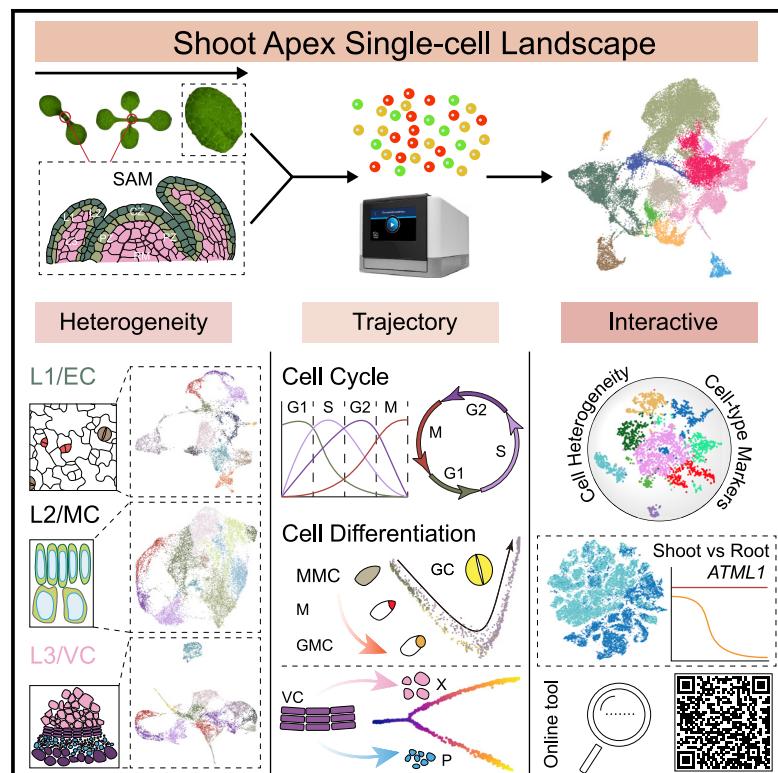


Developmental Cell

A single-cell analysis of the *Arabidopsis* vegetative shoot apex

Graphical abstract



Authors

Tian-Qi Zhang, Yu Chen, Jia-Wei Wang

Correspondence

jwwang@sippe.ac.cn

In brief

Zhang et al. report a comprehensive single-cell compendium of the *Arabidopsis* vegetative shoot apex. The definition of numerous cellular clusters and state changes offers researchers the opportunity to investigate, at unprecedented resolution, how the shoot apical meristem allows for reiterative formation of new aerial structures through cell division and differentiation.

Highlights

- A transcriptome landscape of shoot apex at single-cell resolution
- Cell-cycle continuums in the epidermis, mesophyll cells, and vascular tissues
- Differentiation trajectories of stomata and vascular tissues with high accuracy
- Integrative analysis of shoot and root apex cell populations at single-cell resolution



Resource

A single-cell analysis of the *Arabidopsis* vegetative shoot apex

Tian-Qi Zhang,¹ Yu Chen,^{1,2} and Jia-Wei Wang^{1,3,4,*}

¹National Key Laboratory of Plant Molecular Genetics (NKLPMG), CAS Center for Excellence in Molecular Plant Sciences (CEMPS), Institute of Plant Physiology and Ecology (SIPPE), Chinese Academy of Sciences (CAS), Shanghai 200032, China

²University of Chinese Academy of Sciences, Shanghai 200032, China

³ShanghaiTech University, Shanghai 200031, China

⁴Lead contact

*Correspondence: jwwang@sippe.ac.cn

<https://doi.org/10.1016/j.devcel.2021.02.021>

SUMMARY

The shoot apical meristem allows for reiterative formation of new aerial structures throughout the life cycle of a plant. We use single-cell RNA sequencing to define the cellular taxonomy of the *Arabidopsis* vegetative shoot apex at the transcriptome level. We find that the shoot apex is composed of highly heterogeneous cells, which can be partitioned into 7 broad populations with 23 transcriptionally distinct cell clusters. We delineate cell-cycle continuums and developmental trajectories of epidermal cells, vascular tissue, and leaf mesophyll cells and infer transcription factors and gene expression signatures associated with cell fate decisions. Integrative analysis of shoot and root apical cell populations further reveals common and distinct features of epidermal and vascular tissues. Our results, thus, offer a valuable resource for investigating the basic principles underlying cell division and differentiation in plants at single-cell resolution.

INTRODUCTION

All above-ground parts of the plant, except the hypocotyl and cotyledons, are generated from the shoot apical meristem (SAM). The angiosperm SAM consists of a small dome of cells with specific structural features (Dinneny and Benfey, 2008; Holt et al., 2014). Studies of the vegetative SAM have revealed distinct layers or zones regulating stem cell homeostasis and leaf primordium initiation. On the basis of clonal studies, the *Arabidopsis* SAM can be divided into three layers (Figure 1C). L1-layer derivatives give rise to the epidermis of shoots, leaves, and flowers, whereas the L2 layer provides the mesodermal tissue and germ cells, and the L3 layer contributes to the vascular tissues and pith. According to morphological and histological studies, the *Arabidopsis* SAM can also be divided into distinct domains or zones as well (Figure 1C). The central zone (CZ) is where the stem cells reside. The peripheral zone (PZ), where cells divide more frequently than in the CZ, is responsible for organ initiation, and the rib meristem (RM), which gives rise to central tissues of the shoot axis.

Plant stem cell niches are specified during embryogenesis. As in animals, stem cells reside in stem cell niches, which produce signals that regulate the balance between self-renewal and the generation of daughter cells that differentiate into new tissues (Aichinger et al., 2012; Greb and Lohmann, 2016; Pierre-Jerome et al., 2018; Sablowski, 2011). In the shoot, maintaining the stem cells requires expression of the homeodomain protein WUSCHEL (WUS) in an underlying group of cells known as the

organizing center (OC) and CLAVATA3 (CLV3), which is expressed in the L1 and L2 layers of the CZ (Galloch et al., 2015; Heidstra and Sabatini, 2014; Kitagawa and Jackson, 2019; Perales and Reddy, 2012). The shoot stem cells undergo several rounds of divisions, generating transit-amplifying cells that eventually differentiate and give rise to different types of lateral organs such as leaves and flowers.

Despite extensive studies, a comprehensive analysis of the identity of every cell in the shoot apex is lacking, limiting our ability to prospectively characterize meristematic functions and delineate the differentiation trajectories of transit-amplifying cells. Previous profiling studies used reporter genes to purify bulk cell populations (Tian et al., 2019; Yadav et al., 2009, 2014), an approach that necessarily underestimates cell complexity and restricts analysis in a marker-biased manner. Compared with bulk RNA sequencing, single-cell RNA sequencing (scRNA-seq) achieves a higher degree of resolution, enabling characterization of the many properties of subpopulations of heterogeneous groups of cells in plants (Birnbaum, 2018; Mironova and Xu, 2019; Nelms and Walbot, 2019; Rich-Griffin et al., 2020). For example, several scRNA-seq studies have revealed that *Arabidopsis* root tip cells are highly heterogeneous in their transcriptomes. The spatial distribution and temporal ordering of the individual cells at different developmental stages reveal a continuous differentiation trajectory of root development (Denyer et al., 2019; Jean-Baptiste et al., 2019; Ryu et al., 2019; Shaham et al., 2020; Shulse et al., 2019; Turco et al., 2019; Wendrich et al., 2020; Zhang et al., 2019).



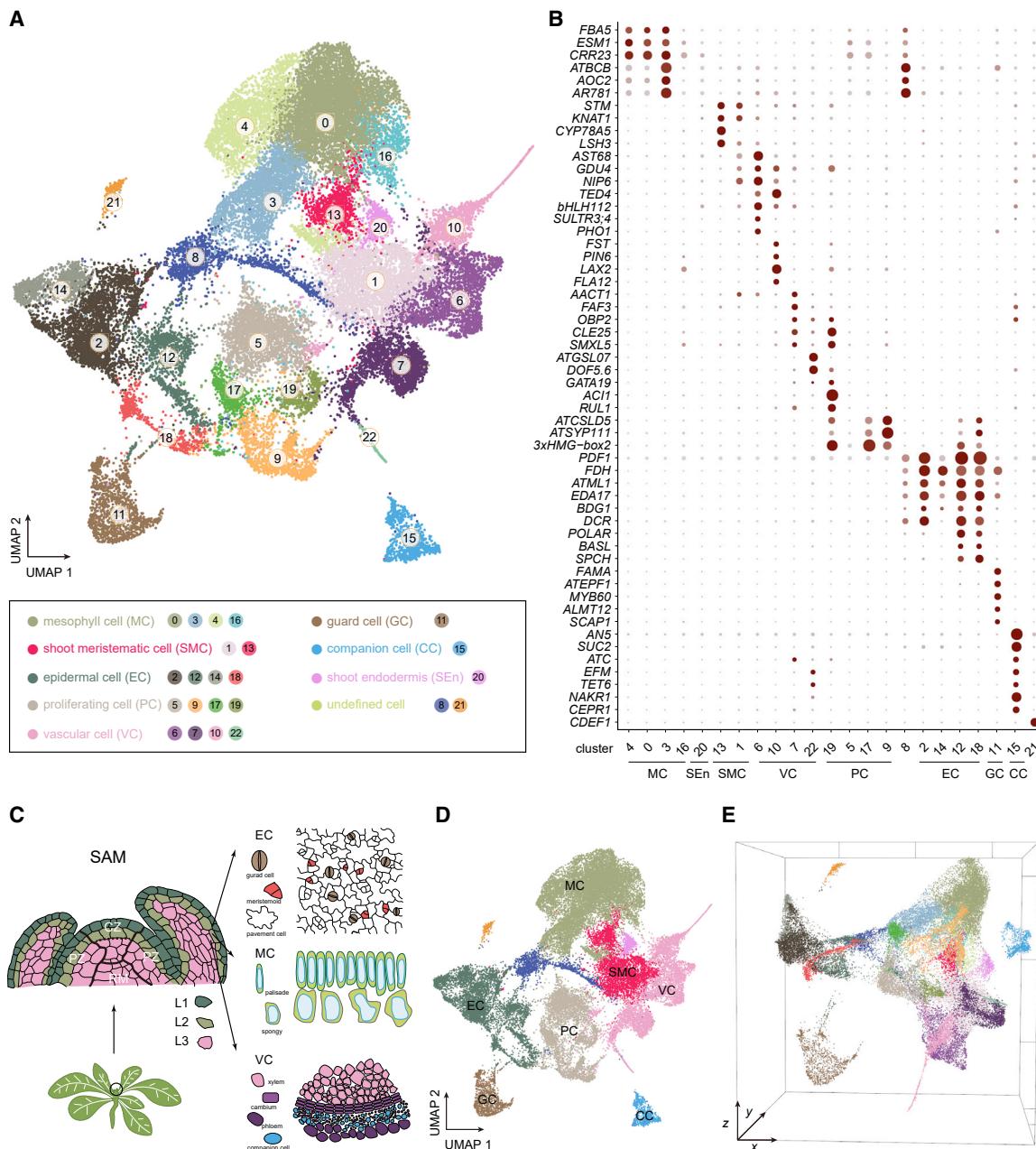


Figure 1. Cell heterogeneity in the vegetative shoot apex

- (A) Visualization of 23 cell clusters using UMAP. Dots, individual cells; $n = 36,643$ cells; color, cell clusters.
- (B) Expression pattern of representative cluster-specific marker genes. Dot diameter, proportion of cluster cells expressing a given gene. The full names of selected genes are given in Table S2.
- (C) Schematic of anatomy and cell types of *Arabidopsis* shoot apex. CZ, central zone; PZ, peripheral zone; OC, organizing center.
- (D) Visualization of 7 broad populations using UMAP. Color, population types.
- (E) Visualization of 7 broad populations by a UMAP 3D scatterplot. Cluster names and colors are the same as in (A).
- See also Figures S1 and S2; Tables S1, S2, and S3.

Here, we performed scRNA-seq of the *Arabidopsis* vegetative shoot apex. In total, 23 cell clusters corresponding to seven broad populations were identified. We inferred gene signatures and specific marker genes for main cell types. The profiling and ordering of individual cells enable us to reconstruct continuous developmental trajectories of epidermal cells and vascular

tissues. In particular, combining cell-type-specific genes revealed by the scRNA-seq dataset and reverse genetics allowed us to uncover new regulators of shoot development or physiology with high accuracy. Moreover, comparative analysis of shoot and root cell apex atlases illustrated conserved and distinct cell types. Finally, the identification of hundreds of

cell-cluster-specific marker genes and implementation of a web server provide a valuable resource for future functional analysis of the *Arabidopsis* vegetative shoot development.

RESULTS

Generation of an *Arabidopsis* vegetative shoot apex cell atlas

The vegetative shoot apex, composed of the SAM and leaf primordia, is deeply embedded within the shoot tip (Figure 1C). Because of the minute nature of the SAM (about 100 μm in diameter), the isolation of the vegetative SAM is technically challenging and transcriptome profiling of the *Arabidopsis* SAM at single-cell resolution has not yet been achieved. To overcome this technical difficulty, we harvested a large number of shoot apices ($n = 500$), thereby increasing the proportion of SAM cells within the tissues (Figures S1A–S1C). To discriminate between the SAM and leaf cell populations, we also collected developing leaves from 18-day-old plants (Figure S1A). The shoot apices and leaf tissues were protoplasted separately and subjected to droplet-based scRNA-seq using the 10x Genomics scRNA-seq platform. Notably, time-lapse imaging revealed that cell wall digestion was nearly complete after 2 h, and all vegetative shoot apex cells except trichomes were captured (Figures S1D and S1E; Video S1). Analyses of two shoot apex samples revealed reproducibility in data quality (Figures S2B and S2C; Table S1). The effect of protoplasting on the transcriptome was removed by regressing out the variations caused by protoplasting-induced genes (Birnbaum et al., 2003) (see STAR methods). Moreover, we mitigated the effects of cell-cycle heterogeneity on cell clustering using a standard pipeline provided by the Seurat (Stuart et al., 2019). We merged two shoot apex samples with one leaf sample after reproducibility analysis (Figure S2). The resultant dataset was used for subsequent cell clustering and annotation.

Principal component analysis and unsupervised analyses of the sequencing data, which did not rely on known markers, uncovered 23 distinct clusters from 36,643 high-quality cells of the shoot apices and leaves (Figures 1A and S2A). To annotate each cluster, we identified cluster-enriched genes whose biological functions or expression patterns have been well studied and performed gene ontology (GO) analysis ($n > 107$ genes, Figure 1B; Tables S1 and S2). Visualization using the uniform manifold approximation and projection (UMAP) algorithm, the *t*-distributed stochastic neighborhood embedding (*t*-SNE) tool, and a 3D scatterplot revealed seven broad populations: mesophyll cell (MC), epidermal cell (EC), shoot meristematic cell (SMC), proliferating cell (PC), vascular cell (VC), guard cell (GC), and companion cell (CC) (Figures 1C–1E and S2). Compared with *t*-SNE, the UMAP offers faster runtime, consistency, meaningful organization of cell clusters, and preservation of continuums (Becht et al., 2018). In contrast, the main advantage of *t*-SNE is the ability to preserve local structure. Therefore, we used both dimensionality reduction techniques for data visualization in the rest of this study.

The MC population consisted of four clusters (clusters 0, 3, 4, and 16) (Figure 1A), in which genes involved in photosynthesis such as CHLORORESPIRATORY REDUCTION23 (CRR23), RUBISCO SMALL SUBUNIT 2B (RBCS2B), and PHOTOSYNTHETIC NDH SUBCOMPLEX L1 (PNSL1) were predominantly

expressed (Figures 1B and 2; Table S1). This annotation was confirmed by the examining the promoter activity of a clusters-enriched gene, AT1G13650 (Figure S2F and S2G). The epidermal-specific genes MERISTEM LAYER1 (ATML1), PROTODERMAL FACTOR1 (PDF1), and FIDDLEHEAD (FDH) were detected in the EC population (i.e., L1 layer; clusters 2, 12, 14, and 18) (Figures 1B and 2). The SMC population consisted of two clusters, 1 and 13 (Figure 1A). Transcripts for SHOOT MERISTEMLESS (STM), which is required for the establishment and maintenance of the *Arabidopsis* SAM (Gallois et al., 2002), and other three related homeodomain genes, KNAT1, KNAT2, and KNAT6, were highly enriched in this population (Figures 1B and 2). We annotated four clusters (5, 9, 17, and 19) as the PC population because cell-cycle-related genes such as HISTONE H4 (HIS4), CYCLIN-DEPENDENT KINASE B2;1 (CDKB2;1), and CYCLIN A1;1 (CYCA1;1) were overrepresented (Figures 1B and 2) (Menges et al., 2002).

The VC population was composed of four clusters (6, 7, 10, and 22) (Figure 1A), in which genes involved in xylem and phloem differentiation were expressed. For instance, transcripts for the phloem gene SMAX1-LIKE5 (SMXL5) and the xylem gene PHLOEM INTERCALATED WITH XYLEM (PXY) were markedly overrepresented in clusters 7 and 10, respectively (Figures 1B and 2) (Miyashima et al., 2013; Wallner et al., 2020). The GC and CC populations, (i.e., clusters 11 and 15, respectively), were separated on the UMAP plot, in agreement with their specific physiological functionalities and unique expression profiles (Figure 1A). The GC marker genes including FAMA (FMA) and EPIDERMAL PATTERNING FACTOR1 (EPF1) were exclusively expressed in cluster 11 cells (Han and Torii, 2019; Lee and Bergmann, 2019), whereas transcripts of CC marker genes such as PHLOEM PROTEIN2-A1 (PP2-A1) and SUCROSE-PROTON SYMPORTER2 (SUC2) highly accumulated in cluster 15 cells (Figures 1B and 2). The assignments of the above cell clusters were further confirmed by the pairwise comparisons between our datasets and published results including the cotyledon scRNA-seq dataset (Liu et al., 2020), stomata lineage transcriptome (Adrian et al., 2015), and shoot domain gene atlas (Tian et al., 2019) (Figure S3).

Cluster 20 was enriched in genes involved in shoot gravitropic response, suggesting the existence of a unique cell population responsible for gravity perception in the shoot (Kawamoto et al., 2020) (Figure 3I). The promoter reporter analysis of two cluster-enriched genes (ACTIN DEPOLYMERIZING FACTOR9 and AT4G11290) revealed that the cluster 20 cells belong to the shoot endodermis (Figures 3A–3H). The remaining two clusters (8 and 21) could not be assigned due to the absence of known marker genes. Moreover, RNA *in situ* hybridization assays of the cluster-specific genes did not generate reliable signals. However, GO analyses revealed that genes annotated to the biological process category “response to stress” was predominantly represented (Table S1). Taken together, the above results indicate that the vegetative shoot apex is composed of highly heterogeneous cells.

Identification of genes involved in controlling shoot branch angle

As described above, the assignment of cell types by scRNA-seq enables us to identify a number of cell-type-specific or enriched

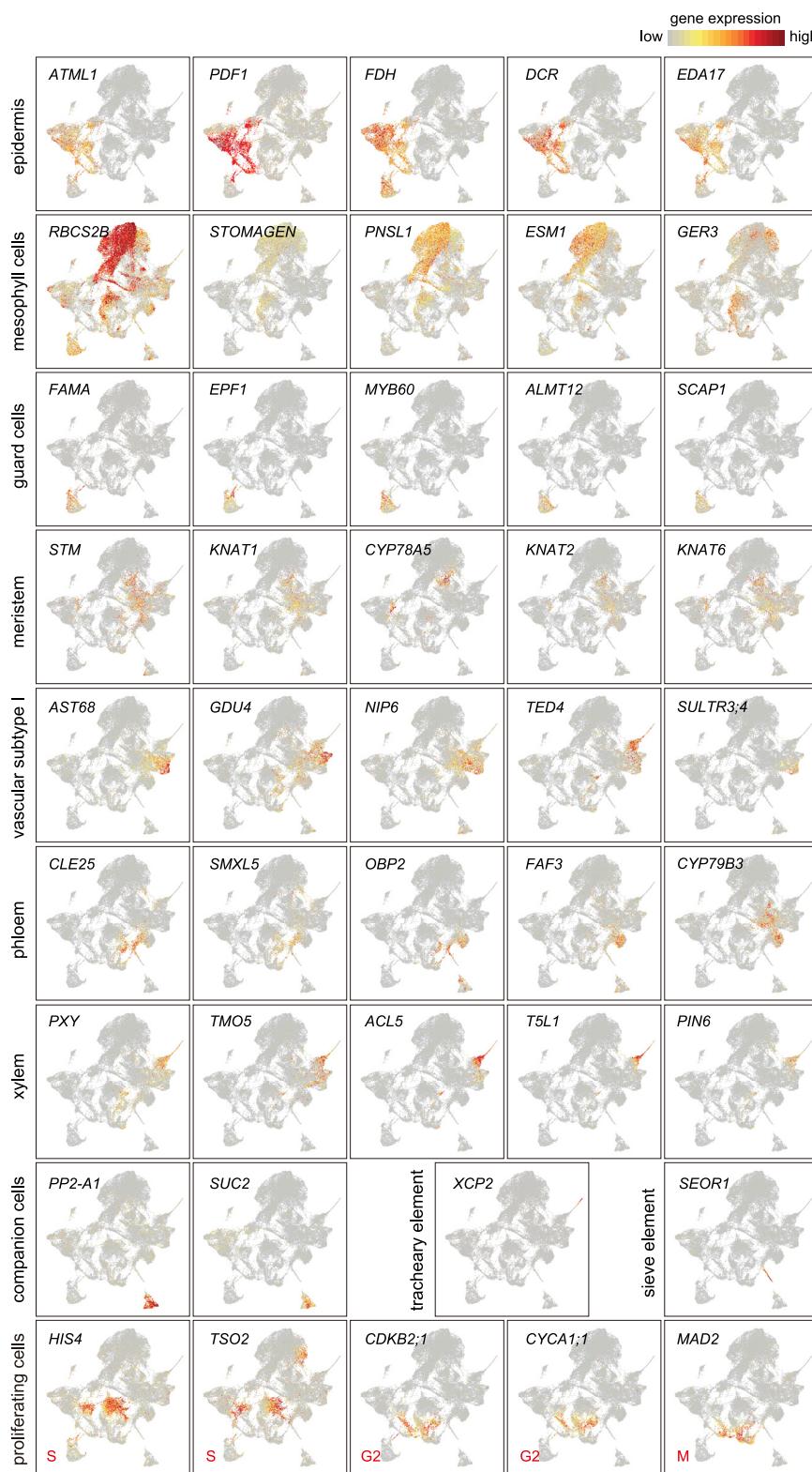


Figure 2. Expression of marker genes

UMAP plot showing the selected top marker genes for different cell types. The full names and referenced expression pattern of selected genes are summarized in Table S2. Three cell-cycle phases (S, M and G2) are shown.

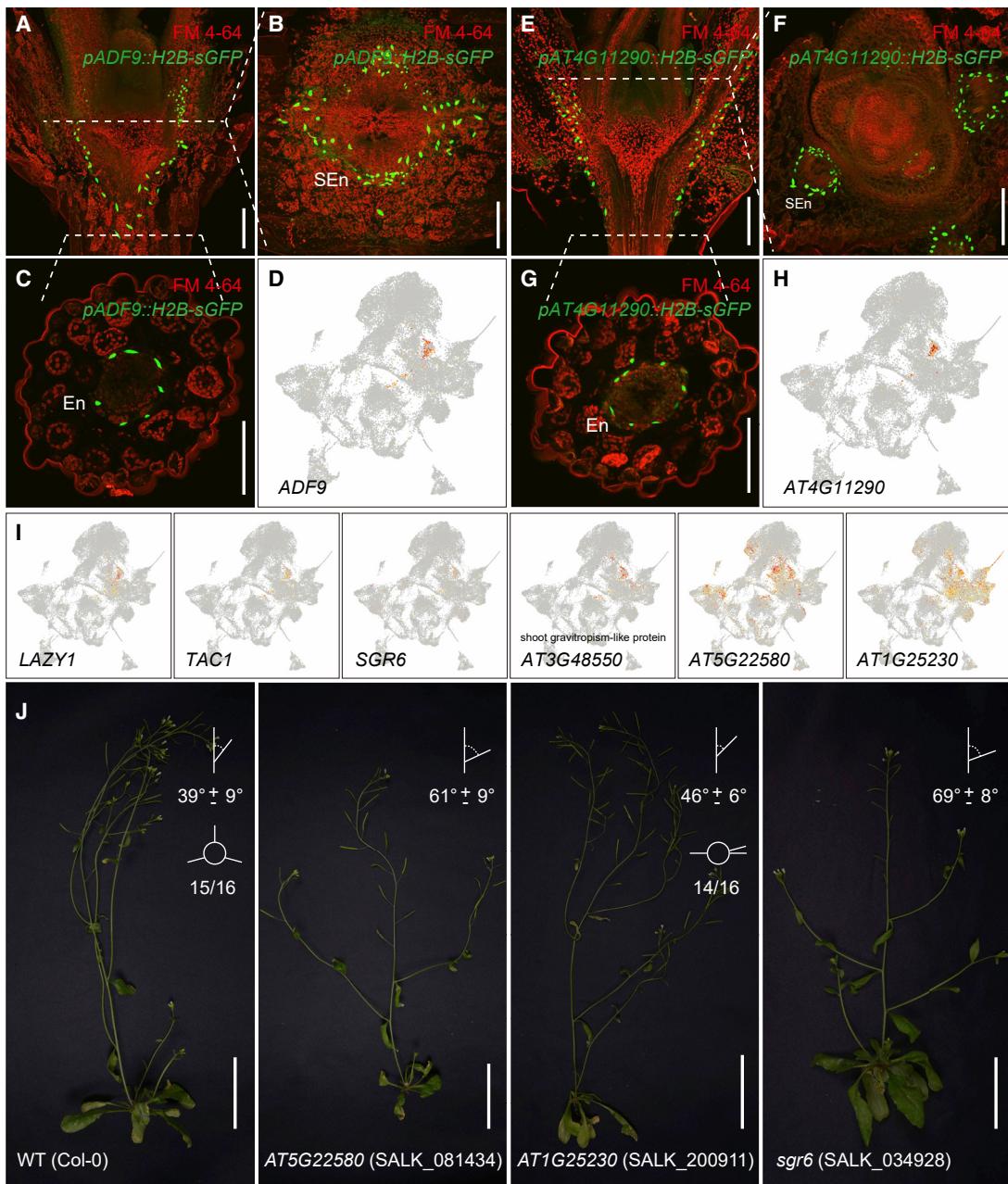


Figure 3. Identification of genes involved in shoot gravitropism

(A–H) Expression of *ACTIN DEPOLYMERIZING FACTOR9* (*ADF9*) and *AT4G11290* in the shoot endodermis (SEn). The gene expression pattern was revealed by promoter Histone2B-sGFP (H2B-sGFP, green) reporter (A–C and E–G) and UMAP plot (D and H). The shoot apices were longitudinally (A and E) or transversely (B, F, C, and G) sectioned. Dashed line marks the position for transverse section. Please note that both reporters were also expressed in the endodermis (En) of the hypocotyls (C and G). The cell outline (red) in (A–C and E–G) is visualized by FM4-64 staining. Scale bar, 100 μ m.

(I) UMAP plots showing the expression pattern of the genes involved in shoot gravitropism.

(J) Characterization of the T-DNA insertion mutants of *AT5G22580*, *AT1G25230*, and *SGR6*. Plants were grown in long days. Please note that all the mutants exhibited agravitropic (increased branch angle) phenotype. The shoot branch angles in mutants are significantly different from wild type ($p < 0.01$, one-way ANOVA was performed by the Turkey's multiple comparisons test). Please also note that the phyllotaxy is largely normal in wild type (15/16 plants) but defective in the *AT1G25230* mutant (14/16 plants). See also Figure S2H.

genes. We attempted to ask whether these results, combined with reverse genetics, could help us uncover new regulators for shoot development and physiology with high accuracy. As a proof-of-concept, we focused on cluster 20, the shoot endo-

dermis enriched in genes involved in shoot gravitropic response. *LAZY1*, *TILLER ANGLE CONTROL1* (*TAC1*), and *SHOOT GRAVITROPISM6* (*SGR6*) are involved in shoot gravitropic response (Guseman et al., 2017; Hashiguchi et al., 2014;

Kawamoto et al., 2020; Taniguchi et al., 2017; Yoshihara et al., 2013). We selected 48 cluster 20-enriched genes and obtained corresponding homozygous transfer-DNA (T-DNA) insertion mutants from the seed stock center (Alonso et al., 2003) (Table S2). As expected, we found that the mutations in *AT5G22580* and *AT1G25230*, which encode an A/B barrel domain-containing protein and a calcineurin-like metallo-phosphoesterase family protein, respectively, exhibited the same large branch angle phenotype as the *sgr6* mutant (Figures 3J and S2H). Thus, these results reveal that the collection of shoot cell-type-specific or enriched genes can accelerate gene discovery and serves as a valuable resource for *Arabidopsis* functional genomics in the future.

Cell-cycle continuum in the epidermis, mesophyll, and vascular tissue

One of the key themes of plant development is organ formation through cell division and differentiation at the shoot apex (Sablowski, 2007). Transit-amplifying cells are defined as an undifferentiated population of cells in transition between a stem cell and differentiated cell state (Hsu et al., 2014). To understand the developmental trajectories of the epidermis, mesophyll, and vascular tissue, we need to identify and characterize corresponding early dividing cells, as well as differentiating cells and terminal cells. The traditional approach to classifying distinct cell-cycle states relies on a few known markers with arbitrary cutoffs in expression levels. Taking advantage of scRNA-seq data, we wanted to establish the cell-cycle continuum in each cell population using the assignment of the sub-cell clusters based on marker gene expression and diffusion pseudotime estimates.

We first focused on the EC population (L1 layer), which contained all the dividing and differentiated epidermis cells. Re-clustering of the EC cells revealed that four sub-cell clusters, E3, E7, E8, and E9, were arranged in a cyclic pattern (Figure 4A). Cluster E3 was assigned to the S phase because high levels of transcripts for the DNA synthesis genes *HIS4* and *TSO2* were found in this cluster (Figure 4D) (Menges et al., 2002). Cluster E9 assigned to the G2 phase as *CDKB2;1* and *CYCA1;1* were highly expressed (Figure 4D) (Van Leene et al., 2010). The M phase cells resided in the sub-cell cluster E7, which had abundant transcripts for *EARLY NODULIN-LIKE PROTEIN15* (*ENOD15*) and the mitotic checkpoint protein *MITOTIC ARREST-DEFICIENT2* (*MAD2*) (Figure 4D) (Menges et al., 2002). Finally, the sub-cell cluster E8 was assigned to the G0 or G1 phase because *MINICHROMOSOME MAINTENANCE7* (*MCM7*), *PROLIFERATING CELL NUCLEAR ANTIGEN2* (*PCNA2*), *CYCLIN D3;2* (*CYCD3;2*), and ribosomal protein coding genes such as ribosomal protein L24e (*RPL24*) and ribosomal protein S6 (*RPS6*) were highly expressed (Figure 4D; Table S1) (Swaminathan et al., 2000). In agreement with studies of the G1/G0 phase in yeast and animals (Pyronnet and Sonenberg, 2001), biological processes related to translation were enriched in this cluster (Table S3). Of particular note, cell continuity estimates obtained by diffusion pseudotime and Palantir and RNA velocity analysis revealed that the cells belonging to these four clusters formed a complete cell cycle, with the key cell-cycle genes being expressed at distinct cell-cycle states (Figures 4G–4J, S4A–S4C, and S4E) (Qiu et al., 2017; Setty et al., 2019; Wolf et al.,

2018). In line with this annotation, the genes enriched in the S and M phases exhibited sporadic expression in the shoot apices as indicated by RNA *in situ* hybridization assays (Figures S4D and S4F–S4M). Taken together, these results indicate that the cells of two successive cell-cycle phases share similarities (overlap) in their transcriptomes and are arranged in a cyclic pattern based on unsupervised clustering.

Next, we sought to dissect the dividing cell populations specific to mesophyll (L2 layer) and vascular tissues (L3 layer). However, the survey of the MC and VC populations did not capture similar cell-cycle clusters as in the EC population. This result suggests that the dividing cell clusters of these two populations may reside in the PC population. Indeed, based on their transcriptome signatures, we were able to divide clusters 5 and 9 into two sub-clusters (Clusters 5a and 5b; Clusters 9a and 9b) (Figures S5A and S5B). Clusters 17, 5a, and 9a exhibited high transcript levels of photosynthesis genes including *RBCS2B* and *PNSL1*, whereas the vascular tissue marker genes *SMXL5* and *GLUTAMINE DUMPER4* (*GDU4*) were highly expressed in clusters 19, 5b, and 9b (Figure S5C). This finding indicates that the PC population is mainly composed of two dividing cell populations, which lead to the differentiation of mesophyll cells and vascular tissues. To separately delineate their developmental trajectories, we grouped the cells of clusters 5a, 9a, and 17 with the VC population, and those of clusters 5b, 9b, and 19 with the MC population. The resultant two populations were named reMC (reconstructed MC) and reVC (reconstructed VC). Intriguingly, similar to the observation for the EC population, we successfully detected complete cell cycles within these two reconstructed populations (Figures 4B, 4C, 4E and 4F; Table S3). The proportion of cells in each cell-cycle state was comparable among the EC, reMC, and reVC populations (Figure 4K).

Finally, on the basis of genes expressed at the same cell-cycle phases in the EC, reMC, and reVC populations, we created a core catalog of *Arabidopsis* genes whose transcript levels vary periodically over the cell cycle (Table S4). The resultant sets of periodic genes offer us a valuable resource for the future identification of new plant cell-cycle components in the shoot apex at single-cell resolution.

A continuum of epidermal cells differentiating toward stomata

The definition of the EC, reMC, and reVC populations paves the way for us to investigate how stem cells differentiate into distinct cell types through transit-amplifying cells. We used stomata and vascular tissues as a proof-of-concept illustration of the power of scRNA-seq in delineation of development trajectories and new genes involved in cell fate determination.

The highly differentiated epidermal cells are derived from epidermal initial cells called protoderm during leaf development. It has been shown that stomata emerge from the meristemoid mother cell through cell divisions and sequential cell fate commitment (Figure 5A) (Han and Torii, 2019; Lee and Bergmann, 2019). To infer the developmental trajectory of stomata, we grouped the corresponding transit-amplifying sub-cell clusters, including dividing sub-cell clusters (E3, E7, E8, and E9) and differentiating sub-cell clusters (E8, E1, and E11) (Figure 4A).

We found that the expression of early stomata genes such as *POLAR LOCALIZATION DURING ASYMMETRIC DIVISION AND*

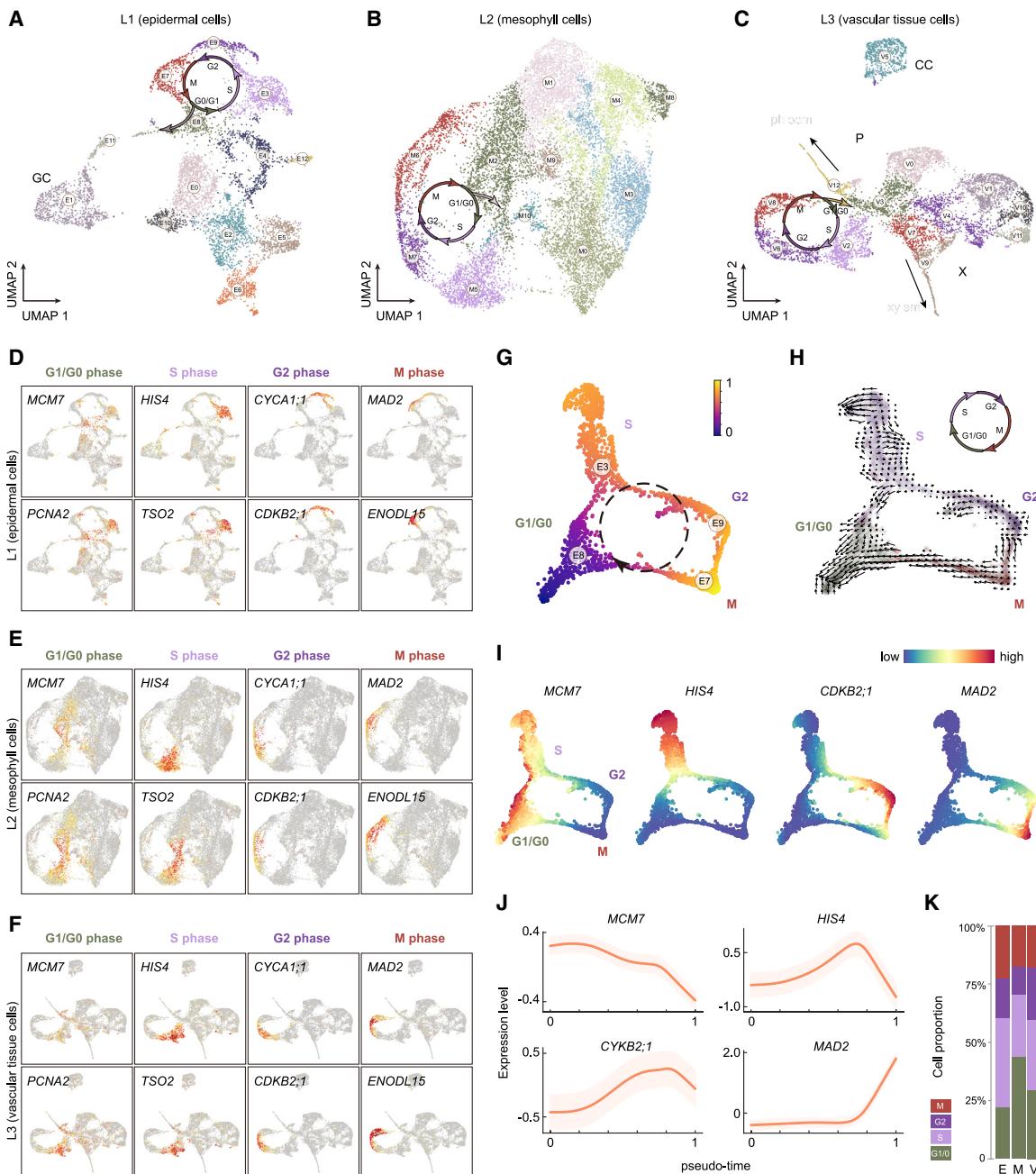


Figure 4. Cell-cycle continuums of transit-amplification cells in the vegetative shoot apex

(A–C) Visualization of the EC ($n = 6,852$ cells), reMC ($n = 15,254$ cells), and reVC ($n = 8,061$ cells) populations by UMAP. Dots, individual cells; color, sub-cell clusters; prefixing sub-cell cluster numbers with E (epidermis), mesophyll (M), or vascular tissue (V).

(D–F) Expression pattern of selected cell-cycle genes across four phases in the EC (D), reMC (E), and reVC (F) populations.

(G) Palantir pseudotime estimate of cell-cycle progression of the E3, E7, E8, and E9 cell clusters. Color bar represents pseudotime level.

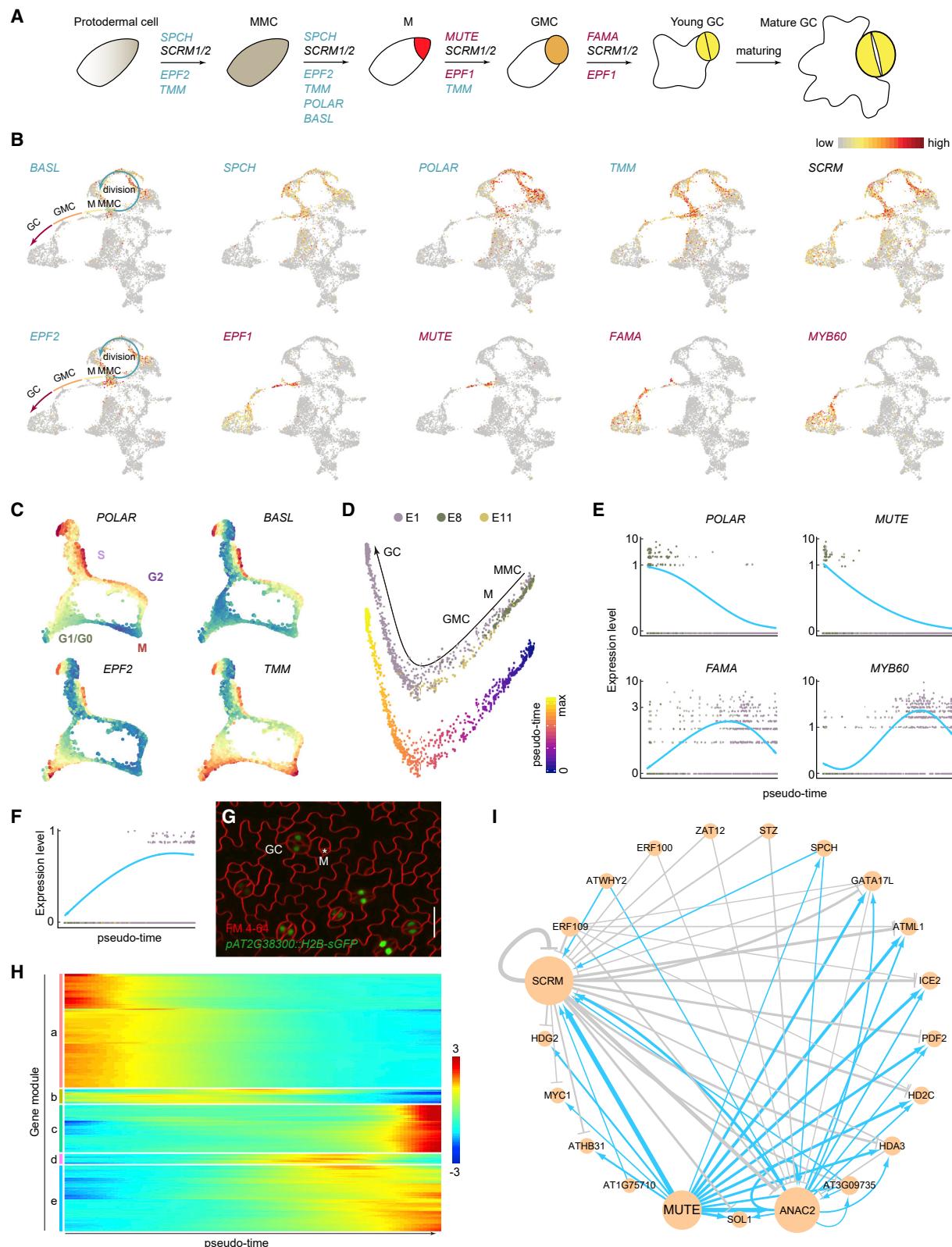
(H) RNA velocity field projected onto a t-SNE map. Colors according to sub-cell clusters (E3, E7, E8, and E9) of the EC population. Arrows represent average velocity and direction.

(I) Visualization of gene expression patterns of *MCM7*, *HIS4*, *CDKB2;1*, and *MAD2* along cell-cycle progression. The cell cycle sub-cell clusters (E3, E7, E8, and E9) of the EC population (Figure 4A) were used ($n = 1,911$ cells).

(J) Variation of gene expression trends of *MCM7*, *HIS4*, *CDKB2;1*, and *MAD2* along S and G2/M pseudotime progression. Line indicates mean expression; shaded area indicates SD.

(K) Proportion of each cell-cycle state in the EC (E), reMC (M), and reVC (V) populations. The proportion of the cells in each cell-cycle phase is calculated by number of the cells in each cell-cycle phase/total number of dividing cells.

See also Figures S4 and S5; Table S4.



(legend on next page)

REDISTRIBUTION (POLAR), SPEECHLESS (SPCH), BREAKING OF ASYMMETRY IN THE STOMATAL LINEAGE (BASL), and TOO MANY MOUTHS (TMM) was coordinated with cell-cycle progression (Figure 5B). In particular, the genes involved in the asymmetric cell division of the meristemoid mother cell such as *POLAR* and *BASL* were highly expressed in the S phase (Figure 5C) (Guo and Dong, 2019), while the expression of signaling elements that restrict stomatal differentiation including *TMM* and *EPF2* mainly peaked in the M or G0/G1 phase (Figure 5C). Thus, the temporal expression pattern of these genes may link the accumulation of their products to the specific time window in which they are required for asymmetric cell division or GC differentiation.

Ordering of the cells of clusters E1, E8, and E9 from the EC population by pseudotime analyses revealed the late-differentiation trajectory of stomata (Figures 5D and S6A–S6D) (Qiu et al., 2017; Setty et al., 2019; Wolf et al., 2018). Consistent with earlier reports (Han and Torii, 2019; Lee and Bergmann, 2019), the key transcription factors governing these cell fate transitions including *MUTE*, *FMA*, and *MYB60* were sequentially expressed along the cell differentiation trajectory (Figures 5E and S6B–S6D). The inferred trajectory was further confirmed by the promoter reporter analysis of a previously uncharacterized late-differentiation gene, *AT2G38300*, which encodes a MYB-like transcription factor (Figures 5F and 5G). The gene module at the beginning of pseudotime (module a) showed gene signatures of the G1 phase with high enrichment of genes involved in translation, whereas the latter modules (c and e) were enriched in genes involved in the response to abiotic stress (Figures 5H and S6E), in agreement with hypersensitive nature of mature GCs to environmental conditions. The secretory peptides *STOMAGEN* and *EPF1/EPF2* function as positive and negative regulators of stomatal development, respectively, by interacting with the receptor-like protein *TMM* (Matos et al., 2014; Sugano et al., 2010). *EPF2* transcripts mainly accumulated in the early cell-cycle clusters E3 and E8, while *EPF1* was highly expressed in late-differentiation cell clusters E1 and E11 (Figure 5B). In contrast, the expression of *STOMAGEN* was largely confined to the MC populations (Figure 2). Therefore, our scRNA-seq dataset reveals the spatial-temporal regulation of transcription factors and intercellular signaling molecules during stomata development with high accuracy.

We inferred the gene regulatory network underlying the stomata differentiation process using SCODE (Table S5) (Matsu-

moto et al., 2017). The incorporation of transcription factor expression dynamics across pseudotime revealed a complex regulatory network. Consistent with previous genetic studies (Han and Torii, 2019; Lee and Bergmann, 2019), *MUTE* and its interactor *SCREAM* served as major and highly connected central regulators (Figure 5I). Intriguingly, *SCREAM* likely represses the expression of most of the components in the network, while *MUTE* functions mainly as an activator. The network also predicted feedback and feed-forward loops at *SCREAM* and the previously unrecognized role of *ANAC2* in stomata development or physiology (Lu et al., 2007) (Figure 5I). Taken together, our results reveal a continuum of epidermal cell differentiation toward stomata and provide fresh insights into how the expression of cell fate determination genes is precisely coordinated with cell-cycle progression during stomata differentiation.

The remaining seven sub-cell clusters of the EC population (Figure 4A, E0, E4, E2, E6, E10, E5, and E12) could not be functionally assigned based on cluster-specific genes and GO terms (Table S3). Since the epidermis is mainly composed of pavement cells, this result suggests that pavement cells are highly heterogeneous. The identity and function of each sub-cell cluster await further investigation.

Differentiation trajectory of vascular tissues

The stem cells in the vascular meristem, also known as the procambium and cambium, continuously produces two conductive tissues, xylem and phloem, via asymmetric periclinal cell division (Miyashima et al., 2013). To infer the developmental trajectory of vascular tissues, we grouped the corresponding transit-amplifying sub-cell clusters, including dividing sub-cell clusters (V2, V3, V6, and V8) and differentiating sub-cell clusters (V3, V7, V9, and V12) (Figure 4C). The transit-amplification cells of the reVC population bifurcated into phloem and xylem lineages (Figures 4C and 6A). Based on gene annotations and *in situ* hybridization assays of cell cluster-enriched genes, the phloem lineage is composed of clusters V0, V5, and V12, whereas clusters V1, V4, V7, and V9–11 belong to the xylem lineage (Figures 6A, 6B, and S7A). For instance, *AT5G61660* and *AT5G46730* were highly expressed in protoxylem, whereas *AT5G24800* was exclusively expressed in phloem (Figure 6B). Cluster V12 and V9 were termed sieve elements (SEs) and tracheary elements (TEs), respectively, because the *SIEVE-ELEMENT-OCCCLUSION-RELATED1* (*SEOR1*), *SEOR2*, *XYLEM CYSTEINE PEPTIDASE1*

Figure 5. A continuum of epidermal cells differentiating toward stomata

- (A) Schematic of the key steps in stomatal development. MMC, meristemoid mother cell; M, meristemoid; GMC, guard mother cell; GC, guard cell.
- (B) UMAP plot showing the expression pattern of early (cyan) and late (magenta) stomata development genes. The EC population (Figure 4A) were used ($n = 6,852$ cells).
- (C) t-SNE map showing the coordinated expression of *POLAR*, *BASL*, *EPF2*, and *TMM* with cell cycle. The cell-cycle sub-cell clusters (E3, E7, E8, and E9) of the EC population (Figure 4A) were used ($n = 1,911$ cells). Color, different expression levels (red, highest expression level).
- (D) Monocle 2 analysis showing stomatal differentiation trajectory from MMC to GC, colored by sub-cell clusters (E1, E8, and E11) or pseudotime. $n = 1,420$ cells.
- (E) Expression of marker genes along pseudotime. Blue line indicates expression tendency. Dots, individual cells.
- (F and G) Expression pattern of a late stomata differentiation trajectory gene *AT2G38300*. The expression pattern was revealed by pseudotime analysis (F) and promoter H2B-sGFP reporter analysis (G). Blue line in (F) indicates expression tendency; Dots, individual cells. The cell outline (red) in (G) is visualized by FM4-64 staining. Scale bar in (G), 20 μ m.
- (H) Heatmap showing the five gene modules (A–E) of 500 top significant differentially expressed genes along pseudotime. The gene list is given in Table S5.
- (I) Gene regulatory network (GRN) inferred by 102 transcription factors that are dynamically expressed across stomata differentiation pseudotime with a parameter cutoff of 2.0. Blue and gray, positive or negative regulation. Node size is equivalent to the number of predicted connections.

See also Figure S6; Table S5.

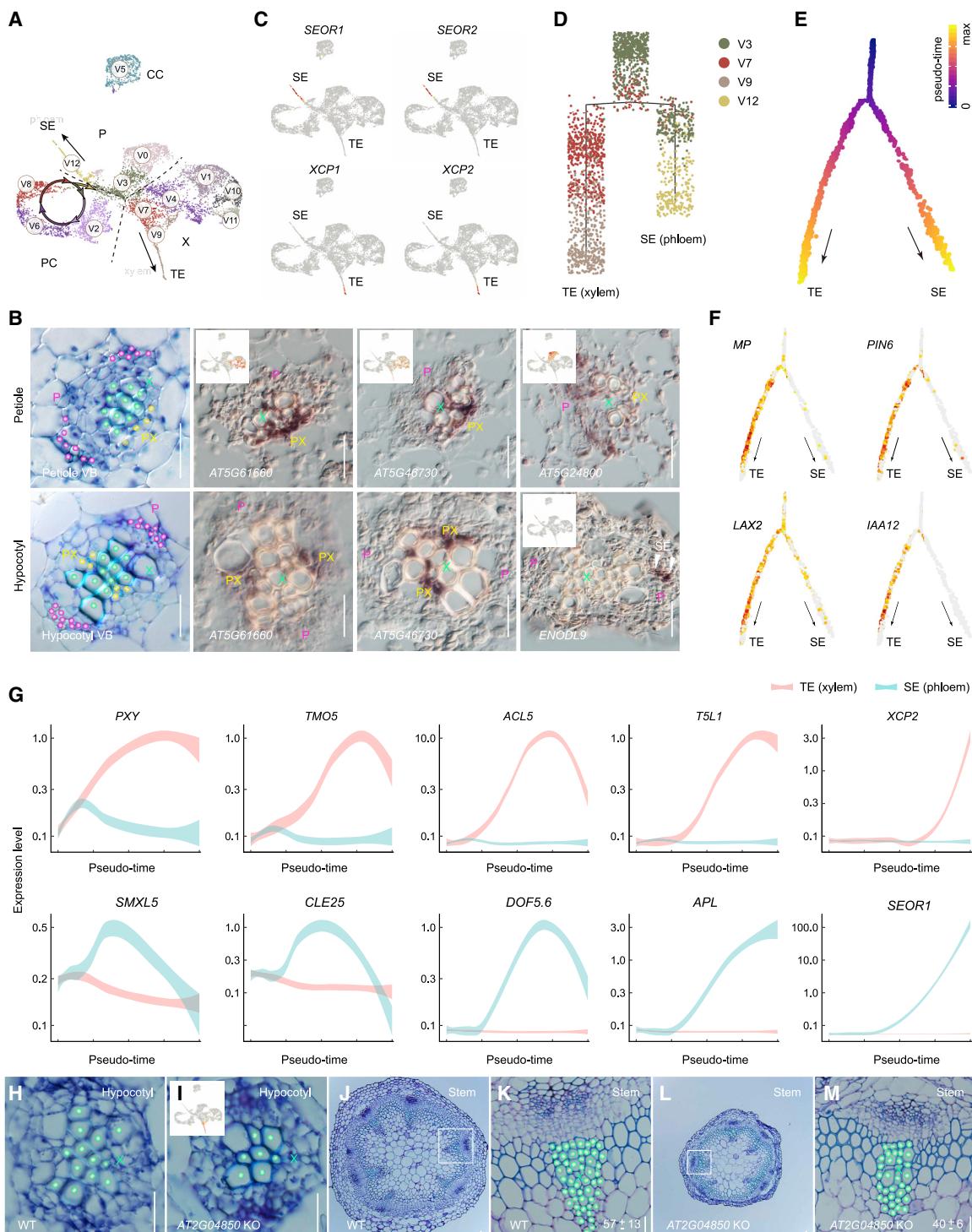


Figure 6. Differentiation trajectories for tracheary elements (xylem) and sieve elements (phloem)

(A) UMAP projection of scRNA-seq showing the reVC population. P, phloem; PC, proliferation cells; SE, sieve elements; X, xylem; TE, tracheary elements. Dash lines mark the boundary among three major cell populations (PC, P, and X).

(B) Expression pattern of cluster-enriched genes in the vascular bundles (VB). The location of protoxylem (PX, yellow dots), xylem (X, cyan dots), and phloem (P, pink dots) cells are shown. Insets, expression pattern of selected genes by UMAP plots. Scale bar, 20 μ m.

(C) UMAP plot showing the expression pattern of *SEOR1*, *SEOR2*, *XCP1*, and *XCP2*.

(D) Monocle 2 inferred hierarchical tree of TE and SE lineages. The clusters V3, V7, V9, and V12 ($n = 2,071$ cells) were used.

(E) Pseudotime estimate of differentiation trajectory of TE (xylem) and SE (phloem) by Monocle 2.

(F) Expression pattern of the genes in the auxin signaling pathway along pseudotime.

(legend continued on next page)

(*XCP1*), and *XCP2* genes were specifically expressed in cells in these clusters (Figure 6C).

Using pseudotime analysis, we inferred continuous differentiation trajectories for TEs (xylem) and SEs (phloem) (Figures 6D and 6E). Consistent with a role of the phytohormone auxin in the promotion of xylem identity (Smetana et al., 2019), genes in the auxin pathway such as *MONOPTEROS* (*MP*), *IAA12*, *LIKE AUXIN RESISTANT2* (*LAX2*), and *PIN-FORMED6* (*PIN6*) were highly expressed in the xylem lineage (Figure 6F). GO enrichment analysis further revealed that biological processes related to auxin response were highly enriched in this branch (Figures S7B and S7C). *PXY*, *TARGET OF MONOPTEROS5* (*TMO5*), *ACAU1S5* (*ACL5*), and *XCP2* were sequentially expressed along the differentiation trajectory (Figure 6G). On the contrary, genes involved in phloem differentiation such as *SMXL5*, *ALTERED PHLOEM DEVELOPMENT* (*APL*), *HIGH CAMBIAL ACTIVITY2* (*HAC2/DOF5.6*), and *SEOR1* were predominantly expressed in the phloem branch (Figure 6G) (Miyashima et al., 2019). Interestingly, our pseudotime analysis predicted a role of *TMO5-LIKE1* (*T5L1*) and *CLAVATA3/ESR-RELATED25* (*CLE25*) in vascular tissue development (Figure 6G). Taken together, these data indicate that scRNA-seq enabled us to reveal new cell types in the vascular tissue and delineate the hidden transition cell states during the differentiation of xylem and phloem.

Having determined the cell types and transcripts specific to the VC population, we wanted to use this knowledge to identify new regulators involved in vascular tissue differentiation. To this end, we chose a previously uncharacterized auxin-responsive gene *AT2G04850*, which was highly enriched in the xylem lineage (Figure 6H). Consistent with its expression pattern, we observed fewer xylem cells in both the hypocotyl and stem of the *AT2G04850* loss-of-function mutant in comparison with those of wild type (Figures 6H, 6I, 6K, 6M, and S7D). In support of this phenotype, the mutant had a thinner stem than wild type (Figures 6J and 6L). Thus, these findings further substantiate the power of scRNA-seq datasets to uncover new cell-type-specific developmental regulators.

Characterization of the SMC population

The SMC population could be divided into 10 sub-cell clusters (Figure 7A). The analysis of cell-cycle genes revealed that SMC transcriptome heterogeneity was not dominated by cell-cycle status (Figure 7B). Due to the low number of cells expressing *WUS* and *CLV3* and their relatively low expression levels, we designated cluster S4 as the OC and CZ using a combination of known OC and CZ marker genes including *PLETHORA7* (*PLT7*), *REPRODUCTIVE MERISTEM1* (*REM1*), *AT3G59270*, and *AT1G26680* (Figure 7C) (Franco-Zorrilla et al., 2002; Yadav et al., 2009).

Analysis of the SMC population did not reveal a clear cell differentiation trajectory. It is well known that *STM* is highly expressed in the SAM and becomes repressed in the leaf primordium (Byrne et al., 2000) (Figure 7E). *STM* expression was observed not only in the SMC population but also in the EC population (cluster E4, Figures 4A and 7D). This finding suggests that the differentiating *STM*-positive (*STM⁺*) cells (i.e., differentiating SAM cells) may reside in differentiated cell populations, rather than in the SMC population. Based on this notion, we grouped all the *STM⁺* cells in the atlas and re-performed unsupervised cell clustering to reveal the heterogeneity of *STM⁺* cells. As a result, we obtained nine cell clusters, named W0 to W8 (Figure 7F). Cluster W6 was comprised ECs with the L1-layer marker genes *ATML1* and *FDH* highly expressed. The phloem lineage PEAR transcription factors *DOF5.6* and *DOF6* were predominantly expressed in W2 (Miyashima et al., 2019), whereas the xylem lineage genes *ACL5* and *PXY* were overrepresented in W5. Therefore, we annotated clusters W2 and W5 as the cells differentiating toward phloem and xylem, respectively.

Cluster W7 showed gene signatures for other vascular tissues since *SUC2* and *PP2-A1* were highly enriched. Cluster W4 was assigned as the boundary domain between the meristem and leaf primordia as revealed by the enrichment of the boundary-specific transcription factor gene *LOB* (*AT5G63090*) (Bell et al., 2012) and by RNA *in situ* hybridization assays of two cluster-enriched genes, *AT5G57785* and *GH3.17* (Figures 7G–7J). The transcripts of two leaf polarity genes, *YABBY3* (*YAB3*) and *FILAMENTOUS FLOWER* (*FIL*) (Braybrook and Kuhlemeier, 2010; Chitwood et al., 2007; Husbands et al., 2009), were detected in cluster W8, suggesting that this cluster may correspond to the abaxial domain of leaf anlagen. The cells corresponding to the adaxial domain of leaf anlagen likely resided in cluster W5 because two adaxial-fate determinator genes *PHABULOSA* (*PHB*) and *PHAVOLUTA* (*PHV*) were highly expressed (McConnell et al., 2001). Clusters W1 and W3 cannot be functionally annotated. However, GO term analysis suggests that these cells are probably involved in stress response and substance transport. Cluster W0 centered on the UMAP, suggesting that the cells within this cluster are undifferentiated pluripotent stem cells. Taken together, the above results not only demonstrate cell heterogeneity of the SAM but also reveal distinct differentiation lineages of the stem cells within the SAM.

Transcription factors have critical roles in cell fate determination. Gene expression analysis across *STM⁺* cells revealed that *STM* was highly co-expressed with *KNAT1* but had low co-expression with the cluster-enriched genes in the differentiating cell clusters (Figures 7K and 7L). This result is consistent with the notion that these homeodomain genes are functionally

(G) Expression trends of the genes involved in TE and SE lineages along pseudotime. The representative genes enriched in specific developmental stages. Shaded area indicates expression levels with 95% confidence interval.

(H and I) Transverse sections of the hypocotyls of wild type (WT, H) and *AT2G04850* knockout mutant (I). X, xylem (cyan dots). Inset, UMAP plot showing the expression pattern of *AT2G04850*.

(J–M) Transverse sections of the stems of WT (J and K) and *AT2G04850* mutant (L and M). Higher magnification of images (K and M) in frames are shown. Four plants for each genotype (H–M) were examined. One representative result is shown. The number of xylem cells in the *AT2G04850* mutant is significantly different from that in WT ($p = 0.0017$, Student's t test). Scale bar in (H–M), 20 μm .

See also Figure S7.

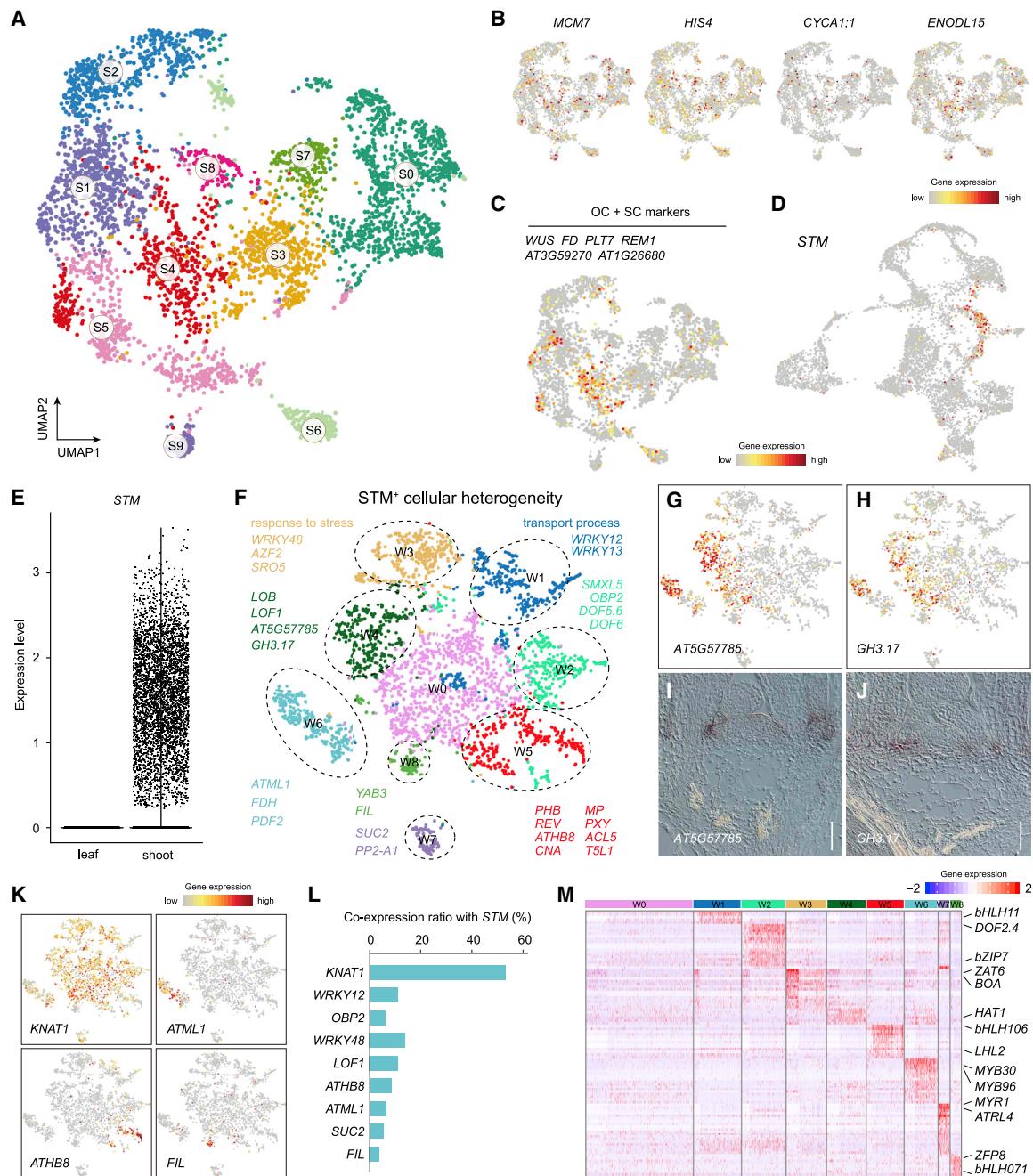


Figure 7. Characterization of the SMC population and STM⁺ cells

- (A) UMAP plot showing ten sub-cell clusters (S0 to S9) of the SMC population. $n = 4,283$ cells.
- (B) Expression of four representative cell-cycle genes on the UMAP. Clustering is the same as in (A).
- (C) Identification of the cells corresponding to the CZ/OC domain of the SAM. The selected marker genes, which are highly expressed in stem cells (SC) and OC, are plotted on the UMAP. Clustering is the same as in (A).
- (D) UMAP plot showing the expression pattern of STM in the EC population. See also Figure 4A.
- (E) Expression of STM in the leaf and shoot samples. Please note that there was no expression of STM in the leaf sample. Each dot represents a cell.
- (F) Cell clustering of the STM⁺ cells. Nine sub-cell clusters (W0 to W8) were revealed. The representative genes and associated biological process of each cluster are shown. $n = 3,456$ cells. The full names of selected genes are summarized in Table S2.
- (G and H) UMAP plots showing the expression pattern of AT5G57785 and GH3.17. Clustering is the same as in (F).
- (I and J) RNA *in situ* hybridization assays of AT5G57785 and GH3.17. Scale bar, 50 μ m.
- (K) UMAP plots showing the expression pattern of KNAT1, ATML1, ATHB8, and FIL genes. Clustering is the same as in (F).

(legend continued on next page)

redundant and have important roles in maintaining division and preventing differentiation of cells in the SAM. We were able to identify a number of transcription factors specifically expressed in clusters W1 to W8 (Figure 7M). The functional analysis of these genes using a reverse genetics approach may shed light on how distinct differentiation lineages are specified in the future.

Characterization of the reMC population

In addition to cell clusters related to the cell cycle, there were seven distinct cell clusters in the reMC population (Figure 4B). Unfortunately, GO term analysis and cluster-specific genes did not identify any functionalities of these cells (Table S3). Although it is generally accepted that mesophyll cells are composed of two types of cells, namely, the spongy and palisade cells (Bray-brook and Kuhlemeier, 2010), the identity and functional specification of mesophyll cells are more complex than we previously thought.

Comparison between shoot and root apical tissues at single-cell resolution

Finally, we tried to dissect the conservation and divergence between shoot and root apices at single-cell resolution. As shown in Figure 8A, combined analyses of root and shoot scRNA-seq datasets, which included 60,099 cells, revealed 27 panoramic cell clusters (P0 to P26) (Denyer et al., 2019; Ryu et al., 2019; Zhang et al., 2019). Most of cell clusters were specific for either the shoot or root (Figure 8B). Cluster P22, which was dominated by the GC cells, harbored 27% root cells. Analyses of the cluster P22-enriched genes revealed high transcript levels of *HEAT SHOCK PROTEIN* (*HSP*) genes (Figure S8A), suggesting that some root cells share transcriptome similarity to the GC, probably due to the same hypersensitive nature to environmental conditions. As such, we annotated cluster P22 as GC/HSP⁺.

The shoot and root apices shared six vascular tissue cell clusters (Figure 8B). These shared vascular clusters were divided into 14 sub-cell clusters, I0 to I13 (Figure 8C). Of particular interest, we found that two clusters, I11 and I12, were specific for the root and shoot, respectively (Figure 8D). Early work has shown that *NITRATE TRANSPORTER1/PEPTIDE TRANSPORTER FAMILY Y3.1* (*NPF3.1*), one of the cluster I12-enriched gene, was exclusively expressed in aerial vascular tissues (Pike et al., 2014). RNA *in situ* hybridization assays revealed that *NPF3.1* was expressed in mesophyll cells and a layer of cells surrounding the vascular bundles (Figures 8E and 8F). GO enrichment analysis further revealed that cluster I12 showed gene signatures for photosynthesis (Figure 8G; Table S6). Hence, these results suggest that I12 consists of bundle sheath cells, which play an important role in the coordination of carbon and nitrogen acquisition (Leegood, 2008). Knowledge of the gene activities of this cluster thus lays the groundwork for engineering crops with C3 photosynthesis into C4 crops in the future.

The shoot EC clusters (P3, P23, and P24) did not overlap with those in roots (P14 and P20) on the t-SNE plot (Figure 8A).

Despite this difference, we noticed that the aerial epidermal marker gene *ATML1* was also expressed in the epidermal cells at the root tip (Figures 8H and 8I). Moreover, a set of genes involved in epidermal cuticle biosynthesis, including *FIDDLEHEAD*, *FATTY ACID REDUCTASE1*, *GLYCEROL-3-PHOSPHATE SN-2-ACYLTRANSFERASE4* (*GPAT4*), *GPAT8*, and *3-KETOACYL-COA SYNTHASE20* was co-expressed in the shoot and root *ATML1*⁺ cells (Kosma et al., 2012; Lee et al., 2009; Li et al., 2007; Pruitt et al., 2000) (Figure 8J, core gene list). This result is in agreement with a recent finding that a cuticle-like cell wall structure exists in the primary root tip and emerging lateral root (Berbin et al., 2019) and suggests an important role of *ATML1* in cuticle biosynthesis in roots. Indeed, the expression domain of *ATML1* spatiotemporally overlapped with the deposition of aliphatic polyesters in the primary and lateral root tips as revealed by fluorol yellow (FY) staining (Figures 8K, S8D, and S8E). Moreover, the comparison between *ATML1*⁺ and *ATML1*⁻ root cells in the root cell atlas indicated that the *ATML1*⁺ cells were associated with high transcript level of cuticle biosynthesis genes, whereas the *ATML1*⁻ root cells are enriched in genes involved in substance transport and defense response, probably due to the lack of a diffusion barrier (Figures S8B and S8C). Taken together, the above results uncover previously unrecognized similarities between shoot and root epidermis and delineate a role of *ATML1* in the EC differentiation in roots.

DISCUSSION

Generation of an *Arabidopsis* vegetative shoot apex cell atlas

In this work, we report a comprehensive single-cell compendium of the *Arabidopsis* vegetative shoot apex. scRNA-seq technology is revolutionizing the identification of cell types and differentiation trajectories of stem cells. We identified numerous cellular clusters and state changes in the shoot apex, each defined either by gene expression markers, gene signatures, or by associated biological processes. This work will facilitate future work on reproductive shoot development at single-cell resolution and serve as a resource to understand cell fate determination during lateral organ formation.

By sequencing the transcriptome of individual cells, we can now measure the output of the regulatory network without the loss of information caused by the signal averaging in bulk RNA sequencing technologies. As such, our results provide three new insights into *Arabidopsis* vegetative shoot development. First, we delineated the cell-cycle continuum in the shoot apex at the tissue level. The generation of the core cell-cycle gene list and the assignment of these genes into distinct cell-cycle phases deepen our understanding of cell-cycle regulation in plants. Moreover, the comparisons of the cell-cycle components across different tissues shed light on the mechanisms by which subtle changes in duration and pattern of cell division contribute to the formation of various organs.

(L) Co-expression analysis. The co-expression ratio between *STM* and other genes across all the *STM*⁺ cells was calculated by number of co-expressed cells/total number of *STM*⁺ cells.

(M) Heatmap of top 105 enriched genes of each cell cluster in the *STM*⁺ cells. There are 1008, 442, 408, 382, 362, 345, 307, 102, and 100 cells from clusters W0 to W8, respectively. The representative transcription factor genes are shown on the left. Color bar indicates the relative expression level. The full names of selected genes are summarized in Table S2.

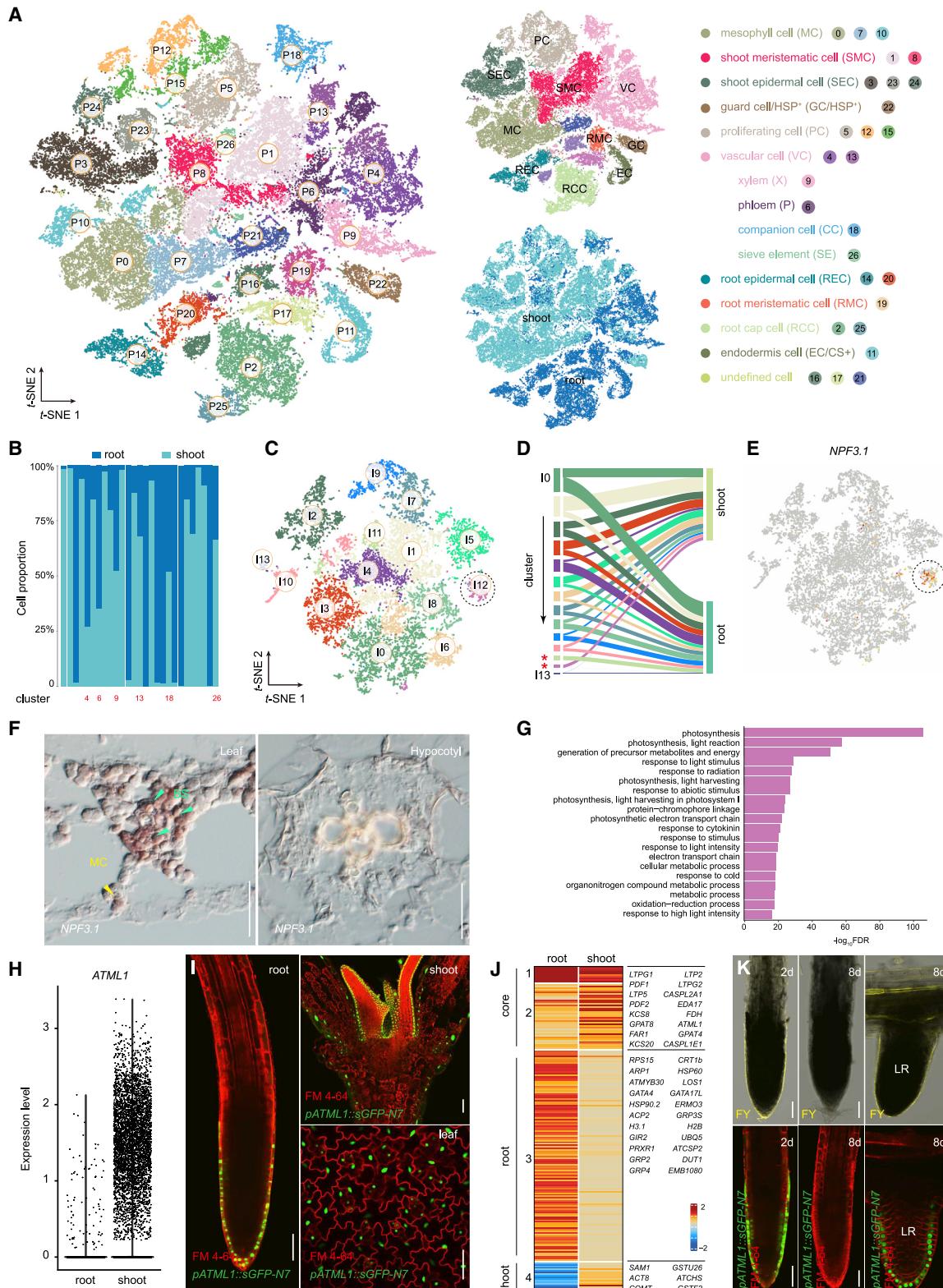


Figure 8. Comparison of shoot and root apical cell populations at single-cell resolution

(A) Visualization of 27 cell clusters using t-SNE. Dots, individual cells; $n = 60,099$ cells (3 root datasets, 2 shoot datasets, and 1 leaf dataset); color, cell clusters. The annotated cell populations and tissues are labeled by different colors (right).

(B) The proportions of tissue contributions to each cluster are shown. Colored cluster number represent shared vascular cell clusters of shoot and root tissues.

(legend continued on next page)

Second, our results demonstrate the accuracy of unsupervised cell clustering and pseudotime analysis in delineation of cell differentiation trajectories in the shoot apex. Similar results have been obtained in scRNA-seq of the *Arabidopsis* roots (Denyer et al., 2019; Jean-Baptiste et al., 2019; Ryu et al., 2019; Shahan et al., 2020; Shulse et al., 2019; Turco et al., 2019; Wendrich et al., 2020; Zhang et al., 2019). Thus, based on these achievements, future work could dissect the differentiation trajectories of other tissues such as epidermal pavement cells and mesophyll cells. More broadly, this strategy can be adapted in major crops, where cell fate determination and pattern formation of the lateral organs are poorly understood.

Third, using *in situ* hybridization assays, previous studies have shown that the expression of the meristematic gene *STM* and leaf identity genes such as *FIL* and *YAB3* are mutually exclusive in the shoot apex (Eshed et al., 2004; Nole-Wilson and Krizek, 2006). It should be noted that, in these experiments, the expression pattern of *STM* and the leaf identity gene were not simultaneously examined on the same slide. Therefore, it is unclear whether some shoot apex cells express both genes. Our scRNA-seq now provides evidence showing that the transcripts of *FIL* and *YAB3* are indeed accumulated in a number of *STM*⁺ cells in the vegetative shoot apex (Figure 7F). These *STM*⁺ *FIL/YAB3*⁺ cells can be considered the transit-amplifying meristematic cells toward leaf identity. Notably, this observation supports the notion that cells often exhibit a continuum of expression states as they differentiate (Stumpf et al., 2017) and further demonstrates a major advantage of using scRNA-seq to capture the hidden cell states along the differentiation trajectory in plants.

scRNA-seq facilitates functional analysis and gene discovery

Our results show the power of scRNA-seq technology in acceleration of gene discovery in *Arabidopsis*. With the development of T-DNA insertion lines for every gene in *Arabidopsis* (Alonso et al., 2003), the current challenge is to understand their biological functions (Harter and Weber, 2013; Koornneef and Meinke, 2010; Provart et al., 2016). However, given the saturation of forward genetics screening in the past three decades, the newly identified single mutants usually lack visible phenotypes, probably due to functional redundancy or phenotypic changes in specific tissues or cells. The cell-type-specific genes revealed by

scRNA-seq may help us clear this hurdle. For instance, we have shown that the mutation in *AT2G04850*, which is specifically expressed in the xylem lineage (Figure 6I), leads to defects in xylem development (Figures 6H–6M). Moreover, our results also uncover the function of *AT5G22580* and *AT1G25230*, two genes enriched in cluster 20, in regulating shoot branch angle (Figure 3J). Thus, the precise and comprehensive gene expression pattern revealed by scRNA-seq provides a useful guide for us to choose the tissue that we should closely examine, thereby markedly increasing the success rate of reverse genetics.

To facilitate the use of our dataset, we have implemented a webserver for investigation of genes of interest (<http://wanglab.sippe.ac.cn/shootatlas/>). The browser provides a user-friendly interface to query individual genes in the shoot apex cell atlas. In addition, it enables exploration of gene expression dynamics along the developmental trajectories of epidermal, vascular, and leaf mesophyll cells.

Limitation of this study

Plant and animal cells share similarities in their structures and functions, but the few structural differences that exist between them are very significant and reflect kingdom-specific differences in their functions. Plant cells are much larger than animal cells. Notably, the size of a plant cell can range from 10 to 100 μm. The large variation in cell size creates a big problem in the cell capture step of plant scRNA-seq experiments, particularly when using commercial droplet-based scRNA-seq platforms including the 10x Genomics Chromium Single Cell Instrument. Therefore, development of cell capture platforms with a wide size range is an important future research direction.

Beyond size, plant cells are surrounded by a cell wall, which differs in composition and thickness according to the plant species, tissue, developmental stage, and environmental conditions. The cell wall creates a hurdle for single-cell preparation. For example, we found that the leaf trichome cells are resistant to protoplasting (Figure S1D), and we were unable to recover single trichome protoplast cells for downstream analysis. More importantly, the protoplasting procedure itself can lead to abrupt changes in transcriptomes. Hence, development of a suitable and universal single-cell preparation protocol is urgently needed.

Single nucleus RNA sequencing (snRNA-seq) has the potential to overcome the above problems (Habib et al., 2017). On the one

- (C) t-SNE plot showing the subclustering of the combined VC population from shoots and roots. $n = 13,659$ cells; dot cycle (I12), unique cell cluster for shoots. Each dot represents a cell.
- (D) Sankey diagram showing the conserved and unique VC populations (asterisk, I11 and I12) in shoots and roots. Colors according to (C).
- (E) t-SNE plot showing the expression pattern of *NPF3.1*. Clustering is the same as in (C).
- (F) Expression pattern of *NPF3.1* in leaf and hypocotyl. The leaf and hypocotyl were transversely sectioned. *NPF3.1* transcripts were barely detected in hypocotyls. MC, mesophyll cells; BS, bundle sheath cells. Scale bar, 20 μm.
- (G) GO term analysis of cluster I12. The top 20 enriched GO biological processes are given.
- (H) Expression of *ATML1* in the root and shoot samples. Please note that *ATML1* was also expressed in some root cells. Each dot represents a cell.
- (I) Expression pattern of the *pATML1::sGPF-N7* reporter (green) reporter. The plasma membrane of root cells was visualized by the FM4-64 staining (red). Scale bar, 50 μm.
- (J) Gene clustering of shoot and root *ATML1*⁺ cells ($n = 4,509$ cells). The core (clusters 1 and 2) and root- or shoot-specific sub-gene clusters (clusters 3 and 4) were identified. The representative genes are shown on the right. The full names of selected genes are summarized in Table S2.
- (K) Expression of *pATML1::sGPF-N7* reporter (green, bottom) and FY staining (overlay bright field and yellow fluorescence, top) in 2-day-old (2 d) and 8-day-old (8 d) seedlings. Cell outlines were visualized by FM4-64 staining (red). Please note that both the FY staining and *sGPF-N7* reporter activity were weak in the primary root tip of 8-day-old seedlings but became evident during the emergence of lateral root (LR). Scale bar, 50 μm.
- See also Figure S8; Tables S2 and S6.

hand, snRNA-seq can be used to analyze the organs that are more vulnerable to the tissue dissociation process and are underrepresented in the final data set. On the other hand, snRNA-seq can, in principle, capture all the cell nuclei regardless of cell size. However, the application of snRNA-seq technology in plants has not yet been reported. In addition, it remains unclear whether individual nuclei contain sufficient diversity and number of transcripts to enable discrimination of closely related cell types at a resolution comparable with whole cells in plants.

Another major challenge of scRNA-seq in plants is to match the messenger RNA profile of a cell with its position within a tissue or organ. One possible solution is spatial transcriptomics that allows for the characterization of gene expression in bar-coded regions of individual tissue sections. This technology, when combined with scRNA-seq or snRNA-seq, may have the potential to answer a wide range of biological questions concerning positional information in plants. Alternatively, the development of toolboxes for *in situ* imaging mRNAs, such as MERFISH, seqFISH⁺, and STARmap (Eng et al., 2019; Su et al., 2020; Wang et al., 2018), will ultimately spatially resolve gene expression pattern in any biological processes (Pichon et al., 2018).

Finally, while the 10x Genomics Chromium system is the most cost effective and time saving method, the sequencing reads are restricted to 3' end of the transcripts and overall sensitivity is reduced compared with "full-length" scRNA-seq methods such as SMART-seq2 and Fluidigm C1 (Ziegenhain et al., 2017). As such, we found that the transcripts encoding the key regulators of stem cells are poorly retrieved, and that our scRNA-seq is not sufficient to faithfully uncover previously identified cell types (i.e., CLV3 domain and the OC) within the SAM. Thus, the combination of scRNA-seq with cell-type marker-based cell-sorting technology will help us to precisely resolve cell heterogeneity of the SAM in the future.

In summary, we have generated a gene expression map of the vegetative shoot apex at single-cell resolution. The definition of cells in distinct layers and functional zones of the vegetative SAM based on unique transcriptomes now offers researchers the opportunity to investigate, at unprecedented resolution, how the SAM is maintained, how stem cells differentiate into distinct cell types through transit-amplifying cells, and how each cell communicates with its neighbors.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

● KEY RESOURCES TABLE

● RESOURCE AVAILABILITY

- Lead contact
- Materials availability
- Data and code availability

● EXPERIMENTAL MODEL AND SUBJECT DETAILS

● METHOD DETAILS

- Plant materials and growth conditions
- Identification of mutant
- Construct
- RNA *in situ* hybridization

- Section, staining and microscope
- Scanning electron microscope
- Sample collection and preparation of single-cell suspensions
- scRNA-seq library construction

● QUANTIFICATION AND STATISTICAL ANALYSIS

- Pre-processing of raw scRNA-seq data
- Doublet detection
- Data integration, clustering, and annotation
- Correlation analysis
- Subclustering
- Differential trajectory and lineage analysis
- RNA velocity analysis
- Gene-ontology enrichment analysis
- Visualization of the transcription factor network
- Arabidopsis cell landscape construction

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.devel.2021.02.021>.

ACKNOWLEDGMENTS

We thank members in J.-W. Wang lab for discussions and comments on the manuscript and Ji-Qin Li (Core Facility Center of CEPMS/SIPPE, CAS) for technical support on the scanning electron microscope. This work was supported by the grants from National Natural Science Foundation of China (31788103, 31525004, and 31721001), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB27030101), the Chinese Academy of Sciences (QYZDB-SSW-SMC002), the Science and Technology Commission of Shanghai Municipality (18JC1415000), the Young Elite Scientists Sponsorship Program by China Association for Science and Technology (2016QNRC001), and the National Postdoctoral Program for Innovative Talents (BX201600178).

AUTHOR CONTRIBUTIONS

J.-W.W. and T.-Q.Z. conceived the project; T.-Q.Z. generated scRNA-seq data; T.-Q.Z., Y.C., and J.-W.W. performed experiments; T.-Q.Z. and J.-W.W. analyzed the data; J.-W.W. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

The author list of this paper includes contributors from the location where the research was conducted, who participated in the data collection, design, analysis, and/or interpretation of the work.

Received: August 7, 2020

Revised: December 6, 2020

Accepted: February 19, 2021

Published: March 15, 2021

REFERENCES

- Adrian, J., Chang, J., Ballenger, C.E., Bargmann, B.O., Alassimone, J., Davies, K.A., Lau, O.S., Matos, J.L., Hachez, C., Lanctot, A., et al. (2015). Transcriptome dynamics of the stomatal lineage: birth, amplification, and termination of a self-renewing population. *Dev. Cell* 33, 107–118.
- Aichinger, E., Kornet, N., Friedrich, T., and Laux, T. (2012). Plant stem cell niches. *Annu. Rev. Plant Biol.* 63, 615–636.

- Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., et al. (2003). Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* 301, 653–657.
- Becht, E., McInnes, L., Healy, J., Dutertre, C.A., Kwok, I.W.H., Ng, L.G., Ginhoux, F., and Newell, E.W. (2018). Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* 37, 38–44.
- Bell, E.M., Lin, W.C., Husbands, A.Y., Yu, L., Jaganatha, V., Jablonska, B., Mangeon, A., Neff, M.M., Girke, T., and Springer, P.S. (2012). Arabidopsis lateral organ boundaries negatively regulates brassinosteroid accumulation to limit growth in organ boundaries. *Proc. Natl. Acad. Sci. USA* 109, 21146–21151.
- Berhin, A., de Bellis, D., Franke, R.B., Buono, R.A., Nowack, M.K., and Nawrath, C. (2019). The root cap cuticle: a cell wall structure for seedling establishment and lateral root formation. *Cell* 176, 1367–1378.e8.
- Birnbaum, K., Shasha, D.E., Wang, J.Y., Jung, J.W., Lambert, G.M., Galbraith, D.W., and Benfey, P.N. (2003). A gene expression map of the Arabidopsis root. *Science* 302, 1956–1960.
- Birnbaum, K.D. (2018). Power in numbers: single-cell RNA-seq strategies to dissect complex tissues. *Annu. Rev. Genet.* 52, 203–221.
- Braybrook, S.A., and Kuhlemeier, C. (2010). How a plant builds leaves. *Plant Cell* 22, 1006–1018.
- Byrne, M.E., Barley, R., Curtis, M., Arroyo, J.M., Dunham, M., Hudson, A., and Martienssen, R.A. (2000). Asymmetric leaves1 mediates leaf patterning and stem cell function in Arabidopsis. *Nature* 408, 967–971.
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890.
- Chitwood, D.H., Guo, M., Nogueira, F.T., and Timmermans, M.C. (2007). Establishing leaf polarity: the role of small RNAs and positional signals in the shoot apex. *Development* 134, 813–823.
- 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.
- Dinneny, J.R., and Benfey, P.N. (2008). Plant stem cell niches: standing the test of time. *Cell* 132, 553–557.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.
- Eng, C.L., Lawson, M., Zhu, Q., Dries, R., Koulena, N., Takei, Y., Yun, J., Cronin, C., Karp, C., Yuan, G.-C., and Cai, L. (2019). Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. *Nature* 568, 235–239.
- Eshed, Y., Izhaki, A., Baum, S.F., Floyd, S.K., and Bowman, J.L. (2004). Asymmetric leaf development and blade expansion in Arabidopsis are mediated by KANADI and YABBY activities. *Development* 131, 2997–3006.
- Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32, 3047–3048.
- Franco-Zorrilla, J.M., Cubas, P., Jarillo, J.A., Fernández-Calvín, B., Salinas, J., and Martínez-Zapater, J.M. (2002). AtREM1, a member of a new family of B3 domain-containing genes, is preferentially expressed in reproductive meristems. *Plant Physiol.* 128, 418–427.
- Gaillochet, C., Daum, G., and Lohmann, J.U. (2015). O cell, where art thou? The mechanisms of shoot meristem patterning. *Curr. Opin. Plant Biol.* 23, 91–97.
- Gallop, J.L., Woodward, C., Reddy, G.V., and Sablowski, R. (2002). Combined SHOOT MERISTEMLESS and WUSCHEL trigger ectopic organogenesis in Arabidopsis. *Development* 129, 3207–3217.
- Greb, T., and Lohmann, J.U. (2016). Plant stem cells. *Curr. Biol.* 26, R816–R821.
- Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32, 2847–2849.
- Guo, X., and Dong, J. (2019). To divide or differentiate: it is about scaffolding. *Trends Plant Sci.* 24, 481–484.
- Guseman, J.M., Webb, K., Srinivasan, C., and Dardick, C. (2017). DRO1 influences root system architecture in Arabidopsis and Prunus species. *Plant J.* 89, 1093–1105.
- Gutierrez, C. (2009). The Arabidopsis cell division cycle. *Arabidopsis Book* 7, e0120.
- Habib, N., Avraham-David, I., Basu, A., Burks, T., Shekhar, K., Hofree, M., Choudhury, S.R., Aguet, F., Gelfand, E., Ardlie, K., et al. (2017). Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* 14, 955–958.
- Han, S.K., and Torii, K.U. (2019). Linking cell cycle to stomatal differentiation. *Curr. Opin. Plant Biol.* 51, 66–73.
- Harter, K., and Weber, A.P. (2013). Arabidopsis 2010 and beyond—big science with a small weed. *Front. Plant Sci.* 4, 18.
- Hashiguchi, Y., Yano, D., Nagafusa, K., Kato, T., Saito, C., Uemura, T., Ueda, T., Nakano, A., Tasaka, M., and Terao Morita, M. (2014). A unique HEAT repeat-containing protein SHOOT GRAVITROPISM6 is involved in vacuolar membrane dynamics in gravity-sensing cells of Arabidopsis inflorescence stem. *Plant Cell Physiol.* 55, 811–822.
- Heidstra, R., and Sabatini, S. (2014). Plant and animal stem cells: similar yet different. *Nat. Rev. Mol. Cell Biol.* 15, 301–312.
- Holt, A.L., van Haperen, J.M., Groot, E.P., and Laux, T. (2014). Signaling in shoot and flower meristems of Arabidopsis thaliana. *Curr. Opin. Plant Biol.* 17, 96–102.
- Hsu, Y.C., Li, L., and Fuchs, E. (2014). Transit-amplifying cells orchestrate stem cell activity and tissue regeneration. *Cell* 157, 935–949.
- Husbands, A.Y., Chitwood, D.H., Plavskin, Y., and Timmermans, M.C. (2009). Signals and prepatterns: new insights into organ polarity in plants. *Genes Dev.* 23, 1986–1997.
- Jean-Baptiste, K., McFadie-Figueroa, J.L., Alexandre, C.M., Dorrit, M.W., Saunders, L., Bubb, K.L., Trapnell, C., Fields, S., Quetsch, C., and Cuperus, J.T. (2019). Dynamics of gene expression in single root cells of *Arabidopsis thaliana*. *Plant Cell* 31, 993–1011.
- Kawamoto, N., Kanbe, Y., Nakamura, M., Mori, A., and Terao Morita, M. (2020). Gravity-sensing tissues for gravitropism are required for "anti-gravitropic" phenotypes of Lzy multiple mutants in Arabidopsis. *Plants* 9, 615.
- Kitagawa, M., and Jackson, D. (2019). Control of meristem size. *Annu. Rev. Plant Biol.* 70, 269–291.
- Koornneef, M., and Meinke, D. (2010). The development of Arabidopsis as a model plant. *Plant J.* 61, 909–921.
- Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.R., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with harmony. *Nat. Methods* 16, 1289–1296.
- Kosma, D.K., Molina, I., Ohlrogge, J.B., and Pollard, M. (2012). Identification of an Arabidopsis fatty alcohol:caffeoyle-coenzyme A acyltransferase required for the synthesis of alkyl hydroxycinnamates in root waxes. *Plant Physiol.* 160, 237–248.
- La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lönnberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. *Nature* 560, 494–498.
- Lee, L.R., and Bergmann, D.C. (2019). The plant stomatal lineage at a glance. *J. Cell Sci.* 132, jcs228551.
- Lee, S.B., Jung, S.J., Go, Y.S., Kim, H.U., Kim, J.K., Cho, H.J., Park, O.K., and Suh, M.C. (2009). Two Arabidopsis 3-ketoacyl CoA synthase genes, KCS20 and KCS2/DAISY, are functionally redundant in cuticular wax and root suberin biosynthesis, but differentially controlled by osmotic stress. *Plant J.* 60, 462–475.
- Leegood, R.C. (2008). Roles of the bundle sheath cells in leaves of C3 plants. *J. Exp. Bot.* 59, 1663–1673.
- Li, Y., Beisson, F., Koo, A.J., Molina, I., Pollard, M., and Ohlrogge, J. (2007). Identification of acyltransferases required for cutin biosynthesis and

- production of cutin with suberin-like monomers. *Proc. Natl. Acad. Sci. USA* **104**, 18339–18344.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079.
- Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930.
- 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.
- Liu, Z., Zhou, Y., Guo, J., Li, J., Tian, Z., Zhu, Z., Wang, J., Wu, R., Zhang, B., Hu, Y., et al. (2020). Global dynamic molecular profiling of stomatal lineage cell development by single-cell RNA sequencing. *Mol. Plant* **13**, 1178–1193.
- Lu, P.L., Chen, N.Z., An, R., Su, Z., Qi, B.S., Ren, F., Chen, J., and Wang, X.C. (2007). A novel drought-inducible gene, ATAF1, encodes a NAC family protein that negatively regulates the expression of stress-responsive genes in Arabidopsis. *Plant Mol. Biol.* **63**, 289–305.
- Matos, J.L., Lau, O.S., Hachez, C., Cruz-Ramírez, A., Scheres, B., and Bergmann, D.C. (2014). Irreversible fate commitment in the Arabidopsis stomatal lineage requires a FAMA and RETINOBLASTOMA-RELATED module. *eLife* **3**, e03271.
- Matsumoto, H., Kiryu, H., Furusawa, C., Ko, M.S.H., Ko, S.B.H., Gouda, N., Hayashi, T., and Nikaido, I. (2017). SCODE: an efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation. *Bioinformatics* **33**, 2314–2321.
- McConnell, J.R., Emery, J., Eshed, Y., Bao, N., Bowman, J., and Barton, M.K. (2001). Role of PHABULOSA and PHAVOLUTA in determining radial patterning in shoots. *Nature* **411**, 709–713.
- McGinnis, C.S., Murrow, L.M., and Gartner, Z.J. (2019). DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst* **8**, 329–337.e4.
- Menges, M., Hennig, L., Gruissem, W., and Murray, J.A. (2002). Cell cycle-regulated gene expression in Arabidopsis. *J. Biol. Chem.* **277**, 41987–42002.
- Mironova, V., and Xu, J. (2019). A single-cell view of tissue regeneration in plants. *Curr. Opin. Plant Biol.* **52**, 149–154.
- Miyashima, S., Roszak, P., Sevilem, I., Toyokura, K., Blob, B., Heo, J.-O., Mellor, N., Help-Rinta-Rahko, H., Otero, S., Smet, W., et al. (2019). Mobile PEAR transcription factors integrate positional cues to prime cambial growth. *Nature* **565**, 490–494.
- Miyashima, S., Sebastian, J., Lee, J.Y., and Helariutta, Y. (2013). Stem cell function during plant vascular development. *EMBO J.* **32**, 178–193.
- Nelms, B., and Walbot, V. (2019). Defining the developmental program leading to meiosis in maize. *Science* **364**, 52–56.
- Nole-Wilson, S., and Krizek, B.A. (2006). ANTIGUMENTA contributes to organ polarity and regulates growth of lateral organs in combination with YABBY genes. *Plant Physiol.* **141**, 977–987.
- Perales, M., and Reddy, G.V. (2012). Stem cell maintenance in shoot apical meristems. *Curr. Opin. Plant Biol.* **15**, 10–16.
- Pichon, X., Lagha, M., Mueller, F., and Bertrand, E. (2018). A growing toolbox to image gene expression in single cells: sensitive approaches for demanding challenges. *Mol. Cell* **71**, 468–480.
- Pierre-Jerome, E., Drapek, C., and Benfey, P.N. (2018). Regulation of division and differentiation of plant stem cells. *Annu. Rev. Cell Dev. Biol.* **34**, 289–310.
- Pike, S., Gao, F., Kim, M.J., Kim, S.H., Schachtman, D.P., and Gassmann, W. (2014). Members of the NPF3 transporter subfamily encode pathogen-inducible nitrate/nitrite transporters in grapevine and Arabidopsis. *Plant Cell Physiol.* **55**, 162–170.
- Provart, N.J., Alonso, J., Assmann, S.M., Bergmann, D., Brady, S.M., Brkljacic, J., Browse, J., Chapple, C., Colot, V., Cutler, S., et al. (2016). 50 years of Arabidopsis research: highlights and future directions. *New Phytol.* **209**, 921–944.
- Pruitt, R.E., Vielle-Calzada, J.P., Ploense, S.E., Grossniklaus, U., and Lolle, S.J. (2000). FIDDLEHEAD, a gene required to suppress epidermal cell interactions in Arabidopsis, encodes a putative lipid biosynthetic enzyme. *Proc. Natl. Acad. Sci. USA* **97**, 13111–13116.
- Pyronnet, S., and Sonenberg, N. (2001). Cell-cycle-dependent translational control. *Curr. Opin. Genet. Dev.* **11**, 13–18.
- Qiu, X., Hill, A., Packer, J., Lin, D., Ma, Y.A., and Trapnell, C. (2017). Single-cell mRNA quantification and differential analysis with census. *Nat. Methods* **14**, 309–315.
- Rich-Griffin, C., Stechemesser, A., Finch, J., Lucas, E., Ott, S., and Schäfer, P. (2020). Single-cell transcriptomics: a high-resolution avenue for plant functional genomics. *Trends Plant Sci.* **25**, 186–197.
- Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. *Nat Biotechnol* **29**, 24–26.
- 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.
- Sablowski, R. (2007). The dynamic plant stem cell niches. *Curr. Opin. Plant Biol.* **10**, 639–644.
- Sablowski, R. (2011). Plant stem cell niches: from signalling to execution. *Curr. Opin. Plant Biol.* **14**, 4–9.
- Setty, M., Kisielovas, V., Levine, J., Gayoso, A., Mazutis, L., and Pe'er, D. (2019). Characterization of cell fate probabilities in single-cell data with Palantir. *Nat. Biotechnol.* **37**, 451–460.
- Shahan, R., Hsu, C.-W., Nolan, T., Cole, B., Taylor, I., Vlot, A.H.C., Benfey, P., and Ohler, U. (2020). A single cell Arabidopsis root atlas reveals developmental trajectories in wild type and cell identity mutants. *bioRxiv* <https://www.biorxiv.org/content/10.1101/2020.06.29.178863v1>.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504.
- Shulse, C.N., Cole, B.J., Ciobanu, D., Lin, J., Yoshinaga, Y., Gouran, M., Turco, G.M., Zhu, Y., O’Malley, R.C., Brady, S.M., and Dickel, D.D. (2019). High-throughput single-cell transcriptome profiling of plant cell types. *Cell Rep* **27**, 2241–2247.e4.
- Smetana, O., Mäkilä, R., Lyu, M., Amiryousefi, A., Sánchez Rodríguez, F., Wu, M.F., Solé-Gil, A., Leal Gavarrón, M., Siligato, R., Miyashima, S., et al. (2019). High levels of auxin signalling define the stem-cell organizer of the vascular cambium. *Nature* **565**, 485–489.
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21.
- Stumpf, P.S., Smith, R.C.G., Lenz, M., Schuppert, A., Müller, F.J., Babtie, A., Chan, T.E., Stumpf, M.P.H., Please, C.P., Howison, S.D., et al. (2017). Stem cell differentiation as a non-markov stochastic process. *Cell Syst.* **5**, 268–282.e7.
- Su, J.H., Zheng, P., Kinrot, S.S., Bintu, B., and Zhuang, X. (2020). Genome-scale imaging of the 3D organization and transcriptional activity of chromatin. *Cell* **182**, 1641–1659.e26.
- Sugano, S.S., Shimada, T., Imai, Y., Okawa, K., Tamai, A., Mori, M., and Hara-Nishimura, I. (2010). Stomagen positively regulates stomatal density in Arabidopsis. *Nature* **463**, 241–244.
- Swaminathan, K., Yang, Y., Grotz, N., Campisi, L., and Jack, T. (2000). An enhancer trap line associated with a D-class cyclin gene in Arabidopsis. *Plant Physiol.* **124**, 1658–1667.
- Taniguchi, M., Furutani, M., Nishimura, T., Nakamura, M., Fushita, T., Iijima, K., Baba, K., Tanaka, H., Toyota, M., Tasaka, M., et al. (2017). The Arabidopsis LAZY1 family plays a key role in gravity signaling within Statocytes and in branch angle control of roots and shoots. *Plant Cell* **29**, 1984–1999.
- Tian, C., Wang, Y., Yu, H., He, J., Wang, J., Shi, B., Du, Q., Provart, N.J., Meyerowitz, E.M., and Jiao, Y. (2019). A gene expression map of shoot domains reveals regulatory mechanisms. *Nat. Commun.* **10**, 141.

- Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., Xu, W., and Su, Z. (2017). agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res.* **45**, W122–W129.
- Turco, G.M., Rodriguez-Medina, J., Siebert, S., Han, D., Valderrama-Gómez, M.Á., Vahldick, H., Shulse, C.N., Cole, B.J., Juliano, C.E., Dickel, D.E., et al. (2019). Molecular mechanisms driving switch behavior in xylem cell differentiation. *Cell Rep.* **28**, 342–351.e4.
- Van Leene, J., Hollunder, J., Eeckhout, D., Persiau, G., Van De Slijpe, E., Stals, H., Van Isterdael, G., Verkest, A., Neirynck, S., Buffel, Y., et al. (2010). Targeted interactomics reveals a complex core cell cycle machinery in *Arabidopsis thaliana*. *Mol. Syst. Biol.* **6**, 397.
- Wallner, E.S., Tonn, N., Shi, D., Jouannet, V., and Greb, T. (2020). SUPPRESSOR of max2 1-LIKE 5 promotes secondary phloem formation during radial stem growth. *Plant J.* **102**, 903–915.
- Wang, F.X., Shang, G.D., Wu, L.Y., Xu, Z.G., Zhao, X.Y., and Wang, J.W. (2020). Chromatin accessibility dynamics and a hierarchical transcriptional regulatory network structure for plant somatic embryogenesis. *Dev. Cell* **54**, 742–757.e8.
- Wang, X., Allen, W.E., Wright, M.A., Sylwestrak, E.L., Samusik, N., Vesuna, S., Evans, K., Liu, C., Ramakrishnan, C., Liu, J., et al. (2018). Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361**, eaat5691.
- Wendrich, J.R., Yang, B., Vandamme, N., Verstaen, K., Smet, W., Van de Velde, C., Minne, M., Wybouw, B., Mor, E., Arents, H.E., et al. (2020). Vascular transcription factors guide plant epidermal responses to limiting phosphate conditions. *Science* **370**, eaay4970.
- Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15.
- Yadav, R.K., Girke, T., Pasala, S., Xie, M., and Reddy, G.V. (2009). Gene expression map of the *Arabidopsis* shoot apical meristem stem cell niche. *Proc. Natl. Acad. Sci. USA* **106**, 4941–4946.
- Yadav, R.K., Tavakkoli, M., Xie, M., Girke, T., and Reddy, G.V. (2014). A high-resolution gene expression map of the *Arabidopsis* shoot meristem stem cell niche. *Development* **141**, 2735–2744.
- Yoshihara, T., Spalding, E.P., and Iino, M. (2013). AtLAZY1 is a signaling component required for gravitropism of the *Arabidopsis thaliana* inflorescence. *Plant J.* **74**, 267–279.
- Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287.
- Zhang, T.Q., Lian, H., Zhou, C.M., Xu, L., Jiao, Y., and Wang, J.W. (2017). A two-step model for *de novo* activation of *WUSCHEL* during plant shoot regeneration. *Plant Cell* **29**, 1073–1087.
- Zhang, T.Q., Xu, Z.G., Shang, G.D., and Wang, J.W. (2019). A single-cell RNA sequencing profiles the developmental landscape of *Arabidopsis* root. *Mol. Plant* **12**, 648–660.
- Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., Leonhardt, H., Heyn, H., Hellmann, I., and Enard, W. (2017). Comparative analysis of single-cell RNA sequencing methods. *Mol. Cell* **65**, 631–643.e4.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-Digoxigenin-AP Fab fragments	Roche	Cat#11093274910; RRID: AB_2734716
Bacterial and virus strains		
<i>Escherichia coli</i> DH5 α	N/A	N/A
<i>Agrobacterium tumefaciens</i> GV3101 (pMP90)	N/A	N/A
Chemicals, peptides, and recombinant proteins		
Murashige & Skoog basal salts medium	Phyto Technology	Cat#M519
Sucrose	ABCONC	Cat#57501
2-(N-morpholino) ethanesulfonic acid (MES)	BBI Life Sciences	Cat#145224948
Phytigel	Sigma-Aldrich	Cat#7101052-1
Agarose G-10	BIOWEST	Cat#EEO015
Kanamycin sulfate	FCNCS	Cat#M211
Rifampicin solution	FCNCS	Cat#M213
Hygromycin B	YEASEN	Cat#60224ES03
Gentamycin sulfate	FCNCS	Cat#M215
Agar Bacteriological Grade	Shanghai Jiafeng	Cat#H8145
KH ₂ PO ₄	Sigma-Aldrich	Cat#7778770
Sodium acetate buffer solution	Sigma-Aldrich	Cat#126965
Dextran Sulfate 50% solution	Millipore	Cat#S4030
Dimethyl sulfoxide (DMSO)	Amresco	Cat#67685
Triton X-100	Sigma-Aldrich	Cat#9002931
Glycerol	Sigma-Aldrich	Cat#56815
Cellulase R10	Yakult Pharmaceutical	Cat#110721-01
Macerozyme R10	Yakult Pharmaceutical	Cat#131126-02
EDTA-Na ₂	Sigma-Aldrich	Cat#6381926
Mannitol	Sigma-Aldrich	Cat#M4125
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)	Sigma-Aldrich	Cat#7365459
Piperazine-N,N'-bis(2-ethanesulfonic acid) (PIPES)	Sigma-Aldrich	Cat#5625376
CH ₃ COOLi · 2H ₂ O	Sigma-Aldrich	Cat#6108174
Na ₂ HPO ₄	Sigma-Aldrich	Cat#7558794
Na ₂ CO ₃	Sigma-Aldrich	Cat#497198
NaH ₂ PO ₄ · H ₂ O	Sigma-Aldrich	Cat#7558807
KCl	Sigma-Aldrich	Cat#7447407
NaCl	Sigma-Aldrich	Cat#7647145
MgCl ₂ · 6H ₂ O	Sigma-Aldrich	Cat#7791186
Tris base	Sigma-Aldrich	Cat#77861
DAPI	AAT Bioquest	Cat#28718903
Paraplast	Sigma-Aldrich	Cat#P3558
BSA	Amresco	Cat#0332
2-mercaptoethanol	Ruibio	Cat#60242
CaCl ₂	Sigma-Aldrich	Cat#C2661
Fluorol Yellow 088	Sigma-Aldrich	Cat#F5520
Aniline Blue diammonium salt	Sigma-Aldrich	Cat#415049

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
FM4-64	Biotium	Cat#70021
Trypan Blue	Leagene	Cat#CT0030
NaOAC	Sigma-Aldrich	Cat#241245
Toluidine blue	Sinopharm Chemical	Cat#71041284
Phenylmethylsulfonyl fluoride (PMSF)	Sigma-Aldrich	Cat# 93482
Phusion High-Fidelity DNA Polymerase	Thermo Fisher Scientific	Cat#022021
PEG 4000	BBI Life Sciences	Cat#25322683
Complete Protease Inhibitor Cocktail	Merck	Cat#04693132001
NBT/BCIP Stock Solution	Roche	Cat#11681451001
Proteinase K	Roche	Cat#03115828001
Critical commercial assays		
Chromium Single Cell 3' Gel Bead and Library Kit v3	10x Genomics	Cat#1000075
TB Green Premix Extaq II	Takara	Cat#RR820A
CloneExpress II One Step Cloning Kit	Vazyme	Cat#C11202
NEBNext Ultra II DNA Library Prep Kit for Illumina	New England Biolabs	Cat#E7645S
NEBNext Multiplex Oligos for Illumina	New England Biolabs	Cat#E7335S
RNeasy Plant Mini Kit	Qiagen	Cat#74904
MinElute Reaction Cleanup Kit	Qiagen	Cat#28204
Dynabeads Protein G for Immunoprecipitation	Invitrogen	Cat#10004D
TruePrep DNA Library Prep Kit v2	Vazyme Biotech	Cat#TD50102
TruePrep Index Kit v2	Vazyme Biotech	Cat#TD202
DIG RNA labeling mix	Roche	Cat#11277073910
TRIzol™ Reagent	Thermo Fisher Scientific	Cat#15596018
DNase I	Thermo Fisher Scientific	Cat#EN0521
T3 RNA Polymerase	Roche	Cat#11031163001
T7 RNA Polymerase	Roche	Cat#10881767001
RevertAid RT Reverse Transcription Kit	Thermo Fisher Scientific	Cat#K1691
Deposited data		
scRNA-seq experiment data	This paper	BioProject PRJCA003094, Beijing Institute of Genomics Data Center (http://bigd.big.ac.cn)
Experimental models: organisms/strains		
<i>A. thaliana</i> : Col-0	N/A	N/A
<i>A. thaliana</i> : AT2G04850 KO	Arabidopsis Biological Resource Center	SALK_012348
<i>A. thaliana</i> : AT5G22580 KO	Arabidopsis Biological Resource Center	SALK_081434
<i>A. thaliana</i> : AT1G25230 KO	Arabidopsis Biological Resource Center	SALK_200911
<i>A. thaliana</i> : sgr6	Arabidopsis Biological Resource Center	SALK_034928
<i>A. thaliana</i> : pATML1::sGFP-N7	This paper	TQ639
<i>A. thaliana</i> : pAT1G13650::H2B-sGFP	This paper	CY117
<i>A. thaliana</i> : pAT4G34970 (ADF9)::H2B-sGFP	This paper	TQ769
<i>A. thaliana</i> : pAT4G11290::H2B-sGFP	This paper	TQ771
<i>A. thaliana</i> : pAT2G38300::H2B-sGFP	This paper	CY112
Oligonucleotides		
See Table S2	N/A	N/A

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Recombinant DNA		
pBSK At5g10400	This paper	TQ707
pBSK At5g59870	This paper	TQ709
pBSK At1g07790	This paper	TQ708
pBSK At2g28740	This paper	TQ736
pBSK At5g25090	This paper	CY030
pBSK At3g51740	This paper	CY031
pBSK At5g56120	This paper	TQ742
pBSK At1g63100	This paper	TQ744
pBSK At3g25710	This paper	CY024
pBSK At3g10080	This paper	TQ748
pBSK At3g28420	This paper	TQ750
pBSK At5g61660	This paper	TQ732
pBSK At5g46730	This paper	TQ733
pBSK At5g60200	This paper	CY025
pBSK At2g18380	This paper	TQ757
pBSK At4g39190	This paper	TQ758
pBSK At3g45610	This paper	TQ756
pBSK At3g20570	This paper	TQ764
pBSK At1g63310	This paper	TQ766
pBSK At5g24800	This paper	TQ739
pBSK At5g57785	This paper	TQ721
pBSK At1g28310	This paper	TQ716
pATML1::sGFP-N7	This paper	TQ639
pAT1G13650::H2B-sGFP	This paper	CY117
pAT2G38300::H2B-sGFP	This paper	CY112
pAT4G34970 (ADF9)::H2B-sGFP	This paper	TQ769
pAT4G11290::H2B-sGFP	This paper	TQ771
AA00::H2B-sGFP	This paper	TQ662
Software and algorithms		
R version 3.6.1	The R Foundation	RRID: SCR_001905 https://www.r-project.org/
GraphPad Prism 8	GraphPad Software	RRID: SCR_002798 http://www.graphpad.com/
Cell Ranger v3.0.1	10x Genomics	RRID:SCR_017344 https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/what-is-cell-ranger
Seurat v3.1.1	(Stuart et al., 2019)	RRID:SCR_016341 https://satijalab.org/seurat/get_started.html
harmony v1.0	(Korsunsky et al., 2019)	https://github.com/immunogenomics/harmony
DoubletFinder v2.0.2	(McGinnis et al., 2019)	RRID:SCR_018771 https://github.com/chris-mcginnis-ucsf/DoubletFinder
Python v3.7.3	Python Foundation	RRID:SCR_008394 http://www.python.org/
Scanpy v1.4.5.1	(Wolf et al., 2018)	RRID:SCR_018139 https://github.com/theislab/scanpy
Palantir v0.2.2	(Setty et al., 2019)	N/A

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
velocityo CLI v.0.17.17	(La Manno et al., 2018)	https://velocityo.org/velocityo.py/index.html
velocityo.R v0.6	(La Manno et al., 2018)	https://github.com/velocityo-team/velocityo.R
Monocle v2.12.0	(Qiu et al., 2017)	RRID:SCR_016339 http://cole-trapnell-lab.github.io/monocle-release/docs/
SCODE	(Matsumoto et al., 2017)	https://github.com/hmatsu1226/SCODE
Adobe Photoshop CC 2018	Adobe Acrobat	N/A
Adobe Illustrator CC 2018	Adobe Acrobat	N/A
Fastp	(Chen et al., 2018)	RRID: SCR_016962 https://github.com/OpenGene/fastp
FastQC v0.11.7	FastQC	RRID:SCR_014583 http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
MulitQC v1.6	(Ewels et al., 2016)	https://multiqc.info/
STAR v.2.5.1b	(Dobin et al., 2013)	RRID:SCR_015899 https://github.com/alexdobin/STAR
Samtools v1.9	(Li et al., 2009)	RRID: SCR_002105 http://www.htslib.org/
featureCounts v1.6.2	(Liao et al., 2014)	RRID: SCR_012919 http://bioinf.wehi.edu.au/featureCounts/
DESeq2 v1.26.0	(Love et al., 2014)	RRID: SCR_015687 https://bioconductor.org/packages/release/bioc/html/DESeq2.html
clusterProfiler v3.14.0	(Yu et al., 2012)	RRID: SCR_016884 https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html
complexHeatmap	(Gu et al., 2016)	RRID: SCR_017270 https://jokergoo.github.io/ComplexHeatmap-reference/book/
Integrative Genomics Viewer	(Robinson et al., 2011)	RRID: SCR_011793 http://software.broadinstitute.org/software/igv/
Other		
Eppendorf realplex2	Eppendorf	Cat#A248709R
Olympus FV3000	Olympus	N/A
Falcon Cell Strainers	Corning	Cat#352340
Lecia Sliding Microtome 1200S	Lecia	N/A
Chromium Single Cell Instrument	10x Genomics	N/A
Agilent 2100 Bioanalyzer	Agilent	N/A
hemocytometer	N/A	N/A
Olympus BX63	Olympus	N/A
Leica ASP200S	Leica	N/A

RESOURCE AVAILABILITY**Lead contact**

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Jia-Wei Wang (jwwang@sippe.ac.cn).

Materials availability

Plasmids and transgenic plant lines generated in this study will be made available on request to the lead contact. This study did not generate new unique reagents.

Data and code availability

The scRNA-seq data (BioProject PRJCA003094) were deposited in Beijing Institute of Genomics Data Center (<http://bigd.big.ac.cn>).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

The *Arabidopsis thaliana* plants were used as the experimental model in the study. The Columbia-0 (Col-0) ecotype was used as wild type.

METHOD DETAILS

Plant materials and growth conditions

The wild-type *A. thaliana* (Col-0 ecotype) were used for all the experiments. Plants were grown on soil at 21°C (day)/19°C (night) in long days (16 h light/8 h dark, with a light intensity of 80 mmol/m²/s using Philips TLD 36W/865 and 36W/830 bulbs).

Identification of mutant

The T-DNA mutants of *AT2G04850* (SALK_012348), *AT5G22580* (SALK_081434), *AT1G25230* (SALK_200911) and *sgr6* (SALK_034928) were ordered from Arabidopsis Biological Resource Center. The homozygous mutants were identified by PCR-based genotyping. The primers were listed in [Table S2](#).

Construct

To generate promoter H2B-sGFP reporters for *AT2G38300*, *AT4G34970* (*ADF9*), *AT4G11290* and *AT1G13650*, the upstream regulatory sequences of these genes was PCR-amplified with Phusion High-Fidelity DNA Polymerase (Thermo Fisher, Cat No./ID: 022021), and cloned into TQ662, in front of Histone 2B-sGFP (H2B-sGFP) coding sequence. Similar approach was used to construct the *ATML1* sGFP-N7 reporter. The primers for cloning were listed in [Table S2](#).

RNA *in situ* hybridization

The selected genes were cloned into pBSK by homologous recombination using ClonExpress II one Step Cloning kit (Vazyme Biotech, Cat No./ID: C112). The primers for amplification were listed in [Table S2](#). For probe synthesis, the linearized vectors were added as templates. *In vitro* transcription was performed with T3 or T7 RNA polymerase (Roche, Cat No./ID: 11031163001/10881767001) and DIG RNA labeling mix (Roche, Cat No./ID: 11277073910). Tissue embedding and RNA *in situ* hybridization were performed as described ([Wang et al., 2020](#); [Zhang et al., 2017](#)). Briefly, the shoot apices, leaves, and stems were fixed with formaldehyde and embedded using a modular automated tissue processor (Leica, ASP200S). Paraffin-embedded samples were sectioned (7-9 µm) with a sliding microtome (Lecia, Sliding Microtome 1200S). The slides were dewaxed, digested with Proteinase K (Roche, Cat No./ID: 03115828001), dehydrated with gradient ethanol and hybridized with probes. After washing, the slides were incubated with anti-digoxigenin-AP Fab fragments (Roche, Cat No./ID: 11093274910). The signals were detected with the NBT/BCIP stock solution (Roche, Cat No./ID: 11681451001).

Section, staining and microscope

The leaves, the hypocotyls, and the basal regions of stems (about 0.5 cm in length) were harvested and fixed with FAA (50% ethanol, 5% glacial acetic acid, 3.7% formaldehyde and 41.3% water). Tissues were dehydrated with a series of ethanol treatments (70%, 90%, 90%, 99.8% and 100%), cleared with xylene, and embedded with paraplast (Sigma-Aldrich, Cat No./ID: P3558). Stems were sectioned at 7 µm using a microtome (Leica, RM2265). Sections were stained with 0.1% Toluidine blue (0.1 M NaOAC, pH 4.5, Sigma-Aldrich, Cat No./ID: 241245) for 3 min, and washed at least three times with water. The waxed sections were then deparaffinized and cleared twice with xylene. The sections were sealed and examined with a microscope (Olympus BX63).

For the analyses of *pAT2G38300::H2B-sGFP*, *pADF9::H2B-sGFP*, *pAT4G11290::H2B-sGFP*, *pAT1G13650::H2B-sGFP*, and *pATML1::sGFP-N7* reporters, the root, leaf, and shoot apex samples were sectioned and stained by FM4-64 [N-(3-triethylammonium-propyl)-4-(p-diethylaminophenyl-hexatrienyl) pyridinium dibromide] (Biotium, Cat No./ID: 70021). The abaxial leaf surface was examined with the Olympus FV3000 confocal microscope. The excitation wavelength was 488 nm (GFP and FM4-64), and the emission wavelengths were 500 to 550 nm (GFP) or 610 to 700 nm (FM4-64).

For the Fluorol Yellow 088 (FY) staining, we used the protocol as described in ([Berhin et al., 2019](#)). Briefly, wild type seedlings were stained with Fluorol Yellow 088 solution (0.01% in methanol, Sigma-Aldrich, Cat No./ID: F5520). Ten roots each were studied in three independent experiments. The fluorescence was examined with the Olympus FV3000 confocal microscope with an excitation at 488 nm and detection at 490-555 nm.

Scanning electron microscope

The 7-day- and 14-day-old shoot apices used for protoplasting were harvested. The samples were immersed in formalin-acetic acid-alcohol (FAA) fixative, vacuum-treated, and fixed in FAA fixative at room temperature overnight. The samples were dehydrated with a series of ethanol treatments (50%, 70%, 80%, 90%, 95% and 100%), and dried in a critical point dryer (Leica, EM CPD300) for ~ 6 h. The samples were coated with gold power using the sputter coater, and observed by the SEM (JEOL, JSM-6360LV).

Sample collection and preparation of single-cell suspensions

According to 10x Genomics User Guide, there are strict requirements on protoplast purity and concentration. Based on the results of our pilot experiments, we found that ~500 vegetative shoot apices can yield enough high quality protoplasts. The single cells were prepared as previously described with modifications (Zhang et al., 2019). Briefly, we harvested the shoot apices from 7-day-old seedlings (shoot #1) and 14-day-old plants (shoot #2), and the 3rd leaf (leaf) from 18-day-old plants, respectively (Figures S1A–S1C). The tissues were separately digested in the RNase-free enzyme solution (1.5% cellulase R10, 1.5% macerozyme R10, 0.4 M mannitol, 0.1 M 4-morpholineethanesulfonic acid, 10 mM KCl, 10 mM CaCl₂ and 0.1% BSA) for 2 h at room temperature (Figure S1D). To increase enzymatic digestion efficiency and enhance protoplast release, we gently dissected shoot apices using syringe needle during protoplasting. 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 of each sample was higher than 85%. The protoplast suspension was counted by a hemocytometer. The concentration of protoplasts was finally adjusted to 1,500 ~ 2,000 cells/µL.

scRNA-seq library construction

scRNA-seq libraries were prepared as previously described (Zhang et al., 2019). Briefly, the shoot apices or leaf cell suspensions were loaded on a Chromium Single Cell Instrument (10x Genomics, Pleasanton, CA) to generate single cell GEMs. scRNA-seq libraries were generated with the Chromium Single Cell 3' Gel Bead and Library Kit v3 (P/N 1000075, 1000073, 10x Genomics). The detailed experimental procedure was performed according to user guide (Chromium Single Cell 3' Reagent Kits v3, CG000183 Rev A). The qualitative analyses of DNA libraries were performed with Agilent 2100 Bioanalyzer. The libraries were sequenced by Illumina sequencer NovaSeq (Genenergy Biotechnology Shanghai Co., Ltd) using two 150 bp paired-end kits. The raw scRNA-seq dataset was comprised of 28 bp Read1, 150 bp Read2 and 8 bp i7 index reads.

QUANTIFICATION AND STATISTICAL ANALYSIS

Pre-processing of raw scRNA-seq data

The raw scRNA-seq data were processed as described (Zhang et al., 2019). Briefly, the raw files were analyzed by Cell Ranger 3.0.1 (10x Genomics). The genome and GTF annotation files of *Arabidopsis* were downloaded from TAIR10 website (<https://www.arabidopsis.org/>). The "cellranger mkref" function with "--genome, --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 94% reads in all the samples were aligned to the TAIR10 reference genome by the aligner STAR (v.2.5.1b) (Dobin et al., 2013). The ratio of the number of fraction reads in cells to total number of reads for each sample was more than 86%. The detailed Cell Ranger reports were given in Table S1. The gene-cell matrices (named 'filtered_gene_bc_matrices' by 10x Genomics) were served as processed raw data for further analyses.

Doublet detection

Doublets in each scRNA-seq dataset were detected with DoubletFinder (v.2.0.2) (McGinnis et al., 2019). Three input parameters are required for doublets prediction: the number of expected real doublets (nExp), the number of artificial doublets (pN) and the neighborhood size (pK). For nExp, the standard Seurat processing pipeline was performed up 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 was a ratio to define the number of generated artificial doublets based on total cell numbers. pN was set to 0.25. An increase in pN value did not lead to a change in DoubletFinder results. To identify optimal pK value, the 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 the defined values of nExp, pN and pK as defined above. The proportion of artificial nearest neighbors (pANN) for every cell was computed. The doublet threshold of pANN was according to the number of expected doublets (nExp) to generate final doublet predictions. Cells which were flagged as singlets were kept for further downstream analysis.

Data integration, clustering, and annotation

Downstream analyses were mainly performed with Seurat (v.3.1.1) package with an approach adapted from our previous study (Zhang et al., 2019). Briefly, analytic procedures included quality control (calculate the percentage of mitochondrial and chloroplast genes, filtering out low-quality cells and genes), data normalization ("NormalizeData" function with LogNormalize method and scaling factor of 10,000), variable gene detection ("FindVariableGenes" function with vst method and 2000 features), scaling the data ("ScaleData"), PCA analysis ("RunPCA" function with variable genes and 100 principal components), constructing the SNN graph and clustering cells based on Louvain ("FindNeighbors" and "FindClusters") and visualization with non-linear dimensional reduction algorithms ("RunTSNE" and "RunUMAP") were successively performed.

For quality control, the low-quality cells and genes were filtered as follows: (1) only the cells in which numbers of expressed genes was 500 to 10,000 were considered; (2) the cells with unique molecular identifiers (UMIs) above 50,000 and below 500 were filtered

out; (3) the genes that were expressed in fewer than 3 cells were removed; (4) the cells with the percentage of mitochondrial genes more than 5% were excluded; (5) the cells with the percentage of chloroplast genes exceeding 10% were removed.

To mitigate the effects of cell cycle heterogeneity and protoplasting on cell clustering, we calculated cell cycle score and the proportion of protoplasting genes in each cell by the "CellCycleScoring" and "PercentageFeatureSet" function provided by Seurat. The cell-cycle and protoplasting genes are provided by two published papers (Gutierrez, 2009; Yadav et al., 2009). The variations due to cell cycle and protoplasting genes were then regressed out by the "ScaleData" function using "vars.to.regress". To assess the effectiveness of data scaling, we compared cell cycle phase before and after regression by PCA plot. After cell clustering, the contribution of protoplasting genes to each cell cluster was assessed by the "RidgePlot" and "VlnPlot" function. The proportion of protoplasting genes was below 3% in all the 23 cell clusters.

The "RunHarmony" function was used to correct batch effects between samples (Korsunsky et al., 2019). The resultant integrated data were clustered and visualized based on harmony dimensionality reductions. To construct a 3D UMAP embedding, three UMAP dimensions were retrieved by "RunUMAP" function with the parameter (n.components = 3). The 3D UMAP scatter graph was then plotted with three UMAP dimensions ('UMAP_1', 'UMAP_2', 'UMAP_3') using the plotly package (v.4.9.1).

The cell type of each cell cluster was manually defined by known marker genes which were listed in Table S2. The cluster-enriched genes were first computed using "FindAllMarkers" function in Seurat with 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 0.25. The well-characterized marker genes among the cluster-enriched genes were then used to annotate cell cluster. The clusters were assigned to a broad cell population with many marker genes which showed consistent expression pattern.

The differentially expressed genes (DEGs) between cluster 17 and cluster 19 were calculated using "FindMarkers" function. Genes with positive values of the log₂ fold change (log2FC) represented higher expression levels in cluster 17 than in cluster 19. Conversely, genes with negative values of log2FC were enriched in cluster 19.

Correlation analysis

The three published datasets (i.e. Arabidopsis cotyledon scRNA-seq, Arabidopsis stomatal lineage transcriptome and Arabidopsis shoot domain TRAP-seq) were download from NCBI (SRP247828, GSE58857 and SRP145572). For bulk RNA-seq datasets, sequencing reads were trimmed by fastp (v.0.20.0), and aligned to the TAIR10 reference genome by STAR (v.2.5.1b) (Dobin et al., 2013). For scRNA-seq datasets, we used Cell Ranger and Seurat workflows. Each single-cell count was pooled by whole cells, cell-type cells and gene-specific cells to generate the pseudo-bulk counts. The bulk RNA-seq counts and pseudo-bulk scRNA-seq counts were transformed into log₂ scale to minimize and normalize the differences in library size. Finally, the Spearman's rank correlation between bulk RNA-seq dataset and scRNA-seq dataset was calculated by the "cor.test" function in R.

Subclustering

To explore cellular heterogeneity, cell cycle continuum and differential trajectory for each SAM cell layer (L1, L2 and L3), the cells belonging to EC, MC and VC populations, and the proliferation cells corresponding to mesophyll or vascular tissue cells (cluster 17, 5a and 9a for mesophyll; cluster 19, 5b and 9b for vascular tissue cells) were extracted. The raw counts were read into Seurat to create a new Seurat object. The analytic procedures were performed as described above. Specifically, the parameters of total number of PCs (npcs), dimensions (dims), the number of neighboring points (n.neighbors), resolution and minimum distance (min.dist) were tuned to represent the best approximation of underlying topology for each subclustering (EC: npcs = 20, dims = 15, n.neighbors = 40, resolution = 0.5, min.dist = 0.3; MC: npcs = 50, dims = 30, n.neighbors = 30, resolution = 0.6, min.dist = 0.25; VC: npcs = 50, dims = 15, n.neighbors = 45, resolution = 0.8, min.dist = 0.25). After subclustering, the cluster-enriched genes were computed using "FindAllMarkers" function with following parameters: a Wilcoxon Rank Sum test; above 0.58-fold difference (logfc.threshold = 0.58); at least 0.1-fold minimum fraction (min.pct = 0.1).

Differential trajectory and lineage analysis

For Palantir pseudotime analysis (Figures 4G–4J, 5C, S6A, and S6B), the raw data slot of the Seurat object (the early stomata sub-cell clusters: E3, E7, E8 and E9; the late stomata sub-cell clusters: E1, E8 and E11) was read into Palantir (v.0.2.2) (Setty et al., 2019). After data pre-processing included principal component analysis, diffusion maps analysis and MAGIC imputation, a start cell was specified before running Palantir. The start cell was defined based on prior and proven information and an appropriate coordinate position in t-SNE map. For instance, the G0/G1 phase is the entry of cell cycle; the stem cell or meristemoid is in the pluripotent or unipotent cell state; *MCM7* is expressed in G0/G1; *BASL* and *POLAR* are expressed in MMC. The terminal states were automatically determined. Each cell along pseudotime and branch probabilities to terminal states were assigned by Palantir. The gene expression trends along Palantir pseudotime or different lineages were modelled based on generalized additive models (GAMs) provided by Palantir.

For visualization of t-SNE, ForceAtlas2 and Diffusion maps in SCANPY (Figure S4), the subset of the Seurat object (the early stomata sub-cell clusters: E3, E7, E8 and E9; the late stomata sub-cell clusters: E1, E8 and E11) was loaded into SCANPY (v.1.4.5.1). Downstream workflow included data normalization (scipy.pp.normalize_per_cell method, scaling factor 10000), log-transformation (scipy.pp.log1p), batch correction (scipy.pp.combat), variable gene selection (scipy.pp.highly_variable_genes), data scaling (scipy.pp.regress_out), principal component analysis (scipy.pp.pca), computing the neighborhood graph (sc.pp.neighbors)

and visualization with *t*-SNE, ForceAtlas2 and Diffusion maps (sc.tl.diffmap, sc.pl.diffmap, sc.tl.tsne, sc.pl.tsne, sc.tl.draw_graph and sc.pl.draw_graph) was performed. The cluster label of each cell was assigned with its subclustering cluster number. For DPT pseudotime analysis, the same start cell in Palantir was used.

For Monocle 2 pseudotime analysis ([Figures 5D–5F, 5H, 6D–6G, and S7B](#)), the subset of the Seurat object (the late stomata sub-cell clusters: E1, E8 and E11; the VC sub-cell clusters: V3, V7, V9 and V12) was converted to a CellDataSet object using "as.CellDataSet". Downstream analysis was performed as previously described ([Zhang et al., 2019](#)). The top 1,000 significant genes were chosen as ordering genes. The gene expression trends were plotted by "plot_genes_in_pseudotime" function. The hierarchical tree of cell lineage was constructed by "plot_complex_cell_trajectory" function. The Monocle 2 pseudotime in each cell was extracted for the downstream transcription factor network analysis.

RNA velocity analysis

To establish the cell continuum of cell cycle, we estimated RNA velocity by velocyto package ([La Manno et al., 2018](#)). Expression matrix of unspliced and spliced mRNA in each sample were generated with velocyto CLI (v.0.17.17) according to the CLI usage guide (velocyto run10x). The output loom files of shoot #1, shoot #2 and leaf were combined using "loompy". The merged loom file was read into velocyto.R (v.0.6). The unspliced and spliced counts of the cells from cluster E3, E7, E8 and E9 were extracted. It should be noted that the reads that mapped confidently to intronic regions were less than 1% in Arabidopsis, which was far less than 35% in human. To filter out lowly variable genes, data normalization, and variable gene detection were performed. The resultant 5,000 variable genes were further filtered using "filter.genes.by.cluster.expression" function (emat: min.max.cluster.average = 0.1; nmat: min.max.cluster.average = 0.01). The final 2,059 genes was used to estimate RNA velocity with "gene.relative.velocity.estimates" function (kCells = 20). To plot individual cell velocities, the *t*-SNE embeddings in Palantir were exported.

Gene-ontology enrichment analysis

All differentially expressed genes including cluster-enriched genes and branch-dependent genes with statistical significance were submitted to AgriGO v2.0 for GO enrichment analysis ([Tian et al., 2017](#)). The top 20 or 30 GO terms with $-\log_{10}\text{FDR}$ values were represented ([Figures 8G, S6E, S7C, and S8C](#)). All GO terms were listed in [Tables S1](#) and [S3](#).

Visualization of the transcription factor network

The transcription factors among the top 2,000 genes whose expression was changed along Monocle 2 pseudotime were selected. The expression levels of 102 transcription factors were fitted into smooth spline curves to depict gene expression trends along pseudotime, based on vector generalized linear and additive models (VGAM). The pseudotime of each cell was normalized from 0 to 1. Gene Regulatory Network (GRN) was inferred by SCODE ([Matsumoto et al., 2017](#)). To obtain reliable relationships, we ran SCODE for 50 times and averaged results. The regulatory relationships between genes were exported to a file named 'meanA.txt'. The positive and negative values represented up-regulation and down-regulation, which inferred the relationship of activation and repression. The diagonal elements represented self-regulation, which meant positive or negative feedback. The arranged files with columns of source, target, SCODE values and regulation were read into Cytoscape for visualization ([Shannon et al., 2003](#)).

Arabidopsis cell landscape construction

To construct the framework of Arabidopsis cell atlas which includes the cells from roots, shoots, and leaves, two published root scRNA-seq datasets from the Schiebelbein lab and the Timmermans lab were downloaded from NCBI (GSE123013 and GSE123818) ([Denyer et al., 2019](#); [Ryu et al., 2019](#); [Zhang et al., 2019](#)). The fastq files were processed by Cell Ranger 3.0.1. After quality control, a total of 60,099 cells from six scRNA-seq datasets (4 datasets from our group) were selected and 2,936 highly variable genes were used for downstream analyses. The Seurat workflow with similar parameters was performed as mentioned above.