Single-cell transcriptome atlas and chromatin accessibility landscape reveal differentiation trajectories in the rice root

Tian-Qi Zhang, Yu Chen, Ye Liu, Wen-Hui Lin & Jia-Wei Wang

Root development relies on the establishment of meristematic tissues that give rise to distinct cell types that differentiate across defined temporal and spatial gradients. Dissection of the developmental trajectories and the transcriptional networks that underlie them could aid understanding of the function of the root apical meristem in both dicots and monocots. Here, we present a single-cell RNA (scRNA) sequencing and chromatin accessibility survey of rice radicles. By temporal profiling of individual root tip cells we reconstruct continuous developmental trajectories of epidermal cells and ground tissues, and elucidate regulatory networks underlying cell fate determination in these cell lineages. We further identify characteristic processes, transcriptome profiles, and marker genes for these cell types and reveal conserved and divergent root developmental pathways between dicots and monocots. Finally, we demonstrate the potential of the platform for functional genetic studies by using spatiotemporal modeling to identify a rice root meristematic mutant from a cell-specific gene cohort.

https://doi.org/10.1038/s41467-021-22352-4

1 National Key Laboratory of Plant Molecular Genetics (NKLPMG), CAS Center for Excellence in Molecular Plant Sciences, Institute of Plant Physiology and Ecology (SIPPE), Chinese Academy of Sciences (CAS), Shanghai, China. 2 University of Chinese Academy of Sciences, Shanghai, China. 3 State Key Laboratory of Crop Genetics and Germplasm Enhancement, Key Laboratory of Landscaping, Ministry of Agriculture and Rural Affairs, College of Horticulture, Nanjing Agricultural University, Nanjing, China. 4 Joint International Research Laboratory of Metabolic & Developmental Sciences, School of Life Sciences & Biotechnology, Joint Center for Single Cell Biology, Shanghai Jiao Tong University, Shanghai, China. 5 ShanghaiTech University, Shanghai 200031, China.

✉ email: tianqizhang@yeah.net; jwwang@sippe.ac.cn
Many dicotyledonous plants such as Arabidopsis have a root system composed of a single primary root and numerous lateral roots. By contrast, cereals such as *Oryza sativa* (rice) develop a dense fibrous root system consisting primarily of numerous postembryonic adventitious roots that develop from the stem (Supplementary Note 1). For both rice and Arabidopsis, root apical meristem (RAM) is formed during embryogenesis which generates the radicle or primary root after seed germination. The RAM is comprised of a stem cell niche (SCN) and undifferentiated small dividing cells in both dicots and monocots (Fig. 1b). In Arabidopsis, the SCN harbors quiescent (SCN) and undifferentiated small dividing cells in both dicots and Arabidopsis, root apical meristem (RAM) is formed during embryogenesis which generates the radicle or primary root after seed germination. The RAM is comprised of a stem cell niche (SCN) and undifferentiated small dividing cells in both dicots and monocots (Fig. 1b). In Arabidopsis, the SCN harbors quiescent center (QC) and surrounding stem cells which give rise to diverse root cell types including epidermis, root hairs, root cap, cortex, endodermis, and stele (pericycle, phloem, xylem, and procambium) through cell division, expansion, and differentiation.

Rice is usually grown under flooding conditions and has developed distinct anatomy to adapt to this environment. Its radial root organization differs greatly from that of dicots such as Arabidopsis (Fig. 1b). The outer cell layers of rice roots are comprised of the epidermis, exodermis, and sclerenchyma. The exodermis (hypodermis with Casparian bands) of plant roots represents the barrier of variable resistance to the radial flow of both water and solutes. Unlike other cereal crops, the rice root features a sclerenchyma layer which is hyper-suberized compared with the exodermis and consists of tightly packed small cells. The exodermis and sclerenchyma together form a barrier that reduces radial oxygen loss. The ground tissue of rice roots is composed of between 7 and 12 layers of cortical cells and one layer of endodermal cells (Fig. 1b). On the contrary, Arabidopsis roots only harbor one cortical and one endodermal cell layer. The increased number of rice cortical layers provides a venue for lysigenous aerenchyma formation and plays an essential role in adaptation to waterlogging conditions. Taken together, although the anatomy of the rice root apical meristem has been intensively investigated, the molecular definition of its cell types is largely incomplete.

![Fig. 1 Generation of a rice radicle cell atlas.](image)

**a** UMAP visualization of 21 cell clusters in rice radicles. Each dot denotes a single cell. Colors denote corresponding cell clusters. EMC, epidermal meristematic cell. **b** Schematic of anatomy of rice radicle. **c** Visualization of cell clusters by 3D UMAP scatterplots. Cluster names and colors are the same as in **a.** **d** Expression patterns of representative cluster-specific marker genes on UMAP. Dot diameter indicates the proportion of cluster cells expressing a given gene. The full names of selected genes are given in Supplementary Data 3.
Recent advances in single-cell RNA sequencing (scRNA-seq) technology provide unprecedented opportunities to systematically identify the entire cellular and molecular differentiation trajectory of plant stem cells at the single-cell level. Several scRNA-seq studies have revealed that Arabidopsis root tip cells have highly heterogeneous transcriptomes. The transcriptome analysis of the individual cells at different developmental stages reveals a continuous differentiation trajectory of root development following their spatial distribution and temporal order.

Results

Generation of a rice radicle cell atlas. To perform scRNA-seq of rice roots, the radicles of wild-type (ZH11) rice seedlings were harvested (~1.0 cm in length from root tip, n = 90), digested into protoplasts (plant cells without cell wall), and subjected to scRNA-seq assay using commercial 10× Chromium platform (10× Genomics, Supplementary Data 1). Two biological replicates were performed. The protoplasting and mitochondrial genes had little effects on the clustering (see Methods; Supplementary Fig. S1a, c). The batch-effect between samples was removed by the Harmony algorithm (Supplementary Fig. S1b, e-g). We used standard computational pipelines (Cell Ranger provided by 10× Genomics) to align the raw sequencing data to the rice genome, and derived a gene expression matrix of 29,919 genes across 27,469 filtered cells.

After selecting highly variable genes, we used principal component analysis (PCA) on a gene expression matrix across 2000 variable genes and identified 100 statistically significant principal components (PCs) (p < 0.05). Using Seurat, these PCs were used to build a k-nearest neighbor graph of the cells, which was then partitioned into 21 transcriptionally distinct clusters. The analysis of cell cycle genes revealed that transcriptome heterogeneity was not dominated by cell cycle status (Supplementary Fig. S1h). To reveal local similarities and global structures of cell populations, the t-distributed stochastic neighborhood embedding (t-SNE) tool and uniform manifold approximation and projection (UMAP) algorithm were used (Fig. 1a; Supplementary Fig. S1d, s). Compared to the t-SNE, the UMAP offers faster runtime and consistency, meaningful organization of cell clusters, and preservation of continuums. In contrast, the main advantage of the t-SNE is the ability to preserve local structure. Three-dimensional (3D) UMAP scatterplots were further constructed to determine the spatial and connective distribution of cell clusters (Fig. 1c; Supplementary Movie 1). By analyzing differentially expressed genes among the clusters, we identified a series of cluster-enriched genes and cluster-specific marker genes (Fig. 1d; Supplementary Data 2).

Because of the lack of specific markers for most cell types, we used the following two strategies to faithfully annotate cell clusters in the rice radicle cell atlas. First, we identified rice genes whose biological functions or expression patterns have been well studied (Supplementary Data 2; Supplementary Data 3; Supplementary Note 2 and 3). The examination of their expression pattern on the UMAP helped us to assign some cell clusters. Second, we performed RNA in situ hybridization assays for over 50 marker genes (see below, Supplementary Fig. S10). The expression pattern of these genes not only confirmed the above annotations, but also enabled us to assign unknown cell clusters. Overall, our results indicate that the rice root tip is composed of highly heterogeneous cells. The single-cell transcriptome atlas enabled us to identify most of the major cell types of the rice radicle (Supplementary Note 2). Clusters 1, 4, and 9 belonged to the epidermis/root hair population (Fig. 1a, see below). These clusters were separated from other cell clusters on the UMAP plot (Fig. 1a), suggesting a unique transcriptome signature. The meristematic cell clusters (Clusters 5, 11, 18, and 19) localized in the center of the UMAP (Supplementary Note 2). Differentiation trajectories such as ground tissue (Clusters 0, 2, 3, 6, 8, and 13), pericycle (cluster 15) and vascular tissue (Clusters 7, 10, 12, 16, and 17) radiated away from the meristematic cell clusters (Fig. 1a, c). Notably, our gene expression and signature analyses provide fresh insights into cell type-specific physiological properties.

The developmental trajectories of epidermal cells. Since cells undergoing transition from one state to another (i.e., intermediate states) can be captured, scRNA-seq enables the exploration of the continuous differentiation trajectory of a developmental process. Therefore, we first aimed to deduce the developmental trajectory of root epidermal cells. Re-clustering of three clusters belonging to the epidermis/root hair population (clusters 1, 4, and 9) revealed ten sub-cell clusters, named E0 to E9 (Fig. 2a). Interestingly, UMAP plotting revealed that cluster 9 topologically bifurcated into two trajectories, dominated by the cells in either cluster 1 or 4 (Fig. 2a). To validate this result, we conducted psuedotime analysis by ordering cells along a reconstructed trajectory using Palantir (Fig. 2b). Consistent with distribution distance on the UMAP, inferred trajectories demonstrated gradual transitions from cells in cluster 9 to early, mid, and late root-hair cells (clusters 4, trichoblasts) or to nonhair cells (clusters 1, atrichoblasts) (Fig. 2a; Supplementary Fig. S2a). Moreover, differentiation potential analysis indicated that the cells in clusters 1 and 4 were more differentiated than those in cluster 9 (Fig. 2c). Thus, these analyses collectively demonstrate that the cells in cluster 9 serve as epidermal meristematic cells. Along with cell division and differentiation, these cells progressively differentiate into nonhair cells (cluster 1) or root hair cells (cluster 4).

We chose Os10g0452700, Os03g0155500, Os01g0248900, and Os10g0578200 as representative genes for clusters 9, 1, and 4, respectively (Fig. 2d, e; Supplementary Fig. S2b). A survey of our scRNA-seq dataset revealed that the expression of Os10g0452700 gradually decreased along pseudotime, followed by increased transcription of Os03g0155500 (Fig. 2e). The expression of Os10g0248900 was elevated at the late developmental stage of atrichoblasts, while the progressive increase of Os10g0578200 expression was only observed in the trichoblast lineage (Fig. 2e).

The analyses of fluorescent (Venus-N7) reporters generally confirmed the expression patterns inferred by the scRNA-seq dataset. For example, expression of the Os10g0452700 reporter was mainly observed in epidermal cells of the root meristematic zone and decreased along with root development (Fig. 2f; Supplementary Fig. S2c). In contrast, the Os03g0155500 reporter only became active in meristematic cells that had started to differentiate (Fig. 2g). The Os01g0248900 and Os10g0578200 reporters showed complementary expression patterns (Fig. 2h, i): Os01g0248900 was strongly expressed in atrichoblasts, whereas the promoter of Os10g0578200 was exclusively activated in trichoblasts. Gene ontology (GO) analyses revealed further differences in the biological processes of atrichoblasts and trichoblasts: cluster 1 (atrichoblasts) was enriched for genes involved in cell wall biosynthesis, whereas cluster 4 (trichoblasts) showed gene signatures for fatty acid metabolic process (Supplementary Fig. S2d, e; Supplementary Data 4). Thus, the combination of scRNA-seq, psuedotime inference and reporter analyses enables us to reconstruct the progression of epidermal cell fate determination during atrichoblast and trichoblast development in rice.

The developmental trajectories of ground tissues. A major difference between monocots and dicots lies in the cell lineages of the ground tissues and epidermis. In monocots, the ground…
tissues and epidermis belong to a single common initial, while in dicots, the lateral root cap and epidermis share a common initial cell (Fig. 1b). The anticlinal division of the rice ground tissue/epidermis and atrichoblasts share a common initial cell. The asymmetrical division of this daughter cell separates the epidermis and ground tissue lineages. Successive periclinal divisions of the ground tissue initial cell generate exodermis, sclerenchyma layers, cortex, and endodermis.

On the basis of functional annotations, clusters 3, 6, and 0 were designated as exodermis, sclerenchyma cell layer, and cortex, respectively (Fig. 1a). Intriguingly, closer examination of the distribution distances on the 3D UMAP and t-SNE plots revealed that all these clusters were connected to cluster 11 (Fig. 1a; Fig. 3a; Supplementary Fig. S3a), which was annotated as meristematic cells (Fig. 1a). Thus, this topology suggests that cluster 11 may serve as common undifferentiated progenitor cells for root ground tissue (Fig. 3a; Supplementary Fig. S3a). In favor of this hypothesis, RNA velocity, ForceAtlas2 and differentiation potential analyses confirmed that the cells in cluster 11 reside in the “source” cell state, and those in clusters 3, 6, and 0 in end states (Fig. 3b–d; Supplementary Fig. S3b–e).

Gene survey analysis uncovered several transcription factors highly expressed in cluster 11 (Fig. 3e). Expression of these genes, for example OsGATA6 and OsERF108 (ETHYLENE RESPONSE FACTOR 108), was progressively decreased along pseudotime (Fig. 3f). In line with this expression pattern, RNA in situ hybridization assay confirmed that OsGATA6 and GROWTH REGULATING FACTOR6 (OsGRF6) transcripts predominantly accumulated in the center of the RAM above the QC, and gradually decreased along the cell division zone until reaching the
The above results indicate that cluster 11 is composed of meristematic cells, giving rise to distinct cell types of ground and vascular tissues in the rice root tip.

To obtain a better understanding of the gene regulatory basis underlying ground tissue differentiation, we re-clustered the cells from clusters 0, 3, 6 and 11, and performed pseudotime and embedded heatmap analyses. We delineated three distinct cell lineages, leading to the formation of the exodermis, sclerenchyma cell layer and cortex (Fig. 3g). GO analyses indicated that the terminally differentiated cells in these lineages exhibited distinct gene signatures (Supplementary Fig. S3f–h; Supplementary Data 4). In addition, the transcription factors associated with their differentiation exhibited temporal expression patterns along pseudotime, and did not overlap (Fig. 3h–j). Taken together, our results reveal a differentiation continuum of meristematic cells toward distinct ground tissues. Importantly, the inferred transcription factors provide a valuable resource for identification of key regulators for ground tissue differentiation.

The above analysis revealed that OsGATA6 is expressed in the center of the RAM above the QC with graded expression along the cell division gradient, suggesting that it may play an important role in the differentiation of ground and vascular tissue. To test this hypothesis, we generated OsGATA6 knockout mutants using clustered regularly interspaced short palindromic repeats (CRISPR)-Cas9 technology (Supplementary Fig. S4a)35–39. The homozygous Osgata6 mutants were dwarfed and developed short roots (Supplementary Fig. S4c). At the tissue level, the mutants had a shorter RAM, and exhibited defects in cell division and differentiation (Supplementary Fig. 4d). Thus, these findings substantiate the precision of scRNA-seq in identification of rice root mutants at the cell-specific level.
Inference of transcriptional regulatory basis underlying cell differentiation by integrative analysis of ATAC-seq and scRNA-seq. To further characterize the molecular basis for cell differentiation in the rice root tip, we performed assays for transposable-accessible chromatin sequencing (ATAC-seq), which probe changes in chromatin accessibility at a genome-wide level. To this end, we harvested the root tissues from the meristematic zone (MZ) and elongation zone (EZ) (Fig. 4a). The comparison between the MZ and EZ samples revealed massive changes in chromatin accessibility (Fig. 4b, c; Supplementary Fig. S5a–c; Supplementary Data 5), suggesting that root meristematic cells undergo chromatin level reprogramming at chromatin level during cell differentiation.

We used the HOMER algorithm to characterize transcription factor binding motifs associated with variations in chromatin accessibility. As shown in Fig. 4d, the MZ and EZ samples showed a distinct enrichment of transcription factor binding motifs and biological processes (Supplementary Fig. Ssd; Supplementary Data 6). The binding motifs for transcription factor families such as bHLH, bZIP and GATA were highly enriched in the MZ, whereas MYB transcription factor binding sites were overrepresented in the EZ (Fig. 4d; Supplementary Fig. S5e). Consistent with these results, OsGATA6 (GATA transcription factor) and BRASSINOESTEROID UPIRGULATED 1-LIKE1 (OsBULL1, a bHLH transcription factor), two genes highly expressed in cluster 11 (meristematic cells corresponding to exodermis, sclerenchyma cell layer, and cortex cells), exhibited open chromatin status only in the MZ (Fig. 4e). In contrast, the regulatory regions of two MYB transcription factor genes (cluster 12), MYB12 and MYB52, were closed in the MZ and became accessible in the EZ (Fig. 4e).

To test whether the chromatin accessibility of a given gene is correlated with its expression level during rice root development, we performed integrative analyses of scRNA-seq and ATAC-seq datasets. To this end, we assigned differentially accessible genes to the cell clusters revealed by our scRNA-seq experiment (“Methods”). The clusters belonging to meristematic cells (e.g., clusters 9 and 11) showed a high proportion of the genes with higher accessibility in the MZ, whereas MYB transcription factor binding sites were overrepresented in the EZ (Fig. 4d; Supplementary Fig. S5e). The comparison between the MZ and EZ samples revealed massive changes in chromatin accessibility (Fig. 4b, c; Supplementary Fig. S5a–c; Supplementary Data 5), suggesting that root meristematic cells undergo chromatin level reprogramming at chromatin level during cell differentiation.

We next grouped Arabidopsis and rice scRNA-seq datasets and performed cell clustering analysis, which resulted in 30 super cell clusters (Fig. 5b; Supplementary Fig. S8). Since Arabidopsis and rice shared the highest one-to-one similarities for root hair, phloem, and xylem supercell clusters (Fig. 5b, c), we characterized them in greater details. For the root hair super cluster, gene clustering analyses revealed a list of core (conserved) genes that were highly expressed in root hairs in both species (Fig. 5d, right panel). Notably, some of these genes are known to play important regulatory roles in root hair development (Supplementary Data 9).

Inter-species comparison reveals conserved and divergent root developmental pathways. Finally, because Arabidopsis root scRNA-seq datasets are already available, we wanted to explore conservation and divergence of the root cell types across the monocot-dicot divide. To this end, we first integrated three published Arabidopsis root scRNA-seq datasets and re-processed cell clustering and assignment. In total, 22 cell clusters were identified (Supplementary Fig. S7). To compare gene expression across Arabidopsis and rice, we defined a set of one-to-one homologous genes (n = 9,727; Supplementary Data 7) across species using OrthoMCL. Pairwise comparisons of root cell clusters of Arabidopsis and rice revealed a relatively high degree of similarities in the clusters corresponding to meristematic, epidermis (nonhair and root hair cells), phloem and xylem cells (Fig. 5a). Interestingly, the rice endodermis cluster showed a low correlation with the corresponding cluster from Arabidopsis but showed a positive correlation to Arabidopsis vascular tissue without phloem and xylem cells (VC; Fig. 5a), suggesting these cell types are functionally aligned. The correlation between rice exodermis and Arabidopsis endodermis suggests that they are homologous tissues (Fig. 5a), consistent with their functionalities as paracellular transport barriers.

Discussion
A fundamental problem in developmental biology is understanding how stem cells give rise to different cell types. To address this question, it is necessary to define the major cell types within a given tissue at both the anatomical and molecular levels. Compared to Arabidopsis, knowledge of rice root development is still limited. Although the anatomy of the rice root apical meristem has been intensively investigated, the molecular definition of its cell types is largely incomplete. Particularly, the inference of rice root cell types based on the expression of orthologous Arabidopsis genes is difficult because of variation in the number and nature of the cell types in monocots and dicots. For instance, the rice root contains exodermis and sclerenchyma cell layers, which are absent from Arabidopsis. Furthermore, the multilayered cortex of the rice root can differentiate into aerenchyma, an anatomical adaptation to water submergence, at late developmental stages, a capacity that is absent in Arabidopsis.

The expression map of the rice root tip at single-cell resolution presented in this study provides a valuable resource for defining different cell types. Based on our scRNA-seq and cell clustering, we successfully identified 21 cell types in the rice root tip. Importantly, in situ hybridization assays and promoter-reporter analysis revealed specific marker genes for most root tip cell types.
factor family genes. The genomic loci are shown, and the representative genes are highlighted in black. The expression patterns in differentiation toward trichoblast (T) and atrichoblast (A) along the pseudotime are shown (right). Color annotation is the same in Fig. 2e.

(Supplementary Fig. S10). For example, we found that the transcripts of Os04g0125700 are highly abundant in the root exodermis (Supplementary Fig. S10), whereas Os01g0248900 and Os10g0578200 are exclusively expressed in atrichoblasts and trichoblasts, respectively (Fig. 2h, i). Cell clusters corresponding to the cortex, exodermis and sclerenchyma layers (cluster 6) were also identified (Fig. 1a). Due to limitations imposed by sequencing depth and gene coverage, lowly expressed genes, which could include important transcription factors for cell fate determination, maybe not faithfully recovered in our scRNA-seq datasets. Nevertheless, the cell-type-specific reporters identified in this work can be used with cell-sorting methods to generate high

Fig. 4 Inference of transcriptional regulatory basis by ATAC-seq and scRNA-seq. a Schematic of tissue sampling. MZ, meristematic zone; EZ, elongation zone. b Volcano plot showing differentially accessible peaks between the MZ and EZ samples. Blue and red, highly accessible peaks in the EZ and MZ samples, respectively; Gray, no difference between the MZ and EZ samples. The number of differential peaks in each sample is given. FDR < 0.05; log2(fold change) > 0.58 or < −0.58. c Pileup of ATAC-seq signals. Heat maps are ranked in decreasing order of ATAC-seq signal. Window size: peak summit ±3.0 kb. d Enrichment of transcription factor binding motifs within differential peaks in the MZ (red, top) and EZ (blue, bottom) samples. −log10(p value) for each binding motif is given. e Representative ATAC-seq tracks for bHLH (OsBUL1), GATA (OsGATA6) and MYB (OsMYB12 and OsMYB52) transcription factor family genes. The genomic loci are shown, and the representative genes are highlighted in black. f Integrative analysis of scRNA-seq and ATAC-seq data. The genes associated with differential peaks identified in e were assigned to 21 cell clusters. The number of genes showing high accessibility in the MZ or EZ in each cell cluster is shown in different colors (top). The ratio (bottom) was calculated by the number of genes showing high accessibility in the MZ or EZ sample/total number of cluster-specific genes. g Representative ATAC-seq tracks for clusters 1, 4, 9 enriched genes. The genomic loci are shown, and the representative genes are highlighted in black. Two representative genes for each cluster are shown. Gene expression patterns during differentiation toward trichoblast (T) and atrichoblast (A) along the pseudotime are shown (right). Color annotation is the same in Fig. 2e.
quality cell-type-specific transcriptomes of the rice root tip, as has been done for Arabidopsis.

A major difference exists in the differentiation of ground tissue and root cap between dicots and monocots; the ground tissue and epidermis belong to a single common initial in rice, whereas lateral root cap and epidermis share a common initial cell in dicots such as Arabidopsis. In monocots, the first asymmetrical division of the epidermis-endodermis initial cell produces two daughter cells, named as the epidermis cell initial and cortex-endodermis initial, with distinct lineages. The epidermis cell initially gives rise to the epidermis through successive periclinal divisions, whereas the cortex-endodermis initial cell subsequently generates the cortex and endodermis progenitor cells through asymmetric anticlinal divisions (Fig. 1c). How these distinct cell fates are specified in a step-wise manner from a common stem cell is largely unknown. Our results identified clusters 1/4/9, 3, 6, 0, and 2/13 as epidermis, exodermis, sclerenchyma cell layer, cortex, and endodermis, respectively (Fig. 1a). Interestingly, pseudotime analysis revealed that the exodermis, sclerenchyma cell layer, and cortex are derived from a common cell cluster (cluster 11). Gene expression analyses indicates that cluster 11 is composed of mixed meristematic cells giving rise to distinct ground and vascular tissue cell types (Supplementary Fig. S4b), although we could not further separate them based on transcriptome differences. In the future, identification of marker genes for the cortex-endodermis initial cell and their use for generation of transgenic plants for cell fate mapping will help to precisely delineate the differentiation trajectories of specific ground tissue cell lineages. Certainly, a similar approach can be used to elucidate how rice root cap cells differentiate from a single columella initial.

Forward genetics and quantitative trait loci mapping have identified a series of mutants showing defects in root meristem maintenance, crown root formation, root hair development, and...
lateral root initiation. Several functional orthologs of Arabidopsis root development mutants have also been studied in rice. However, the number of available rice root mutants remains limited, preventing the accurate modeling of the gene regulatory network underlying rice root development. In this study, we have demonstrated the power of our scRNA-seq resource to accurately guide functional genetic studies of specific cell types. We show that mutation of OsGATA6, which is highly expressed in root meristematic cells (Supplementary Fig. S4), led to defects in cell division and differentiation. Therefore, we envision that exploration of our datasets will accelerate gene discovery and uncover more important regulators of rice root development or physiology in the near future.

Root architecture is a highly dynamic trait and contributes to plant adaptation to its environment. For example, drought tolerance in many plants, including Arabidopsis, rice, and maize, is associated with the ability to generate a deep, wide-spread, highly-branched root system. The comprehensive definition of cell types by scRNA-seq lays the ground for understanding how root development and root system architecture are shaped in response to environmental cues, such as drought, or flooding. Future work that generates roots scRNA-seq and ATAC-seq datasets of plants grown under different conditions or treatments should yield insights into how such environmental signals direct root growth.

Methods

Plant materials and growth conditions. The wild-type O. sativa (ZH11) seedlings were used for root scRNA-seq experiments. Seeds were sown on filter paper soaked with water in petri dishes. After 5 days at 30 °C, the root tips of the seedlings were harvested for scRNA-seq experiment. For promoter Venus-N7 reporter lines, rice plants were grown in a growth chamber at 29 °C in long days. Plant genomic DNA kit (Transgen, Cat No./ID: EE111-02). The regulatory fragments including promoter and 3' UTR regions of candidate genes were PCR amplified into JW1696 using ClonExpress II one Step Cloning kit (Vazyme Biotech). The resultant DNAs were purified according to the protocol. Brieﬂy, the MZ and EZ tissues were harvested and cut into small pieces (0.1 cm in length from root tip) of rice radicles were harvested and digested in the RNase-free enzyme solution (4% cellulase R10, 1.5% macerozyme R10, 0.4 M mannitol, 0.1 M 4-morpholineethanesulfonic acid, 10 mM KCl, 10 mM CaCl$_2$, and 0.1% BSA) for 2 h at room temperature. The protoplasts were filtered 3–4 times with cell strainers (40 μm in diameter, Falcon, Cat No./ID: 352340), concentrated, and washed 3–4 times with 8% mannitol at room temperature. The protoplast viability was determined by trypan blue staining. The ratio of viable cells to total cells for each sample was >85%. The concentration of protoplasts was counted by a hemocytometer, and finally adjusted to 1500–2000 cells/μL.

scRNA-seq library construction and sequencing. The raw scRNA-seq dataset was comprised of 28 bp Read1, 150 bp Read2 and 8 bp i7 index reads.

ATAC-seq experiment. ATAC-seq were performed according to published protocol. Brieﬂy, the MZ and EZ tissues were harvested and cut into small pieces in lysis buffer. The slurry was filtered and resuspended in lysis buffer. The crude nuclei were stained with 4,6-diamidino-2-phenylindole and loaded into a flow cytometer. The doubltess, FACSAria III. The nuclei were sorted once with Tris-Mg buffer. The purified nuclei were tagged with Tn5 transposome (Vazyme Biotech, Cat No./ID: 501-02). The resultant DNAs were purified using a Qiagen MinElute PCR Purification Kit (Qiagen, Cat No./ID: 28004) and amplified using 2x NEBNext High fidelity PCR mix (New England Biolabs, Cat No./ID: M0541L). Amplified libraries were purified with AMPure beads (Beckman Coulter, Cat No./ID: A63880), and sequenced on Illumina sequencer NovaSeq by PE150 strategy. Two biological replicates were performed.

In situ hybridization assays. For probe synthesis, in vitro transcription was performed with T3 or T7 RNA polymerase (ThermoFisher, Cat No./ID: EP0101/EP0111) in which linearized vectors were added as templates. RNA in situ hybridization was performed as described. Brieﬂy, root tissues were harvested and fixed with formaldehyde. 2× Affin-embedded samples were sectioned (7–9 μm) with a Leica sliding microtome (RM2260). The sections were stained, digested with Proteinase K (Roche, Cat No./ID: 03115828001), dehydrated with gradient ethanol, hybridized with corresponding probes and incubated with anti-digoxigenin-AP Fab fragments (Roche, Cat No./ID: 11093274910). After washing, the signals were detected with NBT/BCIP stock solution (Roche, Cat No./ID: 11681451001).

Microscopy. Radicles were stained with FM4-64 (1.0 μg/mL). The fluorescence signal was observed under a Zeiss 880 upright confocal microscope. The 514 nm laser was used for excitation, emission long pass filters between 519 and 573 nm were detected; for FM4-64, emission was 698–759 nm.

Pre-processing of raw scRNA-seq data. The raw files were analyzed by Cell Ranger 3.0.1 (10X Genomics). The rice genome and GTF annotation files which exclude the organelle genomes were downloaded from the Rice Annotation Project Database (RAP-DB) database (https://rapdb.dna.affrc.go.jp/index.html). To calculate the percentage of mitochondrial genes in each cell, the mitochondrial genome was merged with the nuclear genome. The “cellranger mkref” function with “--genomic,” “--fasta” and “--genes” arguments was used to build reference. The “cellranger count” with “--id,” “--transcriptome,” “--fastqs,” “--sample” and “--r2-length” = 98 arguments was performed to generate single-cell gene counts. More than 90% reads in all the samples were aligned to the Nipponbare reference genome (IRGSP 1.0). The ratio of the number of cells to total number of reads for each sample was >77%. The detailed Cell Ranger report was given in Supplementary Data 1. The gene-cell matrices (named “filtered_gene_bc_matrices” by 10X Genomics) were served as processed raw data for further analyses.

Doublet detection. We identified doublets with DoubletFinder (v.2.0.2). Three input parameters, namely the number of expected real doublets (nExp), the number of artificial doublets (pN), and the neighborhood size (pK), were defined as follows. For nExp, the standard processing pipeline was performed to the clustering stage with the low cell cluster number resolution (resolution = 0.5). The cluster labels of cells were used as “annotations” data to model the proportion of homotypic doublets. The doublet ratio was estimated by N/100000 (N, the cell numbers). The nExp value was adjusted according to the proportion of homotypic doublets and doublet ratio, pN, a ratio to define the number of generated artificial doublets based on total cell numbers, was set to 0.25. We found that an increase in pN value did not alter DoubletFinder results. To define an optimal pK value, pre-processed Seurat data were loaded into the “paramSweep, v3 (PCs = 1:15)” function, and subsequently fed into “summarizeSweep” and “find.pK” functions. A single and easily discernible maximum of pK value was selected as an optimal pK parameter.

The doublets were finally predicted with the pre-processed Seurat data using “doubletFinder_V3” function and defined values of nExp, pN, and pK as defined above. The proportion of artificial nearest neighbors (pANN) for every cell was calculated. The doublet threshold was defined according to the results of final doublet predictions. The resultant cells flagged as singlets were used for downstream analyses.
Data integration, clustering, and annotation. Downstream analyses were mainly performed with the Seurat package (v.3.1.1) as previously described \( ^{38} \). Briefly, we performed dimensionality reductions by filtering out low-quality cells and genes, and normalizing data with “NormalizeData” function (LogNormalized method, scaling factor of 10,000). We then detected variable genes with “FindVaria-

bleGenes” function (vst method, 2000 features), scaled data with “ScaleData” function, performed PCA analysis with “RunPCA” function (100 principal components), determined the biological significance of PCA scores by “JackStraw” function, constructed the SNN graph, clustered cells based on Louvain (“FindNeighbors” and “FindClusters”), and visualized data with non-linear dimension reduction algo-

rithms (“RunTSNE” and “RunUMAP”).

The low-quality cells and genes were filtered according to the following four criteria: (1) we only considered the cells with number of expressed genes between 500 and 6000; (2) we ignored the cells with unique molecular identifiers (UMIs) above 40,000 or below 500; (3) we filtered out the genes that were expressed in fewer than three cells; (4) the percentage of mitochondrial UMIs was no more than 10%. We corrected batch effects between samples with “RunHarmony” function. The resulting integrated data were clustered and visualized based on harmonized dimensionality reductions.

To mitigate the effects of cell cycle heterogeneity on cell clustering, the cell cycle score of each single cell was calculated by using the “CellCycleScoring” function with the cycling-orthogonal genes in Arabidopsis. These cell cycle effects were then re-grouped by using the “ScaleData” function using “vars.to.regress”.

To evaluate the effect of protoplasting for cell clustering, the proportion of the protoplasting genes which was identified by the Qian lab (Supplementary Data 2) \( ^{38} \) was calculated and plotted. The proportion of protoplasting genes in each cell cluster was below 1%.

The cluster-enchanced genes were computed with “FindALMarkers” function in Seurat using the following parameters: a Wilcoxon Rank Sum test; above 1.5-fold difference (logFC-threshold = 0.58) between the two groups of cells; test genes that a minimum fraction was at least 0.01. We assigned cells by the resultant cluster-enchanced genes with known functions and expression patterns (Supplementary Data 3). To identify cluster-specific marker genes, the following parameters were applied: the log2 fold change of genes was >0.25 and the proportion of marker genes expressed in cells among all other genes, the following parameters were applied: the log2 fold change of genes was ≥0.1. We assigned cell clusters by the resultant difference (logfc.threshold = 0).

The estimated RNA velocity by velocyto package \( ^{33} \). For integrative analysis of ATAC-seq and scRNA-seq, the accessibilities of cell cluster-specific marker genes were extracted and assigned. For each cell cluster, the correlation between accessibility (ATAC peak Log2 FC) and expression level (scRNA cluster-specific LogFC) of each gene was calculated and visualized by the volcano plot (Supplementary Fig. S6). Genes with positive values for both accessibility and expression level stand for positive correlation (red). Conversely, it indicates negative correlation (blue). The number and proportion of cluster-specific genes with positive or negative correlation were then calculated (Fig. 4).
were removed after integration (Supplementary Fig. S8a). After clustering, 30 super cell clusters were revealed (Supplementary Fig. S8b). The comparable proportion of rice and Arabidopsis cells in a given cell type indicates cell type conservation across species (e.g., RH, P and X in Fig. 5c). To identify core regulators for conserved cell types including RH (I10), P (I21), and X (I22), we used “FindConservedMarkers” function in Seurat. The most differentially expressed genes were clustered (k-means = 4, Fig. Sd-f).

Note added in proof: While this work was under revision, similar research by Liu et al. was published elsewhere.

**Reporting summary**. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

**Data availability**

The scRNA-seq and ATAC-seq data were deposited in NCBI with the accession numbers PRJNA706455 and PRJNA706099. Source data are provided with this paper.

**Received:** 9 September 2020; **Accepted:** 8 March 2021; **Published online:** 06 April 2021

**References**

1. Rebourcail, J. et al. Molecular genetics of rice root development. *Rice* **2**, 15–34 (2009).
2. Coudert, Y., Perin, C., Courtois, B., Khong, N. G. & Gantet, P. Genetic control of root development in rice, the model cereal. *Trends Plant Sci.* **15**, 219–226 (2010).
3. Aichinger, E., Kornet, N., Friedrich, T. & Laux, T. Plant stem cell niches. *Curr. Top. Dev. Biol.* **116**, 35–43 (2016).
4. Benfey, P. N. Designing the path from stem cells to differentiated tissue. *Curr. Top. Dev. Biol.* **116**, 35–43 (2016).
5. Draper, C., Sparks, E. E. & Benfey, P. N. Uncovering gene regulatory networks controlling plant cell differentiation. *Trends Genet.* **33**, 529–539 (2017).
6. Miyashima, S., Sebastian, J., Lee, J. Y. & Helariutta, Y. Stem cell function during plant vascular development. *EMBO J.* **32**, 178–193 (2013).
7. Hochholdinger, F. & Zimmermann, R. Conserved and diverse mechanisms in rice and Arabidopsis cells in a given cell type indicates cell type conservation across species. *NATURE COMMUNICATIONS* 12, 2053 (2021). https://doi.org/10.1038/s41467-021-22352-4

Received: 9 September 2020; Accepted: 8 March 2021; Published online: 06 April 2021
56. Birnbaum, K. et al. A gene expression map of the Arabidopsis root. Science 306, 1956–1960 (2003).
57. Kretzschmar, K. & Watt, F. M. Lineage tracing.
58. Shao, Y. et al. OsSPL3, an SBP-domain protein, regulates crown root development in rice. Plant Cell 31, 1257–1275 (2019).
59. Zhao, Y. et al. Loci and natural alleles underlying robust roots and adaptive domestication of upland ecotype in rice aerobic conditions. PLoS Genet 14, e1007521 (2018).
60. Zheng, H. et al. LATERAL ROOTLESS2, a cyclophilin protein, regulates lateral root initiation and auxin signaling pathway in rice. Mol. Plant 6, 1719–1721 (2013).
61. Kang, B. et al. OsCYP2, a chaperone involved in degradation of auxin-responsive proteins, plays crucial roles in rice lateral root initiation. Plant J. 74, 1036–97 (2013).
62. Li, J. et al. qRT9, a quantitative trait locus controlling root thickness and root length in upland rice. J. Exp. Bot. 66, 2723–2732 (2015).
63. Li, J. et al. A rice glutamate receptor-like gene is critical for the division and survival of individual cells in the root apical meristem. Plant Cell 18, 340–349 (2016).
64. Liu, H. et al. ARLI, a LOB-domain protein required for adventitious root formation in rice. Plant J. 43, 47–56 (2005).
65. Cui, H. et al. An evolutionarily conserved mechanism delimiting SHR movement defines a single layer of endodermis in plants. Science 316, 421–425 (2007).
66. Kamiya, N., Itoh, J., Morikami, A., Nagato, Y. & Matsuoka, M. The SCARECROW gene’s role in asymmetric cell divisions in rice plants. Plant J. 36, 45–54 (2003).
67. Ni, J., Shen, Y., Zhang, Y. & Wu, P. Definition and stabilisation of the quiescent centre in rice roots. Plant Biol. 16, 1014–1019 (2014).
68. Giri, J. et al. Rice auxin influx carrier OsAUX1 facilitates root hair elongation in response to low external phosphate. Nat. Commun. 9, 1408 (2018).
69. Ogura, T. et al. Root system depth in arabidopsis is shaped by EXOCYST70A3 via the dynamic modulation of auxin transport. Cell 178, 400–412 (2019).
70. Shekhar, V., Stickle, D., Thellmann, M. & Vermeer, J. E. M. The role of plant root systems in evolutionary adaptation. Curr. Top. Dev. Biol. 131, 55–80 (2019).
71. Ye, H. et al. Genetic diversity of root system architecture in response to drought stress in grain legumes. J. Exp. Bot. 69, 3267–3277 (2018).
72. Ristova, D., Giovannetti, M., Metesch, K. & Busch, W. Natural genetic variation shapes root system responses to phytohormones in Arabidopsis. Plant J. 96, 468–481 (2018).
73. Nishimura, A., Aichi, I. & Matsuoka, M. A protocol for Agrobacterium-mediated transformation in rice. Nat. Protoc. 1, 2796–2802 (2006).
74. Lu, Y. et al. Genome-wide targeted mutagenesis in rice using the CRISPR/Cas9 system. Mol. Plant 10, 1242–1245 (2017).
75. Wang, F. X. et al. Chromatin accessibility dynamics and a hierarchical transcriptional regulatory network structure for plant somatic embryogenesis. Dev. Cell 54, https://doi.org/10.1016/j.devcel.2020.1007.1003 (2020).
76. Zhang, T. Q. et al. A two-step model for de novo Activation of WUSCHEL during plant shoot regeneration. Plant Cell 29, 1073–1087 (2017).
77. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21 (2013).
78. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. Cell Syst. 8, 329–337 e234 (2019).
79. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 19, 15 (2018).
80. Tian, T. et al. agrigo v2.0: a GO analysis toolkit for the agricultural community, 2017 update. Nucleic Acids Res. 45, W122–W129 (2017).
81. Ludwig, L. S. et al. Transcriptional states and chromatin accessibility movement de WUSCHEL. Cell Rep. 27, 3228–3240 e3227 (2019).
82. Yan, F., Powell, D. R., Curtis, D. J. & Wong, N. C. From reads to insight: a hitchhiker’s guide to ATAC-seq data analysis. Genome Biol. 21, 22 (2020).
83. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359 (2012).
84. Li, H. et al. The Sequence Alignment/Map format and SAM tools. Bioinformatics 25, 2078–2079 (2009).
85. Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast processing of NGS alignment formats. Bioinformatics 31, 2032–2034 (2015).
86. Ross-Innes, C. S. et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature 481, 389–393 (2012).
87. Lawrence, M. et al. Software for computing and annotating genomic ranges. PLoS Comput. Biol. 9, e1003118 (2013).
88. Yu, G., Wang, L. G. & He, Q. Y. ChIPseeker: an R/bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics 31, 2382–2383 (2015).
89. Tosches, M. A. et al. Evolution of pallium, hippocampus, and cortical cell types revealed by single-cell transcriptomics in reptiles. Science 360, 881–888 (2018).
90. Liu, Q. et al. Transcriptional landscape of root cells at the single-cell resolution. Mol. Plant 14, 384–394 (2021).

Acknowledgements
We thank Yi-Qun Gao and Dai-Yin Chao (SIPPE, CAS) for discussion and Yan-Xia Mai (Core Facility Center of CEPMSS/PIPPE, CAS) for technical support on the flow cytometer. This work was supported by the grants from National Natural Science Foundation of China (31788103; 31525004), Strategic Priority Research Program of the Chinese Academy of Sciences (XDB27030101), the Chinese Academy of Sciences (QYZDB-SSW-SMC002), Science and Technology Commission of Shanghai Municipality (18JC1415000), Young Elite Scientists Sponsorship Program by CAST (2016QNRC001), and National Postdoctoral Program for Innovative Talents (BX2016060178).

Author contributions
T.-Q.Z. and J.-W.W. designed the research. T.-Q.Z., Y.C., Y.L., and J.-W.W. performed research. T.-Q.Z. performed bioinformatic analysis. W.-H.L. contributed to rice GATA6 KO #2 plants. T.-Q.Z. and J.-W.W. analyzed the data. T.-Q.Z. and J.-W.W. wrote the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-22352-4.

Correspondence and requests for materials should be addressed to T.-Q.Z. or J.-W.W.

Peer review information Nature Communications thanks the anonymous reviewers for their contribution to the peer review of this work. Peer review reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021