Dynamics of Gene Expression in Single Root Cells of Arabidopsis thaliana

Ken Jean-Baptiste,a José L. McFalone-Figueroa,a Cristina M. Alexandre,a Michael W. Dorrity,a Lauren Saunders,a Kerry L. Bubb,a Cole Trapnell,a Stanley Fields,a,b Christine Queitsch,a and Josh T. Cuperusa,1

a Department of Genome Sciences, University of Washington, Seattle, Washington 98195
b Department of Medicine, University of Washington, Seattle, Washington 98195

INTRODUCTION

Many features of plant organs such as roots are traceable to specialized cell lineages and their progenitors (Irish, 1991; Petricka et al., 2012). The developmental trajectories of these lineages have been based on tissue-specific and cell-type-specific expression data derived from tissue dissection and reporter gene-enabled cell sorting (Birnbaum et al., 2003; Brady et al., 2007; Cartwright et al., 2012). However, tissue dissection is labor-intensive and imprecise, and cell sorting requires prior knowledge of cell-type-specific promoters and genetic manipulation to generate reporter lines. Few such lines are available for plants other than the reference plant Arabidopsis (Arabidopsis thaliana; Rogers and Benfey, 2015). Advances in single cell transcriptomics can replace these labor-intensive approaches. Single cell RNA sequencing (RNA-Seq) has been applied to heterogeneous samples of human, worm, and animal cells, revealing both known and novel genes that are expressed along a developmental trajectory. We identify transcription factor motifs that are enriched in early and late cells, together with the corresponding candidate transcription factors that likely drive the observed expression patterns. We assess and interpret changes in total RNA expression along developmental trajectories and show that trajectory branch points mark developmental decisions. Finally, by applying heat stress to whole seedlings, we address the longstanding question of possible heterogeneity among cell types in the response to an abiotic stress. Although the response of canonical heat-shock genes dominates expression across cell types, subtle but significant differences in other genes can be detected among cell types. Taken together, our results demonstrate that single cell transcriptomics holds promise for studying plant development and plant physiology with unprecedented resolution.

RESULTS

Single-Cell RNA-Seq of Whole A. thaliana Roots Reveals Distinct Populations of Cortex, Endodermis, Hair, Nonhair, and Stele Cells

We used whole Arabidopsis roots from 7d-old seedlings to generate protoplasts for transcriptome analysis using the 10×
IN A NUTSHELL

Background: Plants are made up of many different types of cells with different roles. These cells can be identified based on the types of genes they express, which are important to their roles. Many approaches have been used to study specific types of cells in Arabidopsis roots, including methods to express fluorescence proteins in only some types of cells. Newer technologies can capture thousands of individual cells at the same time, measuring the total RNA molecules in each cell. We have applied this single-cell RNA-seq technology to the Arabidopsis root.

Question: We sought to answer several important questions, but also bring new technological advances to plant systems. We asked the biological questions: Can distinct cell-types be resolved using single-cell RNA-seq of a whole Arabidopsis root? Can developmental information be understood using single-cell RNA-seq?

Findings: We identified all the major cell types in Arabidopsis roots based on each cell type’s unique RNA expression profile. We also visualized the developmental trajectory of many cell types, finding differences between young and old cells within the same tissue, and identifying transcription factors that were expressed specifically in young or old cells. We captured distinct, small populations of cells, including ones actively dividing. Lastly, we applied an abiotic stress, heat shock, to these roots. We found drastic changes in expression, some of which were cell-type specific. Cells in the exterior layers of the root had larger changes in expression than did interior cell types.

Next steps: Our work shows that it is possible to discern different cells with single cell RNA-seq in Arabidopsis roots, so in the future we hope to apply the same technique to Arabidopsis shoots, and other plants such as camelina, sorghum, and maize.
Figure 1. Annotation of Cell and Tissue Types for Single Cell RNA-Seq of Whole Arabidopsis Roots.

(A) Root cells were clustered and projected onto two-dimensional space with UMAP (McInnes and Healy, 2018). Solid circles represent individual cells; colors represent their respective Louvain component. Monocle 3 trajectories (black lines) are shown for clusters in which a trajectory could be identified.

(B) Solid circles represent individual cells; colors indicate cell and tissue type based on highest Spearman’s rank correlation with sorted tissue-specific bulk expression data (Brady et al., 2007; Cartwright et al., 2009).

(C) Known marker genes (Brady et al., 2007; Cartwright et al., 2009) were used to cluster single cell gene expression profiles based on similarity. The expression of 530 known marker genes was grouped into seven clusters, using k-means clustering. Mean expression for each cluster (rows) is presented for each cell (columns). Cells were ordered by their respective Louvain component indicated above by color (see (A), starting at component 1 at left). Number of genes in each cluster is denoted at right.

(D) Single cell RNA-Seq pseudo-bulked expression data are compared with bulk expression data of whole roots (Li et al., 2016).

(E) Single cell pseudo-bulk expression data are compared with bulk-expression data of the three developmental regions of the Arabidopsis root (Li et al., 2016).

(F) Proportions of cells as annotated by either UMAP (A), Spearman’s rank correlation (B), or Pearson’s rank (in Supplemental Figure 2), are compared with proportions determined by microscopy (Brady et al., 2007; Cartwright et al., 2009).
showed high and specific mean expression of cortex marker genes (Figure 1C; Supplemental Figure 3; Supplemental Data Set 3). Both expression correlations and marker gene expression allowed us to assign the Louvain components to five major groups: root hair cells, nonhair cells (containing both an early and late cluster), cortex cells, endodermis cells, and stele cells (containing both xylem and phloem cells; Figure 1A). Although some cells were most highly correlated in expression with the cell-type columnella in Spearman’s rank tests and RNA-Seq Pearson’s correlation, these cells coclustered with nonhair cells (Figure 1B; Supplemental Figure 2). This finding is consistent with bulk RNA-Seq data of sorted cells (Li et al., 2016). Specifically, the PET111 (columnella)-sorted bulk RNA-Seq data are most similar to bulk RNA-Seq data sorted for GLABRA2 (GL2) and WEREWOLF (WER; Li et al., 2016), both of which mark nonhair cells (Petricka et al., 2012). Therefore, these cells were grouped as early nonhair cells with other nonhair cells in Louvain component 8. As their expression values were best correlated with RNA-Seq data for WER-sorted cells, they likely represent a mix of early nonhair and lateral root cap cells, which have very similar expression profiles (Supplemental Figure 2).

We assessed the extent to which combined single cell root expression data resembled bulk whole-root expression data (Figures 1D and 1E; Li et al., 2016). We observed strong correlations between these two data sets (Pearson’s correlation coefficient $R^2 = 0.52$, Spearman’s $\rho = 0.71$). We also compared the combined single cell expression data to three bulk expression data sets representing the major developmental zones in the Arabidopsis root: the meristematic zone, the elongation zone, and the maturation zone (Figure 1E). We observed the highest correlation of single cell and bulk expression in the elongation zone ($R^2 = 0.70$, $\rho = 0.83$) and a lower correlation in the maturation zone ($R^2 = 0.58$, $\rho = 0.70$). This observation is surprising given the more mature developmental stage of the harvested roots (Supplemental Figure 1A), and likely reflects that younger cells are more easily digested during protoplasting and contribute in greater numbers to the gene expression data. As expected, single cell and bulk expression were poorly correlated in the meristematic zone ($R^2 = 0.11$, $\rho = 0.43$), as meristematic tissue accounts for only a small proportion of mature roots. Furthermore, we compared tissue-specific expression (Li et al., 2016) to expression both in the annotated cell clusters and in cells expressing appropriate marker genes. In general, we found strong correlations among these data sets, suggesting that the clusters are annotated correctly (Supplemental Table 1).

We also compared the relative representation of root cell types between our data set and estimates based on microscopy studies (Figure 1F; Brady et al., 2007; Cartwright et al., 2009). Independent of annotation method, we observed the expected numbers of cortex (222 Spearman’s/233 UMAP), endodermis (306/304), nonhair cells (1,201/1,061), and columnella cells (111/no UMAP cluster). Hair cells (565/898) were overrepresented whereas stele cells (508/490) were underrepresented, possibly reflecting a bias in the protoplast preparation of whole root tissue.

Protoplasting, the removal of the plant cell wall, alters the expression of 346 genes (Birnbaum et al., 2003; 76 of these genes were included in the 1,500 genes with the highest variation in expression (Supplemental Data Set 1; Supplemental Figure 1B) that we used for clustering. Some of the 76 genes showed cell-type-specific expression. To exclude the possibility that the expression pattern of these genes produced artificial clusters and cell-type annotations, we removed them from the analysis and reclustered, which resulted in a similar UMAP visualization, with similar numbers of Louvain components and cell types.

### Single Cell RNA-Seq Identifies Novel Genes with Cell-Type- and Tissue-Type-Specific Expression

Some marker genes are not expressed exclusively in a single cell type, making it desirable to identify additional genes with cell-type-specific expression. First, we confirmed the high and cluster-specific expression of well-known marker genes (Figure 2A; Supplemental Figure 4; Li et al., 2016) such as the root-hair-specific COBRA-LIKE 9 (COBL9), the endodermis-specific SCARECROW (SCR) and the three stele-specific genes MYB46 (xylem-specific), ALTERED PHLOEM DEVELOPMENT (APL; phloem-specific), and SUCROSE-PROTON SYMPORTER 2 (SUC2; phloem-specific). The nonspecific expression of the quiescent center cell marker genes WUSCHEL RELATED HOMEOBOX 5 (WOX5) and AGAMOUS-LIKE 42 (AGL42) is likely due to the failure to capture sufficient numbers of these rare cells.

Second, to find novel marker genes, we identified genes with significantly different expression within and among Louvain component clusters by applying the Moran’s I test implemented in Monocle 3. We found 317 genes with cluster-specific expression, 164 of which were novel, including at least one in each cluster (Figure 2A; Supplemental Data Set 4). Using cell-type annotations rather than Louvain clusters, we identified 510 genes with cell-type-specific expression, of which 317 overlapped with the Louvain component cluster-specific expression genes, as well as an additional 125 novel genes, some of which have been implicated in the development of a cell lineage in targeted molecular genetics studies.

For example, the stele-specific gene AT1G68810 (ABNORMAL SHOOT 5; stele; Figure 2A) encodes a basic helix-loop-helix (bHLH) protein that promotes vascular cell division and differentiation as part of a heterodimer with a second bHLH protein, LONESOME HIGHWAY (Ohashi-Ito et al., 2014; Katayama et al., 2015). Another stele-specific gene, AT4G36160 (VASCULAR-RELATED NAC-DOMAIN 2; cluster 7; Figure 2A), encodes a Class IIb Naccent polypeptide-Associated Complex (NAC)-domain transcription factor that contributes to xylem vessel element differentiation by promoting secondary cell wall formation and programmed cell death (Tan et al., 2018). In tissue-specific bulk data (Brady et al., 2007; Winter et al., 2007), both genes show xylem-specific expression consistent with their biological functions; ABNORMAL SHOOT 5 expression is high only in the meristematic and elongation zones, whereas VASCULAR-RELATED NAC-DOMAIN 2 expression starts in the elongation zone and persists throughout the maturation zone. Other genes, not previously implicated in root development, show tissue-specific bulk expression patterns consistent with the single cell data. For example, AT1G54940 (GLUCURONIC ACID SUBSTITUTION OF
XYLAN4), which encodes a xylan glucuronosyltransferase (Mortimer et al., 2010; Lee et al., 2012), was specifically expressed in hair cells (cluster 9/hair; Figure 2A) and is most highly expressed in cells destined to become hair cells in the elongation zone and in differentiated hair cells in the maturation zone (Brady et al., 2007; Cartwright et al., 2009).

Expression of Some Factor Genes Shows High Correlation with Specific Cell Types

We asked whether we could identify transcription factors that may contribute to the cluster-specific expression patterns. To do so, we tested for transcription factor motif enrichments in the proximal regulatory regions of genes with cluster-specific expression, examining 500 bp upstream of the transcription start site (Sullivan et al., 2014; Alexandre et al., 2018) and a comprehensive collection of Arabidopsis transcription factor motifs (O’Malley et al., 2016). This analysis revealed significant transcription factor motif enrichments among clusters and annotated major tissues and cell types (Figure 2B).

As transcription factors in Arabidopsis often belong to large gene families without factor-specific motif information (Riechmann et al., 2000), it is challenging to deduce the identity of the specific transcription factor that drives cluster-specific transcription factor motif enrichment and expression. As an approximation, we examined transcription factor genes that were expressed in the cluster or
tissue in which a significant enrichment of their motif was found, or in neighboring cell layers (some factors move between cells; Petricka et al., 2012; Supplemental Data Set 4). We focused first on the small BRI1-EMS Suppressor (BES)/Brassinazole-Resistant (BZR) Homolog (BEH) gene family whose motif was specifically enriched in cortex cells (cluster 10). Of the six genes (BEH1/AT3G50750, BEH2/AT4G36780, BEH3/AT4G18890, BEH4/AT1G78700, BES1/AT1G19350, and BZR1/AT1G75080), the single recessive beh4, bes1, and bzar1 mutants exhibit altered hypocotyl length (Lachowiec et al., 2018). Double mutant analysis suggests partial functional redundancy, which agrees with our observation of overlapping expression patterns for these genes across cell types (Supplemental Figures 5A and 5B). By contrast, neither beh1 and beh2 single mutants nor the respective double mutant show phenotypic defects (Lachowiec et al., 2018). However, BEH2 was the most highly expressed BZR/BEH family member across clusters and annotated root tissue and cell types (Supplemental Figures 5A and B). Although BEH4, the most ancient family member with the strongest phenotypic impact, showed cortex-specific expression, none of the BZR/BEH genes showed significance for cluster-specific expression, suggesting that combinations of family members, possibly as heterodimers, may result in the corresponding motif enrichment in cortex cells (Supplemental Figures 5A and 5B). In particular, BES1 and BZR expression was highly correlated, consistent with these genes being enriched in cortex (cluster 10), endodermis (cluster 1), and stele (cluster 4). TCP14 (AT3G47620) and TCP15 (AT1G69690) were expressed primarily in stele (clusters 7 and 4), although this cluster-specific expression was not statistically significant (Figure 2B; Supplemental Figures 5D and 5E; Supplemental Data Set 4). TCP14 and TCP15 encode class-I TCP factors thought to promote development. Acting together, TCP14 and TCP15 promote cell division in young internodes (Kieffer et al., 2011), seed germination (Resentini et al., 2015), cytokinin and auxin responses during gynoecium development (Lucero et al., 2015), and repression of endoreduplication (Peng et al., 2015). Both genes are expressed in stele in bulk tissue data (Brady et al., 2007; Winter et al., 2007), with TCP14 expression also observed in the vasculature by in situ hybridization (Tatematsu et al., 2008). TCP14 can affect gene expression in a non-cell–autonomous manner.

To further investigate the co-occurrence of cluster-specific transcription factor motif enrichments with transcription factor expression, we next examined the novel genes with significant cluster-specific expression. Eight of these encode transcription factors with corresponding highly enriched cluster-specific binding motifs. For one of these, BEARSKIN2 (AT4G10350), cluster-specific expression coincided with enrichment of the NAC transcription factor family motif (cluster-8, nonhair, and lateral root cap cells; Figure 2B). BEARSKIN2 encodes a Class IIB NAC transcription factor implicated in root cap maturation together with BEARSKIN1 and SOMBRERO. Class IIB NAC transcription factors are thought to contribute to terminal cell differentiation accompanied by strong cell wall modifications (Bennett et al., 2010). In our data, BEARSKIN2 was most highly expressed in cluster 8 (nonhair and lateral root cap cells) and less so in cluster 6 (Supplemental Data Set 4).

Clustering Stele Cells Identifies Novel Genes with Cell-Type–Specific Expression in the Vasculature

Our initial attempts to annotate and separate cell types within stele tissue with marker gene expression or Spearman’s rank correlations failed. Instead, we separately clustered stele cells to reveal six subclusters upon UMAP visualization, with five subclusters containing more than 40 cells. Their annotation via Spearman’s rank correlation with sorted bulk data was not successful; however, using well-established marker genes expression, we detected cluster-specific expression patterns (Figures 3A and 3B).

Cells closely related to the xylem pole pericycle constituted the largest group of cells (205 cells); phloem pole pericycle cells were the second largest (84 cells). The high number of pericycle cells likely reflects our experimental procedure, as these cells reside on the exterior of the vascular bundle. Both phloem and xylem clusters showed similar numbers of cells (77 cells and 72 cells, respectively); the phloem companion cells formed a distinct cluster. We observed the expected subcluster expression for several known genes and marker genes and identified novel genes with subcluster-specific expression (Figures 3C and 3D; Supplemental Data Set 1). Although there was some discrepancy, especially for the APL gene, which is expressed in both companion and phloem cells (Figure 3C), this is largely due to missing data.

Pseudotime Trajectories Coincide with the Development Stages of Cortex, Endodermis, and Hair Cells

We next sought to visualize the continuous program of gene expression changes that occurs as each cell type in the root differentiates. Because whole roots contain a mix of cells at varying developmental stages, we reasoned that our experiment should have captured a representative snapshot of their differentiation. Monocle not only clusters cells by type but also places them in “pseudotime” order along a trajectory that describes their maturity. To make these trajectories, Monocle 3 learns an explicit principal graph from the single cell expression data through reverse graph embedding, an advanced machine learning method (Trapnell et al., 2014; Qiu et al., 2017a, 2017b). To dissect the developmental dynamics of individual clusters, we first focused on the well-defined root-hair cells, in which combined single cell expression values highly correlated with those from bulk protoplasts sorted for expression of the COB1L9 root-hair marker gene (Supplemental Table 1). To annotate the unsupervised trajectory that Monocle 3 created for hair cells, we used the Spearman’s rank test to compare expression in all cells to bulk expression data representing 13 different developmental stages in root tissues from all the available sorted cell types (Supplemental Figure 6; Brady et al., 2007; Cartwright et al., 2009). Each cell was assigned the developmental stage and cell type most correlated with its expression values (Figure 4A). The hair cells with the earliest developmental stage assignment were designated as the root of...
Figure 3. Reclustering of Stele Cells Yields Distinct Subclusters of Vasculature Cell Types.

(A) Cells initially annotated as stele tissue were reclustered, resulting in six distinct subclusters cells, five of which contained >40 cells.

(B) Mean expression for previously identified cell-type-specific genes (Cartwright et al., 2009) in each cell is shown, allowing annotation of stele subcluster identities as shown in (A).

(C) Proportion of cells (circle size) and mean expression (circle color) of genes with cluster-specific and tissue-specific expression are shown, starting with known marker genes at the top, labeled with their common name (right) and their systematic name (left). Below, novel significant tissue-specific genes are shown with their systematic names, identified by principal graph tests (Moran’s I) as implemented in Monocle 3.

(D) Example expression overlays for cluster-specific genes identified by the principal graph test in (C).
the trajectory. Next, pseudotime was calculated for all other hair
cells based on their distance from the root of the trajectory
(Figure 4B). We compared this calculated pseudotime with the
most highly correlated developmental assignment from bulk data,
finding close agreement (Figure 4B). Examples of genes that are
expressed early or late in pseudotime in the UMAP hair cluster are
shown in Figure 4C.

Hair cells undergo endoreduplication as they mature, resulting
up to 16N genomic copies in the developmental stages assayed
(Bhosale et al., 2018). Although endoreduplication is thought to
increase transcription rates (Bourdon et al., 2012), general tran-
scription might decrease as hair-cell-specific genes become
more highly expressed during hair-cell differentiation. Single cell
RNA-Seq affords us the opportunity to explore whether transcrip-
tion rates differ across development. Single cell RNA-Seq can
measure both relative expression (as in bulk RNA-Seq) and the
total number of RNA molecules per cell. The total amount of
cellular mRNA was drastically reduced across hair-cell de-
velopment (Figure 4D). This result may be due to technical bias; for
example, gene expression in larger, endoreduplicated cells may be
more difficult to assess with this droplet-based method. If so,
the observed reduction in captured transcripts should affect all
genes more or less equally. Alternatively, this observation may
reflect hair-cell differentiation, whereby transcription of hair-
cell-specific genes should remain unaffected or increase over
pseudotime. Our results support the latter scenario as tran-
scription of hair-cell-specific genes appears to increase over
pseudotime, consistent with these cells undergoing differentiation
toward terminally differentiated hair cells (Figure 4E; Supplemental
Figure 7A).

To further explore this transcriptional dynamic, we calculated
RNA velocity (La Manno et al., 2018), a measure of the trans-
scriptional rate of each gene in each cell of the hair-cell cluster.
RNA velocity takes advantage of errors in priming during 3’ end
reverse transcription to determine the splicing rate per gene and
cell. It compares nascent (unspliced) mRNA to mature (spliced)
mRNA; an overall relative higher ratio of unspliced to spliced
transcripts indicates that transcription is increasing. In our data,
only ~4% of reads were informative for annotating splicing rates,
a lower percentage than what has been used in mammalian cells
for velocity analyses, and thus our results may be less reliable.
Based on data for 996 genes, mean RNA velocity increased across
pseudotime (Supplemental Figure 7B, $P = 2.2 \times 10^{-16}$ linear model,
$p = 0.73$). This increase in velocity was associated with the pre-
dicted changes in endoreduplication (Bhosale et al., 2018), es-
specially between the 4N and 8N stages (Supplemental Figure 7C;
Tukey’s multiple comparison $P = 0.0477$).

We also observed developmental signals in other cell types,
including cortex and endodermis (Figures 5A to 5D; Supplemental
Figure 8). Combined single cell expression values for cortex cells
highly correlated with those from bulk protoplasts sorted for
expression of the COR cortex marker gene (Figure 5B; $R^2 = 0.74$,
$p = 0.86$). As Monocle 3 did not identify a trajectory for cortex
cells in the context of all cells, we isolated the cortex cells and
reperformed UMAP dimensionality reduction, clustering, and
graph embedding (Supplemental Data Set 1). Each cortex cell
was assigned a developmental stage based on its Spearman’s rank correlation with bulk expression data (Brady et al., 2007; Cartwright et al., 2009). Cortex cells with the earliest de-
velopmental signal were designated as the root of the cortex
trajectory, and pseudotime was assigned to the remaining cortex
cells based on their distance from the root (Figures 5A to 5D;
Supplemental Figure 6). As pseudotime increased for cortex cells,
so did their age, indicating good agreement of the trajectory with
developmental bulk RNA-Seq data. Although we observed some
decrease in total RNA expression and increased expression in
cell-type-specific genes for endodermis, we did not see a clear
pattern of change in total RNA across cortex pseudotime
(Supplemental Figures 8 and 9).

We asked whether we could assign the transcription factors that
drive gene expression along these developmental trajectories in
early and late hair, cortex, and endodermis cells. As before, we first
analyzed transcription factor motif enrichments and then explored
the expression of the corresponding transcription factor gene families. Indeed, for most developmentally enriched transcription
factor motifs, we could pinpoint candidate transcription factors
that are expressed either early or late. For example, the APE-
TAL2/ethylene-responsive element binding protein transcription
factor family is one of the largest in Arabidopsis (Riechmann et al.,
2000), with nearly 80 covered in our data set; of these, only
four (AT2G25820, AT5G65130, AT1G36060, and AT1G44830)
showed strong expression in late hair cells (Figures 4F and 4G;
Supplemental Figure 10). One of these, AT1G36060 (Translucent
Green), regulates expression of aquaporin genes (Zhu et al., 2014).
Overexpression of this gene confers greater drought tolerance
(Zhu et al., 2014), consistent with its expression in older hair cells.
Similar examples of developmental stage-specific motif enrich-
ments with corresponding transcription factor expression were
also found for cortex and endodermis (Figures 5E and 5F;
Supplemental Figures 8 and 10).

Branch Points in Developmental Trajectories Mark
Developmental Decisions

Although a developmental trajectory that reflects the differenti-
ation from early to late cells within a cell type should be branchless,
we did observe some branch points, for example in Louvain
component 8, affording us the opportunity to explore their bi-
ological relevance. As discussed, Louvain component 8 contains
early nonhair cells and likely some lateral root cap cells. To further
explore the cells within the branch, we performed a principal graph
test, comparing their expression profiles to those of cells else-
where in the cluster (Figure 6A). We found that cells within the
branch were significantly enriched for expression of genes in-
volved in cell plate formation, cytokinesis, and cell cycle. We
explored this enrichment for cell-cycle annotations by comparing
expression of previously identified core cell-cycle genes
(Gutierrez, 2009) in cells within the branch to cells in the rest of
the cluster, finding many core cell-cycle genes, in particular many G2
genes, to be specifically expressed in branch cells (Figure 6B).
Among these genes were several of the cyclin-dependent kinase
(CDK) B family members that direct the G2 to M transition. Two
CDK subunits (encoded by CKS1 and CKS2), thought to interact
with several CDK family members, were also specifically ex-
pressed in branch cells (Vandepoele et al., 2002). Other branch-
cell-specific genes included AURORA1 (AUR1) and AUR2, both
Figure 4. Developmental Trajectory of Hair Cells.

(A) UMAP-clustered hair cells were assigned a developmental time point based on highest Spearman’s rank correlation with bulk expression data of staged tissue (13 developmental stages; Brady et al., 2007; Cartwright et al., 2009). Cell type and developmental time points are indicated in shades of blue (and pink). Graphic illustrates developmental stages in Arabidopsis root (plant illustrations).
involved in lateral root formation and cell plate formation (Figure 6C; Van Damme et al., 2011). Louvain component 9 also showed a strong, but short branching point. We did not find any biological processes enriched in genes expressed specifically in this short branch; however, one gene whose expression is known to be affected by protoplasting was specifically expressed in these cells, perhaps reflecting that cells within this branch were more stressed by our experimental procedure (data not shown).

**Heat-Shocked Root Cells Show Subtle Expression Differences among Cell Types**

A major question in studying plant responses to abiotic stress, such as heat or drought, is the extent to which such responses are nonuniform across cell types. Canonically, the heat stress response is characterized by rapid and massive upregulation of a few loci, mostly encoding heat-shock proteins, with dramatic downregulation of most other loci, in part because of altered mRNA splicing and transport (Yost and Lindquist, 1986, 1988; Saavedra et al., 1996). In plants, a set of 63 genes, most encoding heat-shock proteins, show extreme chromatin accessibility at both promoter and gene body upon heat stress, consistent with their high expression (Sullivan et al., 2014). In mammals and insects, not all cells are competent to exhibit the hallmarks of the heat-shock response (Dura, 1981; Morange et al., 1984); specifically, cells in early embryonic development fail to induce heat-shock protein expression upon stress.

We explored whether all cells within developing roots were capable of exhibiting a typical heat-shock response. To do so, we applied a standard heat stress (45 min, 38°C) to 8-d-old seedlings, harvested their roots along with roots from age- and time-matched control seedlings, and generated protoplasts for single cell RNA-Seq of both samples. For the control sample, we captured 1,076 cells, assaying expression for a median 4,079 genes per cell and a total of 22,971 genes; 82.7% of reads mapped to the TAIR10 genome assembly. The results for these control cells were similar to those described earlier with regard to captured cell types, genome assembly. The results for these control cells were similar to those described earlier with regard to captured cell types, gene expression.

For the heat-shock sample, we captured 1,009 cells, assaying expression for a median 4,384 genes per cell and a total of 21,237 genes; 79.8% of reads mapped to the TAIR10 genome assembly.

Due to global gene expression changes upon heat shock, we could not simply assign cell and tissue types as before for heat-shocked cells. The overwhelming impact of heat shock was also apparent when comparing the first and second highest cell-type and developmental Spearman’s rank correlations for control cells and heat-shocked cells. Upon heat shock, many cells, especially those with nonhair, phloem, and columna as their highest rank, commonly showed as their second highest rank a different cell type instead of another developmental time point of the same cell type as observed in control cells (Supplemental Figure 11A). Unsurprisingly, the drastic changes in gene expression led to cells being embedded in UMAP space primarily as a function of treatment, making direct comparisons of treatment effects on any one cell type impossible (Supplemental Figure 11B). To enable such comparisons, we used a mutual nearest-neighbor to embed control cells, and generated protoplasts for single cell RNA-Seq of both samples. For the control sample, we captured 1,076 cells, assaying expression for a median 4,079 genes per cell and a total of 22,971 genes; 82.7% of reads mapped to the TAIR10 genome assembly. The results for these control cells were similar to those described earlier with regard to captured cell types, proportion of cell types (e.g. 28.8% versus 34% annotated hair cells and 9.7% versus 7.2% endodermis cells), and correlation of gene expression (R^2 = 0.86 for the 21,107 genes captured in both experiments). For the heat-shock sample, we captured 1,009 cells, assaying expression for a median 4,384 genes per cell and a total of 21,237 genes; 79.8% of reads mapped to the TAIR10 genome assembly.

In response to stress, organisms are thought to upregulate stress genes and to specifically downregulate genes involved in growth and development to optimize resource allocation. In response to heat stress, this presumed “dichotomy” in gene expression is mirrored by the rapid localization of RNA polymerase II to the heat-shock gene loci and its depletion elsewhere in the genome (Teves and Henikoff, 2011). Our data provide strong evidence of this regulatory trade-off at the level of individual cells. Using hair cells (Louvain component 2) as an example, we found that hair-cell–specific genes are overwhelmingly repressed and
Figure 5. Developmental Trajectory of Cortex Cells.

(A) Cortex cells were reclustered to create a trajectory, in which each cell was assigned a developmental time point and identity (shades of yellow, brown, and pink) based on the highest Spearman’s rank correlation of a cell’s gene expression with prior sorted bulk data (Brady et al., 2007; Cartwright et al., 2009).
that heat-shock genes are upregulated, often dramatically so (Figures 7B and 7D). Indeed, **HEAT SHOCK PROTEIN 101 (HSP101)**, the most highly expressed and chromatin-accessible gene upon heat shock in previous studies (Sullivan et al., 2014), was strongly expressed across all clusters whereas expression of the hair marker gene **COBL9** decreased dramatically upon stress (Figures 7C and 7D).

Having established comparable clusters, we next identified genes that were differentially expressed as a function of treatment and cluster identity, excluding those with <15 cells in either control or heat-shock conditions. This analysis identified 8,526 genes (false discovery rate [FDR] < 0.1%) whose expression was altered by heat-shock treatment in one or more clusters; of these, 2,627 genes were up- or downregulated at least twofold (Figure 7E; Supplemental Data Set 5; FDR < 0.1% and absolute value of log2-fold change > 1). As for hair cells (Figure 7B), cell-type marker genes for all clusters were enriched among the downregulated genes upon heat shock. To identify cluster-specific differences in the response to heat shock, we compared gene expression of cells within individual clusters to the rest of the cells across treatments. We observed the largest number of cluster-specific gene expression changes in hair, nonhair, and cortex cells (Figure 7F). As these cell types are the three outermost cell layers of the root, they may be exposed more directly to the heat shock and respond more quickly. Genes differentially expressed in hair cells (Louvain component 2) upon heat shock were enriched for ribosome-associated genes and RNA methylation. Stele cells (Louvain component 6) showed differential expression of genes involved in cell wall organization and biogenesis, and endodermis cells (Louvain component 4) showed differential expression of genes involved in response to external, chemical, and stress stimuli as well as nitrate and anion transport (Figure 7F).

The expression of heat-shock proteins protects cells from heat shock and aids their recovery (Parsell et al., 1993; Parsell and Lindquist, 1993; Queitsch et al., 2000). We were interested in whether we could detect cluster- and cell-type–specific differences in the canonical heat-shock response. In principle, such differences could be exploited to alter heat-shock protein expression in a cell-type–specific manner to boost plant heat and drought tolerance without pleiotropically decreasing whole organism fitness. To address such possible differences, we focused on genes that from bulk analyses have differential expression upon heat shock (1,783 genes) or reside near regulatory regions that change in accessibility upon heat shock (1,730 genes; Sullivan et al., 2014; Alexandre et al., 2018). Although these gene sets overlap (942 genes), they contain complementary information, as changes in accessibility do not necessarily translate into altered expression, and vice versa (Alexandre et al., 2018). In our single cell expression analysis, we identified 752 of 1,783 heat-responsive genes as differentially expressed upon heat shock, and 564 of 1,730 genes near dynamic regulatory regions as differentially expressed. We hierarchically clustered control and heat-shock–treated single cell transcriptomes for both gene sets (Supplemental Figures 12A and 12C), resulting in several gene clusters with distinct expression patterns. Overall, cellular responses were dominated by the canonical heat-shock response, as visualized in cluster 4 (Supplemental Figure 12A) and cluster 2 (Supplemental Figure 12C). The 63 genes showing extreme accessibility and high expression upon heat shock (Sullivan et al., 2014) are largely contained in these two clusters (Supplemental Figure 12A, cluster 4, 49 of 63; Supplemental Figure 12C, cluster 2, 42 of 63).

Our analysis also revealed subtle but significant differences among some tissue types (Supplemental Figures 12A and 12B, e.g. clusters 3 and 8; Supplemental Figures 12C and 12D, e.g. clusters 5 and 7; Supplemental Data Set 6). Although most of these gene clusters were not enriched for specific annotations, cluster-8 genes were associated with rRNA metabolic processes (P value = 0.048) and cluster-5 genes (Supplemental Figures 12A and 12B) were enriched for transport genes (P value = 0.045). These results demonstrate both the promise and the challenges inherent in comparing single cell data across different conditions and treatments.

**DISCUSSION**

Here, we use Arabidopsis roots to establish both experimental and analytic procedures for single cell RNA-Seq in plants. Using Monocle 3, we could assign over 3,000 cells to expected cell and tissue types with high confidence. In particular, cortex, endodermis and hair cells were easily identified. However, distinguishing other cell types was challenging. For example, nonhair
and columella cells had high similarity in their expression profiles, consistent with their correlation in bulk expression data (Brady et al., 2007; Cartwright et al., 2009). Similarly, it was difficult to designate cells in Louvain component 8 as early nonhair cells, as these cells showed overlapping expression signatures for early nonhair cells, lateral root caps, and epidermis cells before differentiation to hair and nonhair cells. These Louvain component-8 cells were difficult to distinguish further with the sparse expression data typical for single cell analysis; however, we postulate that the root of the trajectory are cells dividing out of the epidermis/root cap precursor and these cells either become root cap cells or epidermis.

We also could not initially split stele tissue into individual cell types, likely because the difficulty of digesting the cell walls of the tightly packed vascular bundle resulted in fewer cells than expected (Brady et al., 2007; Cartwright et al., 2009). However, analyzing stele cells separately yielded six subclusters, which correspond to known vasculature cell types. Our approach to annotate these subclusters exemplifies the ad hoc nature of current single cell genomics studies, which require all available sources of information to be exploited to interpret the genomic data. Neither Spearman rank correlations with sorted bulk RNA-Seq data nor microarray expression data yielded obvious cluster identities. However, mean expression values of genes known to be expressed in vasculature cell types allowed us to assign the stele subclusters.

We identified hundreds of novel genes with cell-type–specific and tissue-type–specific expression, which may allow the generation of new marker lines for detailed genetic analyses. These genes, together with cluster-specific enriched transcription factor motifs and their corresponding transcription factors, are candidates for driving differentiation and cell-type identity. Similarly, the developmental trajectories we identified highlight the potential of single cell transcriptomics to advance a high-resolution view of plant development. These trajectories can be detected without the use of spatial information because plants have a continuous body plan with new cells continuously arising while older cells persist. Additionally, while this study allowed us to infer transcription factor motifs and candidate transcription factors, future analyses with greater numbers of cells than assayed here may include combinatorial expression of multiple transcription factor family members.

We explored the relationships of endoreduplication, transcriptional rates, and differentiation to find that transcriptional rates, measured as mRNA velocity, increase with increasing ploidy. However, this transcriptional increase appears to be limited to genes specifically expressed in hair cells, as overall levels of RNA decreased over pseudotime. These observations are consistent with hair cells becoming more specialized and moving toward a terminally differentiated state over time. However, this phenomenon of increasing specialization was not as apparent in other cell types. This difference may be due to biological causes,

---

Figure 6. Branch Analysis Reveals Actively Dividing Cells.

(A) The 70 cells that resided in the branch of Louvain component 8 (purple) show significant branch-specific expression of genes enriched for cell-cycle function.

(B) Comparison of all known cell-cycle genes with expression in at least 5% of cells in Louvain component 8. Known cell-cycle expression is denoted for each gene, if unknown ‘?’.

(C) Two kinases, *AUR1* and *AUR2*, were specifically expressed in branch cells. These genes are involved in cell plate formation and lateral root formation.
Figure 7. Single-Cell RNA-Seq Highlights Canonical and Novel Aspects of the Heat-Shock Response.

(A) A nearest-neighbor approach aligns control and heat-shocked cells in a UMAP embedding to allow for concomitant cluster/cell-type assignment. 

(B) Volcano plots of average gene expression change upon heat shock within Louvain component 2 for all genes (black), known hair marker genes (blue), and heat-shock signature genes (red).
such as the higher rates of endoreduplication in hair cells, or to technical causes, such as the better clustering and trajectory of hair cells compared with the other cell types assayed.

By allowing trajectories with side branches, we discovered that branch points can mark developmental decisions. In Louvain component 8, the small but distinct cell-cycle–enriched branch may mark lateral root primordia cells differentiating into epidermal cells or epidermal/lateral root precursor cells. Cells within this branch express many cell-cycle genes, among them members of the CDK B family that govern the G2 to M transition. Moreover, these cells specifically express the AUR1 and AUR2 genes, which function in cell plate formation; plants with mutations in these genes lack lateral roots (Van Damme et al., 2011). Although expression of cell-cycle genes may persist in nondividing cells because of their roles in endoreduplication, AUR1 and AUR2 expression (and cell plate formation) should not persist, consistent with our speculation that the cells within this branch are actively dividing cells in the G2 to M transition (Gutierrez, 2009). We also examined the cells in Louvain component 1 (designated endodermis) that are nearer to Louvain component 10 (designated cortex). The cells residing in this position correlate best with cortex endodermis initial cells.

We explored the Arabidopsis heat-shock response with single cell RNA-Seq because not all cells and tissues are equally competent to respond to stress. By identifying plant cell types that most strongly respond to abiotic stresses such as heat, drought, and nutrient starvation, ultimately we may be able to genetically manipulate relevant cell types to generate stress-tolerant crops without pleiotropically affecting plant fitness and yield. Although all heat-shocked cells showed gene expression changes typical of the canonical heat-shock genes, we detected subtle but highly significant expression differences among cells and tissue types for other genes. Thus, single cell transcriptomics across stress conditions holds potential for future crop breeding and genetic engineering. However, such analyses require much larger numbers of cells than currently accessible by droplet-based methods. More- ever, such analyses should focus on treatments that are less impactful on the analysis of plant development and environmental response.

**METHODS**

**Plant Material and Growth Conditions**

Arabidopsis (Arabidopsis thaliana) Col-0 seedlings were grown vertically at 22°C, on 1× Murashige and Skoog (MS) + 1% Suc (w/v) plates covered with one layer of filter paper. Seven- or 8-d-old seedlings (Long Day, 16-h light/8-h dark, −100 μmol m⁻² s⁻¹, 50% relative humidity) were collected around Zeitgeber Time 3, and the roots/shoots excised with a sharp razor blade. For the heat shock, seedling plates were transferred from 22°C to 38°C for 45 min (Conviron TC-26, −100 μmol m⁻² s⁻¹, 4100 K, 82 CRI, Sylvania Octron F017/84/Eco fluorescence bulbs), and the roots harvested immediately after.

**Protoplast Isolation**

Protoplast isolation was done as described in Bargmann and Bimbaum (2010), with slight modifications. Briefly, 1 g of whole-roots was incubated in 10 mL of protoplasting solution for 1.5 h at 75 rpm. After passing through a 40-μm strainer, protoplasts were centrifuged at 500 g for 5 min and washed once in protoplasting solution without enzymes. Final suspension volume was adjusted to a density of 500 to 1,000 cells/μL. Protoplasts were placed on ice until further processing.

**Single-Cell RNA-Seq Protocol**

On two separate sets of Arabidopsis root protoplasts on separate days, single cell RNA-Seq was performed using the 10× scRNA-Seq platform, the Chromium Single Cell Gene Expression Solution (10× Genomics).

**Data Analysis**

**Estimating Gene Expression in Individual Cells**

Single cell RNA-Seq reads were sequenced and then mapped to the TAIR10 Arabidopsis genome using the software Cellranger (v. 2.1.0; https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/what-is-cell-ranger). Cellranger produces a matrix of UMI counts where each row is a gene and each column represents a cell. The ARAPORT gene annotation was used. For the heat-shock analysis, reads from a control sample and reads from a heat-shocked sample were aggregated using “cellranger aggr” to normalize libraries to an equivalent number of mean reads per cell across libraries.

**Running Monocle 3: Dimensionality Reduction, and Cell Clustering**

The output of the Cellranger pipeline was parsed into R (v. 3.5.0) using the Cellranger R kit (v. 2.0.0) and converted into a CellDataSet (CDS) for further analysis using the software Monocle 3 Alpha (v. 2.99.1; http://cole-trapnell-lab.github.io/monocle-release/monocle3/). All Monocle 3 analysis was...
performed on a High Performance Computing cluster using 128 GB of RAM spread across eight cores. The lower detection limit for the CDS was set at 0.5, and the expression family used set to negbinomial.size() to control for variability in library construction efficiency across cells.

We visualized cell clusters and trajectories using the standard Monocle workflow. Monocle internally handles all normalization needed for dimensionality reduction, visualization, and differential expression via size factors that control for variability in library construction efficiency across cells. After estimating the library size factors for each cell (via estimateSizeFactors) and estimating the dispersion in expression for each gene (via estimateDispersions) in the data set, the top 1,500 genes in terms of dispersion, i.e. 1,500 genes with the most expression variability in our data set, were selected to order the cells into clusters. The expression values of these 1,500 genes for each cell were log-transformed and projected onto the first 25 PCs via Monocle’s data preprocessing function (preprocessCDS). Then, these lower-dimensional coordinates were used to initialize a nonlinear manifold learning algorithm implemented in Monocle 3 called “UMAP” (via reduceDimension; McLinnes et al., 2018). This allows us to visualize the data into two or three dimensions. Specifically, we projected onto two components using the “cosine distance” metric, setting the parameters n_neighbors = 50, and min_dist = 0.1.

The Louvain method was used to detect cell clusters in our two-dimensional representation of the data set (partitionCells); this resulted in 11 cell clusters, or Louvain components. Cells were then clustered into super groups using a method derived from approximate graph abstraction (Wolf et al., 2018) and for each super group, a cell trajectory was drawn atop the projection using Monocle’s reversed graph embedding algorithm, which is derived from SimplePPT (learnGraph; Mao et al., 2015). This yielded six cell trajectories.

To further analyze the clusters we annotated as stele, clusters 3, 4, and 7 were reclustered together and were reanalyzed using Monocle 3 as previously described in this article except the parameter “min_dist” was changed to “0.05” when the “reduceDimension function” was called. This revealed six additional subclusters.

To further analyze the cluster we annotated as cortex, Cluster 10 was reclustered and reanalyzed using Monocle 3 as previously described in this article except the parameters “n_neighbors” was reduced to 25. This did not reveal any subclusters, but a trajectory was generated.

**Estimating Doublets**

Single Cell Remover of Doublets (Scrublet) was used to predict doublets in our scRNA-Seq data (https://github.com/AllonKleinLab/scrublet). Using the software Python 3.5, Scrublet was run using default settings as described by the example tutorial that is available at a Python notebook (https://github.com/AllonKleinLab/scrublet/blob/master/examples/scrublet basics.ipynb). The only significant change was that expected double rate was set to 0.1; in the tutorial it is 0.06.

**Identifying Cell Types**

To categorize the cells into cell types and to apply developmental information, a deconvolved root expression map was downloaded from The Arabidopsis Gene Expression Database (AREX LITE; http://www.arexdb.org/data/decondatamatrix.zip). Using this data matrix, the Spearman’s rank correlation was calculated between each cell in our data set and each cell type and longitudinal annotation in the data matrix (3,121 x 128 Spearman’s rank correlations total). Specifically, we looked at the correlation of 1,229 highly variable genes in our data set. These 1,229 genes represent the overlap between our 1,500 highly variable genes and genes in the root expression map data matrix. Cells in our data set were assigned a cell type and a developmental label based on the annotation with which each cell had the highest correlation (i.e. if a cell correlated highest with the endodermis cells in longitudinal zone 11, then it would be called as endodermis_11).

In addition to using the Spearman’s rank correlation to assign cells their cell type, a set of known marker genes derived from green fluorescent protein (GFP) marker lines of the Arabidopsis root were used to identify cell types based on the high gene expression of these marker genes. These genes were obtained from Brady et al. (2007) and Cartwright et al. (2009).

**Running Monocle 3: Identifying High-Specificity Genes**

To identify differentially expressed genes between cell clusters, the Moran’s I test was performed on our UMAP (principalGraphTest), with the projection being broken up into 25 x 25 spatial units. Then marker genes were identified for each cluster, and each annotated grouping of clusters using a Moran’s I threshold of 0.1 and a qval threshold of 0.05. For a gene to be considered highly specific, it must have had a specificity rating of >0.7.

**Transcription Factor Motif Analysis**

Highly specific genes were identified for each cell cluster, and their promoters were analyzed for presence of transcription factor motifs. Promoters were defined as 500 bp upstream of the start site of each gene. Instances of each motif were identified using the method from Grant et al. (2011) at a P value cutoff of 1e-5 for each match. The input position weight matrices for each motif were enumerated in a previous study of binding preferences for nearly all Arabidopsis transcription factors (O’Malley et al., 2016). Motif frequencies in genes specific to each cell cluster were compared with a background set of motif frequencies across all promoters in the Arabidopsis genome to determine a log2 enrichment score. Transcription factor family genes were pulled from the gene family page of TAIR10 (https://www.arabidopsis.org/browse/genefamily/index.jsp).

**Calculating Total mRNA**

After pseudotime analysis was performed on a cell cluster, cells were binned together such that each bin contained a similar number of cells and each bin represented cells from similar pseudotimes. The median total mRNA and the log of the total mRNA of each bin was then calculated.

**Calculating Significance with the Permutation Test**

The permutation test was used to calculate the significance of observed trends that the total mRNA of hair marker genes and hair-specific genes increased as pseudotime increased in hair cells. To do this, 10,000 random samplings of 441 genes (the number of hair marker genes) and 201 genes (the number of hair specific genes) were taken respectively. Next, the median total mRNA was calculated across pseudotime for each random sampling and the slope of this data was calculated using a generalized linear model. The observed slope of the marker genes and the hair-specific genes was compared with the distribution of slopes generated by 10,000
random samplings. No random sampling of genes had a slope that was higher than the observed slopes generated by the hair marker genes or the hair-specific genes. The significance, or the $P$ value, of the trend seen in the hair marker genes and the hair-specific genes can then be calculated simply as the proportion of sampled permutations that have a slope that is equal to or greater than the slope generated by our genes of interest. This gives us a $P$ value = $1/10,001$, or roughly $10^{-4}$.

**Analyzing Expression Differences between Branches of Louvain Component 8 (Early Nonhair)**

To identify genes responsible for the branching in the pseudotime trajectory of Louvain component 8 (early nonhair), the principal graph test was used to identify genes with expression specific to the side branch versus the main branch. Genes were considered specific if they had a specificity value >0.8. Genes were removed from the analysis if they did not have expression in at least 10% of the cells considered and a mean expression >0.25.

**Calculating RNA Velocity**

We used the Velocyto R and Python packages (v. 0.6 and 0.17, respectively) to estimate RNA velocity for root hair cells (La Manno et al., 2018). Matrices of spliced and unspliced RNA counts were generated from Cellranger outputs using velocyto.py CLI and run10x defaults. We followed the velocyto.py and velocyto.R manuals (http://velocyto.org/) and used spliced (emat) and unspliced (innat) matrices to estimate RNA velocity. With predefined cell-type annotations, we performed gene filtering with the parameter min.max.cluster, average set to 0.2 and 0.05 for emat and innat respectively. RNA velocity using the selected 996 genes was estimated with the defaults to the function gene.relative.velocity.estimates() except parameters kCells and fit.quantile which were set to 5 and 0.05, respectively. Velocity measurements for each cell were calculated as the difference between $\delta T = 1$ results from the estimated velocity output.

**Analysis of Heat-Shock Data**

For each pair of cell types and for each gene cluster, we used a generalized linear model to determine the significance of an interaction between the effects of cell type and heat treatment on the normalized expression level of genes in that cluster. Then, to identify differentially expressed genes specific for every Louvain cluster, we subsetted cells from every cluster that contained 15 or more cells in both control and treated conditions, estimated dispersions for each subset, and tested for differential gene expression identified using the differentialGeneTest function in Monocle specifying a full model of Treatment cluster and a residual model of 1. FDR values per gene were then obtained across all tests using the Benjamini-Hochberg method. The overlap of differentially expressed genes as a function of heat-shock treatment between clusters was visualized using an upsetR plot. Briefly, a binary matrix of differentially expressed genes by cluster was generated where gene-cluster combinations were set to 1 (significant) or 0 (not significant). This matrix was then passed to the upset function from the UpsetR R package specifying nine sets and ordering by frequency. To identify whether clusters contained subtle differences in the expression of previously identified heat-shock–responsive genes, we tested for differential gene expression across all cells and clusters and identified the intersect between differentially expressed genes obtained from single cell profiles and previously identified dynamic changes in DNase I Hypersensitive Sites (DHSs)-linked genes and bulk-differentially expressed genes upon heat shock. Differentially expressed genes as a function of heat-shock treatment for all cells in unison were identified using the differentialGeneTest function in Monocle, specifying a full model of Treatment*UMAP cluster and a residual model of UMAP cluster. Hierarchical clustering of these DHS-linked and bulk-differentially expressed gene sets across control and heat-shock–treated cells was performed using the heatmap function in the heatmapR package (v. 1.0.10) specifying ward.D2 as the clustering method. Genes with similar dynamics across treatment and cell types were recovered using the cutree function from the stats package in R, specifying k = 8 for both DHS-linked genes and bulk differentially expressed genes. To generate signatures from these eight groups of clustered genes, we log-normalized expression values using a pseudocount of 1, and for each cell calculated the mean normalized expression value across genes that belong to one of the eight gene clusters.

**Accession Numbers**

All sequencing data can be found on Gene Expression Omnibus at: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE121619.

**Supplemental Data**

**Supplemental Figure 1.** General tissue and data features.

**Supplemental Figure 2.** Pearson correlation to sorted RNA-seq samples.

**Supplemental Figure 3.** Marker gene expression in cell-type clusters.

**Supplemental Figure 4.** Examples of tissue-specific gene expression.

**Supplemental Figure 5.** Transcription factor family expression patterns.

**Supplemental Figure 6.** Spearman’s rank correlation for each cell’s development and tissue type.

**Supplemental Figure 7.** Changes in transcription across hair cell development.

**Supplemental Figure 8.** Developmental trajectory of endodermal cells.

**Supplemental Figure 9.** Median total RNA in cortex across pseudotime.

**Supplemental Figure 10.** Developmental expression of individual transcription factors.

**Supplemental Figure 11.** Heat-shock clustering and expression profiling.

**Supplemental Figure 12.** Conditional expression in genes with dynamic chromatin accessibility during heat shock.

**Supplemental Table 1.** Bulk RNA-seq comparisons to single cell RNA-seq.

**Supplemental Table 2.** Number of cells in the control-versus-heat-shock analysis.

**Supplemental Data Set 1.** List of ordering/high-dispersion genes.

**Supplemental Data Set 2.** Correlation with bulk expression data.

**Supplemental Data Set 3.** Marker genes.

**Supplemental Data Set 4.** Novel high-specificity genes.

**Supplemental Data Set 5.** Cluster-specific heat-shock differentially expressed genes.

**Supplemental Data Set 6.** Generalized linear model pairwise test of significance among cortex, hair, and nonhair cells.

**ACKNOWLEDGMENTS**

We thank members of the Trapnell Lab for input and discussion. This work was supported by the National Science Foundation (NSF grant MCB-1516701 to C.Q., and NSF grant RESEARCH-PGR 1748843 to C.Q. and...
Brennecke, P., Anders, S., Kim, J.K., Ko Brady, S.M., Orlando, D.A., Lee, J.Y., Wang, J.Y., Koch, J., Bourdon, M., Pirrello, J., Cheniclet, C., Coriton, O., Bourge, M., Cao, J., et al. (2017) Comprehensive single-cell transcriptional pro-
Birnbaum, K., Shasha, D.E., Wang, J.Y., Jung, J.W., Lambert, G.M., Alexandre, C.M., et al. (2018) Complex relationships between chro-
REFERENCES
26, 2019; published March 28, 2019.
Received October 31, 2018; revised February 12, 2019; accepted March 26, 2019.

AUTHOR CONTRIBUTIONS
J.T.C., C.M.A., and K.J.-B. performed and designed most of the experi-
ments; K.J.-B. did most of the bioinformatic analysis; M.W.D. performed transcription factor enrichment analysis; L.S. performed RNA velocity
analysis; J.L.M.-F. performed the heat-shock analysis; K.L.B. performed statistical analysis of heat-shock experiments; J.T.C., C.Q., and K.J.-B.
A

S.F.); the University of Washington National Institutes of Health (Big Data for Genomics and Neuroscience Training Grant T32LM01249 to K.J.-B.); the National Institutes of Health (NIH grants U54DK107979, DP2HD088158, RC2DK114777, and R01HL118342 to C.T.); and the Paul G. Allen Frontiers Group (to C.T.).

Haghiwerdi, L., Lun, A.T.L., Morgan, M.D., and Marioni, J.C. (2018). Batch effects in single-cell RNA-seq sequencing are corrected by matching mutual nearest neighbors. Nat. Biotechnol. 36: 421–427.

Katayama, H., Iwamoto, K., Kariya, Y., Asakawa, T., Kan, T., Fukuda, H., and Ohashi-Ito, K. (2015). A negative feedback loop controlling bHLH complexes is involved in vascular cell division and differentiation in the root apical meristem. Curr. Biol. 25: 3144–3150.

Kieffer, M., Master, V., Waites, R., and Davies, B. (2011). TCP14 and TCP15 affect internode length and leaf shape in Arabidopsis. Plant J. 68: 147–158.

Lachowiec, J., Lemus, T., Thomas, J.H., Murphy, P.J., Nemhauser, J.L., and Queitsch, C. (2013). The protein chaperone HSP90 can facilitate the divergence of gene duplicates. Genetics 193: 1269–1277.

Lachowiec, J., Mason, G.A., Schultz, K., and Queitsch, C. (2018). Redundancy, feedback, and robustness in the Arabidopsis thaliana BZR/BEH gene family. bioRxiv 053447; .

La Manno, et al. (2018) RNA velocity of single cells. Nature 560: 494–498.

Lan, X., and Pritchard, J.K. (2016). Coregulation of tandem duplicate genes slows evolution of subfunctionalization in mammals. Science 352: 1009–1013.

Lee, C., Teng, Q., Zhong, R., and Ye, Z.H. (2012). Arabidopsis GUX proteins are glucuronosyltransferases responsible for the addition of glucuronic acid side chains onto xylan. Plant Cell Physiol. 53: 1204–1216.

Lex, A., Gehlenborg, N., Strobelt, H., Vuilleumot, R., and Plichter, H. (2014). UpSet: Visualization of intersecting sets. IEEE Trans. Vis. Comput. Graph. 20: 1983–1992.

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 lincRNA regulation. Dev. Cell 39: 508–522.

Lucero, L.E., Uberti-Manassero, N.G., Arce, A.L., Colombatti, F., Alemano, S.G., and Gonzalez, D.H. (2015). TCP15 modulates cytokinin and auxin responses during gynoecium development in Arabidopsis. Plant J. 84: 267–282.

Mao, Q., Yang, L., Wang, L., Goodison, S., and Sun, Y. (2015). SimplePPT: A simple principal tree algorithm. In Proceedings of the 2015 SIAM International Conference on Data Mining, pp. 792–800.

McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for dimension reduction. arXiv:1802.03426 [stat.ML]. https://arxiv.org/abs/1802.03426 .

REFERENCES
Alexandre, C.M., et al. (2018) Complex relationships between chro-
matin accessibility, sequence divergence, and gene expression in Arabidopsis thaliana. Mol. Biol. Evol. 35: 837–854.
Bargmann, B.O.R., and Birnbaum, K.D. (2010). Fluorescence acti-
vated cell sorting of plant protoplasts. J. Vis. Exp. 36: 1673.
Bennett, T., van den Toorn, A., Sanchez-Perez, G.F., Campilho, A., Willemsen, V., Snel, B., and Scheres, B. (2010). SOMBRERO, BEARSKIN1, and BEARSKIN2 regulate root cap maturation in Arabidopsis. Plant Cell 22: 640–654.
Bhosale, R., et al. (2018) A spatiotemporal DNA endoploidy map of the Arabidopsis root reveals roles for the endocycle in root development and stress adaptation. Plant Cell 30: 2330–2351.
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.
Blondel, V.D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. J. Stat. Mech. 2008: 1–12.
Bourdon, M., Pirrello, J., Cheniciet, C., Coriton, O., Bourge, M., Brown, S., Moïse, A., Pepelut, M., Rouyer, V., Renaudin, J.P., Chevalier, C., and Frangne, N. (2012). Evidence for karyoplasmic homeostasis during endoreduplication and a ploidy-dependent increase in gene transcription during tomato fruit growth. Development 139: 3817–3826.
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. (2007). A high-resolution root spatiotemporal map reveals dominant expres-
sion patterns. Science 318: 801–806.
Brennecke, P., Anders, S., Kim, J.K., Kołodziejczyk, A.A., Zhang, X., Proserpio, V., Baying, B., Benes, V., Teichmann, S.A., Marioni, J.C., and Helsier, M.G. (2013). Accounting for technical noise in single-cell RNA-seq experiments. Nat. Methods 10: 1093–1095.
Cao, J., et al. (2017) Comprehensive single-cell transcriptional pro-
file of a multicellular organism. Science 357: 661–667.
Cartwright, D.A., Brady, S.M., Orlando, D.A., Sturmels, B., and Benfey, P.N. (2009). Reconstructing spatiotemporal gene expression data from partial observations. Bioinformatics 25: 2581–2587.
Denyer, T., Ma, X., Kiesen, S., Scacchi, E., Nieselt, K., and Timmermans, M.C.P. (2019). Spatiotemporal developmental trajec-

tories in the Arabidopsis root revealed using high-throughput single-cell RNA sequencing. Dev. Cell 48: 840–852.
Dura, J.M. (1981). Stage dependent synthesis of heat shock induced proteins in early embryos of Drosophila melanogaster. Mol. Gen. Genet. 184: 381–385.
Efroni, I., Ip, P.-L., Navy, T., Mello, A., and Birnbaum, K.D. (2015). Quantification of cell identity from single-cell gene expression profiles. Genome Biol. 16: 9.
Efroni, I., Mello, A., Navy, T., Ip, P.-L., Rahni, R., DelRose, N., Powers, A., Satija, R., and Birnbaum, K.D. (2016). Root rein-
generation triggers an embryo-like sequence guided by hormonal interactions. Cell 165: 1721–1733.
Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: Scanning for occurrences of a given motif. Bioinformatics 27: 1017–1018.
Gutierrez, C. (2009). The Arabidopsis cell division cycle. Arabidopsis Book 7: e0120.

A
Packer, J., and Trapnell, C. (2018). Single-cell multi-omics: A genome browser for new quantitative models of gene regulation. Trends Genet. 34: 653–665.

Parsell, D.A., and Lindquist, S. (1993). The function of heat-shock proteins in stress tolerance: degradation and reactivation of damaged proteins. Annu. Rev. Genet. 27: 437–496.

Parsell, D.A., Taulien, J., and Lindquist, S. (1993). The role of heat-shock proteins in thermotolerance. Philos. Trans. R. Soc. London. Ser. B Biol. Sci. 339: 279–286.

Peng, Y., Chen, L., Lu, Y., Wu, Y., Dumenil, J., Zhu, Z., Bevan, M.W., and Li, Y. (2015). The ubiquitin receptors DAR1, DAR1 and DAR2 redundantly regulate endoreplication by modulating the stability of TCP14/15 in Arabidopsis. Plant Cell 27: 649–662.

Petricka, J.J., Winter, C.M., and Benfey, P.N. (2012). Control of Arabidopsis root development. Annu. Rev. Plant Biol. 63: 563–590.

Qiu, X., Hill, A., Packer, J., Lin, D., Ma, Y.A., and Trapnell, C. (2017a). Single-cell mRNA quantification and differential analysis with Census. Nat. Methods 14: 309–315.

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

Queitsch, C., Hong, S.W., Vierling, E., and Lindquist, S. (2000). Heat shock protein 101 plays a crucial role in thermotolerance in Arabidopsis. Cell Repor 7: 482–492.

Resentini, F., Felipo-Benavent, A., Colombo, L., Blázquez, M.A., Alabadi, D., and Masiero, S. (2015). TCP14 and TCP15 mediate the promotion of seed germination by gibberellins in Arabidopsis thaliana. Mol. Plant 8: 482–492.

Riechmann, J.L., et al. (2000) Arabidopsis transcription factors: Genome-wide comparative analysis among eukaryotes. Science 290: 2105–2110.

Rogers, E.D., and Benfey, P.N. (2015). Regulation of plant root system architecture: Implications for crop advancement. Curr. Opin. Biotechnol. 32: 93–98.

Russell, A.B., Trapnell, C., and Bloom, J.D. (2018). Extreme heterogeneity of influenza virus infection in single cells. eLife 7: e32303.

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.

Saavedra, C., Tung, K.S., Amberg, D.C., Hopper, A.K., and Cole, C.N. (1996). Regulation of mRNA export in response to stress in Saccharomyces cerevisiae. Genes Dev. 10: 1608–1620.

Sullivan, A.M., et al. (2014) Mapping and dynamics of regulatory DNA and transcription factor networks in A. thaliana. Cell Reports 8: 2015–2030.

Tan, T.T., Endo, H., Sano, R., Kurata, T., Yamaguchi, M., Ohtani, M., and Demura, T. (2018). Transcription factors VND1-VND3 contribute to cotyledon xylem vessel formation. Plant Physiol. 176: 773–789.

Tatematsu, K., Nakabayashi, K., Kamiya, Y., and Nambara, E. (2008). Transcription factor AtTCP14 regulates embryonic growth potential during seed germination in Arabidopsis thaliana. Plant J. 53: 42–52.

Teves, S.S., and Henikoff, S. (2011). Heat shock reduces stalled RNA polymerase II and nucleosome turnover genome-wide. Genes Dev. 25: 2387–2397.

Trapnell, C. (2015). Defining cell types and states with single-cell genomics. Genome Res. 25: 1491–1498.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelson, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. 32: 381–386.

Van Damme, D., De Rybel, B., Gudesblat, G., Demidov, D., Grunewald, W., De Smet, I., Houben, A., Beeckman, T., and Russinova, E. (2011). Arabidopsis a Aurora kinases function in formative cell division plane orientation. Plant Cell 23: 4013–4024.

Vandepoele, K., Raes, J., De Veylder, L., Rouzé, P., Rombauts, S., and Inzé, D. (2002). Genome-wide analysis of core cell cycle genes in Arabidopsis. Plant Cell 14: 903–916.

Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G.V., and Provart, N.J. (2007). An “Electronic Fluorescent Pictograph” browser for exploring and analyzing large-scale biological data sets. PLoS One 2: e718.

Wolf, F.A., Hamey, F., Plass, M., Solana, J., Dahlin, J.S., Gottgens, B., Rajewsky, N., Simon, L., and Theis, F.J. (2018). Graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. bioRxiv.

Wolock, S.L., Lopez, R., and Klein, A.M. (2018). Scrublet: Computational identification of cell doublets in single-cell transcriptomic data. bioRxiv https://www.biorxiv.org/content/10.1101/357368v1.

Yost, H.J., and Lindquist, S. (1986). RNA splicing is interrupted by heat shock and is rescued by heat shock protein synthesis. Cell 45: 185–193.

Yost, H.J., and Lindquist, S. (1988). Translation of unspliced transcripts after heat shock. Science 242: 1544–1548.

Zhu, D., Wu, Z., Cao, G., Li, J., Wei, J., Tsuge, T., Gu, H., Aoyama, T., and Qu, L.J. (2014). TRANSLUCENT GREEN, an ERF family transcription factor, controls water balance in Arabidopsis by activating the expression of aquaporin genes. Mol. Plant 7: 601–615.