

Transcriptional landscape of rice roots at the single-cell resolution

Qing Liu^{1,5}, Zhe Liang^{2,5}, Dan Feng^{1,5}, Sanjie Jiang^{3,5}, Yifan Wang¹, Zhuoying Du¹, Ruoxi Li⁴, Guihua Hu¹, Pingxian Zhang¹, Yanfei Ma², Jan U. Lohmann^{2,*} and Xiaofeng Gu^{1,*}

¹Biotechnology Research Institute, Chinese Academy of Agricultural Sciences, Beijing 100081, China

²Centre for Organismal Studies, Heidelberg University, Heidelberg 69120, Germany

³BGI Genomics, Shenzhen 518083, China

⁴Fu Foundation School of Engineering and Applied Science, Columbia University, New York, NY 10027, USA

⁵These authors contributed equally.

*Correspondence: Jan U. Lohmann (jan.lohmann@cos.uni-heidelberg.de), Xiaofeng Gu (guxiaofeng@caas.cn)

<https://doi.org/10.1016/j.molp.2020.12.014>

ABSTRACT

There are two main types of root systems in flowering plants, namely taproot systems of dicots and fibrous root systems found in monocots. Despite this fundamental split, our current knowledge of cellular and molecular mechanism driving root development is mainly based on studies of the dicot model *Arabidopsis*. However, the world major crops are monocots and little is known about the transcriptional programs underlying cell-type specification in this clade. Here, we report the transcriptomes of more than 20 000 single cells derived from root tips of two agronomically important rice cultivars. Using combined computational and experimental analyses we were able to robustly identify most of the major cell types and define novel cell-type-specific marker genes for both cultivars. Importantly, we found divergent cell types associated with specific regulatory programs, including phytohormone biosynthesis, signaling, and response, which were well conserved between the two rice cultivars. In addition, we detected substantial differences between the cell-type transcript profiles of *Arabidopsis* and rice. These species-specific features emphasize the importance of analyzing tissues across diverse model species, including rice. Taken together, our study provides insight into the transcriptomic landscape of major cell types of rice root tip at single-cell resolution and opens new avenues to study cell-type specification, function, and evolution in plants.

Key words: single-cell RNA-seq, transcriptional landscape, rice, root, plants

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

INTRODUCTION

Roots are essential for vascular plant growth and development and serve nutrient and water uptake as well as providing firm anchorage to the soil. Quantifying gene expression in single cells or cell types is crucial for understanding the complex gene regulatory networks controlling root development. In recent years, powerful single-cell RNA sequencing (scRNA-seq) techniques have been developed (Tang et al., 2011; Macosko et al., 2015), and have allowed researchers to record robust transcriptional signatures of cell types in a wide spectrum of eukaryotic organisms, including human, mouse, and *Caenorhabditis elegans* (Trapnell et al., 2014; Zeisel et al., 2015; Cao et al., 2017). The technical advances have also made it possible to analyze the transcriptome of plant roots at single-cell resolution, focusing on the reference species *Arabidopsis* (Efroni and Birnbaum, 2016). To date, scRNA-seq of single root

cells has only been reported in dicot *Arabidopsis*, demonstrating high rate of heterogeneity of roots and the expression signatures for the major cell types (Zhang et al., 2019b; Denyer et al., 2019; Jean-Baptiste et al., 2019; Ryu et al., 2019; Shulse et al., 2019; Rich-Griffin et al., 2020). The anatomy and morphology of the monocot species root system (Rebouillat et al., 2009) differs from those of dicots, such as the presence of a lysigenous cortex and additional cell layers of epidermis and cortex (Supplemental Figure 1). Although previous work in transcriptional profiling has been carried out by using bulk RNA isolated from root tissues (Zhang et al., 2010; Takehisa et al., 2012; Gutjahr et al., 2015), the molecular basis of monocot root cell types and their evolution are largely unknown.

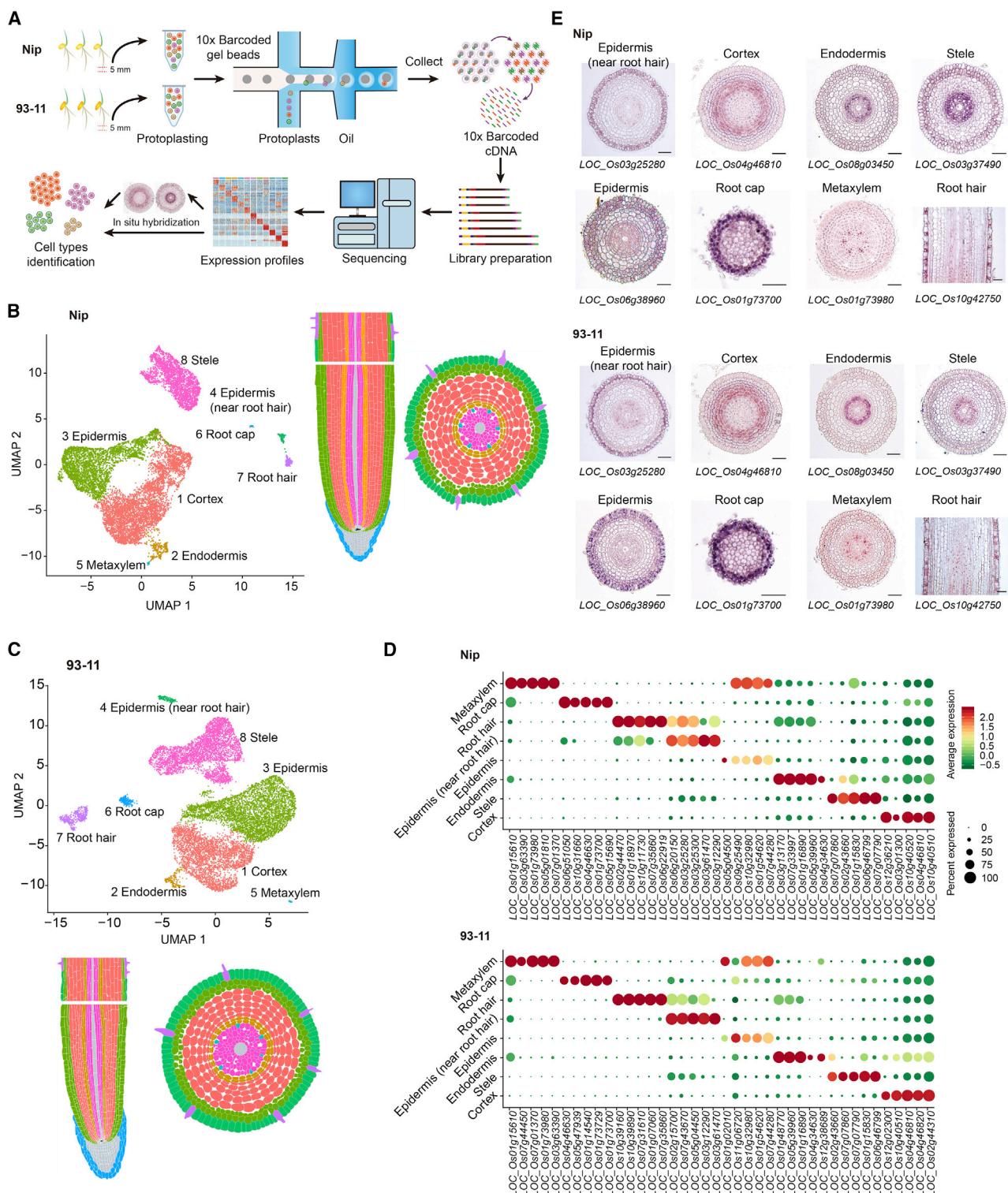


Figure 1. Single-cell RNA-seq and cluster annotation of rice root tips.

(A) Overview of rice scRNA-seq and cluster annotation workflow. Protoplasts were isolated from 5-mm root tips of rice Nip and 93-11, respectively. scRNA-seq libraries were generated using the 10x Genomics platform followed with high-throughput sequencing. Analysis workflow including removal of low-quality cells, gene expression normalization, calculation of variable genes, clustering, and *in situ* hybridization assistant cluster annotation.

(B and C) UMAP visualization of putative clusters from 10 968 cells and 12 564 cells and their spatial distribution in root tips of Nip (B) and 93-11 (C), respectively. Each dot indicates a single cell. Colors in the diagram of root tip indicate corresponding cell clusters. The greyish white and black parts indicate uncaptured cells.

(legend continued on next page)

In this study, we sequenced over 20 000 single cells of root tip of rice, which is a widely used reference monocot species, as well as one of the most important crops worldwide. Combining computational analyses with *in situ* hybridization experiments, we identified nearly all major cell types with novel cell-type-specific marker genes. Comparative analyses of single-cell expression data between rice cultivars, as well as between rice and *Arabidopsis*, allow us to examine the conserved and divergent features of root cell-type transcriptomes.

RESULTS AND DISCUSSION

scRNA-seq and identification of cell-type clusters

We performed scRNA-seq to *Japonica* group cultivar Nipponbare (Nip) and *Indica* group cultivar 93-11, which are the two main cultivated varieties and parental lines used for breeding in Asia (Qian et al., 2016). Protoplasts were isolated from 250 root tips (5 mm in length of crown roots) of 3-day-old Nip and 93-11 seedlings, respectively. About 18 000 protoplasts per sample were initially mixed with 10x Genomics single-cell reaction regents including cell barcodes. The libraries were constructed followed by high-throughput sequencing (Figure 1A). Data were prefiltered at both cell and gene level. In total, we successfully profiled 10 968 cells with 2592 genes/cell and 12 564 cells with 2636 genes/cell for Nip and 93-11, respectively (Supplemental Table 1). To examine the sequencing depth and robustness of the scRNA-seq results, we divided the data into two random subsamples, and found high correlation between them in both Nip and 93-11 (Supplemental Figure 2A and 2B). Further, we found a high correlation between the combined scRNA-seq data with bulk RNA sequencing (RNA-seq) of independently isolated and protoplasted root tips (Supplemental Figure 2C and 2D). In contrast, the correlation between the combined scRNA-seq data with bulk RNA-seq of independently isolated but not protoplasted root tips was lower (Supplemental Figure 2E and 2F), which was expected due to the cellular response during the 3 h of enzymatic digestion for generating protoplasts. Gene ontology (GO) analysis suggested that the protoplasting induced differentially expressed genes (DEGs) ($P < 0.01$, fold change >2) were involved in environmental responses and RNA processing, etc. (Supplemental Figure 2G). All of our results can interactively be mined on the Web using the Root Cell Atlas in Rice (RCAR), which is freely available at <http://www.elabcaas.cn/rkar/index.html>.

After linear dimensional reduction, the uniform manifold approximation and projection (UMAP) algorithm and *t*-distributed stochastic neighborhood embedding (*t*-SNE) tools were used to visualize and explore the datasets (Van der Maaten and Hinton, 2008; Becht et al., 2019). Unsupervised analyses grouped cells into eight major clusters based on UMAP or *t*-SNE (Figure 1B and 1C; Supplemental Figure 3A). Similarly, eight major clusters were also observed when we removed 5914 DEGs in response to protoplasting in both Nip and 93-11 (Supplemental Figure 3B), suggesting that the effect of

cell wall digestion was minor for cell clustering. Moreover, application of RaceID, another independent clustering approach, also supported a similar structure with eight clusters (Supplemental Figure 3C).

Since only very few marker genes for rice root cell types were known, we first performed RNA *in situ* hybridizations using orthologs of *Arabidopsis* marker genes to annotate these clusters (Zhang et al., 2019b; Denyer et al., 2019). This strategy allowed us to identify several genes specifically expressed in stem cells (Supplemental Figure 4A); however, these genes were only represented in very few cells in the scRNA-seq data, suggesting rare cell types were not efficiently captured in our experiment or these genes were not well detected by scRNA-seq technology because of their low expression. Importantly, we found that more than 20 orthologs of known marker genes for diverse cell types of *Arabidopsis* did not show cell-type-specific expression in rice (Supplemental Figure 4B). This finding indicated that a significant number of marker genes may not be conserved between *Arabidopsis* and rice, and hence are not suitable for robust cell-type identification across species.

To alleviate this problem, we mined our own data for potential marker genes that were highly and specifically expressed in one or two clusters (Figure 1D; Supplemental Figure 5; Supplemental Tables 2 and 3), and also took relevant references into account and selected genes with a maximum of available information (Ishimaru et al., 2011; Kim et al., 2007). To validate these candidates *in vivo*, we performed *in situ* hybridization and were able to identify *bona fide* cell-type marker genes for rice root cell types (Figure 1E and Supplemental Figure 4C). For example, *LOC_Os06g38960* was specifically expressed in cells of the epidermis, *LOC_Os03g25280* in the epidermis near root hair, *LOC_Os04g46810* in cortex cells, *LOC_Os01g73700* in the root cap, *LOC_Os08g03450* in the endodermis, *LOC_Os03g37490* in the stele, and *LOC_Os01g73980* in the metaxylem, respectively in both Nip and 93-11 (Figure 1E; Supplemental Figures 4C–4E and 5A–5C). In addition, we selected seven of these maker genes and created reporter lines using fluorescent proteins, and observed cell-type-specific signal in transgenic rice (Supplemental Figure 6), which was in agreement with our *in situ* hybridization results. Together with previously known markers for root hair (Kim et al., 2007; Ding et al., 2009), our novel marker genes allowed us to annotate all clusters containing the major cell types of the root tip in Nip and 93-11 (Figure 1B and 1C; Supplemental Figure 3A–3C). Interestingly, we robustly identified a specific sub-cell type of epidermal cell located near root hair, which was not reported in *Arabidopsis*, suggesting an evolutionary divergence between monocots and dicots. Of note, epidermis, cortex, and stele were the major clusters containing cells ranging from 2328 to 4,833, while clusters of root hair, metaxylem, root cap, and endodermis had only small number of cells (17–435) (Supplemental Table 4), which was

(D) Expression of cell-type marker genes for each cluster. Dot diameter, proportion of cluster cells expressing a given gene; color, mean expression across cells in that cluster.

(E) Representative RNA *in situ* hybridization of cell-type marker genes for the eight putative clusters (B and C) in Nip and 93-11, respectively. Scale bars, 40 μ m.

comparable with the actual ratio of cell numbers in the rice root tip ([Supplemental Figure 1](#)). Next, we tried to identify subclusters within the clusters containing many cells, and identified *LOC_Os01g68589* from the cortex cluster, which turned out to be specifically expressed in the innermost layer of cortex ([Supplemental Figure 4F](#)). In summary, our scRNA-seq dataset revealed eight distinct clusters, which corresponded to eight major cell types. These clusters were validated by identifying a set of novel rice-specific marker genes and using them for *in situ* hybridization and reporter genes.

Highly conserved cell-type clusters between rice cultivars

The similarity in cell cluster separation and cell morphology between Nip and 93-11 promoted us to investigate the conservation of their single-cell transcriptomes. To this end, the single-cell data from Nip and 93-11 were merged and subsequently clustered. We observed that all eight cell-type clusters were well aligned and matched one to one between the two cultivars ([Figure 2A](#)). Cell-type homologies were also supported by our cell-type marker genes and high Pearson's correlation coefficients of the expression profiles between cell-type clusters of both cultivars ([Figure 2B](#) and [Supplemental Figure 7A](#)). Furthermore, we compared conserved and divergently expressed genes for each of the homologous cell types ([Supplemental Table 5](#)). Venn diagram analyses showed that more than 80% of expressed genes were shared between Nip and 93-11 in all big cell-type clusters, including cortex, endodermis, epidermis, and stele. For the cultivar-specific genes, the highest percentage was found in the cluster of root cap ([Figure 2C](#)), which fits well with the fact that the morphology of the root cap is the most divergent, while other cell types are nearly identical between Nip and 93-11 ([Supplemental Figure 1](#)). Since auxin-related genes are important for root cap formation ([Wang et al., 2005](#)), we found that 14 auxin-related genes in root cap were differentially expressed between Nip and 93-11 ([Supplemental Table 6](#)), which may contribute to the difference in their morphology. Next, we identified cultivar-specific marker genes for cell types ([Supplemental Figure 8A](#) and [8B](#)), which may contribute to the understanding of the cellular and physiological differences between rice cultivars ([Zhang et al., 2019a](#)). To provide insights into the molecular functions in each cell type in both cultivars, we performed GO analysis and found matching functional categories for each cell type ([Figure 2D](#)). For example, genes related to water and fluid transport were enriched in cells of the cortex, genes related to cell wall were enriched in the epidermis, while genes related to response to extracellular and external stimuli were enriched epidermis (near root hair) ([Figure 2D](#)). We also examined the function of the cultivar-specific genes for each cell type, and found that most of the resulting overrepresented GO terms were related to response to various environmental stimuli ([Supplemental Figure 8C](#)), suggesting that distinct mechanisms may act in response to external environmental factors between Nip and 93-11.

Next, we analyzed spatio-temporal expression patterns for genes with known functions in cell-type specification, focusing on phytohormones, which are important for root development and stress responses ([Vanstraelen and Benkova, 2012](#)). To this end, we plotted the expression of genes related to major hormone biosynthesis and response pathways and found that, for

several hormones, biosynthesis and response genes were overrepresented in the same cell types ([Figure 2E](#) and [Supplemental Figure 7B](#)). For example, auxin biosynthesis and response genes were overrepresented in the clusters of stele and epidermis, suggesting an autocrine function of auxin in these cells. In contrast, for other hormones, these two types of genes were enriched in distinct clusters. For example, abscisic acid (ABA) responses genes were enriched in the clusters of epidermis and cortex, whereas there was no clear pattern for the ABA biosynthesis genes. Gibberellic acid (GA) biosynthesis genes were overrepresented in the cluster of cortex cells, but GA response genes were overrepresented in the stele cluster. Taken together, our datasets revealed complex transcriptional programs in all cell types and allowed us to predict potential molecular signaling between cell types.

Differentiation trajectory of epidermal and root hair cells

Since cells in both terminal and intermediate developmental states are captured simultaneously ([Trapnell et al., 2014](#)), scRNA-seq enables the exploration of the continuous differentiation trajectory of a developmental process. Epidermis and root hair are the outmost layers in the root, and root hairs are differentiated from the epidermis. To examine the trajectory of these developmentally related cell types, pseudotime analysis was applied to the clusters representing these cell types using Monocle 2 ([Trapnell et al., 2014](#)). Individual cells were projected into two ends of the pseudotime backbone, representing two distinct final states ([Figure 3A](#)). Cells from the two cultivars showed highly consistent pseudotime order, suggesting a conserved developmental trajectory ([Figure 3A](#) and [3B](#)). By labeling the cells according to their types, we found that the developmental trajectory started with cells from the subset of epidermis cluster, after a bifurcation point that marks developmental state transition, the trajectory gradually ended with two differentiated cell types: the root hair cells (branch 1) and the differentiated epidermis cells (branch 2), respectively ([Figure 3A–3C](#)). In the root hair branch, cells from epidermis, epidermis (near root hair), and root hair were ordered continuously ([Figure 3B](#)), suggesting that the pseudotime order derived from single-cell data essentially recapitulates the root hair developmental process. We then identified 534 differentially expressed genes across the pseudotime order. These genes fell into three clusters with distinct gene expression patterns reflecting transcriptional rewiring during root development ([Figure 3D](#)). Gene expression patterns differ greatly across the branch point. Genes in the first cluster were preferentially expressed in the prebranch, which consists of epidermal cells. These genes were enriched in GO terms related to response to various stimuli, which is consistent with the function of the epidermis in environmental sensing and communication ([Figure 3D](#) and [Supplemental Figure 8C](#)). Branch 1 included more mixed cell clusters, and genes related to epidermis development and root hair differentiation were found, suggesting that they may regulate this cell state transition. For example, compared with root hair-specific markers, we found *LOC_Os07g43670* encoding a ribonuclease T2 family domain-containing protein in the third-cluster genes expressed in both epidermis near root hair and root hair cells. *LOC_Os07g43670* expression was validated by *in situ* hybridization in both Nip and 93-11 ([Figure 3E](#)), supporting the ordered

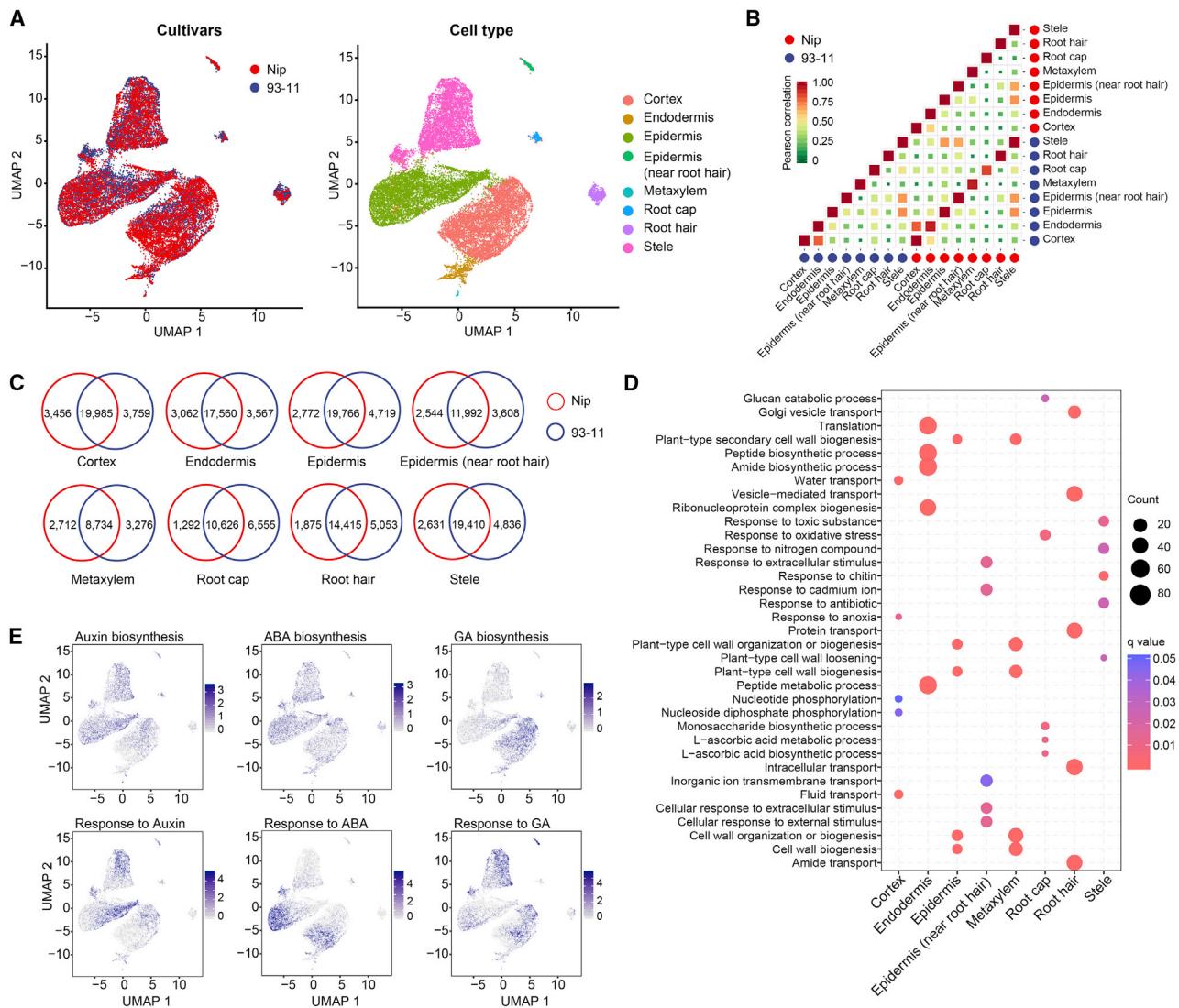


Figure 2. Highly conserved cell-type clusters between rice cultivars.

(A) UMAP visualization of Nip and 93-11 clusters after alignment. Each dot indicates a single cell. The colors indicate cultivar (left panel) or cell type (right panel).

(B) Heatmap showing high correlations between cell-type transcriptomes between Nip and 93-11. The single-cell data were merged together for each cell type prior to comparison.

(C) Venn diagram showing the number of shared and cultivar-specific expressed genes for each cell-type cluster between Nip and 93-11. For each comparison, statistical significance of the overlap was calculated with exact hypergeometric probability. All overlaps are significant.

(D) Scatter plots of GO enrichment analysis of shared genes for each cell type between Nip and 93-11 (C). Only enriched categories with an adjusted P value <0.05 are shown. Overrepresentation analysis and visualization performed using ClusterProfiler R package.

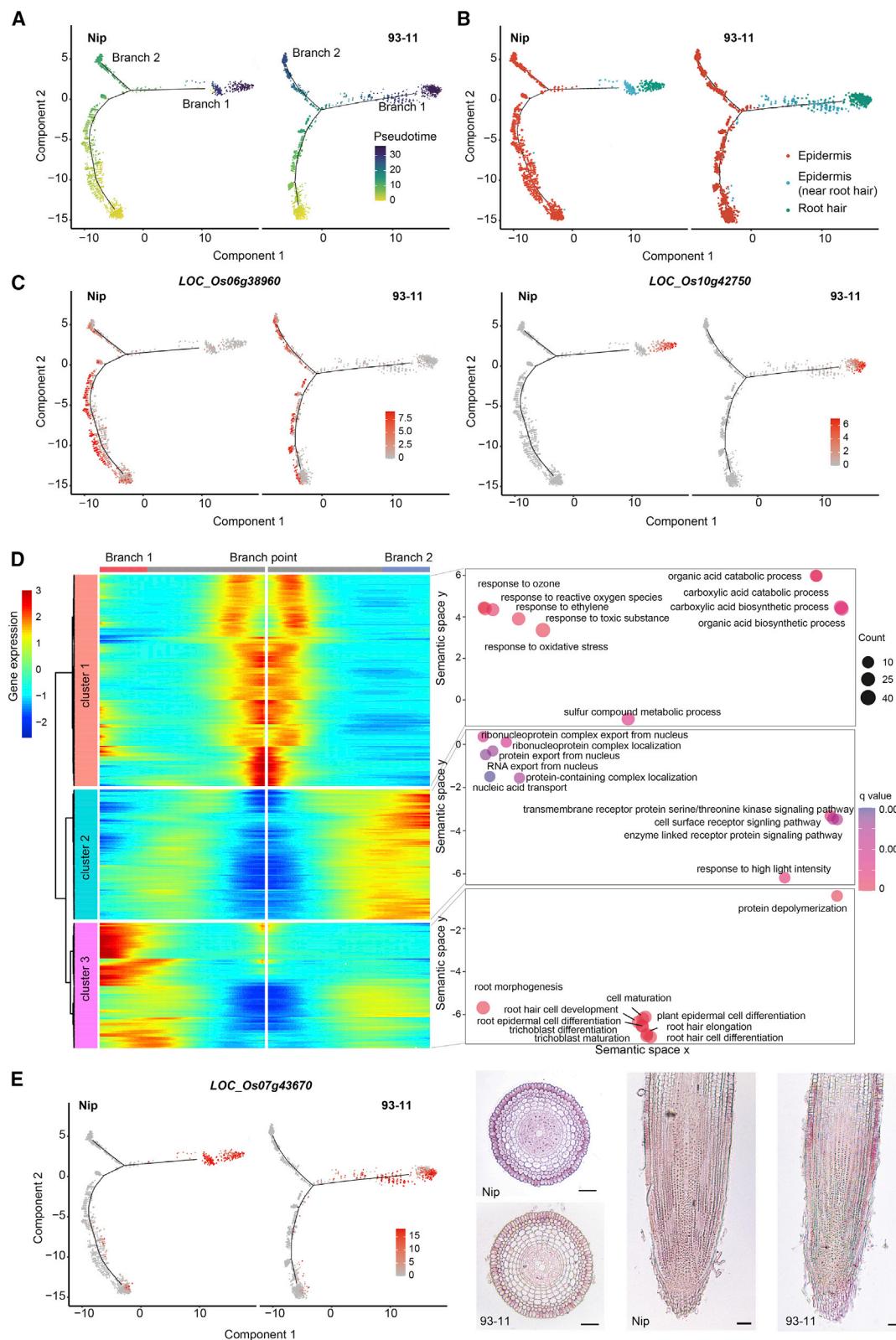
(E) UMAP visualization of expression patterns of the genes related to auxin, ABA, and GA biosynthesis and response. The colors represent expression levels of these genes in individual cells.

developmental process from epidermis to root hair. Collectively, pseudotime analysis provided insight into the dynamic process of root cell-type development and essential gene expression rewiring during the transition of cell states.

Conserved and divergent cell-type gene expression between rice and *Arabidopsis*

To determine the evolutionary conservation of root cell-type development in monocots and dicots, we compared scRNA-seq datasets from rice and *Arabidopsis*. To this end, we used a

recently published scRNA-seq dataset, also derived from protoplasts of 5-mm root tips using the 10x Genomics platform (Zhang et al., 2019b). First, we used the previously described *Arabidopsis* cell-type marker genes and obtained similar cell-type clusters to the reported ones (Supplemental Figure 9A) (Denyer et al., 2019; Zhang et al., 2019b), suggesting that our analysis pipeline is robust. Next, we combined rice and *Arabidopsis* scRNA-seq datasets, aligned them by using orthologous genes (Supplemental Figure 9B), and then clustered these cells by dimension reduction. Consistent cell-type homologies were identified based on shared cluster membership using Seurat (Butler

**Figure 3. Differentiation trajectory of epidermal and root hair cells.**

(A) Single-cell transcriptome data of epidermis and root hair from Nip and 93-11 analyzed by Monocle 2, and revealing a key branch point in the trajectory. Each dot indicates a single cell. Color on the dots indicates the pseudotime score.

(B) Cell types were labeled on the differentiation trajectory for Nip and 93-11, respectively.

(legend continued on next page)

et al., 2018) second alignment (Figure 4A–4D; Supplemental Figure 9C–9E). UMAP visualization and Pearson's correlation coefficient analysis revealed that clusters of root hair cells were highly correlated and matched one to one, while clusters of cortex cells were more moderately correlated between rice and *Arabidopsis* (Figure 4E). We also analyzed another scRNA-seq dataset from *Arabidopsis* root tip (Ryu et al., 2019), and our cross-species comparison revealed similar results (Supplemental Figure 10A–10E). Next, we extracted the scRNA-seq data of root hair and cortex from rice and *Arabidopsis*, and subsequent alignment and clustering further suggest that these two cell types are correlated between rice and *Arabidopsis* (Supplemental Figure 11A and 11B). Moreover, these homologies were also supported by shared marker gene expression between species (Figure 4F; Supplemental Figure 11C) and may help to predict properties and conservation of homologous cell types (Supplemental Figure 11D), as well as the function of marker genes. For example, *Proline-rich protein-like 1* (*PRPL1*) was demonstrated as key regulator for *Arabidopsis* root hair development (Boron et al., 2014). We found the rice ortholog also specifically expressed in root hair cells (Supplemental Figure 11C), suggesting a conserved function for root hair development for this gene, although further experiments would be needed to test this prediction. Of note, cells of *Arabidopsis* cortex matched to a specific subset of rice cortex cells (Figure 1D and Supplemental Figures 9E, 10B, and 10D), maybe reflecting a partially conserved function of cortex in the two species. In contrast, clusters of epidermis, endodermis, stеле, and root cap cells seemed to diverge more substantially between rice and *Arabidopsis* (Figure 4D–4G and Supplemental Figures 9D, 9E, and 10A–10E). For these cell types, cell-type clusters were not spatially overlapped in UMAP, and generally more genes showed divergent expression (Supplemental Table 7), which may contribute to anatomical differences between rice and *Arabidopsis*. Additionally, marker genes for these tissues were generally not shared between rice and *Arabidopsis* (Figure 4F and Supplemental Figure 4B). Consistently, for epidermis, epidermis (near root hair), endodermis, stèle, and root cap, we performed *in situ* hybridization on the orthologs of rice marker genes in *Arabidopsis*, and found that these genes were not cell-type-specifically expressed in *Arabidopsis* (Supplemental Figure 12A).

We performed GO analysis to investigate the function of the divergent genes for each cell type, and found the top GO terms were related to catabolic process and response to various environmental stimuli (Supplemental Figure 12B). Of note, several GO terms, such as “response to cadmium ion”, were found in almost all cell types, suggesting that rice and *Arabidopsis* may have differences in basic cell functions, even though they are anatomically similar for some cell types. The most divergent gene families included those encoding membrane proteins and

primary pumps (Figure 4H), whose differential activities may contribute to the divergence of cell function and root morphology between rice and *Arabidopsis* (Hoffmann et al., 2019).

Taken together, our single-cell transcriptomic analysis has identified cell-type-specific regulatory programs for almost all cell types of rice root tip and robust cell-type markers for two cultivars, Nip and 93-11. Hence, our results represent a valuable resource to study the developmental and physiological function of cell types at the molecular level and at single-cell resolution. The comparison of rice and *Arabidopsis* scRNA-seq data revealed evolutionary conserved and divergent cell-type and species-specific features of gene expression, emphasizing the importance of expanding scRNA-seq approaches to multiple species.

METHODS

Rice growth conditions and protoplast isolation

Rice seeds were soaked in water in the dark until germination, and then cultured in Hoagland's Complete Nutrient Solution for 3 days (28°C with 10-h-light/14-h dark cycles, with light intensity of 300 lux). Two-hundred and fifty tips (5 mm) of crown roots were cut into small pieces of about 1 mm and incubated in digest solution (2% cellulase RS, 0.75% macerozyme R10, 1% hemicellulase, 0.5% pectolyase Y-23, 10 mM MES (2-(N-morpholino)ethanesulfonic acid), 0.6 M mannitol, 1 mM CaCl₂, 0.1% BSA, and 0.04% 2-mercapto-ethanol) at vacuum pump (~60 kPa) at room temperature for 30 min, then the mixture was placed in a 28°C incubator at 60–70 rpm for 2.5 h to release root cell protoplasts. The protoplasts were filtered through 80-μm nylon mesh and centrifuged at 130 g for 5 min, and washed three times with 8% mannitol at room temperature. After discarding the supernatant, a small amount of 8% mannitol solution was added to resuspend the protoplasts. Trypan blue staining was applied to confirm the quality (<10% stained) of the protoplasts. The concentration of the protoplasts was adjusted to 1000–1200 cells/μl and then processed with 10x Genomics Single Cell Protocol (CG00052, RevC).

scRNA-seq library preparation and sequencing

The isolated protoplasts were processed through the 10x Chromium 3' Single Cell Platform according to the manufacturer's instructions (10x Genomics). Briefly, 18 000 cells in one batch were loaded on a single-cell chip for library construction using Chromium Single Cell 30 v2 Reagent Kit. The protoplasts partitioned into single-cell GEMs (gel beads in emulsion) followed by cell lysis and barcoded reverse transcription of RNA in the droplets. Quantification of DNA library was done by an Agilent 2100 Bioanalyzer, and sequencing was performed with an Illumina HiSeq2000 sequencer.

Cell clustering and identification of marker genes

scRNA-seq data were aligned to the Nip and 93-11 reference genomes (Zhang et al., 2018), respectively, and counted using the Cell Ranger pipelines (version 2.0, 10x Genomics). The raw count matrix data were imported into R using the Seurat (V3.1.0) package for further data analysis (Butler et al., 2018). Cells with fewer than 200 detected genes and genes detected in fewer than 20 cells were removed. Each

(C) Expression of cell-type marker genes on the pseudotime trajectory. The colors represent expression levels of these genes in individual cells.

(D) Heatmap showing the expression of 534 most significant DEGs in three clusters across the pseudotime. Each row represents one gene, the DEGs were clustered into three clusters with distinct expression patterns. Color bar indicates the relative expression level. Cluster 1 genes are enriched in cells on the prebranch of the trajectory, cluster 2 genes are enriched in cells on branch 2 of the trajectory, and cluster 3 genes are enriched in cells on the branch 1 end of the trajectory. Scatter plots of representative GO terms for each cluster are shown on the right.

(E) Expression of *LOC_Os07g43670* on the pseudotime trajectory (left panel), and *in situ* hybridization validation of its expression in Nip and 93-11 (right panel). Scale bars, 40 μm.

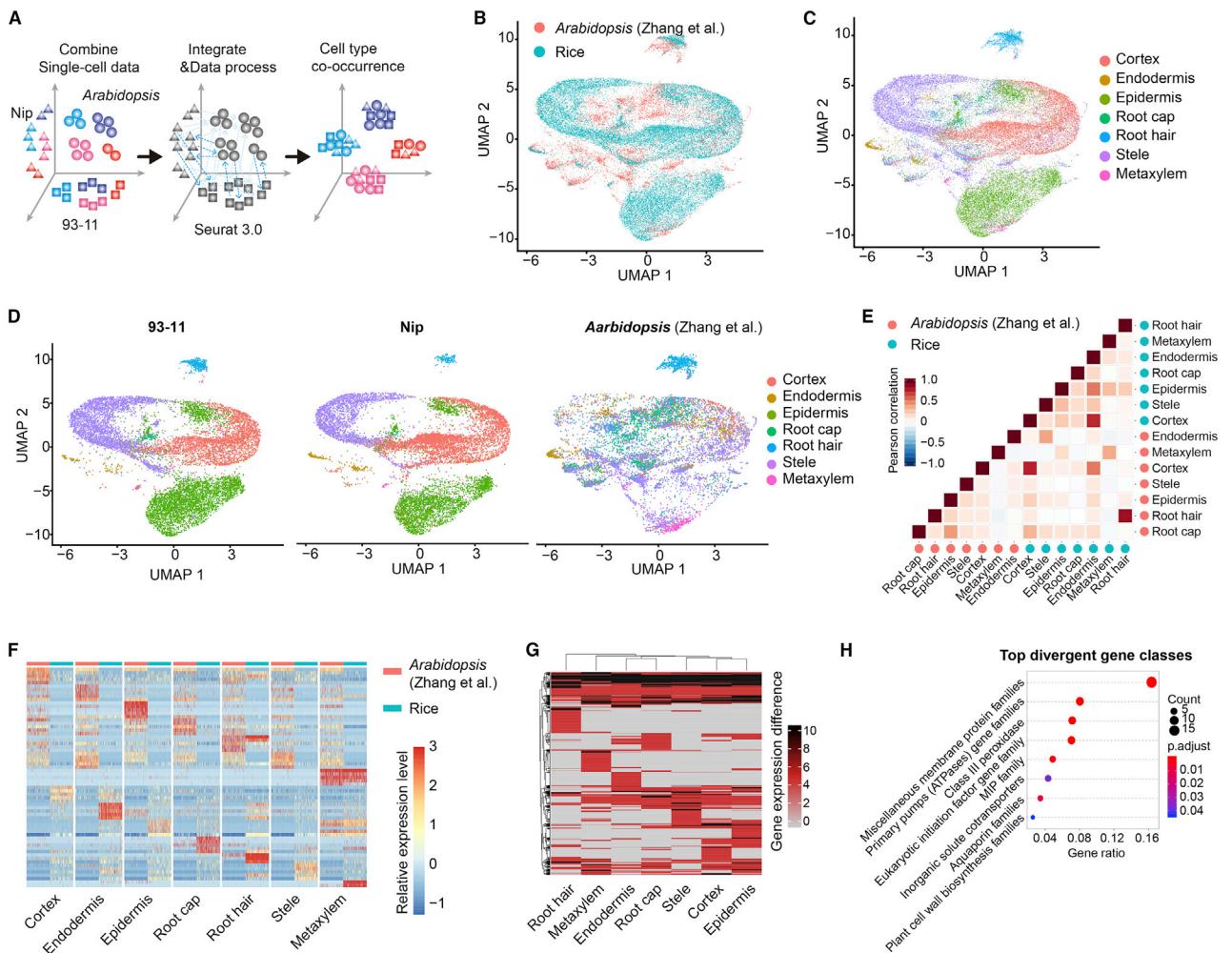


Figure 4. Evolutionary conservation and divergent of cell types between rice and *Arabidopsis*.

- (A) Schematic of alignment and clustering of combined rice (Nip and 93-11) and *Arabidopsis* root samples using Seurat 3.0.
- (B and C) UMAP visualization of Nip, 93-11, and *Arabidopsis* clusters after alignment. The colors indicate species (B) or cell types (C).
- (D) UMAP visualization of separated of *Arabidopsis*, rice Nip, and 93-11 single cells.
- (E) Heatmap showing Pearson's correlations between rice and *Arabidopsis* cell-type transcriptomes. The single-cell data were merged together for each cell type.
- (F) Ortholog marker gene expression for each cell type reveals similar and divergent patterns between rice and *Arabidopsis*.
- (G) Patterns of expression changes between rice and *Arabidopsis* for 1437 divergent genes (15% of the total number of orthologous genes). Divergent genes are defined as gene expression difference >5-fold change.
- (H) Gene families with the most divergent expression patterns. Hypergeometric distribution test was used as a significance enrichment test.

individual dataset was scaled and normalized by the ScaleData and NormalizeData functions. The top 2000 highly variable genes were used for principal component analysis (PCA) dimensionality reduction. The first 20 principal components (PCs) were selected according to the PCA elbow plot and used for clustering with resolution parameter 0.3. The clusters were visualized and explored by t-SNE (Van der Maaten and Hinton, 2008) and UMAP (Becht et al., 2019). Marker genes (cluster-enriched genes) were identified using function FindAllMarkers in the Seurat package. Differential expression analysis was performed based on the Wilcoxon rank-sum test. The cluster-enriched genes were detected under the parameters of min.pct = 0.5 and min.diff.pct = 0.3. The final cluster annotation was validated by *in situ* hybridization experiments.

RaceID was used as an independent clustering approach to analyze rice scRNA-seq data. RaceID has some improvements in detection of cell types with small cell number, offers batch effect removal utilities, and of-

fers optional imputing of gene expression. RaceID was run with the following parameters: mintotal = 200, minexpr = 5, minnumber = 5, samp = 1000, clustnr = 30, FUNcluster = "hclust", perplexity = 60.

Bulk RNA-seq and analysis

Bulk RNA-seq was applied to RNA extracted from 3-day-old root tips (5 mm) of crown roots and protoplasts, which were independently isolated from 3-day-old root tips (5 mm) of crown roots from Nip and 93-11, respectively. RNA-seq was performed as previously described (Liang et al., 2018). Briefly, total RNA was extracted using the RNeasy Plus Mini Kit (Qiagen). The amount and quality of RNA were tested by Qubit RNA Assay Kit (Life Technologies), gel electrophoresis, and Agilent Bioanalyzer 2100 system. Oligo(dT)-attached magnetic beads were used to purify mRNA. Nextera XT DNA Library Preparation Kit (Illumina) was used to construct the library, and the amplified library was sequenced

on the HiSeq X Ten platform. Sequence reads were trimmed using SOAPnuke (version 1.5.6) (Chen et al., 2018) and aligned to the Nip and 93-11 reference genomes, respectively, with HISAT2 (version 2.0.4) (Kim et al., 2019). Gene expression values were calculated on uniquely mapped reads using HTSeq (version 0.11.2) (Anders et al., 2015), and DEseq2 (version 1.4.5) was used to calculate differentially expression (fold change >2 and $P < 0.01$) (Love et al., 2014).

Toluidine blue staining

The 3-day-old root tips (5 mm) of the crown roots from Nip and 93-11 were cut and then fixed in the formaldehyde acetic acid solution (50% ethanol, 5% acetic acid, and 3.7% formaldehyde). The fixed tissue was dehydrated through an ethanol series and HistoClear. Then samples were embedded in paraffin, cut into paraffin sections with a thickness of 8 μm , and mounted on microscope slides (Fisher Scientific). Sections of primary roots from Nip and 93-11 were stained with 0.01% toluidine blue. Sections were first deparaffinized and hydrated to distilled water. After staining in toluidine blue solution for 90 s, sections were washed in distilled water and dehydrated quickly through an ethanol series, and the coverslip covered with resinous mounting medium. After coloration, the samples were sealed by cover slips and observed under light microscope (Leica DM6 B).

RNA *in situ* hybridization

The specific regions of marker genes were cloned into the pGEM-TEasy (Promega) vector, then Digoxigenin RNA labeling kit (Roche) was used for *in vitro* transcription and labeling. The hybridization and immunological detection were performed as described previously (Zhao et al., 2009; Liang et al., 2015), and microscopy was carried out in bright-field mode using Leica DM6 B. The primers were listed in *Supplemental Table 8*.

Vector construction, plant transformation, and microscopy

To create transcriptional reporter lines, around 2-kb promoter fragments of selected genes were amplified and followed by cloning into pCAMBIA1300-GFP (Pu et al., 2017) or pCAMBIA1300-3xVenus-NLS. All constructs were sequenced before being introduced into rice Nip via *Agrobacterium tumefaciens* EHA105-mediated transformation. The primer sequences are listed in *Supplemental Table 8*. More than 20 positive transgenic rice seedlings were obtained from each line.

The 3-day-old crown roots from the transcriptional promoter:3xVenus-NLS/GFP reporter lines were used for fluorescent proteins signal observation. Briefly, 5-mm root tips were embedded in 3% agarose and then cut into sections with a thickness of 40 μm and mounted on microscope slides. The GFP/3x Venus were imaged on a Leica DM6 B microscope excited with a white light laser at 488 nm and 515 nm, respectively.

Integrative analysis of the Nip and 93-11 scRNA-seq datasets

Normalized scRNA-seq data from Nip and 93-11 were combined into one object. Variable genes were selected using the SCTransform function with default parameters. The FindIntegrationAnchors function command used default parameters (dims = 1:30) to discover integration anchors across all samples. The IntegrateData function was run on the anchor set with default additional arguments. ScaleData and RunPCA were then performed on the integrated assay to compute 20 PCs. UMAP dimensionality reduction was carried out and a shared nearest neighbor (SNN) graph constructed using dimensions 1:20 as input features and default PCA reduction. Clustering analysis was performed on the integrated assay at a resolution of 0.5.

Pseudotime analysis

The pseudotime analysis of cell differentiation and the determination of cell fate was performed by the Monocle2 (v.2.10.0) R package (Trapnell et al., 2014). To explore the developmental trajectory of specific cell types, a subset of raw data with target clusters was analyzed. The detailed process was as follows: firstly, the variance in each gene's expression

across cells were calculated by the dispersionTable function. Variable genes were chosen based on average expression level to define a developmental trajectory. Second, we reduced the dimensionality of the data to two (set max_components = 2, method = 'DDRTree'). With the expression data in a lower dimensional space, the state transition of single cells was described by orderCells. The trajectory was plotted by plot_cell_trajectory in Monocle 2. To specify *a priori* the beginning of the trajectory, run orderCells again to set the root_state argument. Genes dynamically expressed along the pseudotime were clustered and visualized using the plot_pseudotime_heatmap function. To determine genes that contribute to the branching of developmental trajectory, branch point was selected. The differentialGeneTest function was used to analyze the pseudotime-dependent or branch-dependent genes. The genes that were significantly branch-dependent were visualized by the plot_genes_branches_heatmap function.

Comparative analysis of root scRNA-seq data from rice and *Arabidopsis*

Arabidopsis root scRNA-seq data were obtained from Zhang et al. (2019b). Diamond was used to calculate pairwise similarity scores between *Arabidopsis* and rice for constructing orthology groups (Buchfink et al., 2015), and the homology relationships were assessed using InParanoid (Sonhammer and Orlund, 2015). InParanoid uses a clustering method based on genome-wide pairwise sequence similarity matches to identify putative orthologous proteins between two species and predicts ortholog groups, where each group contains one (the highest sequence similarity matches) or more (with high pairwise similarity matches relative to the best pair) pairs, including those in-paralogs within a defined cutoff value. Finally, a total of 9919 one-to-one orthologous gene pairs were kept for next analysis. An intersection set of the top 1000 highly variable genes from each dataset that are expressed in all samples was used by FindIntegrationAnchors function command with other default parameters (dims = 1:20) to discover integration anchors across all samples. The IntegrateData function was run on the anchor set with default additional arguments. ScaleData and RunPCA were then performed on the integrated assay to compute 20 PCs. UMAP dimensionality reduction was carried out and an SNN graph constructed using dimensions 1:20 as input features and default PCA reduction. Clustering was performed on the integrated assay at a resolution of 0.2. Average expression values of each cluster were log2-transformed and scatter plots and Pearson's correlations were calculated to compare *Arabidopsis* and rice. The other approach was performed by using the Reference-based method in the Seurat package. In the Reference-based method, a subset of the scRNA-seq dataset (or a single dataset) are listed as a reference. This method provides advantages in both efficiency and in finding anchors. The *Arabidopsis* scRNA-seq results from two independent studies (Zhang et al., 2019b; Ryu et al., 2019) were used as references. The functions FindIntegrationAnchors and IntegrateData were used for integrating and combining the datasets before scaling, PCA, and clustering.

Seurat was used to identify differentially expressed genes across homologous cell types between rice and *Arabidopsis*. The fold difference was the threshold to identify divergent (>5-fold), moderate (2- to 10-fold), and small (<2-fold) differences. A heatmap was generated showing expression differences across cell types, and hierarchical clustering using ward.D method was applied to group genes with similar patterns of expression change. To gain insight into the functional divergence of genes, functional gene families ([ftp://ftp.arabidopsis.org/home/tair/Genes/Gene_families/](http://ftp.arabidopsis.org/home/tair/Genes/Gene_families/)) enrichment was analyzed with the divergent genes.

Functional analysis

Cluster-enriched genes with statistical significance were applied to clusterProfiler (Yu et al., 2012). The R package was used for GO enrichment analysis. Rice phytohormone-related genes were manually selected for

visualization of expression profiles, the expression levels of the selected genes were summarized, and the top 20% of the cells with high expression levels were visualized on UMAP.

ACCESSION NUMBERS

The raw data of Illumina sequencing, including single-cell RNA-seq and bulk RNA-seq, have been deposited in the NCBI GEO with the accession number GSE146035. All the other data are available from the corresponding authors upon request.

SUPPLEMENTAL INFORMATION

Supplemental Information is available at *Molecular Plant Online*.

FUNDING

This work was supported by National Transgenic Major Program (2019ZX08010-002) (to X.G.), Central Public-interest Scientific Institution Basal Research Fund (no.Y2020PT06 to X.G.), the DECODE ERC Synergy grant (to J.U.L.), and the intramural research support from Chinese Academy of Agricultural Sciences, Centre for Organismal Studies, Heidelberg University.

AUTHOR CONTRIBUTIONS

Z.L., J.U.L., and X.G. conceived and designed the study. Q.L., Z.L., D.F., Z.D., Y.W., G.H., and P.Z. performed the experiments. Q.L., Z.L., S.J., Y.W., Y.M., R.L., J.U.L., and X.G. analyzed data. Z.L., J.U.L., and X.G. wrote the paper.

ACKNOWLEDGMENTS

No conflict of interest declared.

Received: June 17, 2020

Revised: October 30, 2020

Accepted: December 16, 2020

Published: December 19, 2020

REFERENCES

- Anders, S., Pyl, P.T., and Huber, W.** (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**:166–169.
- Becht, E., McInnes, L., Healy, J., Dutertre, C.A., Kwok, I.W.H., Ng, L.G., Ginhoux, F., and Newell, E.W.** (2019). Dimensionality reduction for visualizing single-cell data using t-SNE. *Nat. Biotechnol.* **37**:38–44.
- Boron, A.K., Van Orden, J., Nektarios Markakis, M., Mouille, G., Adriaensen, D., Verbelen, J.P., Hofte, H., and Vissenberg, K.** (2014). Proline-rich protein-like PRPL1 controls elongation of root hairs in *Arabidopsis thaliana*. *J. Exp. Bot.* **65**:5485–5495.
- Buchfink, B., Xie, C., and Huson, D.H.** (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**:59–60.
- 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., Packer, J.S., Ramani, V., Cusanovich, D.A., Huynh, C., Daza, R., Qiu, X., Lee, C., Furlan, S.N., Steemers, F.J., et al.** (2017). Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357**:661–667.
- Chen, Y., Chen, Y., Shi, C., Huang, Z., Zhang, Y., Li, S., Li, Y., Ye, J., Yu, C., Li, Z., et al.** (2018). SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Gigascience* **7**:1–6.
- 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.
- Ding, W., Yu, Z., Tong, Y., Huang, W., Chen, H., and Wu, P.** (2009). A transcription factor with a bHLH domain regulates root hair development in rice. *Cell Res.* **19**:1309–1311.
- Efroni, I., and Birnbaum, K.D.** (2016). The potential of single-cell profiling in plants. *Genome Biol.* **17**:65.
- Gutjahr, C., Sawers, R.J.H., Marti, G., Andres-Hernandez, L., Yang, S.Y., Casieri, L., Angliker, H., Oakeley, E.J., Wolfender, J.L., Abreu-Goodger, C., et al.** (2015). Transcriptome diversity among rice root types during symbiosis and interaction with arbuscular mycorrhizal fungi. *Proc. Natl. Acad. Sci. U S A* **112**:6754–6759.
- Hoffmann, R.D., Olsen, L.I., Ezike, C.V., Pedersen, J.T., Manstretta, R., Lopez-Marques, R.L., and Palmgren, M.** (2019). Roles of plasma membrane proton ATPases AHA2 and AHA7 in normal growth of roots and root hairs in *Arabidopsis thaliana*. *Physiol. Plant* **166**:848–861.
- Ishimaru, Y., Kakei, Y., Shimo, H., Bashir, K., Sato, Y., Sato, Y., Uozumi, N., Nakanishi, H., and Nishizawa, N.K.** (2011). A rice phenolic efflux transporter is essential for solubilizing precipitated apoplastic iron in the plant stele. *J. Biol. Chem.* **286**:24649–24655.
- Jean-Baptiste, K., McFaline-Figueroa, J.L., Alexandre, C.M., Dorrrity, 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.
- Kim, C.M., Park, S.H., Il Je, B., Park, S.H., Park, S.J., Piao, H.L., Eun, M.Y., Dolan, L., and Han, C.D.** (2007). OsCSLD1, a cellulose synthase-like D1 gene, is required for root hair morphogenesis in rice. *Plant Physiol.* **143**:1220–1230.
- Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L.** (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**:907–915.
- Liang, Z., Brown, R.C., Fletcher, J.C., and Opsahl-Sorteberg, H.G.** (2015). Calpain-mediated positional information directs cell wall orientation to sustain plant stem cell activity, growth and development. *Plant Cell Physiol.* **56**:1855–1866.
- Liang, Z., Shen, L., Cui, X., Bao, S., Geng, Y., Yu, G., Liang, F., Xie, S., Lu, T., Gu, X., et al.** (2018). DNA N⁶-adenine methylation in *Arabidopsis thaliana*. *Dev. Cell* **45**:406–416.e3.
- Love, M.I., Huber, W., and Anders, S.** (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**:550.
- Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al.** (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**:1202–1214.
- Pu, C.X., Han, Y.F., Zhu, S., Song, F.Y., Zhao, Y., Wang, C.Y., Zhang, Y.C., Yang, Q., Wang, J., Bu, S.L., et al.** (2017). The rice receptor-like kinases DWARF AND RUNTISH SPIKELET1 and 2 repress cell death and affect sugar utilization during reproductive development. *Plant Cell* **29**:70–89.
- Qian, Q., Guo, L.B., Smith, S.M., and Li, J.Y.** (2016). Breeding high-yield superior quality hybrid super rice by rational design. *Natl. Sci. Rev.* **3**:283–294.
- Rebouillat, J., Dievart, A., Verdeil, J.L., Escoute, J., Giese, G., Breitler, J.C., Gantet, P., Espeout, S., Guiderdoni, E., and Perin, C.** (2009). Molecular genetics of rice root development. *Rice* **2**:15–34.
- Rich-Griffin, C., Stechemesser, A., Finch, J., Lucas, E., Ott, S., and Schafer, P.** (2020). Single-cell transcriptomics: a high-resolution avenue for plant functional genomics. *Trends Plant Sci.* **25**:186–197.
- 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.

- 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., et al.** (2019). High-throughput single-cell transcriptome profiling of plant cell types. *Cell Rep.* **27**:2241–2247.e4.
- Sonnhammer, E.L.L., and Ostlund, G.** (2015). InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. *Nucleic Acids Res.* **43**:234–239.
- Takehisa, H., Sato, Y., Igarashi, M., Abiko, T., Antonio, B.A., Kamatsuki, K., Minami, H., Namiki, N., Inukai, Y., Nakazono, M., et al.** (2012). Genome-wide transcriptome dissection of the rice root system: implications for developmental and physiological functions. *Plant J.* **69**:126–140.
- Tang, F., Lao, K., and Surani, M.A.** (2011). Development and applications of single-cell transcriptome analysis. *Nat. Methods* **8**:6–11.
- Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L.** (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**:381–386.
- Van der Maaten, L., and Hinton, G.** (2008). Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**:2579–2605.
- Vanstraelen, M., and Benkova, E.** (2012). Hormonal interactions in the regulation of plant development. *Annu. Rev. Cell Dev. Biol.* **28**:463–487.
- Wang, J.W., Wang, L.J., Mao, Y.B., Cai, W.J., Xue, H.W., and Chen, X.Y.** (2005). Control of root cap formation by microRNA-targeted auxin response factors in *Arabidopsis*. *Plant Cell* **17**:2204–2216.
- Yu, G., Wang, L., Han, Y., and He, Q.** (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* **16**:284–287.
- Zeisel, A., Munoz-Manchado, A.B., Codeluppi, S., Lonnerberg, P., La Manno, G., Jureus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., et al.** (2015). Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**:1138–1142.
- Zhang, G., Guo, G., Hu, X., Zhang, Y., Li, Q., Li, R., Zhuang, R., Lu, Z., He, Z., Fang, X., et al.** (2010). Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. *Genome Res.* **20**:646–654.
- Zhang, J., Liu, Y., Zhang, N., Hu, B., Jin, T., Xu, H., Qin, Y., Yan, P.X., Zhang, X., Guo, X., et al.** (2019a). NRT1.1B is associated with root microbiota composition and nitrogen use in field-grown rice. *Nat. Biotechnol.* **37**:676–684.
- Zhang, Q., Liang, Z., Cui, X., Ji, C., Li, Y., Zhang, P., Liu, J., Riaz, A., Yao, P., Liu, M., et al.** (2018). N⁹-methyladenine DNA methylation in *Japonica* and *Indica* rice genomes and its association with gene expression, plant development, and stress responses. *Mol. Plant* **11**:1492–1508.
- Zhang, T., Xu, Z., Shang, G., and Wang, J.** (2019b). A single-cell RNA sequencing profiles the developmental landscape of *Arabidopsis* root. *Mol. Plant* **12**:648–660.
- Zhao, Y., Hu, Y., Dai, M., Huang, L., and Zhou, D.** (2009). The WUSCHEL-related homeobox gene *WOX11* is required to activate shoot-borne crown root development in rice. *Plant Cell* **21**:736–748.