A Stroll Through the Developmental Landscape of Plants

Warre Dhondt[1], Michael Van de Voorde[2], and Prof. Dr. Ir. Steven Maere[3]

[1]Universiteit Gent

[2, 3]Evolutionary Systems Biology, VIB-Ugent Center for Plant Systems Biology

March 28, 2023

Contents

1 Results

1.1 Single-cell transcriptomics of the *Arabidopsis* root

The entire analysis used the dataset of a single-cell RNA-Seq (scRNA-Seq) experiment by [**Denyer2019**], who profiled 4727 *Arabidopsis thaliana* root cells. Plant cells have a rigid polysaccharide cell wall that can make it hard to isolate single cells. The principal method to make release single plant cells and make them amenable for scRNA-Seq is to release them from their cell wall by means of enzymatic digestion [**Seyfferth2021**]. To prevent systematic biases during the analysis, 6063 genes that showed significant differential expression in a bulk transcriptomic analysis of plant cell protoplasting [**Denyer2019**] were removed prior to further analysis. Initially, putative high quality cells were identified by filtering the raw library for cells with < 5 percent of genes mapping to mitochondrial and chloroplast genes, as well as ad hoc cutoffs for library size and number of counts. However, mapping the filtered cells to the final processed and annotated dataset via mutual nearest neighbors showed that cells that were removed during QC belonged predominantly to the trichoblast cell type, which would have led to the removal of approximately 30 percent of all trichoblasts (χ^2-test, $p = 2 \times 10^{-4}$). We therefore opted not to filter the cells based on the number of counts and/or features in order to prevent a systematic bias against one cell type during the analysis. The raw count data were normalized using a variance stabilizing regularized negative binomial regression [**Hafemeister2019**]. The resulting library contained information on the expression of 15313 features across 4727 cells. Cells that are similar in terms of gene expression were clustered into 19 groups via the Louvain algorithm on the shared nearest-neighbor graph. To assign cell types to these 19 clusters, the top differentially expressed genes per cluster were mapped to bulk RNA-Seq studies of isolated tissues of the Arabidopsis root ([**Wendrich2020**, **Brady2007**, **Seyfferth2021**]A). The annotated cells were embedded in two dimensions via UMAP for visualization (1B). Central clusters 3, 4, 10 and 16 all mapped to meristematic xylem tissue, but closer inspection of the genes used in this annotation showed that these were predominantly markers for cells with high mitotic/proliferative activity, without presence of markers relating to xylem development. For example, clusters 3, 4 and 16 showed high expression of genes related to DNA replication or nucleolar functionality, including those for histone family proteins H2A (AT1G51060, AT3G54560, AT4G27230) and H2B (AT1G07790), a subunit of the H/ACA complex (AT3G03920) that is involved in pseudouridinylation, and ribosomal proteins including RPL16A (AT2G42740) ([**Bernstein2004**, **Bernstein2007**]), as well as high expression of mitotic regulators *CYCB1;1* and *AUR1* (AT4G32830) in cluster 16, both of which are commonly used cell cycle markers ([**Schnittger2018**, **Weimer2016**]). Cluster 10 additionally showed expression of QC marker GLV6 ([**Fernandez2013**]) as well as GA3, which is involved in GA biosynthesis, a process that has been associated with QC identity ([**Nawy2005**]). Cluster 10 therefore likely consists of QC cells and its closely related initials. The UMAP shows a disconnected star-like structure consisting of a central mass of meristematic cells extending out to different developmental lineages, including trichoblast and atrichoblast lineages with their characteristic expression of *COBL9* (AT5G49270) and *GL2* (AT1G79840), respectively (1C). Additional identified lineages include

the cortex-endodermis differentiation axis and that of the root cap, consisting of columella and lateral root cap cells. Many of the genes that are involved in the regulation of root cap development showed differential expression upon deprotoplasting and were therefore removed during the preprocessing of the data, which is reflected by the fact that the evidence for the identity of these cells is rather low compared to e.g. the cortex-endodermis lineage (**??**). Further, the resolution between the different vascular cell types is rather low. This lack of resolution is a known problem in plant single-cell transcriptomics and can be partly attributed to a high overlap in gene expression between vascular cells and its associated pericycle ([**Parizot2012**]), as well as issues with accessibility, causing to vascular cells to be underrepresented in single-cell atlases ([**Otero2022**]).

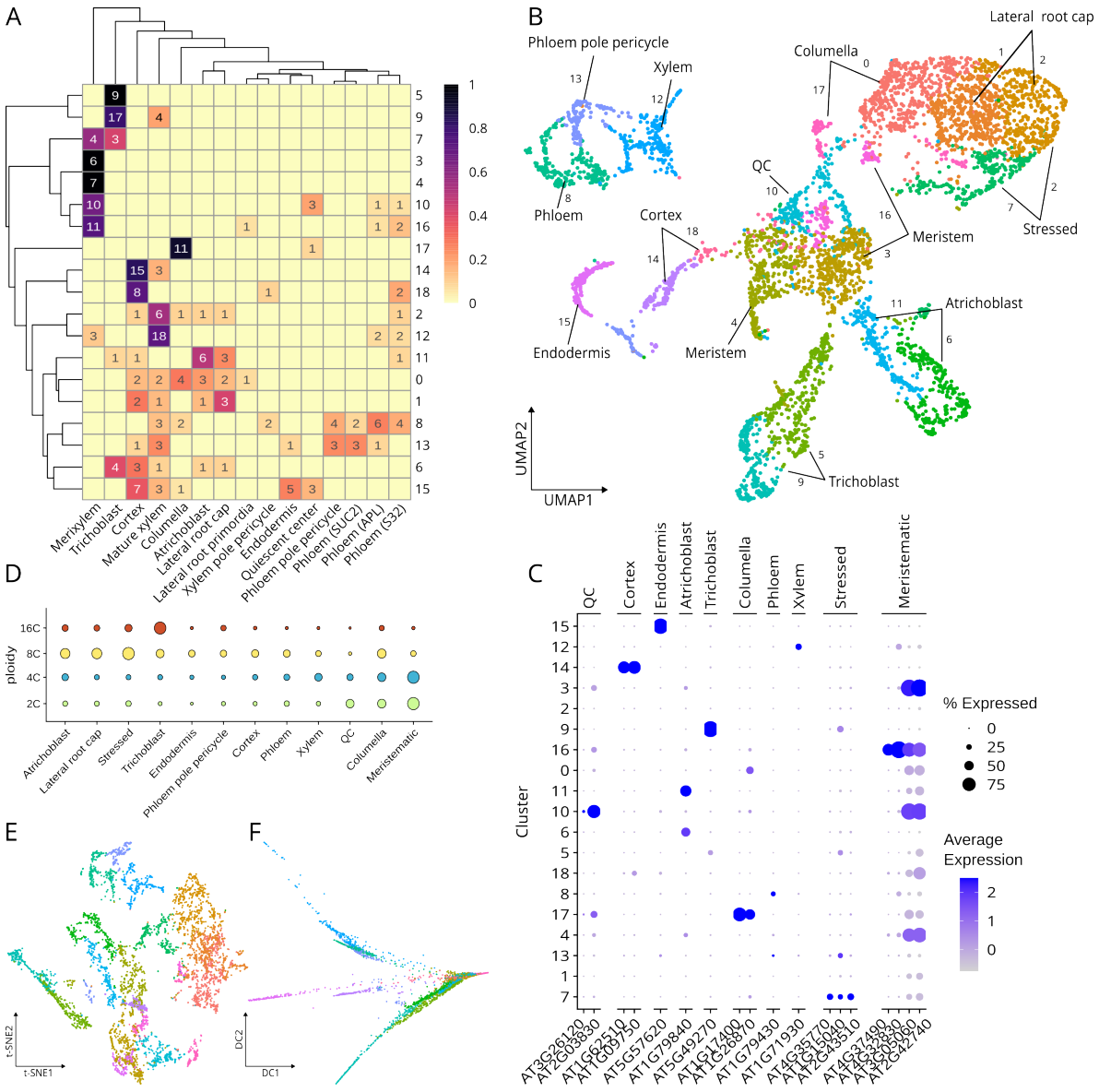

Figure 1: **Cell types of the *Arabidopsis* Root Identified by scRNA-Seq.**
(A) Heatmap showing the contribution of marker genes of different cell types to the cluster identity. The top 50 differentially expressed genes per cluster were mapped to marker genes of bulk RNA-Seq data of plant root tissue sections. The numbers in the matrix represent the number of DE genes of the top 50 of that cluster mapping to a certain tissue. The row colors are the normalized contribution of the cell types to the top DE genes of a given cluster. (B) UMAP of the 4727 *Arabidopsis* root cells. The (sub)clusters are colored and named according to their cell type annotation. Numbers besides the arrows indicate the cluster numbering as referred to in the text. (C) Dot plot showing expression of known cell type marker genes by the annotated cell types in the scRNA-Seq dataset. Dot sizes represent percentage expression in a cell type, color intensity reflects average normalized and scaled expression. (D) Dot plot showing the contribution of different ploidy levels to the cell types identified in the scRNA-Seq dataset. The ploidy level of each cell was identified by correlation with bulk RNA-seq reference expression profiles. (E) and (F): t-SNE and diffusion map, respectively, of the scRNA-Seq dataset. Cells are color-coded according to cell type annotation.