Cell Reports, Volume *27*

Supplemental Information

High-Throughput Single-Cell Transcriptome Profiling of Plant Cell Types

Christine N. Shulse, Benjamin J. Cole, Doina Ciobanu, Junyan Lin, Yuko Yoshinaga, Mona Gouran, Gina M. Turco, Yiwen Zhu, Ronan C. O'Malley, Siobhan M. Brady, and Diane E. Dickel

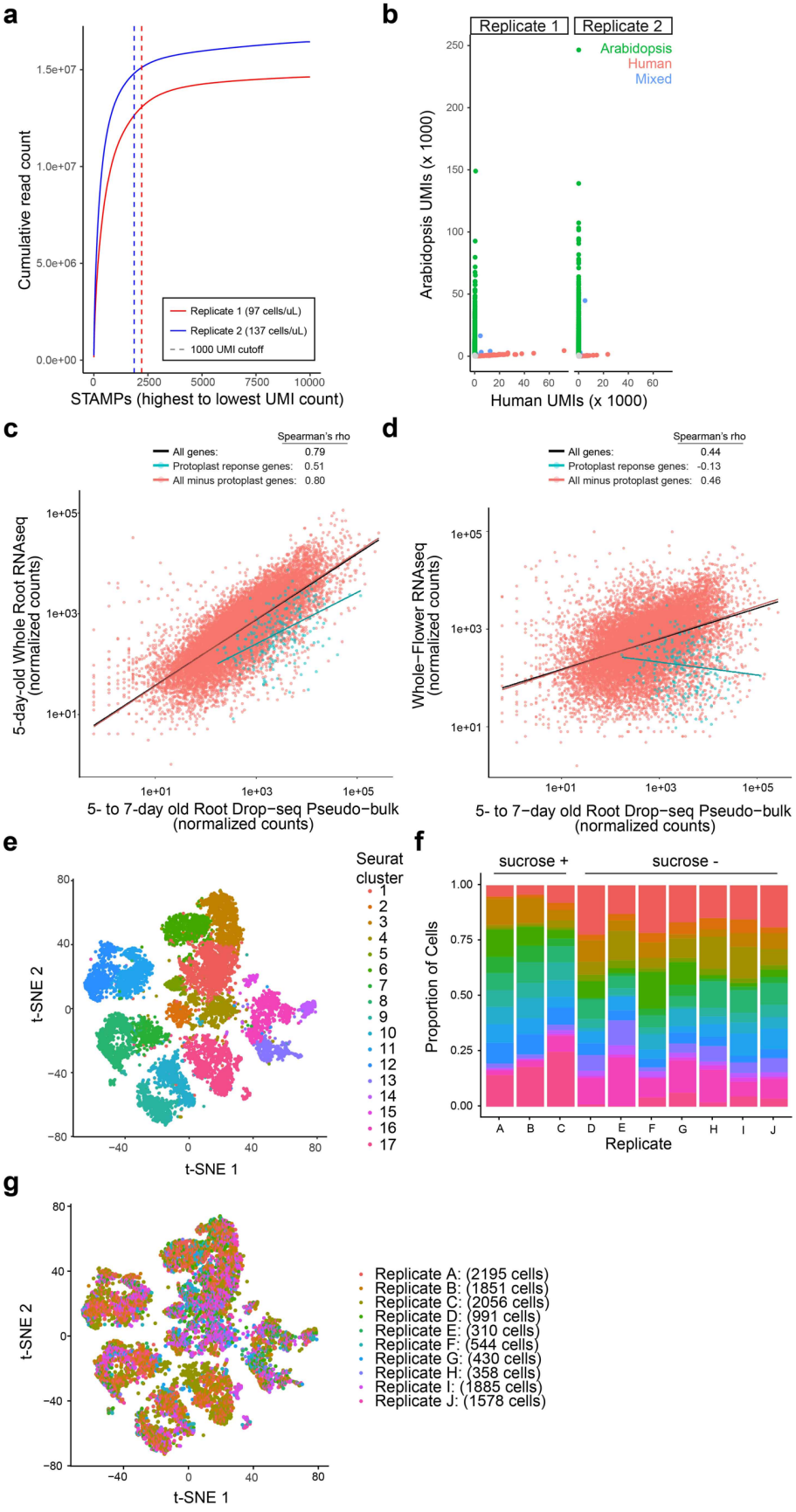
a
Cumulative read count
1.5e+07
1.0e+07
5.0e+06
0.0e+00
0
2500
5000
7500
10000
STAMPs (highest to lowest UMI count)
Replicate 1 (97 cells/uL)
Replicate 2 (137 cells/uL)
1000 UMI cutoff
b
Replicate 1
Replicate 2
Arabidopsis
Human
Mixed
Arabidopsis UMIs (x 1000)
250
200
150
100
50
0
0 20 40 60
0 20 40 60
Human UMIs (x 1000)
c
Spearman's rho
All genes: 0.79
Protoplast reponse genes: 0.51
All minus protoplast genes: 0.80
5-day-old Whole Root RNAseq (normalized counts)
1e+05
1e+03
1e+01
1e+01
1e+03
1e+05
5- to 7-day old Root Drop−seq Pseudo−bulk (normalized counts)
d
Spearman's rho
All genes: 0.44
Protoplast reponse genes: -0.13
All minus protoplast genes: 0.46
Whole−Flower RNAseq (normalized counts)
1e+05
1e+03
1e+01
1e+01
1e+03
1e+05
5- to 7−day old Root Drop−seq Pseudo−bulk (normalized counts)
e
t-SNE 2
80
40
0
−40
−80
−40
0
40
80
t-SNE 1
Seurat cluster
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
f
sucrose +
sucrose -
Proportion of Cells
1.00
0.75
0.50
0.25
0.00
A B C D E F G H I J
Replicate
g
t-SNE 2
80
40
0
−40
−80
−40
0
40
80
t-SNE 1
Replicate A: (2195 cells)
Replicate B: (1851 cells)
Replicate C: (2056 cells)
Replicate D: (991 cells)
Replicate E: (310 cells)
Replicate F: (544 cells)
Replicate G: (430 cells)
Replicate H: (358 cells)
Replicate I: (1885 cells)
Replicate J: (1578 cells)

Figure S1. Five- to seven-day root Drop-seq quality control and unsupervised clustering, related to Figure 1 and STAR Methods. Single-cell RNA-seq by Drop-seq was performed on protoplasts from 5 or 7-day whole root tissue. **a)** Reads per Single-cell Transcriptomes Attached to MicroParticles (STAMPs, representative of cells) from two representative runs, sorted by the number of reads (high to low) versus the cumulative read count for each replicate. The vertical dashed lines indicate the 1000 unique molecular identifier (UMI) cutoff; cell barcodes with at least 1,000 UMIs were considered true STAMPs. **b)** Number of human and *Arabidopsis* UMIs associated with each STAMP captured by Drop-seq from two representative runs. Human cells had been spiked into the *Arabidopsis* samples to estimate capture purity. Green dots indicate individual STAMPs classified as *Arabidopsis* (>98% *Arabidopsis* transcripts), pink dots indicate STAMPs classified as human, and blue dots indicate mixed STAMPs. **c)** Comparison of gene expression between Drop-seq and a bulk root mRNA-seq library (see **STAR Methods**). All individual cell profiles from Drop-seq were combined into a single pseudo-bulk profile. Each point represents a unique gene, with lines showing the correlation between the bulk and pseudo-bulk expression. **d)** Comparison of gene expression between Drop-seq root pseudo-bulk and an mRNA-seq library prepared from bulk unopened floral bud tissue (GEO accession GSM2616967). For **c** and **d**, teal dots show genes known to be induced by protoplasting (Birnbaum et al., 2003). **e)** t-SNE representation of 12,198 single *Arabidopsis* root cells clustered into 17 populations. Each point represents a single cell and is colored by cluster, as assigned by the FindClusters function within the Seurat R package (Satija et al., 2015). **f)** Proportion of cells in each of the 17 clusters by replicate, colored as in **e**. Labels along the top indicate the growth conditions for the plants (sucrose+ or -). **g)** t-SNE representation as in **e**, colored by replicate. Within each growth condition group (sucrose+ or -), cells from all 17 populations are similarly represented among replicates.

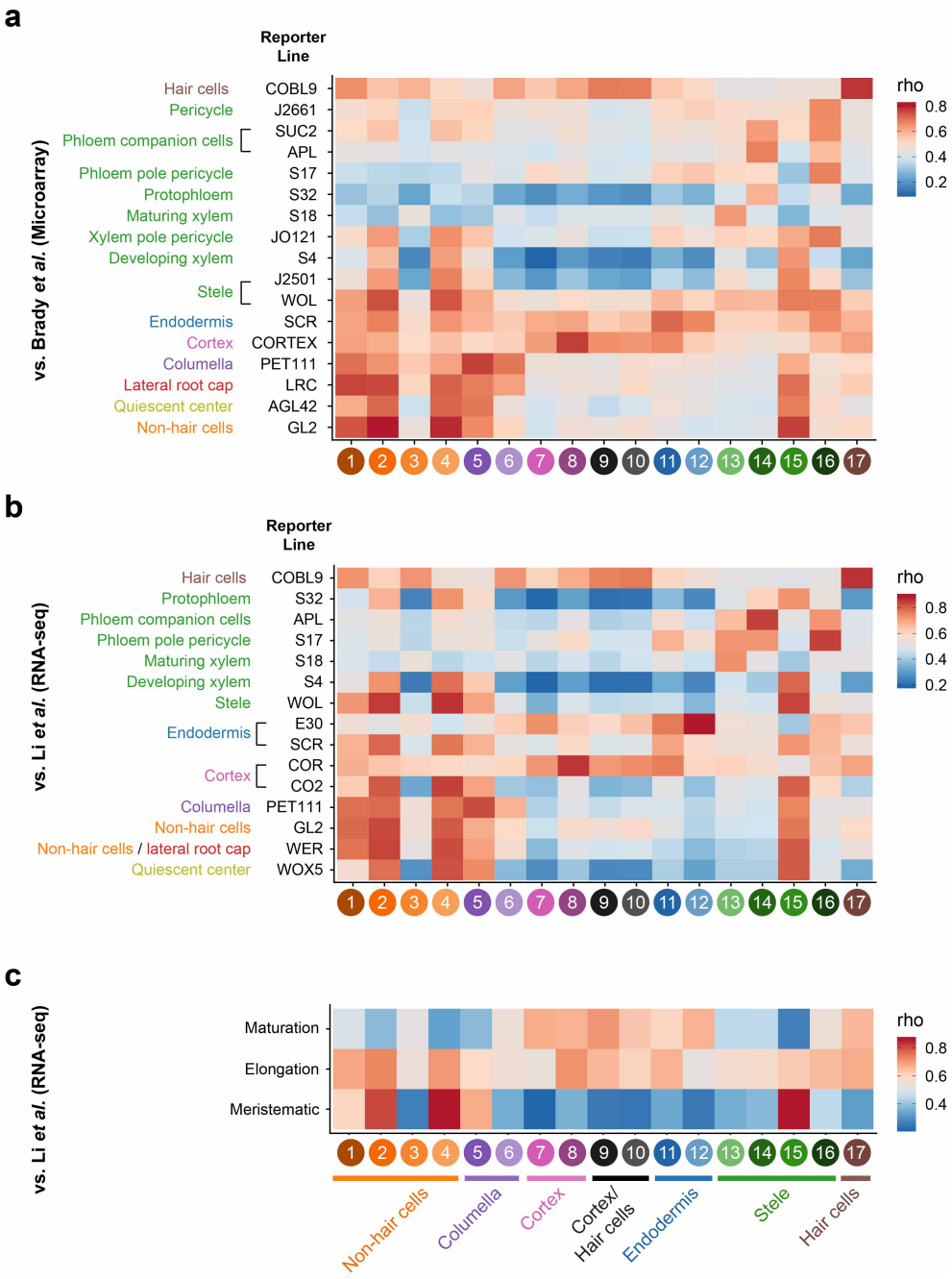

Figure S2. Similarity to transcriptional profiles from reporter lines, related to Figure 1. Heatmaps show the correlation between transcriptional profiles from each Drop-seq-defined cell cluster (x-axis) and transcriptional profiles of cell populations captured from reporter lines or dissected tissue (y-axes). Shown are comparisons to microarray expression data from Brady *et al.* (Brady et al., 2007) (**a**), and RNA-seq data from Li *et al.* (Li et al., 2016) (**b**: cell type transcriptomes, **c**: developmental stage transcriptomes). For **a**-**b**, the text on the left indicates the name of each reporter line and the cell or tissue type the reporter is described to capture.

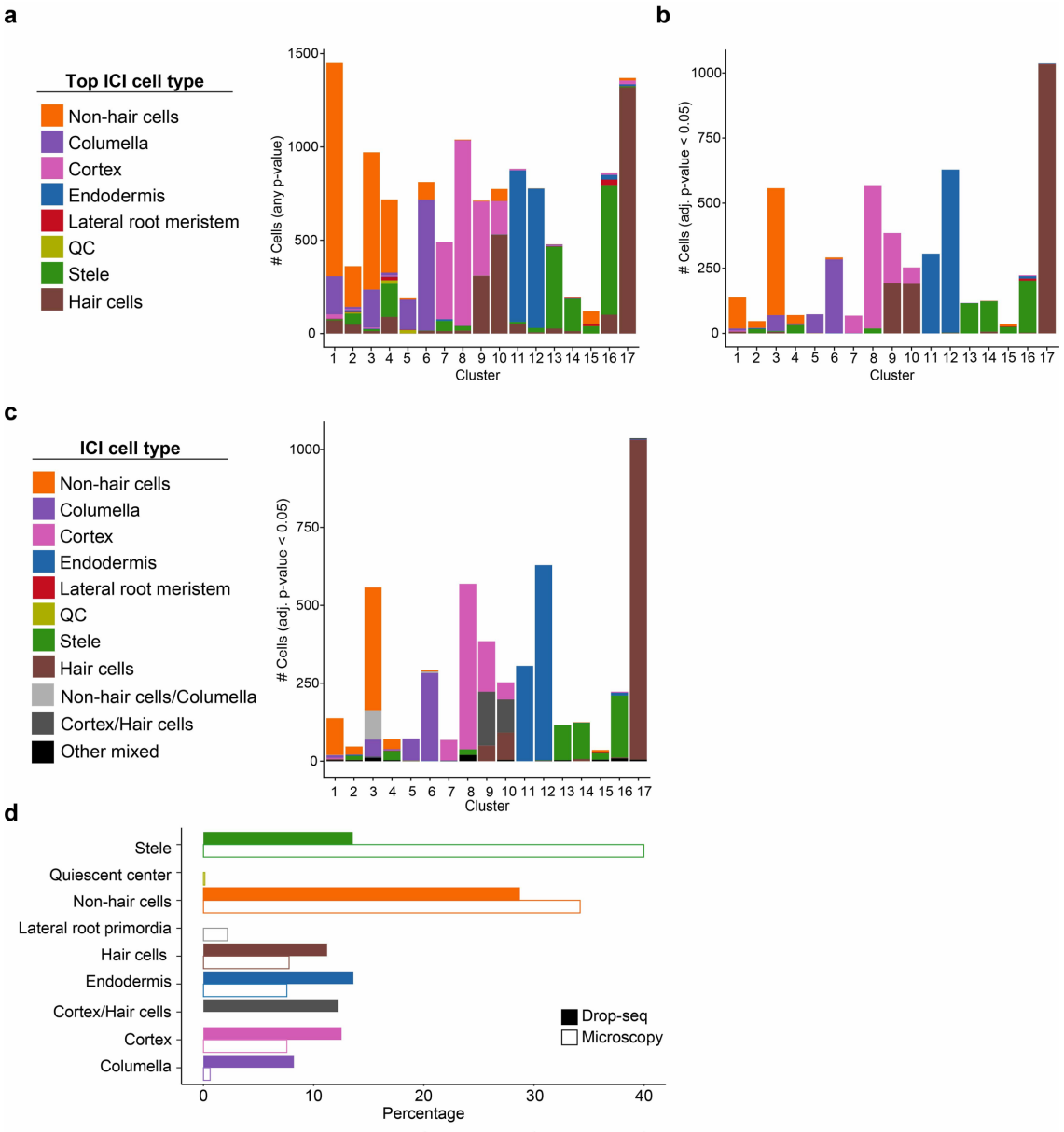

Figure S3. Comparison against existing root cell type classification methods, related to Figure 1. a) Quantification of cells with each ICI classification (Efroni et al., 2015) per cluster. Cluster numbers correspond to the unsupervised Seurat-generated clusters show in **Figure 1b** and **Figure S2**. Bar color-coding is according to ICI cell type, as indicated in the legend. **b)** Same as **a**, showing only those with a statistically significant adjusted *p*-value ($p < 0.05$). **c)** Same as **b** but allowing for cells with mixed identities (more than one identity at an adjusted *p*-value < 0.05). **d)** Percentage of cells recovered by Drop-seq of each type versus microscopy-based surveys (Cartwright et al., 2009). For all comparisons, stele includes pericycle, phloem, xylem, and procambium cells. Lateral root cap is not included in the ICI model, and this cell type was grouped with non-hair cells for microscopy results shown in **d**. The samples from this study (whole root) contain more mature tissue than samples from the microscopy study (root tips to the first lateral root primordium). QC, quiescent center.

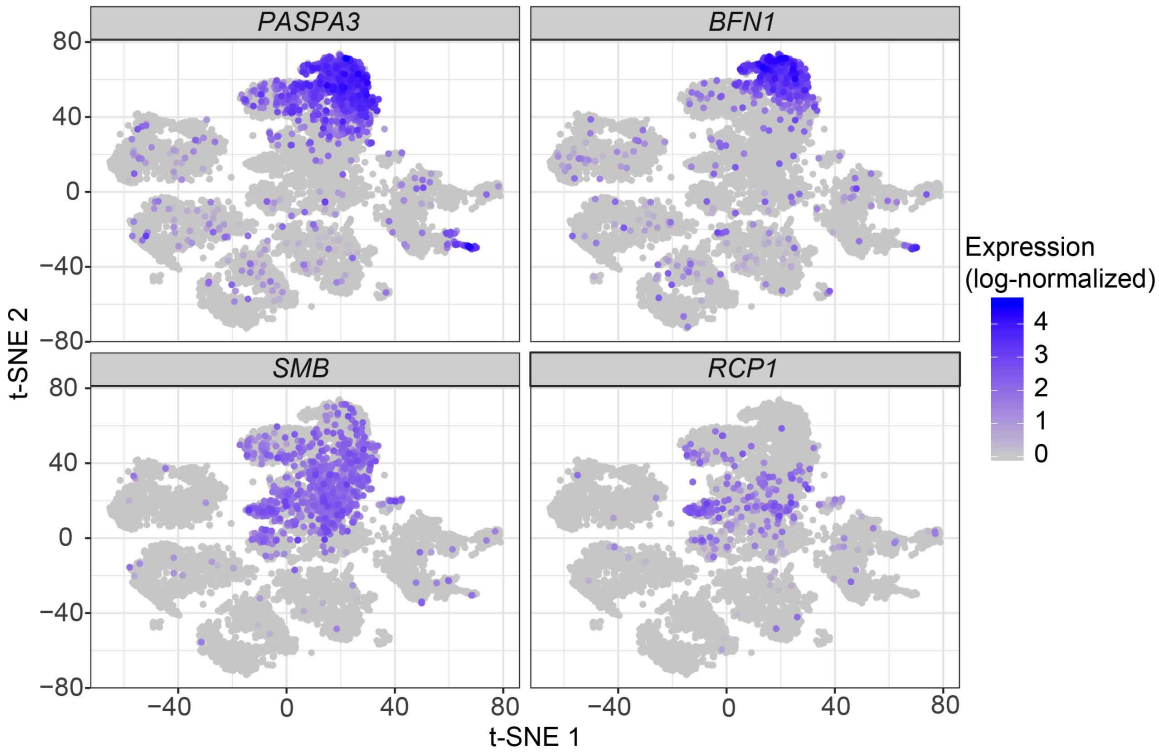

Figure S4. Expression of lateral root cap (LRC)-associated genes, related to Figure 1. t-SNE representations, as in **Figure 1b**, of all captured single-cells showing Seurat-normalized (natural log transform of normalized and scaled counts) expression of the indicated LRC marker genes.

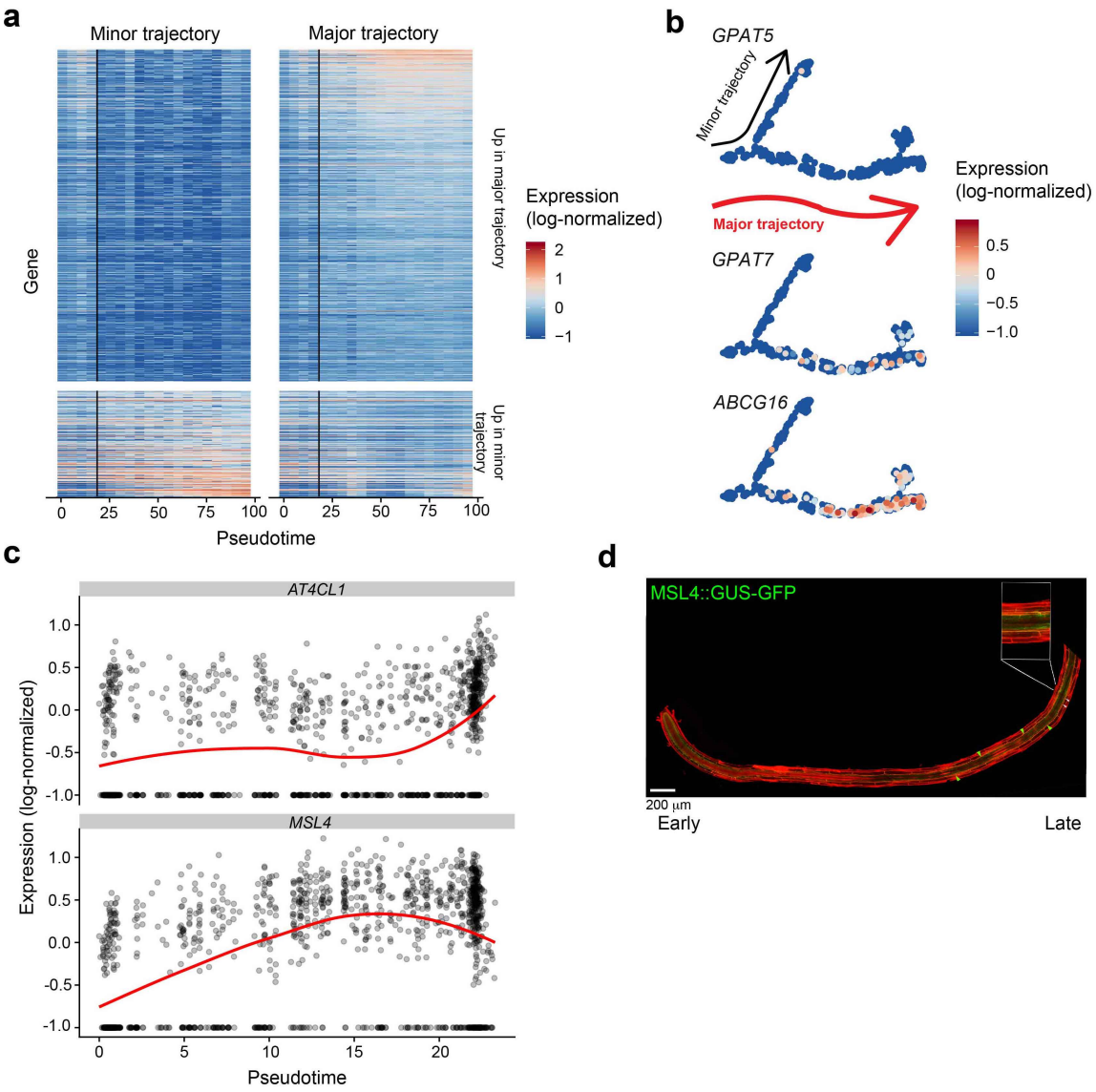

Figure S5. Gene expression changes over endodermis development, related to Figure 5. a) Heat maps showing the log-normalized expression of genes with differential expression between the major and minor branches of endodermis pseudotime (see **Figure 5b** and **STAR Methods**). Vertical black line represents bifurcation point between the two branches. Genes up in the major trajectory are shown in the top, with genes up in the minor trajectory below. **b**) Seurat-normalized expression (calculated using the function, NormalizeData, which computes the natural log of normalized and scaled data) of genes associated with suberin synthesis over endodermis pseudotime. Each point is a unique cell, color coded by expression level of the indicated gene. "Major trajectory" indicates the pseudotime trajectory with the majority of cells. **c)** Lasso regression plots of *AT4CL1* (top) and *MSL4* (bottom) expression over the major endodermis pseudotime trajectory. Points indicate individual cells. **d)** *MSL4* promoter-driven GFP(GUS) in the endodermis of the root. Green arrowheads highlight expression in the late endodermis. White arrowheads indicate zoomed-in region in inset.

Table S1. Drop-seq datasets generated in this study, related to Figure 1.

Library	Age (days)	Growth Conditions	Initial Concentration (cells/µl)	# Arabidopsis STAMPs (cells) captured	# UMIs/STAMP (mean ± standard deviation)	# UMIs/STAMP (median)	# genes with expression information/STAMP (mean ± standard deviation)	# genes with expression information/STAMP (median)
A	5	Sucrose+	97	2195	5607 ± 8313	2622	1796 ±1432	1227
B	5	Sucrose+	137	1851	7515 ± 12686	2941	2008 ±1652	1375
C	7	Sucrose+	56	2056	7698 ± 14146	2976	2118 ± 1659	1497
D	5	Sucrose-	100	991	3450 ± 3375	2391	1529 ± 940	1272
E	5	Sucrose-	100	310	3590 ± 3241	2608	1582 ± 937	1335
F	5	Sucrose-	100	544	2570 ± 2659	1726	1201 ± 770	994
G	5	Sucrose-	100	430	1878 ± 1231	1454	1068 ± 468	919
H	5	Sucrose-	100	358	3347 ± 5586	1918	1445 ± 1060	1119
I	5	Sucrose-	100	1885	3484 ± 5465	2039	1513 ± 1020	1188
J	5	Sucrose-	100	1578	3635 ± 4369	2235	1531 ± 981	1235

Libraries A and B are technical replicates made from the same protoplast preparation. Libraries F and G, and libraries I and J, respectively, are separate protoplast preparations made from the same seedling events. Otherwise, all libraries are independent biological samples. STAMPs: Single-cell Transcriptomes Attached to MicroParticles; UMIs: unique molecular identifiers