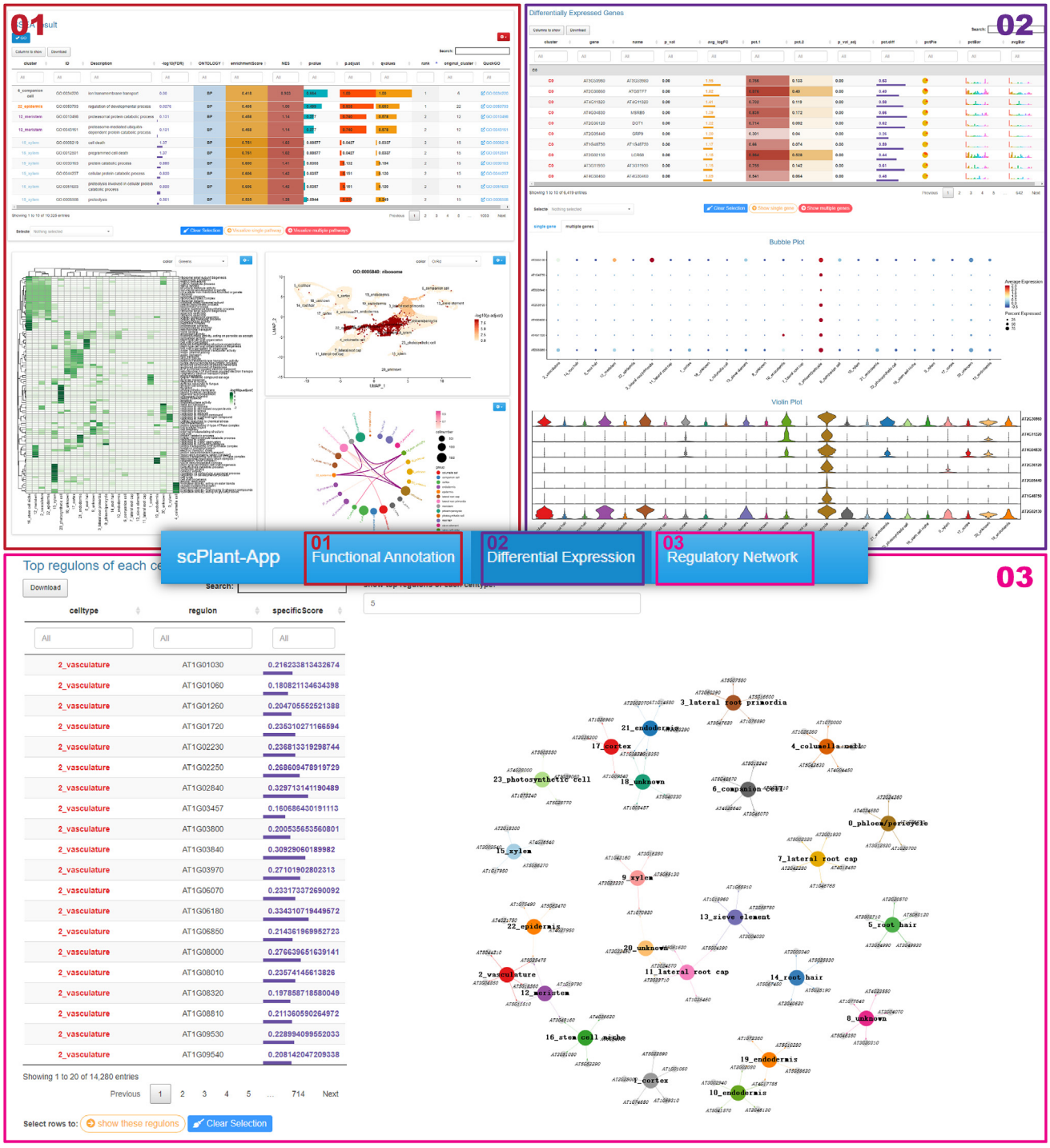

Figure 7. scPlant provides a built-in Shiny application (scPlant-App) for on-the-fly exploration of single-cell atlases.
Screenshots show the three major panels of scPlant-App.

SlingShot (Street et al., 2018), are included in scPlant for inference of cell trajectory and cell fate determination.

GRN inference and regulon discovery

TF DNA-binding motifs for different plants have been compiled into our ChIP-Hub database (Fu et al., 2022) after redundancy filtering. The cisTarget databases have been constructed according to the SCENIC (Aibar et al., 2017) protocol (https://github.com/aertslab/create_cisTarget_databases) for several model plant species. To speed up analysis, the python version (pySCENIC; https://github.com/aertslab/pySCENIC) (Aibar et al., 2017) is used to infer the GRN. SCENIC involves three major steps for gene network inference: (1) identify co-expression modules between TF and potential TGs; (2) for each co-expression module, infer direct TGs based on motif enrichment of the corresponding TF; each regulon is then defined as a TF and its direct TGs; and (3) calculate the regulon activity score in each single cell via the area under the recovery curve.

MINI-EX (Ferrari et al., 2022) is also used to infer the GRN based on scRNA-seq data. Initially, an expression-based network is inferred from gene expression profiles, associating TFs with putative TGs. Next, predicted TGs for each TF are filtered based on TF-binding site enrichment, resulting in refined regulons (TF + TGs). In the third step, each regulon is assigned to a specific cell type through enrichment analysis based on expression of the TGs and further filtered based on expression of the TF. We used the default parameters of MINI-EX (https://github.com/VIB-PSB/MINI-EX) to infer the GRN. To compare the results obtained from MINI-EX and SCENIC, we calculated the Jaccard index, which measures the degree of overlap between the top 30 regulons inferred by both tools for each cell cluster.

Regulon–regulon co-association analysis

We incorporated PMET (https://github.com/kate-wa/PMET-software) (Rich-Griffin et al., 2020) into scPlant to detect the co-localization of pairs of TF-binding motifs within the promoters of cell-type-specific differentially expressed genes. To reduce false-positive results, only regulon TFs are retained for analysis.

Gene set enrichment analysis

scPlant provides annotations for several model plants from knowledge databases such as GO (Gene Ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes). GSEA is performed using the GSEA function from the *clusterProfiler* (Wu et al., 2021) package.

Functional annotation of cell types

We use *CellFunTopic* (https://github.com/compbioNJU/CellFunTopic) to perform functional annotation of cell clusters. Specifically, *CellFunTopic* scores the activity of each functional gene set defined by GO vocabularies using the gene set–scoring method of GSEA (Subramanian et al., 2005). The resulting enrichment scores in terms of a clusters-by-gene-set scoring matrix is subjected to topic modeling. *CellFunTopic* applies latent Dirichlet allocation (Blei et al., 2003) with variational expectation maximization (Nasios and Bors, 2006) to factorize the scoring matrix into a clusters-by-topics matrix θ (the probability of a cluster belonging to a topic) and a topics-by-gene-set matrix φ (the contribution of a topic within a cell cluster). The results are highly interpretable: the top topics under each cell cluster directly reveal the specificity of the biological programs for a particular cell type.

Code availability

The scPlant pipeline is freely available at GitHub (https://github.com/compbioNJU/scPlant) for non-commercial purposes.

DATA AVAILABILITY

All example datasets in this study were taken from previous studies. The Arabidopsis root scRNA-seq datasets are publicly available at the NCBI SRA database with accession numbers PRJNA517021 and PRJNA507252. The maize scRNA-seq dataset is available at the GEO database under accession number GEO: GSE172302. The rice dataset is available at NCBI with the accession number PRJNA706435. The bread wheat (*Triticum aestivum*) snRNA-seq dataset (CRR602489) is available in the Genome Sequence Archive at the Beijing Institute of Genomics (BIG) Data Center under the accession number CRA008788. The raw 10× Genomics fastq files were aligned, filtered, and counted using the Cell Ranger pipeline v.3.1.0 (10× Genomics), and the resulting count matrix files were subjected to the scPlant pipeline for downstream analysis.

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Plant Communications Online*.

FUNDING

This work is supported by the National Natural Science Foundation of China (no. 32070656) and the Nanjing University Deng Feng Scholars Program.

AUTHOR CONTRIBUTIONS

D.C. conceived the study. D.C. and C.-Y.Z supervised the study. S.C. implemented the computational pipeline with support from D.C., X.Z., and L.-Y.F. S.C. performed the data analysis with contributions from Z.H., R.C., Y.L., C.H., and W.Y. D.C. wrote the manuscript. All authors reviewed and approved the manuscript.

ACKNOWLEDGMENTS

No conflict of interest is declared.

Received: January 5, 2023
Revised: February 13, 2023
Accepted: May 24, 2023
Published: May 29, 2023

REFERENCES

Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.-C., Geurts, P., Aerts, J., et al. (2017). SCENIC: single-cell regulatory network inference and clustering. Nat. Methods **14**:1083–1086.

Aran, D., Looney, A.P., Liu, L., Wu, E., Fong, V., Hsu, A., Chak, S., Naikawadi, R.P., Wolters, P.J., Abate, A.R., et al. (2019). Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat. Immunol. **20**:163–172.

Baccin, C., Al-Sabah, J., Velten, L., Helbling, P.M., Grünschläger, F., Hernández-Malmierca, P., Nombela-Arrieta, C., Steinmetz, L.M., Trumpp, A., and Haas, S. (2019). Combined single-cell and spatial transcriptomics reveal the molecular, cellular and spatial bone marrow niche organization. Nat. Cell Biol. **22**:38–48.

Baran, Y., Bercovich, A., Sebe-Pedros, A., Lubling, Y., Giladi, A., Chomsky, E., Meir, Z., Hoichman, M., Lifshitz, A., and Tanay, A. (2019). MetaCell: analysis of single-cell RNA-seq data using K-nn graph partitions. Genome Biol. **20**:206–219.

Bezrutczyk, M., Zöllner, N.R., Kruse, C.P.S., Hartwig, T., Lautwein, T., Köhrer, K., Frommer, W.B., and Kim, J.-Y. (2021). Evidence for phloem loading via the abaxial bundle sheath cells in maize leaves. Plant Cell **33**:531–547.

Blei, D.M., Ng, A.Y., and Jordan, M.I. (2003). Latent dirichlet allocation. J. Mach. Learn. Res. **3**:993–1022.

Blondel, V.D., Guillaume, J.L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. J. Stat. Mech. Theor. Exp. **2008**:P10008.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. **36**:411–420.

Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F.J., et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis. Nature **566**:496–502.

Cervantes-Pérez, S.A., Thibivilliers, S., Laffont, C., Farmer, A.D., Frugier, F., and Libault, M. (2022). Cell-specific pathways recruited for symbiotic nodulation in the Medicago truncatula legume. Mol. Plant **15**:1868–1888.

Chen, D., Fu, L.Y., Hu, D., Klukas, C., Chen, M., and Kaufmann, K. (2018). The HTPmod Shiny application enables modeling and visualization of large-scale biological data. Commun. Biol. **1**:89–98.

Chen, H., Yin, X., Guo, L., Yao, J., Ding, Y., Xu, X., Liu, L., Zhu, Q.H., Chu, Q., and Fan, L. (2021). PlantscRNAdb: a database for plant single-cell RNA analysis. Mol. Plant **14**:855–857.

Clarke, Z.A., Andrews, T.S., Atif, J., Pouyabahar, D., Innes, B.T., MacParland, S.A., and Bader, G.D. (2021). Tutorial: guidelines for annotating single-cell transcriptomic maps using automated and manual methods. Nat. Protoc. **16**:2749–2764.

Collin, A., Daszkowska-Golec, A., and Szarejko, I. (2021). Updates on the role of ABSCISIC ACID INSENSITIVE 5 (ABI5) and ABSCISIC ACID-RESPONSIVE ELEMENT BINDING FACTORs (ABFs) in ABA signaling in different developmental stages in plants. Cells **10**:1996.

Denyer, T., Ma, X., 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.e5.

Dorrity, M.W., Alexandre, C.M., Hamm, M.O., Vigil, A.L., Fields, S., Queitsch, C., and Cuperus, J.T. (2021). The regulatory landscape of Arabidopsis thaliana roots at single-cell resolution. Nat. Commun. **12**:3334.

Farmer, A., Thibivilliers, S., Ryu, K.H., Schiefelbein, J., and Libault, M. (2021). 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.

Ferrari, C., Manosalva Pérez, N., and Vandepoele, K. (2022). MINI-EX: integrative inference of single-cell gene regulatory networks in plants. Mol. Plant **15**:1807–1824.

Fu, L.-Y., Zhu, T., Zhou, X., Yu, R., He, Z., Zhang, P., Wu, Z., Chen, M., Kaufmann, K., and Chen, D. (2022). ChIP-Hub provides an integrative platform for exploring plant regulome. Nat. Commun. **13**:3413–3415.

Gu, Z. (2022). Complex heatmap visualization. iMeta **1**:e43.

Gulati, G.S., Sikandar, S.S., Wesche, D.J., Manjunath, A., Bharadwaj, A., Berger, M.J., Ilagan, F., Kuo, A.H., Hsieh, R.W., Cai, S., et al. (2020). Single-cell transcriptional diversity is a hallmark of developmental potential. Science **367**:405–411.

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell **184**:3573–3587.e29.

Jean-Baptiste, K., McFaline-Figueroa, J.L., Alexandre, C.M., Dorrity, M.W., Saunders, L., Bubb, K.L., Trapnell, C., Fields, S., Queitsch, C., and Cuperus, J.T. (2019). Dynamics of gene expression in single root cells of Arabidopsis thaliana. Plant Cell **31**:993–1011.

Jha, S.G., Borowsky, A.T., Cole, B.J., Fahlgren, N., Farmer, A., Huang, S.S.C., Karia, P., Libault, M., Provart, N.J., Rice, S.L., et al. (2021). Vision, challenges and opportunities for a plant cell atlas. Elife **10**:e66877.

Jin, J., Lu, P., Xu, Y., Tao, J., Li, Z., Wang, S., Yu, S., Wang, C., Xie, X., Gao, J., et al. (2022). PCMDB: a curated and comprehensive resource of plant cell markers. Nucleic Acids Res. **50**:D1448–D1455.

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

Liu, Z., Zhou, Y., Guo, J., Li, J., Tian, Z., Zhu, Z., Wang, J., Wu, R., Zhang, B., Hu, Y., et al. (2020). Global dynamic molecular profiling of stomatal lineage cell development by single-cell RNA sequencing. Mol. Plant **13**:1178–1193.

Liu, Q., Liang, Z., Feng, D., Jiang, S., Wang, Y., Du, Z., Li, R., Hu, G., Zhang, P., Ma, Y., et al. (2021a). Transcriptional landscape of rice roots at the single-cell resolution. Mol. Plant **14**:384–394.

Liu, H., Hu, D., Du, P., Wang, L., Liang, X., Li, H., Lu, Q., Li, S., Liu, H., Chen, X., et al. (2021b). Single-cell RNA-seq describes the transcriptome landscape and identifies critical transcription factors in the leaf blade of the allotetraploid peanut (Arachis hypogaea L.). Plant Biotechnol. J. **19**:2261–2276.

Lopez-Anido, C.B., Vatén, A., Smoot, N.K., Sharma, N., Guo, V., Gong, Y., Anleu Gil, M.X., Weimer, A.K., and Bergmann, D.C. (2021). Single-cell resolution of lineage trajectories in the Arabidopsis stomatal lineage and developing leaf. Dev. Cell **56**:1043–1055.e4.

Luecken, M.D., and Theis, F.J. (2019). Current best practices in single-cell RNA-seq analysis: a tutorial. Mol. Syst. Biol. **15**:e8746.

Luecken, M.D., Büttner, M., Chaichoompu, K., Danese, A., Interlandi, M., Mueller, M.F., Strobl, D.C., Zappia, L., Dugas, M., Colomé-Tatché, M., and Theis, F.J. (2021). Benchmarking atlas-level data integration in single-cell genomics. Nat. Methods **19**:41–50.

Maaten, L. van der, and Hinton, G. (2008). Visualizing Data using t-SNE. J. Mach. Learn. Res. **620**:267–284.

Marand, A.P., Chen, Z., Gallavotti, A., and Schmitz, R.J. (2021). A cis-regulatory atlas in maize at single-cell resolution. Cell **184**:3041–3055.e21.

McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction Advance.

Nasios, N., and Bors, A.G. (2006). Variational learning for Gaussian mixture models. IEEE Trans. Syst. Man Cybern. B Cybern. **36**:849–862.

Neumann, M., Xu, X., Smaczniak, C., Schumacher, J., Yan, W., Blüthgen, N., Greb, T., Jönsson, H., Traas, J., Kaufmann, K., and Muino, J.M. (2022). A 3D gene expression atlas of the floral meristem based on spatial reconstruction of single nucleus RNA sequencing data. Nat. Commun. **13**:2838–2911.

Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. Nat. Methods **12**:453–457.

Ortiz-Ramírez, C., Guillotin, B., Xu, X., Rahni, R., Zhang, S., Yan, Z., Araujo, P.C.D., Demesa-Arevalo, E., Lee, L., van Eck, J., et al. (2021). Ground tissue circuitry regulates organ complexity in maize and Setaria. Science **374**:1247–1252.

Pawittra, P., Suzuki, T., Kawabe, H., Takebayashi, A., Demura, T., and Ohtani, M. (2020). Isolation of dominant Arabidopsis seiv mutants defective in VND7-induced xylem vessel cell differentiation. Plant Biotechnol. **37**:311–318.

Pisco, A.O., Tojo, B., and McGeever, A. (2021). Single-cell analysis for whole-organism datasets 4:207–226. https://doi.org/10.1146/annurev-biodatasci-092820-031008.

Pliner, H.A., Shendure, J., and Trapnell, C. (2019). Supervised classification enables rapid annotation of cell atlases. Nat. Methods **16**:983–986.

Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H.A., and Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell trajectories. Nat. Methods **14**:979–982. https://doi.org/10.1038/nmeth.4402.

Rhee, S.Y., Birnbaum, K.D., and Ehrhardt, D.W. (2019). Towards building a plant cell atlas. Trends Plant Sci. **24**:303–310.

Rich-Griffin, C., Eichmann, R., Reitz, M.U., Hermann, S., Woolley-Allen, K., Brown, P.E., Wiwatdirekkul, K., Esteban, E., Pasha, A., Kogel, K.H., et al. (2020). Regulation of cell type-specific immunity networks in arabidopsis roots. Plant Cell **32**:2742–2762.

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.

Ryu, K.H., Zhu, Y., and Schiefelbein, J. (2021). Plant cell identity in the era of single-cell transcriptomics 55:479–496. https://doi.org/10.1146/annurev-genet-071719-020453.

Saelens, W., Cannoodt, R., Todorov, H., and Saeys, Y. (2019). A comparison of single-cell trajectory inference methods. Nat. Biotechnol. **37**:547–554.

Sakuraba, Y., Jeong, J., Kang, M.Y., Kim, J., Paek, N.C., and Choi, G. (2014). Phytochrome-interacting transcription factors PIF4 and PIF5 induce leaf senescence in Arabidopsis. Nat. Commun. **5**:4636–4713.

Satterlee, J.W., Strable, J., and Scanlon, M.J. (2020). Plant stem-cell organization and differentiation at single-cell resolution. Proc. Natl. Acad. Sci. USA **117**:33689–33699.

Seyfferth, C., Renema, J., Wendrich, J.R., Eekhout, T., Seurinck, R., Vandamme, N., Blob, B., Saeys, Y., Helariutta, Y., Birnbaum, K.D., et al. (2021). Advances and opportunities in single-cell transcriptomics for plant research 72:847–866. https://doi.org/10.1146/annurev-arplant-081720-010120.

Shahan, R., Hsu, C.W., Nolan, T.M., Cole, B.J., Taylor, I.W., Greenstreet, L., Zhang, S., Afanassiev, A., Vlot, A.H.C., Schiebinger, G., et al. (2022). A single-cell Arabidopsis root atlas reveals developmental trajectories in wild-type and cell identity mutants. Dev. Cell **57**:543–560.e9.

Shao, X., Liao, J., Lu, X., Xue, R., Ai, N., and Fan, X. (2020). scCATCH: automatic annotation on cell types of clusters from single-cell RNA sequencing data. iScience **23**:100882.

Shulse, C.N., Cole, B.J., Ciobanu, D., Lin, J., Yoshinaga, Y., Gouran, M., Turco, G.M., Zhu, Y., O'Malley, R.C., Brady, S.M., and Dickel, D.E. (2019). High-throughput single-cell transcriptome profiling of plant cell types. Cell Rep. **27**:2241–2247.e4.

Street, K., Risso, D., Fletcher, R.B., Das, D., Ngai, J., Yosef, N., Purdom, E., and Dudoit, S. (2018). Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. BMC Genom. **19**:477–516.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. Cell **177**:1888–1902.e21.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA **102**:15545–15550.

Tasic, B., Yao, Z., Graybuck, L.T., Smith, K.A., Nguyen, T.N., Bertagnolli, D., Goldy, J., Garren, E., Economo, M.N., Viswanathan, S., et al. (2018). Shared and distinct transcriptomic cell types across neocortical areas. Nature **563**:72–78.

Toledo-Ortiz, G., Johansson, H., Lee, K.P., Bou-Torrent, J., Stewart, K., Steel, G., Rodríguez-Concepción, M., and Halliday, K.J. (2014). The HY5-PIF regulatory module coordinates light and temperature control of photosynthetic gene transcription. PLoS Genet. **10**:e1004416.

Viola, I.L., Alem, A.L., Jure, R.M., and Gonzalez, D.H. (2023). Physiological roles and mechanisms of action of class I TCP transcription factors. Int. J. Mol. Sci. **24**:5437.

Wang, Y., Huan, Q., Li, K., and Qian, W. (2021). Single-cell transcriptome atlas of the leaf and root of rice seedlings. Journal of Genetics and Genomics **48**:881–898.

Wendrich, J.R., Yang, B.J., Vandamme, N., Verstaen, K., Smet, W., Velde, C. van de, Minne, M., Wybouw, B., Mor, E., Arents, H.E., et al. (2020). Vascular transcription factors guide plant epidermal responses to limiting phosphate conditions. Science, 370.

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. Innovation **2**:100141.

Xu, X., Crow, M., Rice, B.R., Li, F., Harris, B., Liu, L., Demesa-Arevalo, E., Lu, Z., Wang, L., Fox, N., et al. (2021). Single-cell RNA sequencing of developing maize ears facilitates functional analysis and trait candidate gene discovery. Dev. Cell **56**:557–568.e6.

Ye, Q., Zhu, F., Sun, F., Wang, T.-C., Wu, J., Liu, P., Shen, C., Dong, J., and Wang, T. (2022). Differentiation trajectories and biofunctions of symbiotic and un-symbiotic fate cells in root nodules of Medicago truncatula. Mol. Plant **15**:1852–1867.

Zhang, T.Q., Xu, Z.G., Shang, G.D., and Wang, J.W. (2019a). A single-cell RNA sequencing profiles the developmental landscape of arabidopsis root. Mol. Plant **12**:648–660.

Zhang, A.W., O'Flanagan, C., Chavez, E.A., Lim, J.L.P., Ceglia, N., McPherson, A., Wiens, M., Walters, P., Chan, T., Hewitson, B., et al. (2019b). Probabilistic cell-type assignment of single-cell RNA-seq for tumor microenvironment profiling. Nat. Methods **16**:1007–1015.

Zhang, T.Q., Chen, Y., Liu, Y., Lin, W.H., and Wang, J.W. (2021a). Single-cell transcriptome atlas and chromatin accessibility landscape reveal differentiation trajectories in the rice root. Nat. Commun. **12**:2053.

Zhang, T.Q., Chen, Y., and Wang, J.W. (2021b). A single-cell analysis of the Arabidopsis vegetative shoot apex. Dev. Cell **56**:1056–1074.e8.

Zhang, L., He, C., Lai, Y., Wang, Y., Kang, L., Liu, A., Lan, C., Su, H., Gao, Y., Li, Z., et al. (2023). Asymmetric gene expression and cell-type-specific regulatory networks in the root of bread wheat revealed by single-cell multiomics analysis. Genome Biol. **24**:1–25.

Zong, J., Wang, L., Zhu, L., Bian, L., Zhang, B., Chen, X., Huang, G., Zhang, X., Fan, J., Cao, L., et al. (2022). A rice single cell transcriptomic atlas defines the developmental trajectories of rice floret and inflorescence meristems. New Phytol. **234**:494–512.