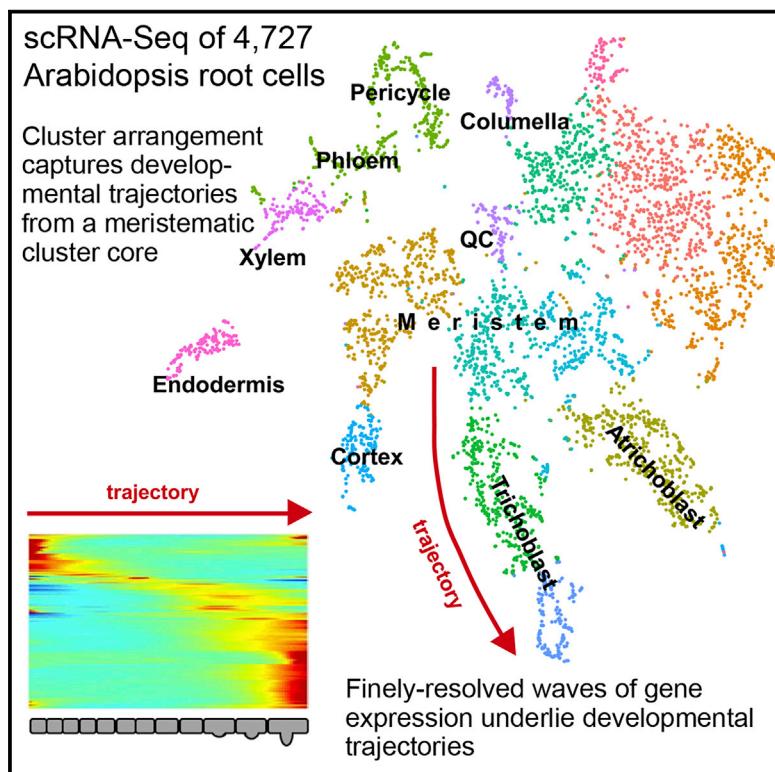


Developmental Cell

Spatiotemporal Developmental Trajectories in the *Arabidopsis* Root Revealed Using High-Throughput Single-Cell RNA Sequencing

Graphical Abstract



Authors

Tom Denyer, Xiaoli Ma, Simon Klesen, Emanuele Scacchi, Kay Nieselt, Marja C.P. Timmermans

Correspondence

marja.timmermans@zmbp.uni-tuebingen.de

In Brief

Denyer and Ma et al. generate a single-cell RNA expression atlas of the *Arabidopsis* root that captures spatiotemporal information for all major cell types and uncovers new regulators. Pseudotime-analysis-derived developmental trajectories depict a cascade of developmental progressions between stem cell and final differentiation mirrored by waves of transcription factor expression.

Highlights

- scRNA-seq of *Arabidopsis* root cells captures precise spatiotemporal information
- Defining expression features for cell types identify new developmental regulators
- Cluster arrangement reflects developmental time with a centrally localized niche
- Intricate waves of gene expression finely resolve developmental trajectories



Spatiotemporal Developmental Trajectories in the *Arabidopsis* Root Revealed Using High-Throughput Single-Cell RNA Sequencing

Tom Denyer,^{1,3} Xiaoli Ma,^{1,3} Simon Klesen,¹ Emanuele Scacchi,¹ Kay Nieselt,² and Marja C.P. Timmermans^{1,4,*}

¹Center for Plant Molecular Biology, University of Tübingen, Auf der Morgenstelle 32, Tübingen 72076, Germany

²Center for Bioinformatics, University of Tübingen, Sand 14, Tübingen 72076, Germany

³These authors contributed equally

⁴Lead Contact

*Correspondence: marja.timmermans@zmbp.uni-tuebingen.de

<https://doi.org/10.1016/j.devcel.2019.02.022>

SUMMARY

High-throughput single-cell RNA sequencing (scRNA-seq) is becoming a cornerstone of developmental research, providing unprecedented power in understanding dynamic processes. Here, we present a high-resolution scRNA-seq expression atlas of the *Arabidopsis* root composed of thousands of independently profiled cells. This atlas provides detailed spatiotemporal information, identifying defining expression features for all major cell types, including the scarce cells of the quiescent center. These reveal key developmental regulators and downstream genes that translate cell fate into distinctive cell shapes and functions. Developmental trajectories derived from pseudotime analysis depict a finely resolved cascade of cell progressions from the niche through differentiation that are supported by mirroring expression waves of highly interconnected transcription factors. This study demonstrates the power of applying scRNA-seq to plants and provides an unparalleled spatiotemporal perspective of root cell differentiation.

INTRODUCTION

In recent years, high-throughput single-cell transcriptomics has developed to a point of becoming a fundamental, widely used method in mammalian research (Potter, 2018). Thousands of cells can be profiled simultaneously and analyzed accurately, revealing unique insights into developmental progressions, transcriptional pathways, and the molecular heterogeneity of tissues. The increasingly high-throughput nature of single-cell RNA sequencing (scRNA-seq) has been facilitated by the development of droplet technology (Macosko et al., 2015; Klein et al., 2015) and increased automation (Zheng et al., 2017). In brief, a cell is encapsulated within an oil droplet and lysed, and its transcripts reverse transcribed on barcoded beads. Following library production and sequencing, transcripts from individual cells can be identified from the bead-derived barcode and individual transcripts accounted for using unique molecular identifiers (UMIs)

(Prakadan et al., 2017). However, while commonly used in animal systems, additional technical demands such as the necessity to break down cell walls (with subsequent transcriptional effects), high osmotic pressure sensitivities, and high cell size variability present potential challenges when applying this technology to plants.

The *Arabidopsis* root provides an ideal tissue for analyzing the promise of scRNA-seq. The transcriptomes of key cell types have been well profiled, and the root shows a strict spatiotemporal organization. Radially, the root is organized in concentric rings of endodermis, cortex, and epidermis that surround a central stele, comprising the pericycle, phloem, and xylem (Figure S1A). These cell types originate from a specialized stem cell niche in which initials, surrounding the quiescent center (QC), divide in a predictable manner, giving rise to long cell files that capture their developmental trajectory along the length of the root (Figure S1B). Several gene expression atlases of the *Arabidopsis* root have been produced (Birnbaum et al., 2003; Brady et al., 2007a; Li et al., 2016). These, however, have focused primarily on describing either radial or temporal expression profiles and typically relied on reporter lines to assess select cell types. scRNA-seq, on the other hand, allows the simultaneous, unbiased sampling of every type of cell at every developmental stage in one experiment.

Here, we present a high-resolution scRNA-seq expression atlas of the *Arabidopsis* root that captures its precise spatiotemporal information, revealing key regulators and defining features for all major cell types. We show how QC cells and meristematic cells are distinguished and resolve intricate developmental trajectories that cells undergo during their transition from stem cell through differentiation. The precise waves of gene expression characterizing this process are mirrored by similar expression changes of highly interconnected transcription factors (TFs). Our atlas offers an unparalleled spatiotemporal perspective of root cell-type differentiation at a resolution not previously achievable.

RESULTS

Single-Cell RNA Sequencing Is Highly Sensitive and Highly Reproducible

4,727 *Arabidopsis* root cells from two biological replicates were isolated and profiled using droplet-based scRNA-seq. At



~87,000 reads per cell, the median number of genes and transcripts detected per cell was 4,276 and 14,758, respectively (Figure S1C; Table S1). In total, transcripts for 16,975 genes were detected ($\text{RPM} \geq 1$), which, after correction for read depth, represents ~90% of genes detected by bulk RNA sequencing (RNA-seq) of protoplasted root tissue. Further, the global gene expression profiles of pooled scRNA-seq and bulk RNA-seq are highly correlated ($r = 0.9$; Figure S1D), indicating that plant scRNA-seq is highly sensitive. This methodology is also highly reproducible, as demonstrated by the facts that ~96% of genes expressed ($\text{RPM} \geq 1$) in one scRNA-seq replicate are detectable in the second and that expression across the two replicates is highly correlated ($r = 0.99$; Figure S1E).

Clusters Comprise the Major Cell Types in the Root

To identify distinct cell populations based on gene expression profiles, an unbiased, graph-based clustering was performed on the 4,727 single-cell transcriptomes using the Seurat software package (Satija et al., 2015; Butler et al., 2018) (Figure 1A). Genes induced by protoplasting (≥ 2 -fold; $q < 0.05$) were identified by standard RNA-seq and dismissed prior to analysis (Figure S1F; Table S1). 15 distinct clusters were identified, each containing between 81 and 596 cells. These clusters harbored similar numbers of cells from each replicate, and their gene expression profiles were highly correlated across the replicates (r between 0.95 and 1; Table S2), highlighting again the impressive reproducibility of this technique.

In order to attribute cell identities to these clusters, expression of cell-type-specific marker genes, either well established or identified from a curated collection of root transcriptomic datasets (Table S2; Efroni et al., 2015), was compared across clusters. This allowed cell identities to be confidently assigned to 8 of the 15 clusters in the cluster cloud (Figure 1B; Table S2). Expression of key root development genes among these markers, such as *PLT1*, *SCR*, *SHR*, *APL*, *COBL9*, and *GL2*, shows high specificity to particular clusters (Figure S2). Cluster identities were confirmed with a complementary approach, whereby transcription profiles of differentially expressed (DE) genes governing the clusters were harvested from microarray datasets (Brady et al., 2007a) and analyzed for tissue specificity (Figure S1G). Together, these approaches revealed that, with the exception of lateral root cap cells (for which limited marker data are available), all known major tissue types in the root were captured and are represented by identifiable clusters.

Clusters 9 and 13 comprise cortex and endodermal cells, respectively (Figures 1B and S1G; Table S2). The identity of the endodermal cluster was further validated by the localized accumulation of GFP transcripts in one of the replicates generated from the *pMIR166A:erGFP* reporter line (see STAR Methods; Figures 2A and S3A). In addition, when cells were re-clustered incorporating scRNA-seq data of *shortroot* mutants (*shr-3*), which lack a defined endodermis (Helariutta et al., 2000; Figure S3A), otherwise well-dispersed *shr-3* cells were absent from a cluster comprising endodermal cells of both wild-type replicates (Figures 2B and S3). This cluster analysis also shows that *shr* cells, while present in all other clusters, localize on the outskirts of some (Figure 2B). This points to subtler effects of *SHR* on cell types other than the endodermis;

although some of this phenomenon may also be attributable to the fact that *shr-3* is in the Ler background. Irrespective, this observation nicely highlights the potential of applying scRNA-seq to identify hidden phenotypic changes, whether stemming from natural variation or mutations.

Clusters 10 and 3 comprise trichoblast and atrichoblast cells, respectively (Figures 1B and S1G; Table S2). Cluster 5 also contains trichoblast cells (Figure S1G). Although cells in this cluster show low expression of a number of atrichoblast marker genes, crucially, the trichoblast marker *COBL9* is expressed in this cluster, whereas the atrichoblast marker, *GL2*, is not (Figures 1B and S2). The co-expression of atrichoblast marker genes hints at a degree of commonality between this subset of trichoblasts and its epidermal counterparts, perhaps reflecting a distinction in developmental stage to the trichoblast cells contained in cluster 10.

Cluster 4 comprises stele cells while a neighboring cluster (12) comprises maturing xylem cells (Figures 1B and S1G; Table S2). Consistent with the tissue complexity of the stele, subclustering reveals cell heterogeneity within cluster 4. Particularly, phloem and pericycle cells are separated into two discrete subclusters (Figure S4), as highlighted by the highly subcluster-specific expression of genes such as *APL* (4.2), *LBD29*, and *TIP2-3* (4.1) (Figure 2D; Bonke et al., 2003; Porco et al., 2016; Gattolin et al., 2009).

Finally, cluster 11 comprises both columella and QC cells (Figures 1B and S1G; Table S2), which can be separated into two subclusters. Subcluster 11.2 contains columella cells that express marker genes such as *COBL2*, *NCED2*, and *ATL63* (Figure 2E; Brady et al., 2007b; Efroni et al., 2015). In contrast, transcripts for the QC-expressed genes *AGL42*, *BBM*, and *TEL1* are largely limited to cells in subcluster 11.1 (Figure 2E; Navy et al., 2005; Efroni et al., 2015). Given the small number of QC cells per root, this cluster may well contain other transcriptionally similar cells, perhaps the adjacent initials in the niche. However, importantly, the fact that QC cells are captured illustrates well the possibilities of this methodology for studying rare cell types or elucidating transcriptional subtleties affecting small numbers of cells within a tissue.

Meristematic Cells Cluster Independently of Tissue Identity

The identity of cells in the remaining clusters is less obvious. Overall gene expression in cells within clusters 0, 1, and 14 is comparatively low (Figure S5A), likely masking their identity at this level of sequencing resolution. However, expression values extracted from a longitudinal microarray dataset (Brady et al., 2007a) for the top DE genes defining these clusters suggest that they comprise mature cells of mixed identity (Figure S5B). In contrast, cells in the final four clusters (2, 6, 7, and 8) show markedly meristematic-based expression profiles (Figure S5B). Notable histone and cytokinesis-linked genes, such as *KNOLLE*, *ENODL14*, and *ENODL15*, are among the most prominently DE genes for these clusters (Figure S2; Table S2; Lauber et al., 1997; Adrian et al., 2015). Subclustering revealed some cell-type identities, albeit that they are generally less distinct than those of the clusters described above. For example, subcluster 2.4 shows a distinct cortex identity (Figure S4). Curiously, this subcluster is positioned adjacent to the main cortex cell cluster.

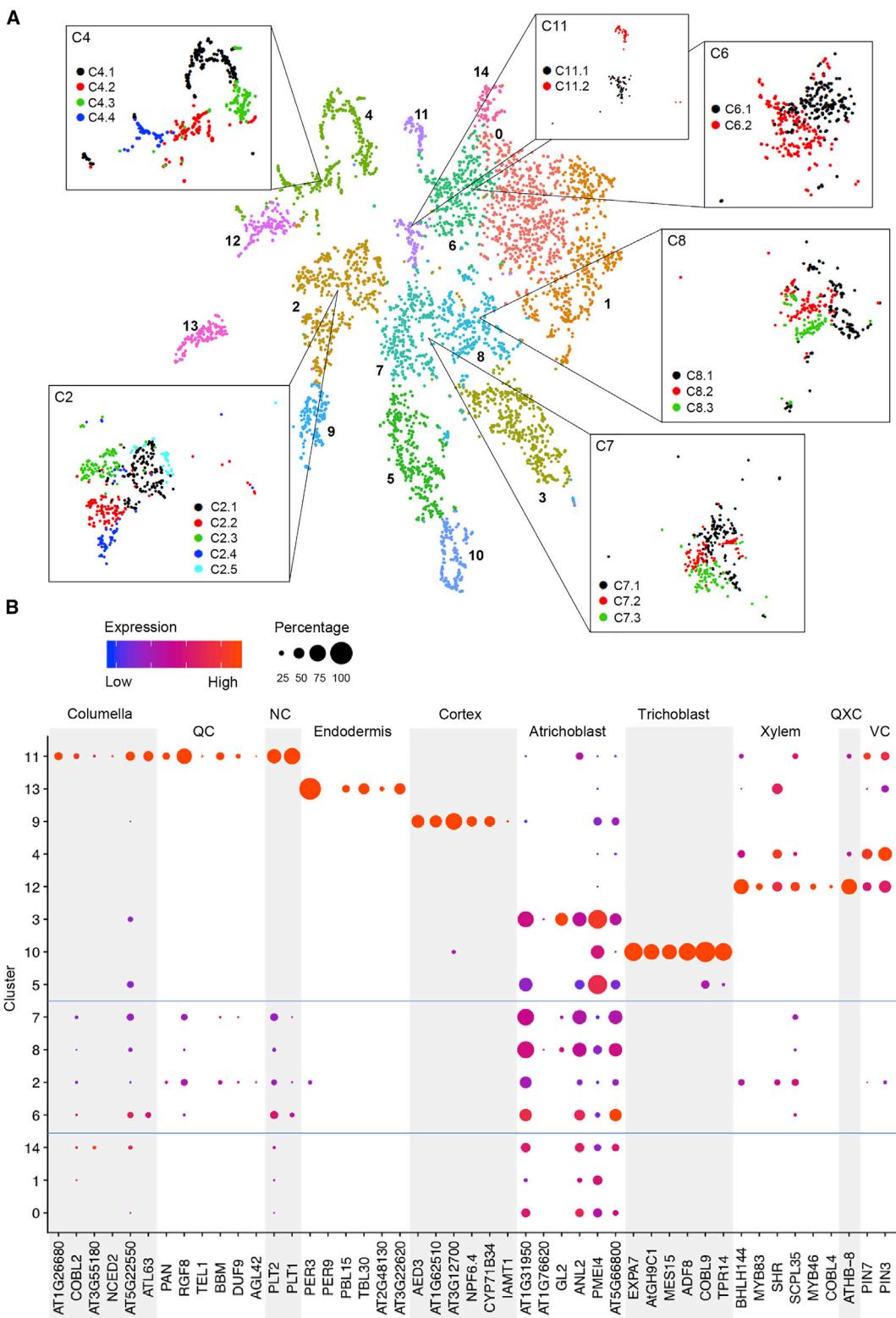
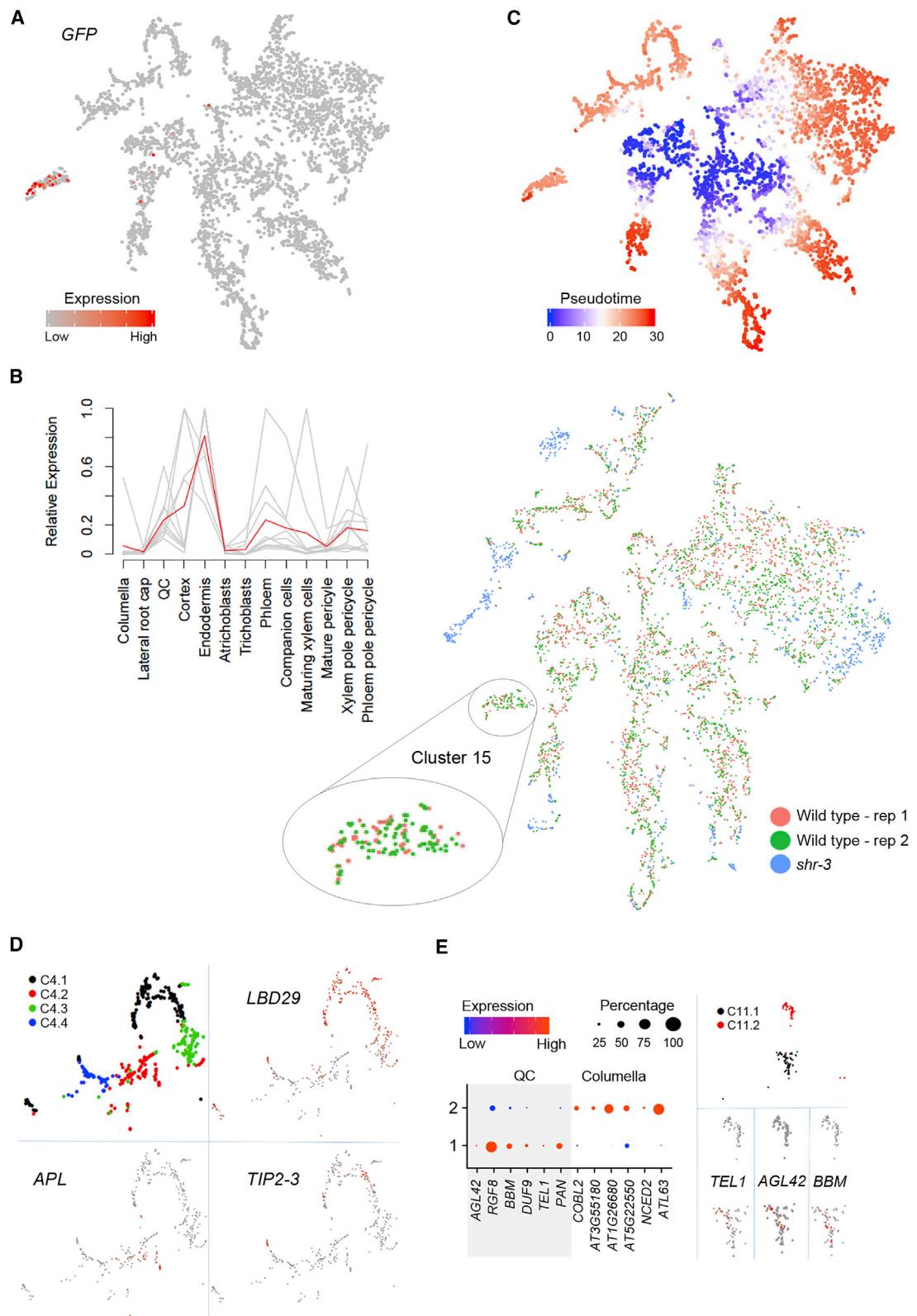


Figure 1. Sequenced Single Cells Cluster by Identity

(A) t-SNE (t-distributed stochastic neighbour embedding) plot of 4,727 *Arabidopsis* root cells shows these group into 15 clusters with additional subclusters. (B) Expression of known cell-type marker genes across cells reveals the identity of clusters. Dot diameter, proportion of cluster cells expressing a given gene; color, mean expression across cells in that cluster. NC, niche and columella; QXC, QC, xylem, and columella; VC, vasculature and columella. See Table S2 for details of all marker genes assessed.



(legend on next page)

A comparable pattern is seen in clusters 7 and 8 with trichoblast cell identity apparent in those subclusters (7.3 and 8.3) closest to the adjacent defined trichoblast clusters (Figure S4).

It is interesting to note that when comprehending all the clusters together, the meristematic clusters are closely localized in the center of the cluster cloud with the subcluster containing QC cells (11.1) at the heart of this. Meanwhile, those clusters with distinct, mature cell identities span out from the meristematic clusters (Figure 1A; Video S1), suggesting an overall cluster arrangement that reflects developmental time. Subclustering of the meristematic clusters refines this idea, showing a degree of closeness of mature and developing cells of the same eventual fate. This notion is further supported by pseudotime analysis across all cells, which reveals that genes DE in cells of the central clusters describe the beginning of cell fate progressions (Figure 2C).

Likewise, the cluster cloud reveals an organization that captures the “lineage” relationships between cell and tissue types. For instance, the trichoblast and atrichoblast clusters, as well as the xylem, vasculature, and the cortex and endodermis clusters, are positioned next to each other within the cluster cloud. The position of the columella in a cluster with QC cells indicates a higher degree of transcriptional accord between these cell types than between these cell types individually and others. This is reflected in the fact that key developmental regulators, such as *PLT2*, *PLT3*, and *PIN4* are co-expressed in the columella and QC (Galinha et al., 2007; Feraru and Friml, 2008). This way of contemplating clusters, along with pseudotime visualization, thus offers a valuable director for early comprehension of developmental trajectories, particularly in the absence of *a priori* knowledge, such as a reference atlas.

Unique Marker Genes Define Cluster Identity in an Unbiased Manner

Given that detailed reference datasets are available only for select tissue and organ types in very few plant species, we developed an unbiased approach to assign cell type identities to scRNA-seq-generated cell clusters. Genes DE in a given cluster compared to all other clusters ($q < 0.01$; average log fold change [FC] ≥ 0.25) were identified using “biomod” on Seurat (McDavid et al., 2013). DE genes were further narrowed down by applying the criteria that cluster-specific marker genes must be expressed in $\geq 10\%$ of cells within the cluster (PCT1), and $\leq 10\%$ of cells across all other clusters (PCT2). Applying these criteria, we uncovered expected marker genes alongside hundreds of additional genes diagnostic for a given developmental stage or cell type that encompass every cluster (Table S3).

The top two cluster-specific genes (based on average log FC) for each cluster are expressed across a substantial proportion of cells specifically within one cluster, with the exception of genes

for clusters 0 and 1, which show substantial co-expression in cluster 14 (Figure 3A). In addition, Gene Ontology (GO) overrepresentation analysis on cluster-specific gene sets reveal GO terms appropriate to their biology (Table S3). For example, the meristematic clusters 2 and 8 show an abundance of marker genes implicated in processes related to cell proliferation and DNA replication, respectively. Further, markers for the root-hair-cell cluster 10 are enriched in trichoblast differentiation and maturation terms; for the QC- and columella-containing cluster 11, in root development and starch biosynthesis; and for cluster 12, in xylem development and secondary cell wall biogenesis. Finally, genes required for the formation and suberization of the Caspary strip are among the markers for cluster 13, which comprises endodermal cells. However, a notable outcome of this analysis is the number of marker genes for which a root function has yet to be assigned. This illustrates the potential of scRNA-seq for identifying new developmental regulators.

To further validate this strategy for marker gene calling and for assigning cell identity to clusters without other references, we assessed the spatiotemporal patterns of expression for select genes using transcriptional *promoter:3xVenus-NLS* reporter lines. Prioritizing by a balance of high log-fold change, high PCT1, low PCT2, and a lack of prior biological information relating to cell-type specificity and root development, we selected ten genes from across clusters. Expression for eight of the ten genes tested localized to specific cell types and/or root zones in line with predictions. Specific expression in the cortex (*AT1G62510*) and maturing trichoblasts (*MES15*) was observed for marker genes for clusters 9 and 10, respectively (Figures 3B and 3C), while genes selected from cluster 4 revealed highly specific phloem (*PME32*) and pericycle (*ATL75*) expression (Figures 3D and 3E). *MLP34* is expressed in the atrichoblasts, as expected for a marker for cluster 3 (Figure 3F). However, expression is also seen in cells of the lateral root cap (Figure 3F), a cell type to which a cluster could not be assigned. *MLP34* shows expression in some cells in cluster 1 (Figure S2), indicating that this cluster may in fact contain cells of the lateral root cap, although further analysis is needed to confirm this. Finally, expression of genes selected from the meristematic cluster 2 was found to localize to the meristematic cortex and endodermis (*AT3G22120*), the meristematic cortex (*AT1G62500*), or the meristematic vasculature (*PIP2-8*) (Figures 3G–3I).

Given the common occurrence of *cis*-regulatory motifs in the introns of genes, the fact that promoter fusions for eight out of the ten marker genes tested confirm predictions is notable. This unbiased approach for assigning identities to cell clusters could prove invaluable when no reference data are available. Moreover, our results reveal a level of sensitivity beyond that of assigning whole cluster identity. This is typified by *PME32* and

Figure 2. Cell Identity and Developmental Stage Are Reflected in the (Sub)clustering

- (A) t-SNE visualization of the cluster cloud (both replicates) shows GFP transcripts localize specifically to *pMIR166:erGFP* cells in the endodermal cluster.
- (B) Wild-type and *shr-3* cells were combined and clustered. (Left) The expression profiles for the top 10 DE genes from cluster 15, taken from a microarray root atlas (Brady et al., 2007a), reveal an endodermal identity. Red line, mean expression profile; gray lines, individual expression profiles. (right) t-SNE visualization of the wild-type and *shr-3* cell cluster cloud shows endodermal cluster 15 lacks *shr-3* cells.
- (C) Pseudotime analysis of all wild-type cells reveals cells in the central clusters are earlier in the pseudotime trajectory, consistent with their meristematic identity.
- (D) t-SNE visualization of cluster 4 subclusters. Expression profiles of selected genes reveal pericycle and phloem identities for subclusters 4.1 and 4.2, respectively.
- (E) Expression of known cell-type marker genes across cells of cluster 11 reveals the identity of subclusters. Dot diameter, proportion of cluster cells expressing a given gene; color, mean expression across cells in that cluster. t-SNE visualization of expression profiles of selected QC marker genes.

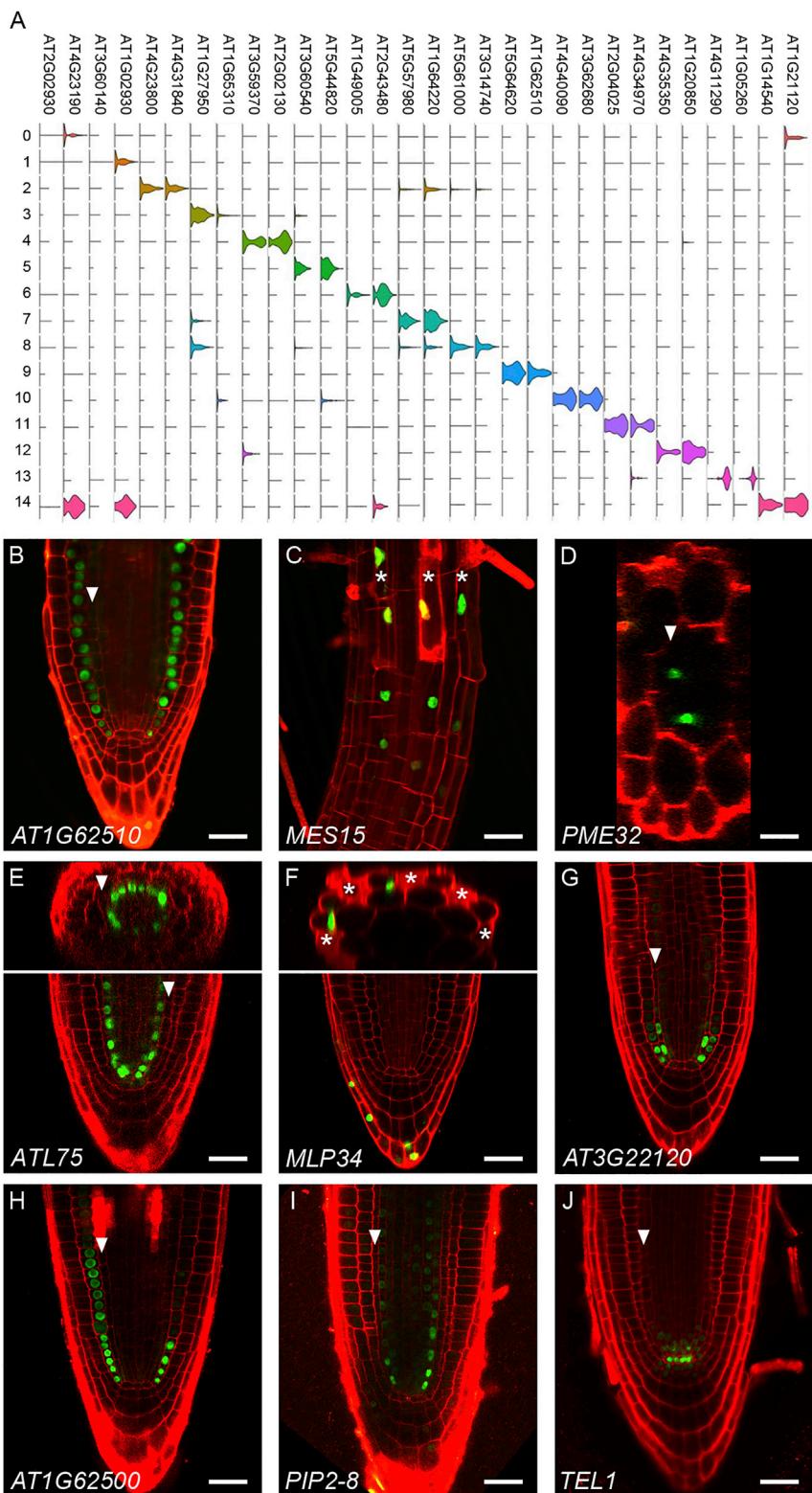


Figure 3. An Unbiased, Marker-Gene-Selection Method Is Validated with Transcriptional Reporter Fusions

(A) Cluster specificity revealed by violin plots depicting expression level (length) and proportion of cells expressing (width) for the top two marker genes by differential expression (x axis) across each cluster (y axis).

(B–J) Spatiotemporal expression patterns for promoter fusions for the following genes reveal the predicted cell-type specificities: (B) *AT1G26510*, cortex; (C) *MES15*, differentiating trichoblasts; (D) *PME32*, phloem; (E) *ATL75*, pericycle; (F) *MLP34*, atrichoblasts and lateral root cap, (G) *AT3G22120*, meristematic cortex and endodermis; (H) *AT1G62500*, meristematic cortex; (I) *PIP2-8*, meristematic vasculature; (J) *TEL1*, QC. White arrowheads indicate endodermal cells; asterisks indicate trichoblasts. Scale bars, 20 μ m.

tively (Figure S2). Furthermore, the meristematic vasculature marker, *PIP2-8*, is primarily expressed in cells of subcluster 2.3, positioned adjacent to the vasculature cluster, while the meristematic cortex and endodermis marker, *AT3G22120*, is mainly expressed in cells of subcluster 2.2 positioned near these mature cell types (Figure S2). The expression patterns observed for the latter reporters thus further reinforce the hypothesis that meristematic subclusters share expression features with the mature cell types they are closest to. Moreover, a gene's expression profile across the cluster cloud is a confident predictor of its localized expression *in planta*.

QC Marker Genes Identified from scRNA-Seq Data

The fact that QC cells are captured offers a unique opportunity to study this rare cell type, a point quite pertinent given that RNA-seq analysis shows the established marker *WOX5* to be induced upon protoplasting (Table S1). Within subcluster 11.1, 36 cells express at least half of 15 proposed QC genes (Figure 4A; Table S2; Efroni et al., 2015; Navy et al., 2005), which is in line with the sampling depth of scRNA-seq and the relatively low expression of most QC genes. The high number of QC cells captured likely reflects a bias in the methodology toward capturing small cells (see STAR Methods), which may also account for an overrepresentation of meristematic cells. Reinforcing our

ATL75 whose promoter fusions show expression in the phloem and pericycle, respectively, in accordance with their expression being predominant in cells of subclusters 4.2 and 4.1, respec-

QC cell calling, it is of note that genes marking initial cells directly neighboring the QC, such as *AT3G22120*, *AT1G62500*, and *PIP2-8* (Figures 3G–3I), show no (33 cells), or negligible (3 cells)

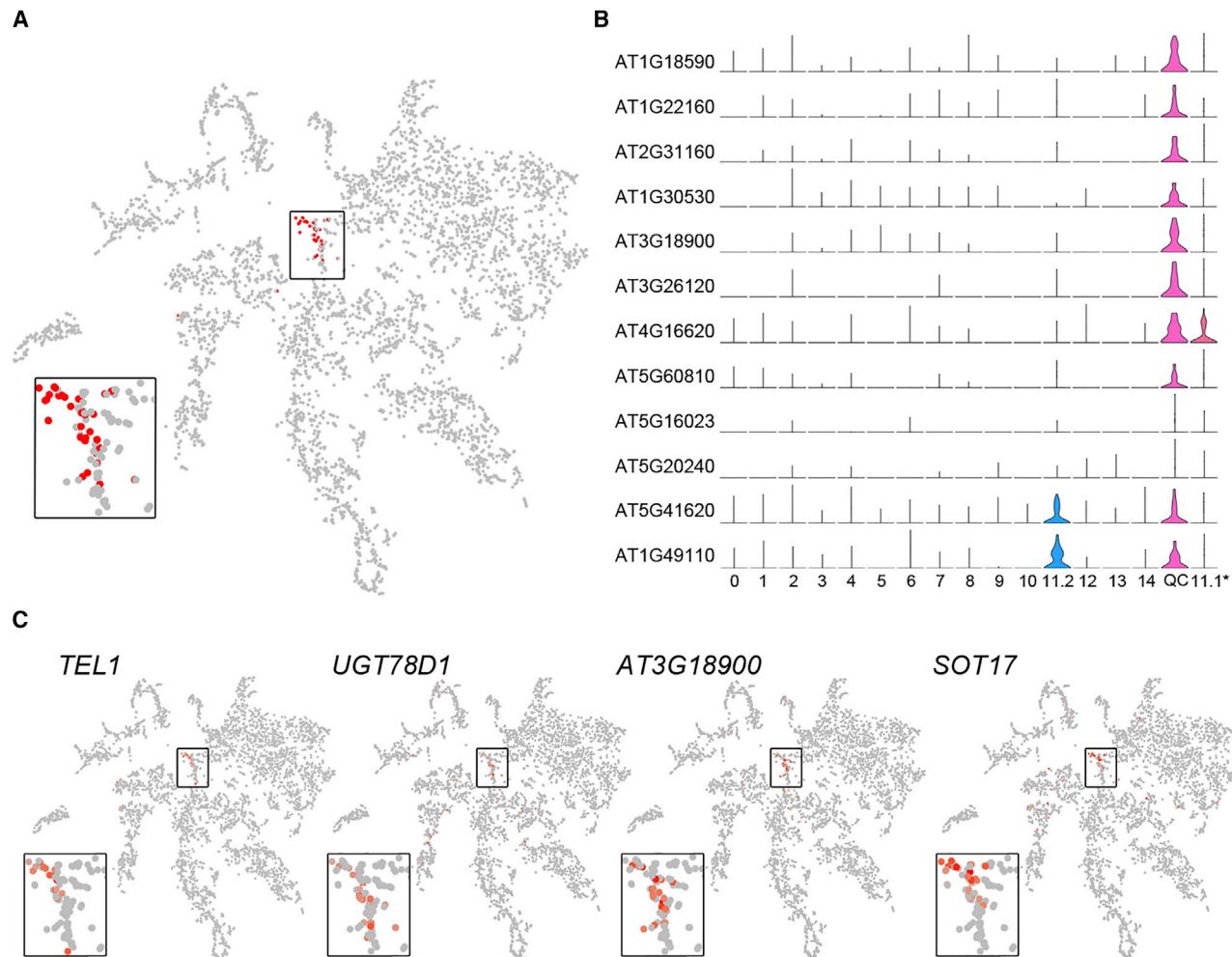


Figure 4. scRNA-Seq Data of QC Cells Identify Unique QC Marker Genes

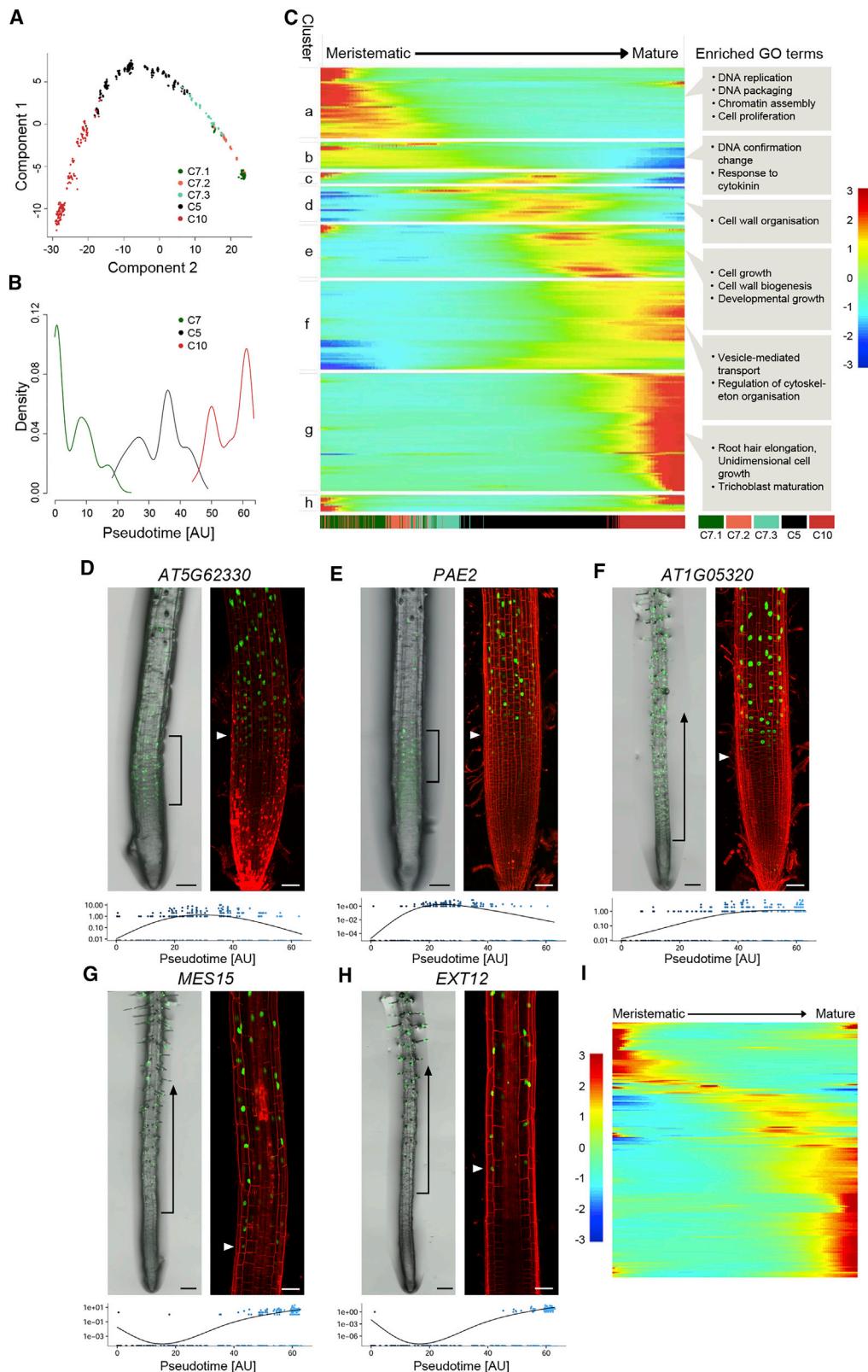
(A) t-SNE visualization indicating the position of 36 QC cells at the center of the cluster cloud.
(B) QC cell specificity revealed by violin plots depicting expression level (length) and proportion of cells expressing (width) for 10 QC cell marker genes (y axis) across each cluster (x axis). Violin plots for two genes showing expression in the QC cells and the columella cluster (11.2) are also shown. *11.1, subcluster 11.1 with QC cells removed.
(C) t-SNE visualization of expression profiles of selected QC marker genes. Subcluster 11.1 is magnified in each case.

expression in the QC cells. Additionally, cells expressing such genes cluster away from the QC, in localized regions of meristematic cluster 2, adjacent to their mature-cell counterparts (Figure S2).

Transcriptomic comparison between the QC cells and undifferentiated cells of the meristem (cells in clusters 2, 6, 7, and 8), identified 254 genes preferentially expressed in the QC (Table S3). While meristematic cells are distinguished by expression of genes involved in cell division and DNA replication, cells of the QC are not. Instead, transcription is an enriched GO term, as is auxin biosynthesis, which is fitting given the role of auxin in QC specification (Sabatini et al., 1999; Galinha et al., 2007). Further, unexpectedly, genes with functions in glucosinolate biogenesis and callose deposition are overrepresented among those genes DE in the QC (Table S3). This finding, in particular, is intriguing. Both processes are characteristic of a defense response, which seems curious given the QC's internal location,

insulated from external stimuli. Their prominence instead points toward a biology of QC cells not previously appreciated. The recent finding that 3-hydroxypropylglucosinolate acts as a reversible inhibitor of root growth (Malinovsky et al., 2017) is, in this regard, intriguing. Likewise, that small RNAs are prevented from moving in and out of the QC (Skopelitis et al., 2018) points to a unique regulation of cell-cell communication via plasmodesmata in the QC.

Among the genes DE between the QC and meristematic cells of the root, 47 show a particularly strong expression bias to the QC cells ($\log FC \geq 0.25$; $PCT1 \geq 10$; $PCT2 \leq 10$) (Table S3). Many of these genes are also expressed in mature cell types, predominantly columella cells, further reinforcing a certain shared biology not present in the apical root meristem. However, ten genes clearly mark the QC cells (Figures 4B and 4C). Reporter lines for one of them (*TEL1*) revealed high expression in the QC cells and minimal expression elsewhere (Figure 3J).

**Figure 5. Trichoblast Development Is Guided by Progressive Waves of Gene Expression**

(A) Pseudotime reconstruction of trichoblast development reveals a linear ordering of cells, reflecting cluster and subcluster arrangement.

(B) Density distribution of cells across pseudotime.

(legend continued on next page)

This expression contrasts strongly with that of genes marking various cell-type initials (Figures 3G–3I). Among the eleven QC markers are several genes not previously described in connection to the QC, including *AT3G18900* and *SOT17* (Figure 4C; Table S3). Intriguingly, the latter functions in the glucosinolate pathway (Klein et al., 2006). This reinforces the prospect of there being another layer to QC function and highlights the potential for this technology in identifying developmental regulators and the downstream genes that translate cell fate into distinctive cell shapes and functions.

The limited number of QC-specific genes, in comparison to the number of specific markers identified for cell types such as the endodermis, xylem, and trichoblasts, would indicate that QC identity reflects the integrative outcome of multiple overlapping expression signatures. This idea is supported by the position of the QC at the intersection of an auxin maximum and SCARECROW activity (Shimotohno et al., 2018). However, an alternative, non-mutually exclusive interpretation of this finding is that QC identity reflects a “subtractive” expression signature. In this scenario, the absence of expression of drivers of tissue identity and differentiation forms a key feature of QC identity. The facts that WOX5 acts as a transcriptional repressor (Forzani et al., 2014; Pi et al., 2015) and that genes, such as *AT3G22120*, *AT1G62500*, and *PIP2-8* are mostly undetectable in the QC, support this idea.

Cell Differentiation Reflects Finely Resolved Waves of Gene Expression

One of the most exciting benefits of scRNA-seq is that it allows for the simultaneous, unbiased analysis of every type of cell at every developmental stage in one experiment. This is broadly illustrated by pseudotime analysis across all cells, which shows how central clusters are defined by expression of genes at the beginning of cell fate progressions, whereas mature cell types are peripheral in the cluster cloud (Figure 2C). However, although the cluster cloud represents a coarse landscape of developmental cell states, it does not reveal how individual cells traverse these states. To resolve the progressions that cells undergo during their transition from stem cell to mature trichoblast, we performed pseudotime analysis on cells of clusters 7, 5, and 10. This revealed a linear ordering of cells that reflects the cluster and sub-cluster arrangement (Figures 5A and 5B). In addition, we identified 3,657 highly dispersed, DE genes that fall into 8 distinct gene clusters (a–h) and depict successive waves of gene expression across pseudotime (Figures 5C and S6A; Table S4).

Reporter lines generated for representative genes of select clusters precisely reflect their pseudotime profiles (Figures 5D–5H). *AT5G62330* and *PAE2* (cluster d) show a distinct peak of expression near the center of the pseudotime trajectory. Reporter lines for both genes capture this expression dynamic. *pAT5G62330:3xVenus-NLS* expression initiates in the distal meristem and persists into the elongation zone, whereas *PAE2* is expressed slightly later and shows strong expression, particu-

larly in the early elongation zone (Figures 5D and 5E). Expression of *AT1G05320* (cluster f) overlaps with that of *AT5G62330* and *PAE2* in elongating cells, but expression persists into the maturation zone and differentiated trichoblasts (Figure 5F). Finally, reporter activity for *MES15* and *EXT12* (cluster g), whose expression initiates late in the pseudotime trajectory, is first detected in cells exiting the elongation zone, with *MES15* expression starting slightly earlier with respect to the first visible protrusion of the emerging hair (Figures 5G and 5H), in accordance with its slightly earlier pseudotime projection.

The pseudotime trajectory thus reflects with great precision the temporal expression changes of individual genes along the length of the root (Figure 5C), providing a refined view of the changes a cell undergoes during its transition from stem cell through to full differentiation. Genes predominantly expressed at the beginning of the developmental trajectory show an overrepresentation of DNA replication, cell proliferation, and ribosomal functions, as is expected for meristematic cells (clusters a and b; Table S4). At the other end of the trajectory, cluster g captures expression of genes involved in unidimensional growth and root hair elongation and maturation. A previous gene regulatory network (GRN) for epidermal cell differentiation had identified 154 core root hair genes from which a temporal progression could be deduced (Bruex et al., 2012). scRNA-seq provides another dimension of temporal resolution. Of the subset (98) of core root hair genes that show a dynamic pattern of expression across the pseudotime, the vast majority (84) are expressed late in the trajectory (cluster g; Table S4). In contrast, early cell fate determinants, including *GL2*, *TRY*, and *WER*, are expressed in cluster a (meristem). Our analysis thus reveals stepwise temporal progressions more dynamic than previously appreciated that connect early cell-fate decisions to morphological and cellular changes.

These stepwise progressions are primarily captured by the central gene expression clusters. Clusters d and e reveal cell growth and cell wall biogenesis among their enriched GO terms, while those of cluster f indicate a burst of cell morphogenesis activity with an overrepresentation of genes involved in cytoskeleton reorganization, vesicle trafficking, and a plethora of transport processes. Cell expansion and cell reorganization are processes known to occur during root hair development (Balcerowicz et al., 2015), and the pseudotime analysis identifies specific genes that could drive processes such as these during root hair differentiation (Table S4). Also found in cluster d is an abundance of flavonoid-associated genes, suggesting that this signaling is occurring during cell elongation, downstream of *GL2* and *WER* but upstream of many auxin or ethylene-responsive genes that are overwhelmingly found in the later gene clusters (f and g).

A Highly Interconnected TF Gene Regulatory Network Coordinates Cell Differentiation

Expression profiles of the 230 TFs among the dynamically expressed genes mirror their waves of expression (Figure 5I);

(C) Expression heatmap of 3,657 highly dynamically expressed genes ordered across pseudotime reveals trichoblast differentiation reflected in multiple progressive waves of gene expression. Significantly enriched GO terms for clusters are labeled. Lower bar, cell density distribution across clusters.

(D–H) Spatiotemporal expression patterns for promoter fusions for the following genes reveals the pseudotime-predicted temporal localization: (D) *AT5G62330*, (E) *PAE2*, (F) *AT1G05320*, (G) *MES15*, and (H) *EXT12*. White arrowheads indicate the start of expression. Expression dynamics for single genes plotted across trichoblast pseudotime are under the corresponding root images.

(I) Expression heatmap of 230 transcription factors (TFs) extracted from Figure 4C shows similar waves of TF expression. See Table S4 for full data.

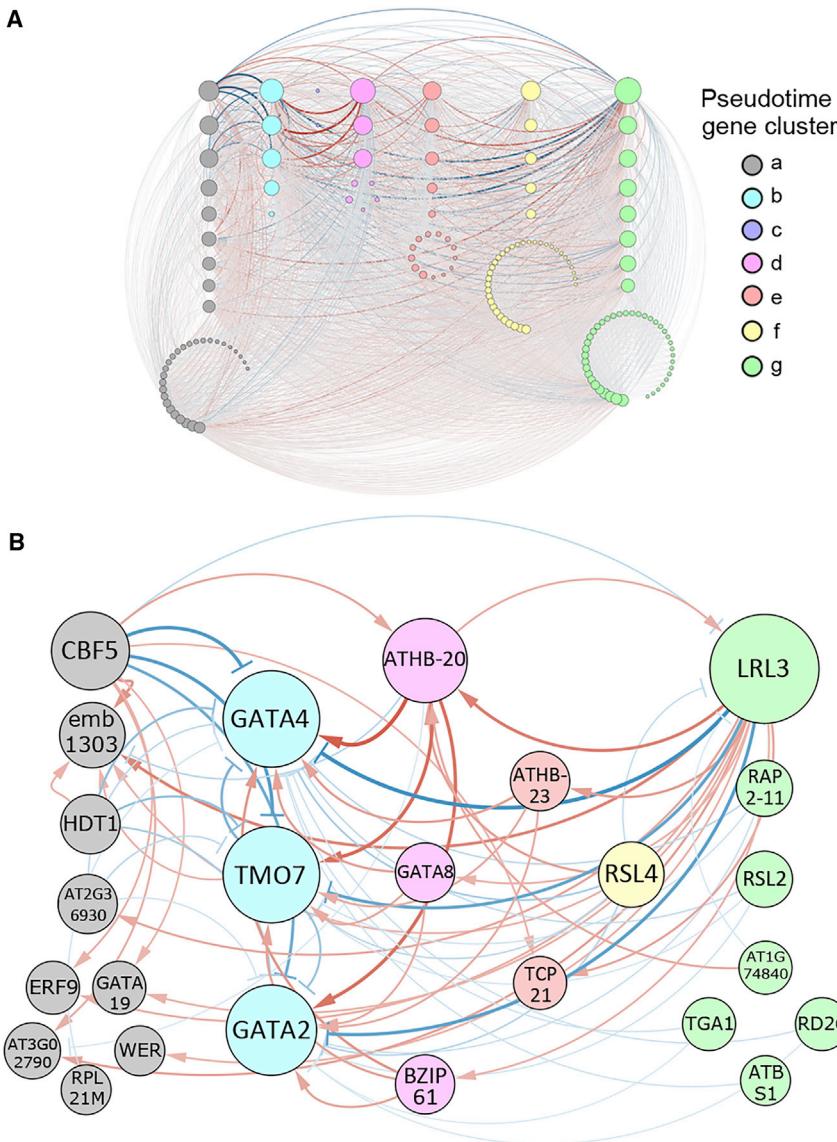


Figure 6. A Gene Regulatory Network (GRN) Predicts Key Regulators during Trichoblast Cell Differentiation

(A) A GRN built of 229 TFs expressed dynamically across trichoblast pseudotime with a parameter cutoff of 0.1 (Table S4).

(B) The same network with a parameter cutoff of 1.5. Node size is equivalent to the number of predicted connections. Edge color represents activation (red) or repression (blue). Edge width represents the strength of the predicted connection.

dynamics across pseudotime, the resultant network reveals key players in this process and their regulatory interactions. What is immediately clear from the network is the presence of major, highly connected central regulators along the whole trajectory (Figure 6A), the majority of which have not previously been implicated in root development. Filtering the network down to its 25 core components (see STAR Methods), we see several distinct passages of feedback regulation along each step of the trichoblast differentiation trajectory (Figure 6B). Notable among these is considerable negative feedback from TFs at the end of the trajectory back to major nodes in the meristem, especially LRL3 (Figure 6B). This suggests that while positive and negative feedback loops in the meristem (also distinct) might maintain meristematic identity, the progression to differentiation gives rise to dominant components that influence the meristematic master players.

ATHB-20 and, to a lesser extent, ATHB-23 stand out as positive regulators of the core network components in the neighboring, upstream, meristematic cluster, including the aforementioned TMO7.

Table S4), pointing toward a causative relationship and an intricate regulation of cell fate progressions. Several of these are known to regulate specific stages of root hair development, including TRY, WER, and many basic-helix-loop-helix (bHLH) TFs (notably, RSL2, RSL4, and LRL1). The expression clusters in which they reside accurately reflect their biological roles (Schellmann et al., 2002; Diet et al., 2006; Bruex et al., 2012; Balcerowicz et al., 2015). Likewise, TMO7, previously reported as a central cell-to-cell communicator and regulator of root meristem activity (Lu et al., 2018), is found early on in the trajectory, while the central clusters contain TCP TFs connected to the exit of cell proliferation (Nicolas and Cubas, 2016) and ARR1, which regulates root meristem size (Dello Ilio et al., 2008).

To elucidate further the genetic coordination along the trichoblast differentiation process, we inferred a GRN using a pipeline integrating the transformation of linear ordinary differential equations (ODEs) and linear regression (SCODE) (Matsumoto et al., 2017; Table S4; see STAR Methods). Incorporating TF expression

While it has been linked to ABA sensitivity in the root (Barrero et al., 2010), no defined role has yet been assigned to ATHB-20 in root development; yet, a connection to a major player such as TMO7 suggests this to be the case. Pertinent to this, it is important to note that, with the exception of those in the late developmental cluster, the 25 core genes displayed in the network are not specific to trichoblast cell development (Figure S2; Table S4). This would indicate that they fulfill a similar role in other tissue layers, possibly coordinating growth and differentiation across the root. This notion is supported by pseudotime analysis of the cortex, which reveals an equally dynamic gene expression cascade mirrored by waves of TF expression (Figures S6B and S6C; Table S4). With the notable exception of TMO7, many of the core nodes identified in the trichoblast network, including ATHB-20, ATHB-23, CBF5, GATA4, and GATA2, are also present in equivalent positions along cortex pseudotime.

While negative feedback from differentiating root hair cells toward the meristem is notable in the simplified network, forward

regulation of cell differentiation can be found as a culmination of many, weaker interactions from the meristematic and central nodes in the extended network (Figure 6A; Table S4). The well-established regulators of trichoblast maturation, expressed at the end of the trajectory, are regulated by a combination of additive and opposing effects of broadly expressed (for example, ATHB-20 and ATHB-23) and locally restricted (such as GL2 and WER) TFs. Network configurations in which tissue-specific expression reflects the combinatorial output of many broadly expressed and locally restricted TFs are emerging as a general feature underlying development (Sparks et al., 2016; Reiter et al., 2017; Niwa, 2018; Barolo and Posakony, 2002), buffering the system and providing robustness.

Our network serves to particularly highlight the complex wiring between components mediating the transition of cells from the stem cell niche to the root maturation zone. Altogether, the pseudotime analysis of scRNA-seq data indicates that these transitions are more gradual than previous data suggested. We see evidence of intricate dynamic transcriptional regulation, particularly across cells of the elongation zone. Termination of meristem activity and the initiation of the differentiation process appear coordinated across tissues, while the trichoblast maturation process relies on tissue-specific TFs whose activity comes about from the combinatorial regulation of dozens of temporally upstream players.

DISCUSSION

While scRNA-seq in plants has previously been limited to analysis of cells in their hundreds, the data presented here show that profiling of developmentally complex tissues using high-throughput scRNA-seq of thousands of cells offers an unparalleled view of the spatiotemporal expression dynamics cells undergo between exiting the niche and their final differentiation.

scRNA-seq of the *Arabidopsis* root proved to be sensitive and highly reproducible. Select genes induced by protoplasting, as well as other plant physiology-based challenges, are not prohibitive but must be considered. For instance, the technology has an inherent bias toward smaller cells, such as those from meristems. Mature cells, while captured in sufficient numbers for analysis, are underrepresented in the atlas as their identities appear to become partially clouded during the protoplasting and cell capture processes. Likewise, epidermal cells are captured more readily than cells of the central stele. These trends, however, do not affect the types of analyses presented here.

The nearly five thousand root cells profiled form a high-resolution atlas that captures all major cell types and developmental stages, including the scarce cells of the QC and niche. With this atlas, we can predict with precision the spatiotemporal patterns of gene expression within the root, as demonstrated by our reporter lines. Further, the atlas offers valuable insights into cellular processes characterizing distinct cell and tissue types. An unbiased approach to marker gene calling identified expression features defining each cluster and subcluster. These negate the need for *a priori* knowledge in assigning cluster identities and identify unique developmental regulators and downstream genes that give cells their distinctive forms and functions. As an example, SOT17 shows a QC-specific pattern of expres-

sion, and genes implicated in glucosinolate biosynthesis and callose deposition more generally show preferential expression in QC cells. Although a connection to cell-specific defense responses cannot formally be excluded, the latter finding more likely relates to plasmodesmata-mediated signaling, given the recent finding that small RNAs are blocked from moving in and out of the QC (Skopelitis et al., 2018).

It is remarkable just how well the arrangement of cell clusters and subclusters described in our atlas reflects developmental time. In the very center are QC cells with a transcriptome distinctive from that of the surrounding initials and meristematic cells. Differentiating cells are on the periphery of the cluster cloud, adjacent to relevant meristematic subclusters. This cluster arrangement is likely a general feature of tissues that capture developmental trajectories, as a similarly developmentally informed cluster orientation has been recorded in a study of mouse spermatogenesis (Lukassen et al., 2018). In the case of the *Arabidopsis* root, it indicates that expression signatures linked to cell fate are not as strong as those defining stem cell or meristem identity.

The progressions that cells undergo during differentiation are, however, far more dynamic than captured in just the cluster arrangement. As illustrated by the successive waves of gene expression revealed by pseudotime analysis, cells transitioning from the niche through differentiation follow finely resolved developmental trajectories, with progression steps beyond the commonly described meristematic, elongation, and maturation zones. Our scRNA-seq data offer the required resolution to distinguish the stepwise temporal progressions connecting early cell-fate decisions to morphological and cellular changes. While the beginning and end of the trichoblast developmental trajectory have been described (Bruex et al., 2012; Balcerowicz et al., 2015), our data not only add to this but reveal additional gene expression dynamics occurring particularly in cells in the elongation zone. The number of progression steps observed, compared to the number of cells along the root, from the meristem to maturation, implies that the distinct progressions are characteristic of few, perhaps even individual, cells along the root elongation zone.

Recent studies elaborate the idea that cell fate reflects the output of intricate GRNs in which numerous TFs control gene expression in a combinatorial manner (Sparks et al., 2016; Reiter et al., 2017; Niwa, 2018; Barolo and Posakony, 2018). This notion is reinforced by the pseudotime-derived trichoblast GRN, which shows that expression of major regulators of trichoblast maturation is linked to many weak interactions from meristematic and central nodes. Interestingly, the network also implies a high degree of feedback regulation toward the meristem, not only from the elongation zone but also from genes in the maturation zone, such as LRL3. Normal root development requires that growth and differentiation be coordinated across tissue layers. In line with this, the central highly integrated nodes predicted to coordinate the transitions between zones show similar pseudotime profiles in both the cortex and trichoblast lineages.

An additional key application for the scRNA-seq technology will be the profiling of mutants to more precisely define the cellular processes, cell types, or developmental stage affected. Cluster analysis of *shr* and wild-type scRNA-seq data revealed an expected absence of endodermal cells in the mutant

(Helariutta et al., 2000) but also points to more extensive SHR-dependent cell-fate changes. The *shr* scRNA-seq data thus nicely exemplify the enticing prospect this technology offers to discern phenotypes not easily recognized by standard RNA-seq, physiological, or even reporter-based approaches, whether stemming from natural variation, mutations, localized stress responses, or plant-microbe interactions.

In summary, the atlas of the *Arabidopsis* root described here provides a unique spatiotemporal perspective of root cell-type differentiation, increasing the number of discernible developmental domains along the length of the root and pointing to countless candidate developmental regulators that orchestrate this process. scRNA-seq will rapidly become a central technique in the plant sciences as it already is in mammalian studies, providing previously unobservable developmental insights.

STAR★METHODS

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

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
 - Protoplast Isolation
 - Bulk RNA-seq Library Preparation and Sequencing
 - Single Cell RNA-seq Library Preparation and Sequencing
 - Generation and Confocal Imaging of Reporter Lines
- QUANTIFICATION AND STATISTICAL ANALYSIS
 - Bulk RNA-seq Analysis
 - Generation of Single Cell Expression Matrices
 - Dimensionality Reduction, t-SNE Visualization, and Cell Clustering Analysis
 - Identification of Differentially Expressed Genes and Cluster-Specific Marker Genes
 - Identification of Cluster Identities
 - Correlation Analysis
 - Single Cell Developmental Trajectory Analysis
 - GO Enrichment Analysis
 - Gene Regulatory Network Analysis
- DATA AND SOFTWARE AVAILABILITY
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental Information can be found with this article online at <https://doi.org/10.1016/j.devcel.2019.02.022>.

ACKNOWLEDGMENTS

We thank members of the Timmermans lab for their insightful feedback, Chong Lu for input into computational analyses, George Janes for providing *shr-3* seed, and Felicity Jones for access to the 10x chromium single-cell controller. This work was supported by an Alexander von Humboldt Professorship to M.C.P.T.

AUTHOR CONTRIBUTIONS

T.D., X.M., and M.C.P.T. designed the project and experiments. T.D. generated the scRNA-seq and bulk RNA-seq libraries and, together with S.K.,

generated and analyzed reporter lines. X.M. performed bioinformatics analyses. X.M. and E.S. carried out the GRN analysis. K.N. provided statistical and bioinformatic support. T.D. and M.C.P.T. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 30, 2018

Revised: January 30, 2019

Accepted: February 22, 2019

Published: March 25, 2019

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.
- Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169.
- Assenov, Y., Ramírez, F., Schelhorn, S.E., Lengauer, T., and Albrecht, M. (2008). Computing topological parameters of biological networks. *Bioinformatics* 24, 282–284.
- Balcerowicz, D., Schoenaers, S., and Vissenberg, K. (2015). Cell fate determination and the switch from diffuse growth to planar polarity in *Arabidopsis* root epidermal cells. *Front. Plant Sci.* 6, 1163.
- Barolo, S., and Posakony, J.W. (2002). Three habits of highly effective signaling pathways: principles of transcriptional control by developmental cell signaling. *Genes Dev.* 16, 1167–1181.
- Barrero, J.M., Millar, A.A., Griffiths, J., Czechowski, T., Scheible, W.R., Udvardi, M., Reid, J.B., Ross, J.J., Jacobsen, J.V., and Gubler, F. (2010). Gene expression profiling identifies two regulatory genes controlling dormancy and ABA sensitivity in *Arabidopsis* seeds. *Plant J.* 61, 611–622.
- 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.
- Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120.
- Bonke, M., Thitamadee, S., Mähönen, A.P., Hauser, M.T., and Helariutta, Y. (2003). APL regulates vascular tissue identity in *Arabidopsis*. *Nature* 426, 181–186.
- Brady, S.M., Orlando, D.A., Lee, J.-Y., Wang, J.Y., Koch, J., Dinneny, J.R., Mace, D., Ohler, U., and Benfey, P.N. (2007a). A high-resolution root spatiotemporal map reveals dominant expression patterns. *Science* 318, 801–806.
- Brady, S., Song, S., Dhugga, K., Rafalski, A., and Benfey, P. (2007b). Combining expression and comparative evolutionary analysis. The COBRA gene family. *Plant Phys.* 143, 172–187.
- Breiman, L. (2001). Random forests. *Machine Learning* 45, 5–32.
- Bruex, A., Kainkaryam, R.M., Wieckowski, Y., Kang, Y.H., Bernhardt, C., Xia, Y., Zheng, X., Wang, J.Y., Lee, M.M., Benfey, P., et al. (2012). A gene regulatory network for root epidermis cell differentiation in *Arabidopsis*. *PLoS Genet.* 8, e1002446.
- Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* 36, 411–420.
- Dello Ioio, R., Nakamura, K., Moubayidin, L., Perilli, S., Taniguchi, M., Morita, M.T., Aoyama, T., Costantino, P., and Sabatini, S. (2008). A genetic framework for the control of cell division and differentiation in the root meristem. *Science* 322, 1380–1384.
- Diet, A., Link, B., Seifert, G.J., Schellenberg, B., Wagner, U., Pauly, M., Reiter, W.D., and Ringli, C. (2006). The *Arabidopsis* root hair cell wall formation mutant *Irx1* is suppressed by mutations in the *RHM1* gene encoding a UDP-L-rhamnose synthase. *Plant Cell* 18, 1630–1641.

- 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.
- Efroni, I., Ip, P.L., Nawy, T., Mello, A., and Birnbaum, K.D. (2015). Quantification of cell identity from single-cell gene expression profiles. *Genome Biol.* 16, 9.
- Feraru, E., and Friml, J. (2008). PIN polar targeting. *Plant Physiol.* 147, 1553–1559.
- Forzani, C., Aichinger, E., Sornay, E., Willemsen, V., Laux, T., Dewitte, W., and Murray, J.A. (2014). WOX5 suppresses cyclin D activity to establish quiescence at the center of the root stem cell niche. *Curr. Biol.* 24, 1939–1944.
- Galinha, C., Hofhuis, H., Luijten, M., Willemsen, V., Blilou, I., Heidstra, R., and Scheres, B. (2007). PLETHORA proteins as dose-dependent master regulators of *Arabidopsis* root development. *Nature* 449, 1053–1057.
- Gattolin, S., Sorieul, M., Hunter, P.R., Khonsari, R.H., and Frigerio, L. (2009). In vivo imaging of the tonoplast intrinsic protein family in *Arabidopsis* roots. *BMC Plant Biol.* 9, 133.
- Helariutta, Y., Fukaki, H., Wysocka-Diller, J., Nakajima, K., Jung, J., Sena, G., Hauser, M.T., and Benfey, P.N. (2000). The SHORT-ROOT gene controls radial patterning of the *Arabidopsis* root through radial signaling. *Cell* 101, 555–567.
- Klein, M., Reichelt, M., Gershenson, J., and Papenbrock, J. (2006). The three desulfoglucosinolate sulfotransferase proteins in *Arabidopsis* have different substrate specificities and are differentially expressed. *FEBS J.* 273, 122–136.
- Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* 161, 1187–1201.
- Lauber, M.H., Waizenegger, I., Steinmann, T., Schwarz, H., Mayer, U., Hwang, I., Lukowitz, W., and Jürgens, G. (1997). The *Arabidopsis* KNOLLE protein is a cytokinesis-specific syntaxin. *J. Cell Biol.* 139, 1485–1493.
- Li, S., Yamada, M., Han, X., Ohler, U., and Benfey, P.N. (2016). High-resolution expression map of the *Arabidopsis* root reveals alternative splicing and lncRNA regulation. *Dev. Cell* 39, 508–522.
- 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.
- Lu, K.J., De Rybel, B., van Mourik, H., and Weijers, D. (2018). Regulation of intercellular TARGET of MONOPTEROS 7 protein transport in the *Arabidopsis* root. *Development* 145.
- Lukassen, S., Bosch, E., Ekici, A.B., and Winterpacht, A. (2018). Characterization of germ cell differentiation in the male mouse through single-cell RNA sequencing. *Sci. Rep.* 8, 6521.
- van der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. *Mach. Learn. Res.* 9, 2579–2605.
- Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161, 1202–1214.
- Malinovsky, F.G., Thomsen, M.F., Nintemann, S.J., Jagd, L.M., Bourgine, B., Burow, M., and Kliebenstein, D.J. (2017). An evolutionarily young defense metabolite influences the root growth of plants via the ancient TOR signaling pathway. *Elife* 6, e29353.
- Mathieu, J., Yant, L.J., Mürdter, F., Küttner, F., and Schmid, M. (2009). Repression of flowering by the miR172 target SMZ. *PLoS Biol.* 7, e1000148.
- 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.
- McDavid, A., Finak, G., Chattopadhyay, P.K., Dominguez, M., Lamoreaux, L., Ma, S.S., Roederer, M., and Gottardo, R. (2013). Data exploration, quality control and testing in single-cell qPCR-based gene expression experiments. *Bioinformatics* 29, 461–467.
- Miyashima, S., Koi, S., Hashimoto, T., and Nakajima, K. (2011). Non-cell-autonomous microRNA165 acts in a dose-dependent manner to regulate multiple differentiation status in the *Arabidopsis* root. *Development* 138, 2303–2313.
- Nawy, T., Lee, J.Y., Colinas, J., Wang, J.Y., Thongrod, S.C., Malamy, J.E., Birnbaum, K., and Benfey, P.N. (2005). Transcriptional profile of the *Arabidopsis* root quiescent center. *Plant Cell* 17, 1908–1925.
- Nicolas, M., and Cubas, P. (2016). TCP factors: new kids on the signaling block. *Curr. Opin. Plant Biol.* 33, 33–41.
- Niwa, H. (2018). The principles that govern transcription factor network functions in stem cells. *Development* 145.
- Pi, L., Aichinger, E., van der Graaff, E., Llavata-Peris, C.I., Weijers, D., Hennig, L., Groot, E., and Laux, T. (2015). Organizer-derived WOX5 signal maintains root columella stem cells through chromatin-mediated repression of *CDF4* expression. *Dev. Cell* 33, 576–588.
- Porco, S., Larrieu, A., Du, Y., Gaudinier, A., Goh, T., Swarup, K., Swarup, R., Kuempers, B., Bishopp, A., Lavenus, J., et al. (2016). Lateral root emergence in *Arabidopsis* is dependent on transcription factor LBD29 regulation of auxin influx carrier LAX3. *Development* 143, 3340–3349.
- Potter, S.S. (2018). Single-cell RNA sequencing for the study of development, physiology and disease. *Nat. Rev. Nephrol.* 14, 479–492.
- Prakadan, S.M., Shalek, A.K., and Weitz, D.A. (2017). Scaling by shrinking: empowering single-cell ‘omics’ with microfluidic devices. *Nat. Rev. Genet.* 18, 345–361.
- Reiter, F., Wienerroither, S., and Stark, A. (2017). Combinatorial function of transcription factors and cofactors. *Curr. Opin. Genet. Dev.* 43, 73–81.
- Sabatini, S., Beis, D., Wolkenfelt, H., Murfett, J., Guilfoyle, T., Malamy, J., Benfey, P., Leyser, O., Bechtold, N., Weisbeek, P., et al. (1999). An auxin-dependent distal organizer of pattern and polarity in the *Arabidopsis* root. *Cell* 99, 463–472.
- Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* 33, 495–502.
- Schellmann, S., Schnittger, A., Kirk, V., Wada, T., Okada, K., Beermann, A., Thumfahrt, J., Jürgens, G., and Hülskamp, M. (2002). TRPTYCHON and Caprice mediate lateral inhibition during trichome and root hair patterning in *Arabidopsis*. *EMBO J.* 21, 5036–5046.
- Shimotohno, A., Heidstra, R., Blilou, I., and Scheres, B. (2018). Root stem cell niche organizer specification by molecular convergence of PLETHORA and SCARECROW transcription factor modules. *Genes Dev.* 32, 1085–1100.
- Skopelitis, D.S., Hill, K., Klesen, S., Marco, C.F., von Born, P., Chitwood, D.H., and Timmermans, M.C.P. (2018). Gating of miRNA movement at defined cell-cell interfaces governs their impact as positional signals. *Nat. Commun.* 9, 3107.
- 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.
- Sparks, E.E., Drapek, C., Gaudinier, A., Li, S., Ansariola, M., Shen, N., Hennacy, J.H., Zhang, J., Turco, G., Petricka, J.J., et al. (2016). Establishment of expression in the SHORTROOT-SCARECROW transcriptional cascade through opposing activities of both activators and repressors. *Dev. Cell* 39, 585–596.
- Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* 32, 381–386.
- Villani, A.C., Satija, R., Reynolds, G., Sarkizova, S., Shekhar, K., Fletcher, J., Griesbeck, M., Butler, A., Zheng, S., Lazo, S., et al. (2017). Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* 356, 4573.
- Zheng, G.X., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* 8, 14049.

STAR★METHODS**KEY RESOURCES TABLE**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and Virus Strains		
<i>Agrobacterium tumefaciens</i>	N/A	gv3101
<i>E.coli</i>	N/A	TOP10
Biological Samples		
<i>Arabidopsis thaliana</i> shr-3 mutant	Helariutta et al. (2000)	N/A
<i>Arabidopsis thaliana</i> pMIR166A:erGFP	Miyashima et al. (2011)	N/A
<i>Arabidopsis thaliana</i> Columbia ecotype	N/A	N/A
Chemicals, Peptides, and Recombinant Proteins		
Murashige and Skoog basal medium	SERVA	Cat#M0221.0005
Agar	Duchefa Biochem.	11396.03
Cellulase R-10	Duchefa Biochem.	Cat#C8001.0010
Pectolyase Y-23	Duchefa Biochem.	Cat#P8004.0001
Bovine serum albumin	Sigma-Aldrich	Cat#A7907-50G
Mannitol	Duchefa Biochem.	Cat#M0803.1000
Propidium Iodide	Sigma-Aldrich	Cat#P4864-10ML
Critical Commercial Assays		
Chromium i7 Multiplex Kit v3	10X Genomics	Cat#PN-120262
Chromium Single Cell 3' Library & Gel Bead Kit v2	10X Genomics	Cat#PN-120237
Chromium Single Cell A Chip Kit Kit v2	10X Genomics	Cat#PN-100009
DNA High Sensitivity Bioanalyzer kit	Agilent	Cat#5067-4626
DynaBeads® MyOne™ Silane Beads	Thermo Fisher Scientific	Cat#37002D
GeneJET Plasmid Miniprep. Kit	Thermo Fisher Scientific	Cat#K0502
Monarch DNA Gel Extraction Kit	New England Biolabs	Cat#T1020L
Monarch PCR and DNA Cleanup Kit	New England Biolabs	Cat#T1030L
NEBNext Library Quantification Kit	New England Biolabs	Cat#E76305
NEBNext Poly(A) mRNA Magnetic Isolation Module	New England Biolabs	Cat#E74905
NEBNext Ultra II RNA Library Prep Kit	New England Biolabs	Cat#E77605
RNA Bioanalyzer kit	Agilent	Cat#5067-1511
Spectrum Plant Total RNA Extraction Kit	Sigma-Aldrich	Cat#STRN250-1KT
Deposited Data		
Single Cell and mRNA Sequencing data	This Study	GSE123818
Single Cell RNA-Seq wild type Replicate 1 Sequencing data	This Study	GSM3511858
Single Cell RNA-Seq wild type Replicate 2 Sequencing data	This Study	GSM3511859
Single Cell RNA-Seq shr-3 mutant Sequencing data	This Study	GSM3511860
mRNA Unprotoplasted wild type Replicate 1 Sequencing data	This Study	GSM3511861
mRNA Unprotoplasted wild type Replicate 2 Sequencing data	This Study	GSM3511862
mRNA Protoplasted wild type Replicate 1 Sequencing data	This Study	GSM3511863
mRNA Protoplasted wild type Replicate 2 Sequencing data	This Study	GSM3511864

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental Models: Organisms/Strains		
<i>Arabidopsis thaliana</i>		
Oligonucleotides		
AT1G62510 fw	This Study	GCGGTACCCAAGCCATTGGTGTCTGTTGTT
AT1G62510 rev	This Study	TCCCCGGGGTTATAGAGGAGAGGTTTC
MES15 fw	This Study	GCGGTACCTGAAACCGAGGAGAGTACGG
MES15 rev	This Study	TCCCCGGGTAAGGTAGACACGTTGTAAG
PME32 fw	This Study	GCGGTACCTGCATGGAAAGTGATTGCG
PME32 rev	This Study	TCCCCGGGTCAAAGGTAGTGGAAAGTTGA
AT3G22120 fw	This Study	GCGGTACCGCCAGGTTACGGTGAGAACAA
AT3G22120 rev	This Study	TCCCCGGGAGGTACGTGTACCTTTATAG
ATL75 fw	This Study	GCGGTACCTGTACATGACCCATCTCGGTG
ATL75 rev	This Study	TCCCCGGGCTGCTTGCTTGGCTTGT
MLP34 fw	This Study	GCGGTACCGGAGAACATGGGCCACA
MLP34 rev	This Study	TCCCCGGGTATCTGGAAACAGTAGGG
AT1G62500 fw	This Study	GCGGTACCGTGGTCAACGTTGATTA
AT1G62500 rev	This Study	TCCCCGGGTATCGTTAACTAGGGTTC
PIP2-8 fw	This Study	GCGGTACCACGACCGTCTCTTTATCC
PIP2-8 rev	This Study	TCCCCGGGCTTGATCTGTGTGTTGCT
AT3G53980 fw	This Study	GCGGTACCAACTTACCAACTCTCTGGCGA
AT3G53980 rev	This Study	TCCCCGGGTGTGGAAATTGAGGCAGT
AT3G15357 fw	This Study	GCGGTACCGATGTGTGATTGGGGCTTTGTTT
AT3G15357 rev	This Study	TCCCCGGGACTGATAAGAGTTAGGACGGC
TEL1 fw	This Study	GCGGTACCCATGGAGGTTGCAAAGTGC
TEL1 rev	This Study	TCCCCGGTAATACCAGAGTTGATATTTCCG
JM164 fw	This Study	ACGCTTACAATTCCATTG
Recombinant DNA		
JM164 binary vector	Mathieu et al. (2009)	N/A
Software and Algorithms		
Trimmomatic - version 0.36	Bolger et al. (2014)	http://www.usadellab.org/cms/?page=trimmomatic
STAR	Dobin et al. (2013)	https://github.com/alexdobin/STAR/wiki
HTSeq - version 0.7.2	Anders et al. (2015)	https://htseq.readthedocs.io/en/release_0.11.1/
DEseq2	Love et al. (2014)	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
Cell Ranger 2.0.2	10X Genomics	https://support.10xgenomics.com/single-cell-gene-expression/software/release-notes/2-0
Seurat - version 2.3.4	Satija et al., 2015	https://satijalab.org/seurat/
Monocle2 - version 2.8.0	Trapnell et al. (2014)	http://cole-trapnell-lab.github.io/monocle-release/docs/
SCODE	Matsumoto et al. (2017)	https://github.com/hmatsu1226/SCODE
Cytoscape	Shannon et al. (2003)	https://cytoscape.org/
NetworkAnalyzer	Assenov et al. (2008)	http://apps.cytoscape.org/apps/networkanalyzer

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Marja C.P. Timmermans (marja.timmermans@zmbp.uni-tuebingen.de).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All analyses were performed in the *Arabidopsis* Columbia (Col-0) ecotype with the exception of the *shr-3* mutant line, which is in the Ler. ecotype. The *pMIR166A:erGFP* reporter and *shr-3* mutant have been described previously (Miyashima et al., 2011; Helariutta

et al., 2000, respectively). Plants were grown at 22°C on 1% agar plates containing 0.5x Murashige and Skoog (MS) medium (Duchefa Biochem.).

METHOD DETAILS

Protoplast Isolation

Seedlings were grown vertically on nylon mesh on agar plates. Roots of 6-day-old seedlings were cut approximately one centimetre from their tip, broadly diced with a scalpel blade, and treated with 7 ml protoplasting solution optimised for scRNA-seq from a protocol in Birnbaum et al. (2003). Immediately before use, 1.5% Cellulase R-10 and 0.1% Pectolyase (Duchefa Biochem.) were added to fresh protoplast buffer (0.1 M KCl, 0.02 M MgCl₂, 0.02 M CaCl₂, 0.1% BSA (Sigma Aldrich), 0.08 M MES, and 0.6 M Mannitol, adjusted to pH 5.5 with 0.1M Tris HCl), and mixed thoroughly. Root tissues were protoplasted for 2 hours at 20°C on an orbital shaker set at 200 revolutions/minute. The mixture was subsequently filtered through a 100 µm nylon filter and rinsed with 1-5 ml of root protoplast buffer. Protoplasts were then centrifuged for 10 minutes (500 g – 4°C), the supernatant gently removed, and the pellet resuspended in 10 ml root protoplast buffer containing 0.4 M Mannitol and no CaCl₂. This wash procedure was repeated once more, the protoplasts centrifuged as before, and resuspended in ~500 µl or less protoplast buffer without CaCl₂ and with 0.4 M Mannitol. Protoplasts were validated under a light microscope, and if necessary any excess debris or un-protoplasted tissues removed with an additional washing step. Cells were filtered through a 40 µm cell strainer (Flowmi Bel Art SP Scienceware), quantified using a haemocytometer, and adjusted to a density of approximately 800-900 cells per µl.

Bulk RNA-seq Library Preparation and Sequencing

RNA was extracted using the Spectrum Plant RNA Extraction Kit (Sigma) from protoplasted and equivalent un-protoplasted root tissue collected at completion of the protoplasting procedure. RNA samples were quantified by Nanodrop, and quality assured based on Agilent RNA Bioanalyzer chip traces. mRNA was enriched by oligo-dT pull-down using the NEBNext Poly(A) mRNA Magnetic Isolation Module, and RNA libraries constructed using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina with NEB Multiplex oligos. Final library size and quality was checked on a DNA High Sensitivity Bioanalyzer chip (Agilent), and libraries were quantified using the NEBNext Library Quantification Kit for Illumina.

Single Cell RNA-seq Library Preparation and Sequencing

Single cell RNA-seq libraries were prepared from fresh protoplasts according to the 10x Genomics Single Cell 3' Reagent Kit v2 protocol. For each replicate, 12,200 cells were loaded in the 10x Genomics Chromium single cell microfluidics device with the aim of capturing 7,000 cells. 11 cycles were used for cDNA amplification, as well as for final PCR amplification of the adapter-ligated libraries. Final library size and quality was checked on a DNA High Sensitivity Bioanalyzer chip (Agilent), and libraries were quantified using the NEBNext Library Quantification Kit for Illumina. ScRNA-seq library sequencing was performed on the NextSeq (Illumina) platform, using the sequencing parameters 26,8,0,98.

Generation and Confocal Imaging of Reporter Lines

To verify select marker genes in vivo, promoter:3xYFP-NLS reporter lines were generated for the following genes: AT1G62510, MES15, PME32, TEL1, AT3G22120, ATL75, MLP34, AT1G62500, PIP2-8, AT6G53980, AT3G15357, AT5G62330, AT1G57590, AT1G05320 and EXT12. Full names for all genes referenced in this paper can be found in Table S5. Promoter fragments between approx. 1.2 - 3.5 kb were amplified using PCR primers containing *Kpn*I and *Xma*I restriction sites, and introduced by classical cloning into binary vector JM164 (Mathieu et al., 2009) to generate transcriptional fusions to a nuclear-localised triple Venus tag. All reporter constructs were transformed into the Col-0 background, and multiple independent events per construct ($n \geq 3$) analysed. Roots of 7-day-old seedlings were mounted in 10 µg/mL Propidium Iodide (PI) (Sigma-Aldrich) and imaged using a Zeiss LSM880 laser-scanning confocal microscope. Excitation for YFP was at 514 nm and images were acquired at 517 - 571 nm. For PI, the excitation wavelength was 561nm, and images were collected at 589 - 718nm.

QUANTIFICATION AND STATISTICAL ANALYSIS

Bulk RNA-seq Analysis

Sequence reads (pair-end, 75 bp) were trimmed using Trimmomatic (version 0.36; Bolger et al., 2014), and aligned to the *Arabidopsis* TAIR10 reference genome with STAR (Dobin et al., 2013). Gene expression values were calculated on uniquely mapped reads using HTSeq (version 0.7.2; Anders et al., 2015), and DEseq2 (Love et al., 2014) was used to calculate differentially expression (absolute log₂FC ≥ 1 and q < 0.05) on genes with expression levels ≥ 1 RPM in either replicate. For correlation analysis of gene expression between protoplasted and un-protoplasted root tissues, the Log₂ (mean RPM+1) expression values were calculated for each gene and the Pearson-correlation coefficient determined in R.

Generation of Single Cell Expression Matrices

Cell Ranger 2.0.2 (10X Genomics) was used to process scRNA-seq data. Cell Ranger Count aligned the sequencing reads to the *Arabidopsis* TAIR10 reference genome using STAR (Dobin et al., 2013). For the mapping of GFP-derived transcripts, the sequence

and gene structure for GFP were added to the reference fastq and gtf files, respectively. Aligned sequence reads with a valid cell barcode and UMI that mapped to exons (Ensembl GTFs TAIR10.37) were used to generate gene expression matrices from which PCR duplicates were removed. Valid cell barcodes were defined based on UMI distribution with the cutoff: cell read count > 5% of 99th percentile of 7000 cells (Zheng et al., 2017). The output files for the two replicates were aggregated into one gene-cell expression matrix using Cell Ranger aggr with the mapped read depth normalization option.

Dimensionality Reduction, t-SNE Visualization, and Cell Clustering Analysis

The Seurat R package (version 2.3.4) (Satija et al., 2015; Butler et al., 2018) was used for dimensionality reduction analysis. Highly variable genes were identified across the single cells, after controlling for the relationship between average expression and dispersion. Genes were placed into 20 bins based on their average expression, and genes with an average expression value <0.011 removed. Within each bin, a z-score of log transformed dispersion measure (variance/mean) was calculated. A z-score cut-off of 1 was applied to identify the highly variable genes. PCA was then performed using the variable genes as input. 50 PCs were selected as input for a graph-based approach to cluster cells by cell type (Villani et al., 2017) and used as input for t-distributed stochastic neighbour embedding (t-SNE; van der Maaten and Hinton, 2008) for reduction to two- or three-dimensional visualization. A resolution value of 0.8 was used in all clustering analyses. Additionally, we used a random forest classifier (Breiman, 2001; Butler et al., 2018) to examine cluster distinctness and merged any clusters where the out-of-bag error (OOBE) of the classifier was >10%.

Identification of Differentially Expressed Genes and Cluster-Specific Marker Genes

Genes differentially expressed across clusters or subclusters were identified by comparing average expression values in cells of a given cluster to that of cells in all other clusters using the Seurat package likelihood ratio test (Bimod). The following cutoffs were applied: average expression difference ≥ 0.25 natural Log and $q < 0.01$. Cluster-specific marker genes were selected from among the differentially expressed genes based on the criteria that marker genes must be expressed in $\geq 10\%$ of cells within the cluster (PCT1), and $\leq 10\%$ of cells across all other clusters (PCT2).

Identification of Cluster Identities

The top 10 DE genes ($q < 0.01$) by fold change were identified for each cluster and expression profiles harvested from a cell-type specific and longitudinal microarray dataset (Brady et al., 2007a). In the case that microarray data was not available, the next best DE gene was selected. Average normalised expression for 10 DE genes across cell types and developmental stages was calculated and visualised in R. In a complementary approach, marker genes for key cell types were identified from Efroni et al. (2015), which integrates root expression data from multiple independent studies. Genes with high normalised expression in a particular cell type (spec. score ≥ 0.6 as detailed in Efroni et al., 2015) were filtered for specificity by applying a <0.2 spec. score cutoff for all other cell types. This latter criterion was not applied to phloem/phloem companion cells, and phloem/protophloem comparisons, as these cell types show considerable co-expression of most genes. See Table S2 for the list of marker genes. Expression of these genes was extracted from the combined single cell expression matrix and visualised using Seurat's SplitDotPlot GG function. Genes with well-defined expression patterns were considered similarly.

Correlation Analysis

For correlation analysis of merged single cell and bulk RNA sequence data, $\text{Log}_2(\text{mean RPM}+1)$ expression values for each gene from two replicates of pooled single cell and bulk RNA sequencing were quantile normalized and the Pearson-correlation coefficient calculated in R. For correlation analysis between single cell RNA-seq replicates, the single cell data replicates were simulated as bulk RNA sequencing data, the $\text{Log}_2(\text{mean RPM}+1)$ expression values calculated for each gene, and the Pearson-correlation coefficient between replicates calculated in R. For correlation analysis between the single cell replicates across individual clusters, the average expression of cells within a cluster was calculated for each replicate using the Seurat command AverageExpression(object, use.raw=T). The Pearson-correlation coefficient between the replicates was then determined for each cluster using Seurat CellPlot.

Single Cell Developmental Trajectory Analysis

Pseudotime trajectory analysis was performed using the Monocle2 R package (version 2.8.0) algorithm (Trapnell et al., 2014) on genes with a mean expression value ≥ 0.011 , and dispersion empirical value larger than the dispersion fit value. Cells were ordered along the trajectory and visualized in a reduced dimensional space. Significantly changed genes along the pseudotime were identified using the differential GeneTest function of Monocle2 with $q\text{-value} < 0.01$. Genes dynamically expressed along the pseudotime were clustered using the 'plot_pseudotime_heatmap' function with the default parameters. Transcript factors were annotated based on information from AtTFDB (<https://agris-knowledgebase.org/AtTFDB/>). Gene description information was downloaded from (<https://www.arabidopsis.org>).

GO Enrichment Analysis

Gene ontology (GO) biological process enrichment analyses (<http://pantherdb.org>) were performed on cluster-grouped differently expressed genes along the pseudotime (average expression ≥ 0.011) via fisher exact test ($q < 0.01$, Fold enrichment >1).

Gene Regulatory Network Analysis

The gene expression levels of transcription factors (Table S4) without genes with dual-polar expression (cluster h) were normalized using the Monocle2 R package (version 2.8.0) genesmoothcurve function (Trapnell et al., 2014). The pseudotime of each cell assigned by Monocle2 was normalized from 0 to 1. Gene regulatory network inference was calculated on dynamic TFs using SCODE (Matsumoto et al., 2017) with parameter z setting as 4, averaging 50x results to obtain reliable relationships. Gene regulatory inference was filtered using various cutoffs on parameter value, the results visualized using Cytoscape (Shannon et al., 2003), and the network topological parameters obtained with NetworkAnalyzer (Assenov et al., 2008).

DATA AND SOFTWARE AVAILABILITY

All high-throughput sequencing data, both raw and processed files, have been deposited in NCBI's Gene Expression Omnibus and are accessible under accession number GEO: GSE123818.

ADDITIONAL RESOURCES

Data deposition: <https://www.ncbi.nlm.nih.gov/geo/>