doi:10.1093/plphys/kiab489 PLANT PHYSIOLOGY 2022: 188: 898–918

Plant Physiology®

Research Article

Single-cell transcriptomics sheds light on the identity and metabolism of developing leaf cells

Rubén Tenorio Berrío ,[1,2] Kevin Verstaen ,[3,4] Niels Vandamme ,[3,4] Julie Pevernagie ,[1,2] Ignacio Achon ,[1,2] Julie Van Duyse,[4,5] Gert Van Isterdael ,[4,5] Yvan Saeys ,[3,4] Lieven De Veylder ,[1,2] Dirk Inzé ,[1,2,*,†] and Marieke Dubois [1,2,†]

1 Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium
2 VIB Center for Plant Systems Biology, Ghent, Belgium
3 Department of Applied Mathematics, Ghent University, Computer Science and Statistics, Ghent, Belgium
4 VIB Center for Inflammation Research, Ghent, Belgium
5 Department of Biomedical Molecular Biology, Ghent University, Ghent, Belgium

*Author for communication: dirk.inze@psb.vib-ugent.be
These authors contributed equally (D.I., M.D.).
†Senior authors.

R.T.B., L.D.V., D.I., and M.D. designed the experiments; Y.S. established the single-cell pipeline; R.T.B., I.A., J.V.D., G.V.I., and M.D. performed the experiments; K.V., N.V., and Y.S. analyzed the single-cell sequencing data; R.T.B. and M.D. further analyzed and interpreted the data; R.T.B., J.P., K.V., and M.D. implemented the online tool; D.I. and M.D. supervised the research; R.T.B., D.I., and M.D. wrote the manuscript with the help of all authors.
The author responsible for the distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (https://academic.oup.com/plphys/pages/general-instructions) is: Dirk Inzé (dirk.inze@psb.vib-ugent.be).

Abstract

As the main photosynthetic instruments of vascular plants, leaves are crucial and complex plant organs. A strict organization of leaf mesophyll and epidermal cell layers orchestrates photosynthesis and gas exchange. In addition, water and nutrients for leaf growth are transported through the vascular tissue. To establish the single-cell transcriptomic landscape of these different leaf tissues, we performed high-throughput transcriptome sequencing of individual cells isolated from young leaves of Arabidopsis (*Arabidopsis thaliana*) seedlings grown in two different environmental conditions. The detection of approximately 19,000 different transcripts in over 1,800 high-quality leaf cells revealed 14 cell populations composing the young, differentiating leaf. Besides the cell populations comprising the core leaf tissues, we identified subpopulations with a distinct identity or metabolic activity. In addition, we proposed cell-type-specific markers for each of these populations. Finally, an intuitive web tool allows for browsing the presented dataset. Our data present insights on how the different cell populations constituting a developing leaf are connected via developmental, metabolic, or stress-related trajectories.

Introduction

Plants exhibit remarkable developmental adaptability and plasticity conferred by the postembryonic development of the plethora of tissues composing different plant organs. The plasticity and functionality of plant tissues are, in turn, determined by the orchestrated interplay of distinct cell populations (Dolan et al., 1993; Tsukaya, 2002). The root, for example, is composed of 14 distinct populations (Wendrich et al., 2020), some of which provide the solidity for downwards growth, while others ensure cell production and

Received April 27, 2021. Accepted October 05, 2021. Advance access publication October 23, 2021

growth. Above-ground, the leaf is the primary photosynthetic organ (reviewed in Tsukaya, 2002; Kalve et al., 2014) composed of a diverse set of cell types organized in three main tissues: mesophyll, vasculature, and epidermis (reviewed in Tsukaya, 2013; Supplemental Figure S1). The mesophyll is composed of the irregularly shaped spongy parenchyma cells and the palisade parenchyma cells, located in the abaxial and adaxial sides of the leaf, respectively. As the main leaf photosynthetic tissue, mesophyll cells are characterized by the expression of genes that are crucial for the electron transport chain of the photosystems (Fett and Coleman, 1994; Moroney et al., 2001; Inada and Wildermuth, 2005; Shitov et al., 2018). The mesophyll cells are flanked by the adaxial and abaxial epidermal layers, which shield the internal cell layers from the environment via the production of the cuticula. Both epidermal layers are composed of three very different cell types: pavement cells (PCs), stomata, and trichomes (reviewed in Tsukaya, 2013). PCs and trichomes are not photosynthetically active, but the stomata participate in photosynthesis by mediating carbon dioxide uptake. Lastly, the vascular tissue, mainly composed of phloem and xylem, is embedded by the bundle sheath (BS) within the mesophyll (reviewed in Turner and Sieburth, 2003). Water is transported via the xylem consisting of parenchyma, fibers, and tracheary elements (TEs); elongated, dead cells forming the actual vessel. Parenchyma cells are also found in the phloem, the sugar-transporting tissue, which also contains sieve tube elements accompanied by the closely associated phloem companion cells (CCs).

The specialized leaf tissues are gradually established during leaf development. In Arabidopsis (*Arabidopsis thaliana*), upon its protrusion from the shoot apical meristem at an auxin maximum, a new leaf primordium is composed of undifferentiated cells (Tsukaya, 2013). One of the first identities acquired by the cells is related to their position within the leaf primordium and is critical for the establishment of dorsiventrality. In a simplified model (reviewed in Kidner and Timmermans, 2010), genes expressed on each side of the leaf counteract each other by driving the expression of different small RNAs that target the degradation of transcripts of the opposite side. Following the early events during leaf initiation, the epidermal layers and inner tissues acquire different properties. Young leaves grow in a distal direction, driven by a massive cell proliferation that is coordinated between the different tissues (Andriankaja et al., 2012; Tsukaya, 2013). This exclusive cell proliferation phase is coupled to the establishment of the venation pattern. The emergence of the primary vein is accompanied by the initiation of secondary and higher-order veins and originates asynchronously from mesophyll cells of the inner layer (Biedroń and Banasiak, 2018). Subsequently, epidermal and mesophyll cells gradually exit the mitotic cell division and start differentiation, a process that in Arabidopsis is coupled to the onset of endoreduplication (DNA synthesis without karyokinesis) and cell expansion. The shift from cell proliferation to cell differentiation (further referred to as the transition phase) first occurs in the distal part of the leaf, but gradually progresses to the base of the leaf (Kazama et al., 2010; Andriankaja et al., 2012). The duration of cell proliferation is an important factor in the determination of leaf size and is very sensitive to environmental stimuli. For example, during water limitation, developing leaves display highly dynamic and specific adaptive responses, reprogramming their growth and metabolism depending on the stress level (Claeys and Inzé, 2013).

While in the last decade, an extensive progress in the elucidation of the intrinsic genetic mechanisms of leaf growth and its responses to a changing environment has been achieved, our knowledge about the molecular mechanisms coordinating growth within and between different leaf tissues is limited. Some studies have focused on a specific tissue or leaf cell type, such as a proteome analysis of the three main leaf tissues (Svozil et al., 2015) and the expression analysis of specific genes of laser-dissected epidermal and mesophyll cells (Inada and Wildermuth, 2005). More recently, by tagging the ribosomes of a desired cell type, sequencing of the translatome at organ- or tissue-specific level was reported in plants (Mustroph et al., 2013; Wang and Jiao, 2014; Reynoso et al., 2015; Thellmann et al., 2020). Alternative methods made use of fluorescence-activated nuclear sorting (Slane and Bayer, 2017). Yet, whole-transcriptome studies have been limited to leaf cell types that can easily be isolated, such as trichomes and stomatal guard cells (GCs) (Jin et al., 2013; Yang and Ye, 2013). Recently, important methodological advances have allowed the development of high-throughput single-cell RNA-sequencing (scRNA-seq), offering unprecedented opportunities to explore gene expression profiles at single-cell resolution in multiple tissues simultaneously (Potter, 2018; Rich-Griffin et al., 2020). In the recent years, single-cell and single-nucleus transcriptomics have been successfully used to uncover and profile the different cell types composing an Arabidopsis root tip (Denyer et al., 2019; Jean-Baptiste et al., 2019; Ryu et al., 2019; Shulse et al., 2019; Zhang et al., 2019; Wendrich et al., 2020; Farmer et al., 2021; Long et al., 2021; reviewed in Seyfferth et al., 2021), the Arabidopsis leaf vascular system (Kim et al., 2021), the rice root (Liu et al., 2021), and the maize root and ear (Satterlee et al., 2020; Bezrutczyk et al., 2021), laying the foundations for revealing the transcriptomic landscape of other plant tissues or species.

In this study, we aimed to produce a single-cell transcriptomic atlas of developing leaves and to define the genetic basis of the identity and functions of the different leaf cell types under two different environmental conditions. We performed droplet-based scRNA-seq on individual Arabidopsis leaf cells obtained from plants under well-watered (WW) and mild drought (Dr) conditions during the developmental transition phase, thus, containing both proliferating and differentiating cells. We present a comprehensive transcriptomic map with 14 different cell populations and marker genes identified in each of them. We further

explored the stress sensitivity of each main tissue, laying the basis for targeted follow-up experiments to explore stress-responsive networks at cellular level. Ultimately, we integrated the whole dataset in an open-access online Plant SC Atlas (Wendrich et al., 2020), enabling to explore the leaf single-cell transcriptome, visualize the expression of any gene of interest, find correlated genes, and perform co-expression analyses.

Results

Unsupervised clustering of the single-cell transcriptome of the leaf delineates distinct cell populations

To explore cell heterogeneity during leaf development, we isolated plant cells from young, actively growing, wild-type Arabidopsis leaves. We reasoned that the robustness of our dataset could be increased by harvesting leaves from plants exposed to different growth conditions, and, therefore, applied a WW regime to part of the seedlings, and a lower soil water potential to others (Dubois et al., 2017). For this purpose, plants were grown on an automated watering platform (WIWAM, http://www.wiwam.be; Skirycz et al., 2011) (for details see "Materials and Methods"). Following cell wall digestion, single-cell suspensions were processed to generate scRNA-seq libraries using the Chromium technology (10X Genomics). After data filtering, we retained 1,887 high-quality protoplasts expressing on average (median) 2,017 genes.

To identify preliminary cell populations, the Louvain clustering method (Blondel et al., 2008) was applied for an unsupervised clustering, which was visualized under uniform manifold approximation and projection (UMAP) and t-distributed stochastic neighbor-embedding projections (Amir et al., 2013; McInnes et al., 2018). Because of the limited amount of cells in the dataset, the clustering resolution was kept low (0.9), but still high enough to identify the numerous cell populations that we expected in young leaves. This unbiased clustering approach yielded at this stage a total of 11 distinct cell populations, denominated #0–#10 (Figure 1; Supplemental Figure S2).

Of the sequenced protoplasts, 1,330 came from the WW leaves and 557 from the leaves of plants exposed to lower water potential. To verify that the two environmental conditions are reflected at transcriptome level, differential expression (DE) analysis between both samples was performed. In total, 343 and 305 genes were found significantly down and upregulated by drought, respectively (Supplemental Dataset S1). This included well-known drought-upregulated genes, as aquaporins, *EARLY RESPONSE TO DEHYDRATION* genes (here, *ERD4* and *ERD6*) and *LIPOXYGENASEs* (*LOX2*, *LOX3*, and *LOX4*) as well as genes downregulated when the drought is mild, like *COPPER/ZINC SUPEROXIDE DISMUTASEs* and other oxidative stress genes (Supplemental Dataset S1 and Supplemental Figure S3A; Dubois et al., 2017; Chen et al., 2021). These differences confirmed that the applied watering regimes were effective and slightly altered the leaves transcriptome. Importantly, however, none of the identified cell populations exclusively contained cells of one of both samples (Supplemental Figure S3B). In addition, each of the identified clusters was represented to a similar level in the two samples (Supplemental Figure 3C). Thus, the identified cell populations are robustly conserved across the two tested growth conditions.

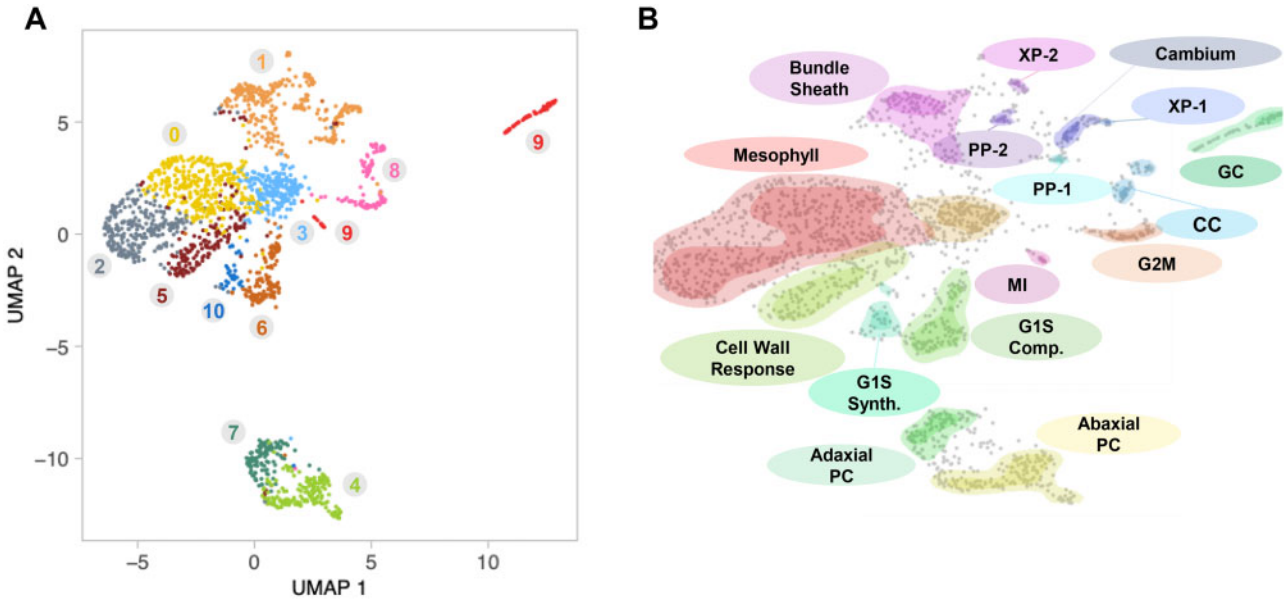

Figure 1 Arabidopsis leaf transcriptomic landscape profiled by scRNA-seq and final annotation. A, Unsupervised clustering distinguished 11 preliminary cell populations (0–10) visualized in a color-coded UMAP plot. In a UMAP plot, the distance between cells correlates with the transcriptomic divergence. As the global structure of the data is maintained, this also holds true for the distance between different clusters. B, Color-coded UMAP plot depicting a schematic view of the final annotation performed in this dataset.

scRNA-seq identifies the main leaf tissues: mesophyll, vasculature, and epidermis

First, we aimed at delineating the populations forming the three main leaf tissues: epidermis, vasculature, and mesophyll (Figure 2). To do so, we visualized the expression in our dataset of marker genes reported in the literature (Table 1). The epidermis is characterized by the expression of genes related to cuticular wax production, such as genes of the *GLYCOSYL-PHOSPHATIDYLINOSITOL-ANCHORED LIPID PROTEIN TRANSFER* and the *3-KETOACYL-CoA-SYNTHASE* (*KCS*) families, encoding enzymes for long-chain lipid biosynthesis (Kim et al., 2013). We visualized the expression pattern of these genes as well as other epidermal markers in the UMAP plot (Figure 2, A and B; Supplemental Figure S4). As a result, we defined clusters #4 and #7 as epidermal populations (Figure 2C). Additionally, cluster #9 was separated into two visibly distinct subclusters (Figure 1A), one of them (top right of UMAP plot) was also defined as epidermal cell population (Figure 2C). Similarly, we used genes known to be expressed in vascular tissue to define the cell populations of the vascular bundle. As the vasculature is a very heterogeneous tissue, multiple reported markers were visualized (Figure 2B; Supplemental Figure S4), each highlighting different small cell populations. The expression of *PHYTOCYSTATIN1* (*CYS1*, AT5G12140) was found to accurately delineate all the vasculature populations (Hwang et al., 2010; Figure 2A). Vascular marker genes were expressed in the subpopulation of cluster #9 without epidermal identity, as well as in parts of clusters #1 and #8, but not in the entire cluster. These clusters were, therefore, split up based on the expression of *CYS1* (Figure 2C). In comparison to the epidermis and vasculature, the mesophyll tissue is highly photosynthetically active. To delineate the mesophyll, we used genes related to photosynthesis, such as genes of the *LIGHT-HARVESTING COMPLEX* or *RUBISCO* subunits (Supplemental Figure S4). However, these genes were not restricted to the remaining, yet unidentified clusters, but were also expressed at the basal level in epidermis and vasculature (Supplemental Figure S5). Therefore, to define the clusters that specifically contain differentiated mesophyll cells, we selected genes that were previously reported to be functional in mature chloroplasts (Figure 2B; Supplemental Figure S4) as well as *EPIDERMAL PATTERNING FACTOR LIKE-9* (AT4G12970), specifically reported in mesophyll cells (Sugano et al., 2010; Figure 2A). This led to the attribution of mesophyll identity to the remaining parts of clusters #1 and #8, which were not identified as the vasculature population, as well as clusters #0, #2, #6, and #10 (Figure 2A). Finally, upon expression analysis of these typical marker genes, clusters #3 and #5 remained unidentified. Both clusters, in general, expressed to very low level many tested genes that were ubiquitously expressed in all other clusters. To unravel their identity, we performed DE analysis, comparing the transcriptome of each of these clusters with that of all other cells. Cluster #3 showed a very low number of significantly enriched genes (Supplemental Dataset S2), with no clear cell type identity emerging from these genes. In addition, no expression of early or late developmental marker genes, such as *ANGUSTIFOLIA3* (*AN3*) or *SIAMESE-RELATED1*, respectively, was observed. We, therefore, considered that these cells did not differentiate into one of the three main tissues at this time-point and further refer to them as "initials" (Figure 2C). In contrast, cluster #5 was enriched for genes related to cell wall metabolism (Supplemental Dataset S2) and shared multiple genes with other differentiated populations, such as *BASIC REGION/LEUCINE ZIPPER MOTIF60* or *FRIABLE1*, which were not present in cluster #3. This suggests that cluster #5 contains cells that are more differentiated than those of cluster #3. Based on the expression of mesophyll markers in cluster #5, albeit to a much lower level than in the mesophyll clusters, and on the position of cluster #5 in the UMAP plot, showing that its transcriptomic profile is more similar to mesophyll cells than to any other cell population (Figure 2C), we included cluster #5 in the mesophyll main tissue.

Next, we analyzed the DE genes in each population (Supplemental Dataset S2) and used them to identify marker genes (for details see "Materials and Methods"). A total of 39, 35, and 10 marker genes were obtained for the mesophyll, epidermis, and vasculature, respectively (Figure 2D; Supplemental Dataset S3). As expected, several additional photosynthesis-related genes were found as good mesophyll-marker genes (Supplemental Dataset S3). Moreover, the *EXTENSIN-LIKE PROTEIN* (*ELP*) and *GERMIN3* (*GER3*) transcripts, previously reported with a possible role in cell expansion and tolerance to stress (Membré et al., 2000), were among the non-photosynthesis-related genes specifically captured in the mesophyll. For the epidermis, *PROLINE-RICH PROTEIN4*, encoding a cell-wall protein with abundant transcripts in aerial plant organs (Fowler et al., 1999), or *SMALLER WITH VARIABLE BRANCHES*, involved in trichome morphogenesis (Marks et al., 2009), appeared to be good marker genes. Finally, *WINDHOSE1*, a downstream effector of *WUSCHEL* and *NOZZLE* in megasporogenesis (Lieber et al., 2011), was revealed as the gene that most accurately delineated the vasculature population in this study (Supplemental Dataset S3). As expected, no marker genes were identified for the undifferentiated cluster of initials in our dataset, as they barely contained any enriched genes.

To validate these results, the transcript profiles of tissue-specific proteins annotated by Svozil et al. (2015) were compared to our dataset (Supplemental Figure S6). Overall, the level of transcripts of the markers of that study corresponded to the main-tissue annotation given in this dataset. Altogether, among the 1,887 profiled cells, 1,137 belonged to the mesophyll population, 349 to the epidermis, 176 to the vasculature, and 225 cells were classified as initials.

Finally, we plotted all these cells against the two molecular principal components (dimensions, Dim.) explaining the diversity of this dataset, Dim. 1 and Dim. 2, which we identified as dimensions related to photosynthetic activity and differentiation, respectively (Figure 2E). This suggests that the annotated mesophyll cells are highly photosynthetic,

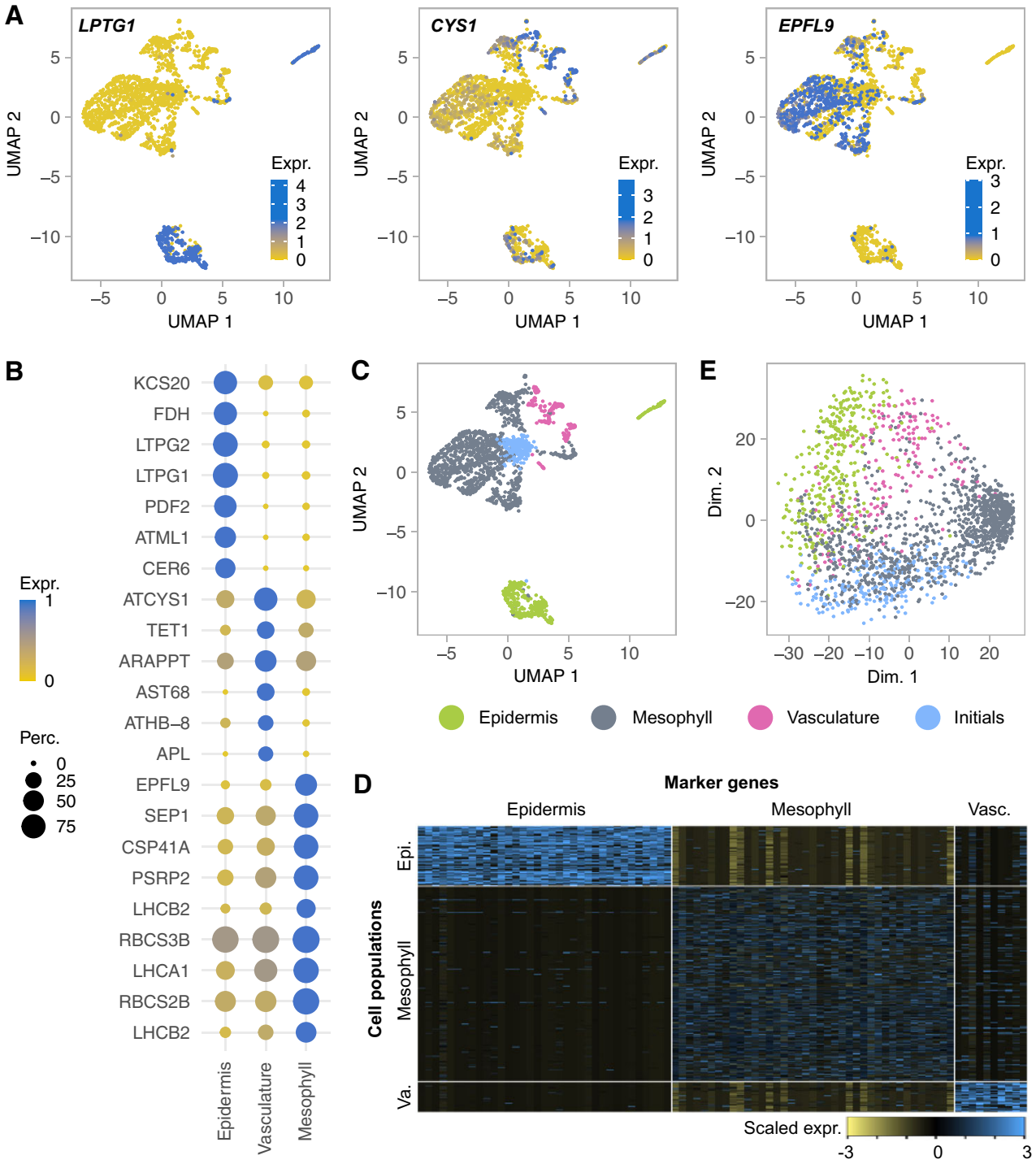

Figure 2 Identification of main leaf tissues from scRNA-seq in developing Arabidopsis leaves. A, UMAP plots showing the normalized expression of selected tissue marker genes for the epidermis (*LPTG1*), vasculature (*CYS1*), and mesophyll (*EPFL9*). B, Dot plot showing the normalized expression of the marker genes used to annotate each tissue. The size of the dots represents the percentage of cells in the clusters that express the gene. The colors represent the relative expression of the gene (yellow = low, blue = maximal expression of the gene). C, UMAP plot of the 1,887 cells colored based on their attribution to a main tissue. D, Heat map illustrating the scaled expression of discriminative genes for each cell. Tissue-specific markers selected for each tissue from this dataset are on the horizontal axis. Vertically, the population (Va., vasculature; Mesophyll; Epi., epidermis) to which each cell was attributed. The colors represent the scaled expression of the gene (yellow = low, blue = high). E, Scatter plot of the 1,887 cells plotted against the first and second main Dimension (Dim.) explaining the largest variance in this dataset. Genes involved in photosynthesis-related functions (based on biological function in GO) mainly contribute to Dim.1, genes involved in differentiation (based on biological function in GO) contribute to Dim.2. *LPTG1 = GLYCOSYLPHOSPHATIDYLINOSITOL-ANCHORED LIPID PROTEIN TRANSFER, (AT)CYS1 = PHYTOCYSTATIN1, EPFL9 = EPIDERMAL PATTERNING FACTOR LIKE9; KCS20 = 3- KETOACYL-CoA-SYNTHASE, FDH = FIDDLEHEAD (=KCS10), PDF1 = PROTODERMAL FACTOR1, ATML1 = MERISTEM LAYER1, CER6 = ECERIFERUM6, TET1 = TETRASPANIN1, ARABIDOPSIS THALIANA PHOSPHATE/PHOSPHOENOLPYRUVATE TRANSLOCATOR, AST68 = ARABIDOPSIS SULFATE TRANSPORTER68, ATHB-8 = ARABIDOPSIS HOMEOBOX8, SEP1 = STRESS-ENHANCED PROTEIN1, CSP41A = CHLOROPLAST STEM-LOOP BINDING PROTEIN OF 41 KDA, PSRP2 = PLASTID-SPECIFIC RIBOSOMAL PROTEIN2, LHCB2/1 = LIGHT-HARVESTING COMPLEX B1/2, RBCS2/3B = RUBISCO SUBUNIT 2/3B.*

Table 1 List of marker genes used to define each population from published literature

Gene	TAIR	Population	Source
LPTG1	AT1G27950	Epidermis	Kim et al. (2013)
CYS1	AT5G12140	Vasculature	Hwang et al. (2010)
EPFL9	AT4G12970	Mesophyll	Sugano et al. (2010)
ESP	AT1G54040	PC	Burow et al. (2007)
YAB5	AT2G26580	Abaxial PC	Tameshige et al. (2013)
CHS	AT5G13930	Adaxial PC	Mahroug et al. (2006)
MYB60	AT1G08810	GC	Rusconi et al. (2013)
HIC	AT2G46720		Gray et al., (2000)
SPCH	AT5G53210		Pillitteri et al. (2007)
YAB1	AT2G45190	Spongy	Uemoto et al. (2018)
TAS3	AT3G17185	Palisade	Chitwood et al. (2009)
H2B	AT1G18370	Late S phase	Cools et al. (2010)
HTA1	AT5G54640		
PCNA1	AT1G07370	Early S phase	
PCNA2	AT2G29570		
NACK1	AT1G18370	G2/M phase	
SULTR2;2	AT1G77990	BS	Kirschner et al. (2018)
SWEET11	AT3G48740	PP1	Chen et al. (2012)
SWEET12	AT5G23660		
AHA3	AT5G57350	Companion	DeWitt and Sussman (1995)
APL	AT1G79430	cell	Bonke et al. (2003)
SEOR1	AT3G01680	SE	Anstead et al. 2012)
SEOR2	AT3G01670		
FLA12	AT5G60490	Xylem	Endo et al. (2018)
ATHB-8,	AT4G32880		Baima et al. (2001)
ACL5	AT5G19530		Muñiz et al. (2008)
PXY	AT5G61480	Cambium	Agusti et al. (2011)
WOX4	AT1G46480		Hirakawa et al. (2010)
CYP83A1	AT4G13770	PP2	Nintemann et al. (2018)
TGG1	AT5G26000	MI	Barth and Jander (2006)
TGG2	AT5G25980		Barth and Jander (2006)

LPTG = GLYCOSYL-PHOSPHATIDYLINOSITOL-ANCHORED LIPID PROTEIN TRANSFER; ESP = EPITHIOSPECIFYING SENESCENCE REGULATOR; ACL5 = ACAULIS 5; PXY = PHLOEM INTERCALATED WITH XYLEM; WOX4 = WUSCHEL-RELATED HOMEOBOX4; EPFL9 = EPIDERMAL PATTERNING FACTOR LIKE-9; HIC = HIGH CARBON DIOXIDE; TAS = TRANS-ACTING SIRNA3.

and that the cells classified as initials are lowly photosynthetic, undifferentiated cells. Interestingly, the long stretch of epidermal and vasculature cells along Dim. 2 suggests that a developmental gradient might be found for these populations, which is what we aimed for by harvesting young leaves in the transition phase.

Developmental and polarity gradients in epidermal cell populations

Three cell populations have been described to constitute the epidermal tissue in Arabidopsis leaves: epidermal PCs, GCs, and trichomes (Glover and Martini, 2000). After enzymatic digestion of the leaves, the large, single-cell trichomes could not be protoplasted, explaining why no cells expressing trichome markers, such as *GLABROUS/GLABRA1* (*GL1*, AT3G27920), *GL2* (AT1G79840), and *GL3* (AT5G41315), were found in our dataset. Subsequently, we aimed to identify which of the epidermal cell populations are the PCs and GCs (Figure 3 and Table 1).

To define the PC population(s), we used the expression of *EPITHIOSPECIFYING SENESCENCE REGULATOR* (AT1G54040), which is transcribed in the epidermal PCs but not in the GCs (Burow et al., 2007). We found this transcript in clusters #4 and #7 (Figure 3B), which contain 171 and 110 cells, respectively. Other DE genes in these clusters include members of the *LTP* and *KCS* families, involved in fatty acid elongation for wax biosynthesis, a process that occurs in the PCs (Kunst and Samuels, 2003; Supplemental Dataset S4). Next, we determined what separated these two clusters by DE analysis, comparing specifically cluster #4 with cluster #7 (Supplemental Dataset S5). We observed that *YABBY* genes, a family of transcription factors involved in determining abaxial cell fate (Tameshige et al., 2013), were significantly enriched in cluster #4 compared to cluster #7 (Supplemental Dataset S5), as illustrated for *YAB5* (AT2G26580) (Figure 3C). In addition, the marker gene *CHALCONE SYNTHASE* (*CHS*, AT5G13930), which is involved in flavonoid biosynthesis and reported to be expressed in the adaxial epidermal layer in *Catharanthus roseus* (Mahroug et al., 2006), was expressed in cluster #7 (Figure 3G). Interestingly, the markers for differentiated PCs were expressed at one side of these populations (Figure 3, B and C), while transcripts present in undifferentiated epidermis, such as photosynthetic gene transcripts, were present on the other side (Supplemental Figure S7). Based on these observations, we defined clusters #4 and #7 as the abaxial and adaxial PC populations, respectively, and a gradient toward a clear epidermal identity for both clusters (Figure 3D). Next, similarly to what we did for the main tissues, we identified polarity marker genes for the abaxial and adaxial PC population, respectively (Figure 3E; Supplemental Dataset S6).

Subsequently, we unraveled the identity of the epidermal part of cluster #9 (Figure 3A). All these cells highly expressed the stomatal markers *MYB60* (AT1G08810) and *HIGH CARBON DIOXIDE* (AT2G46720), and are, therefore, considered GCs (Figure 3F; Supplemental Figure S8C; Gray et al., 2000; Rusconi et al., 2013). An asymmetric cell division of an undifferentiated epidermal cell and successive amplification divisions are required to form a guard mother cell, processes which are regulated by *SPEECHLESS* (*SPCH*, AT5G53210) and *MUTE* (AT3G06120), respectively (Pillitteri et al., 2007; Pillitteri and Dong, 2013; Supplemental Figure S8A). Next, *FAMA* (AT5G53210) promotes GC differentiation as the final step of stomata formation (Pillitteri et al., 2007). Interestingly, in a subset of abaxial and adaxial PCs, we captured the expression of *SPCH* and *MUTE* along with other genes, such as *BREAKING OF ASYMMETRY IN THE STOMATAL LINEAGE* (AT5G60880) and *POLAR* (AT4G31805), associated with cellular asymmetry of meristemoids (Facette and Smith, 2012; Houbaert et al., 2018; Supplemental Figure S8, B and C). In addition, a differential gene expression analysis in the few cells that express *MUTE* revealed transcripts of other genes regulating asymmetric division such as *CYCLIN D7;1* (*CYCD7;1*, AT5G02110) and *MUSTACHES* (AT1G75640) (Keerthisinghe et al., 2015; Weimer et al., 2018). Cells expressing *FAMA* were found throughout the entire GC cluster

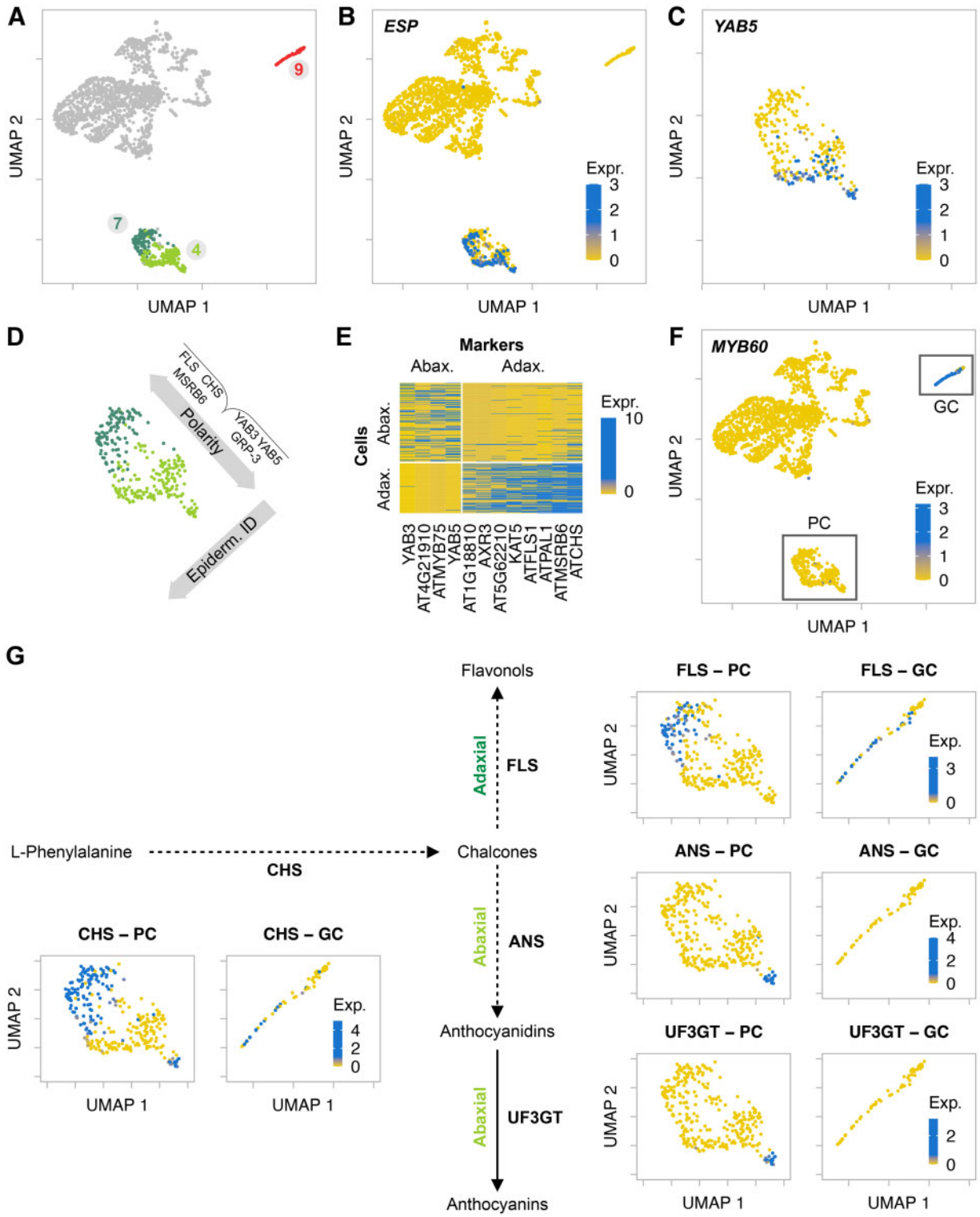

Figure 3 Identification of PC and GC populations. A, Highlights of the epidermal clusters in the UMAP plot. B, UMAP plot with the normalized expression of the epidermal marker gene *ESP*. C, UMAP plot with the normalized expression of the abaxial polarity marker *YAB5*. D, Schematic representation of the PC populations, the polarity gradient, and some polarity-specific marker genes, as well as the differentiation gradient toward clear epidermal identity (ID). E, Heat map illustrating the expression level of polarity-specific genes identified in this study. On the vertical axis, the annotated cells from the PC clusters. On the horizontal axis, the tissue-specific markers from this dataset. The colors represent the normalized expression of the gene (blue = high, yellow = low). F, UMAP plot with the normalized expression of the GC marker *MYB60*. G, Tracking the anthocyanin and flavonol biosynthetic pathways in the GC and PC populations indicated in (F). The UMAP plot shows the normalized expression of selected genes encoding enzymes for anthocyanin and flavonol biosynthesis. C and G, Full UMAP plots can be found in Supplemental Figure S28., *YAB5 = YABBY5, (AT)CHS = CHALCONE SYNTHASE, (AT)FLS = FLAVONOL SYNTHASE, (AT)MSRB6 = METHIONINE SULFOXIDE REDUCTASE B6, GRP-3 = GLYCINE-RICH PROTEIN3, KAT5 = 3-KETO-ACYL-CoA THIOLASE, ATPAL1 = PHE-AMMONIA LYASE1ANS = ANTHOCYANIDIN SYNTHASE, UF3GT = UDP-GLUCOSE: FLAVONOID 3-O-GLUCOSYLTRANSFERASE.*

(Supplemental Figure S8B). Finally, *SCREAM1* and *2* (*SCRM1*, AT3G26744; *SCRM2*, AT1G12860), which both interact with *SPCH*, *MUTE*, and *FAMA* (Kanaoka et al., 2008), were expressed in the cells in which these three transcription factors were expressed, further supporting the expression pattern. Altogether, the expression of well-described stomatal lineage genes allowed us to trace the stomatal lineage across the GC population (Supplemental Figure S8B).

Finally, our attention was attracted to a small group of abaxial PCs that expressed the typical adaxial PC marker gene *CHS* (Figure 3G). As *CHS* encodes one of the enzymes of the flavonols and anthocyanins biosynthesis, we visualized the expression of other genes of this pathway (Figure 3G). Interestingly, the gene encoding the enzyme acting downstream of CHS to convert the CHS product into flavonols, *FLAVONOL SYNTHASE* (AT5G08640), was expressed in the adaxial PC and GC population, but not in the abaxial cells (Figure 3G). In contrast, genes encoding enzymes acting downstream of CHS for the biosynthesis of anthocyanins, such as *ANTHOCYANIDIN SYNTHASE* (AT4G22880) or *UDP-GLUCOSE:FLAVONOID 3-O-GLUCOSYLTRANSFERASE* (AT5G54060), were exclusively expressed in this abaxial PC subpopulation (Figure 3G). Together, these expression patterns suggest that chalcone, as an intermediary product in both pathways, is synthesized at both sides of the leaf, while more downstream the production of flavonols and anthocyanins is restricted to the adaxial side and a small cluster of abaxial cells, respectively.

Mesophyll populations are defined by stress level, cell division stage, and polarity

Embedded between the abaxial and adaxial epidermis, mesophyll cells are defined as all photosynthetic cells enclosed by these two layers. Based on the expression of photosynthetic genes and mesophyll differentiation markers, the clusters #0, #2, #5, #6, #10, and a part of clusters #1 and #8 were previously identified as the mesophyll populations (Figure 4 and Table 1). To uncover the identity of each of these clusters, we performed a differential gene expression analysis by comparing the transcriptome of each of these clusters to the rest of the dataset.

Clusters #0 and #2 both seemed to be composed of the typical photosynthetic cells in the mesophyll, as they were almost exclusively defined by genes related to the photosynthetic activity (Supplemental Dataset S4). We visualized the expression of *YABBY1* (*YAB1*, AT2G45190), a known spongy parenchyma marker gene (Uemoto et al., 2018) involved in abaxial cell-type specification, and of *TRANS-ACTING SIRNA3* (*TAS3*, AT3G17185), encoding a regulator of auxin response factor genes at the adaxial side of the mesophyll (Chitwood et al., 2009; Figure 4, B and C). Notably, the expression pattern of these genes did not overlap with that of the delineated clusters. Instead, *YAB1* was expressed in the upper part of both clusters #0 and #2, while *TAS3* transcripts were present at the lower side, also in both clusters (Figure 4, B and C). Although these results suggest a polarity gradient along the vertical axis of the UMAP plot, it is not the major determinant of the unsupervised clustering in the mesophyll population, which separated cluster #0 from cluster #2. To explore what does make clusters #0 and #2 different, we performed a differential gene expression analysis, comparing specifically cluster #0 with cluster #2. Gene ontology (GO) enrichment analysis of the top 100 DE genes (Supplemental Dataset S7) between these two clusters suggested that cells in the cluster #2 are responsive to different kinds of stresses: GO terms like "response to chitin," "wounding," "hypoxia," "fungal response," and "jasmonic acid biosynthetic process" were enriched. In addition, visualization of jasmonic acid (JA)-related genes, like *JASMONATE-ZIM DOMAIN 5* (AT1G17380), supported this (Figure 4D). Therefore, clusters #0 and #2 were labeled, together, as the spongy and palisade mesophyll population, which is characterized by a stress gradient and, to a lower extent, a polarity gradient.

Next, clusters #6, #8, and #10 were enriched for genes related to cell division, suggesting that these clusters group mesophyll cells that are undergoing division. Interestingly, the expression of G2/M phase marker genes was restricted to cluster #8, suggesting that this population is undergoing mitotic cell division, while, as seen by the expression of G1/S-specific genes, both cluster #6 and #10 contained cells going through different stages of DNA replication in the mitotic cell cycle or endocycle (Figure 4G). This was further supported by *HISTONE* genes that were consistently expressed in the G1/S clusters (#6 and #10) (Figure 4E) and by cytokinesis-related genes (e.g. *NPK1-ACTIVATING KINESIN 1* (AT1G18370)) expressed in the G2/M cluster #8 (Figure 4F). Subsequently, several genes related to different cell cycle checkpoints enabled to make the distinction between clusters #6 and #10. Cluster #6 was enriched in chromatin organization or nucleosome assembly genes, such as *HISTONE H2B* (AT5G22880) and *H2A* (*HTA1*, AT5G54640) (Supplemental Figure S9). On the other hand, cluster #10 cells were enriched in genes related to DNA biosynthesis and strand elongation, such as *PROLIFERATING CELLULAR NUCLEAR ANTIGEN 1* and *2* (*PCNA1*, AT1G07370; *PCNA2*, AT2G29570) (Supplemental Figure S9). Altogether, these results suggest that dividing mesophyll cells that are in the progress of DNA synthesis are grouped in cluster #10. Cluster #6 groups mesophyll cells that are compacting the synthesized DNA, which occurs in a later stage of the S phase (Cools et al., 2010). Finally, mesophyll cells that are undergoing cytokinesis are grouped in cluster #8. Thus, while the cells of these three clusters express typical spongy and palisade mesophyll marker genes, the cell division phases appeared to be the major determinants for these clusters, dominating over their identity of spongy or palisade mesophyll cell type.

Subsequently, we explored the identity of cluster #1. As this cluster was close to the vasculature tissue, we hypothesized that it could contain the BS cells, chloroplast-containing cells that surround the vascular bundle. We

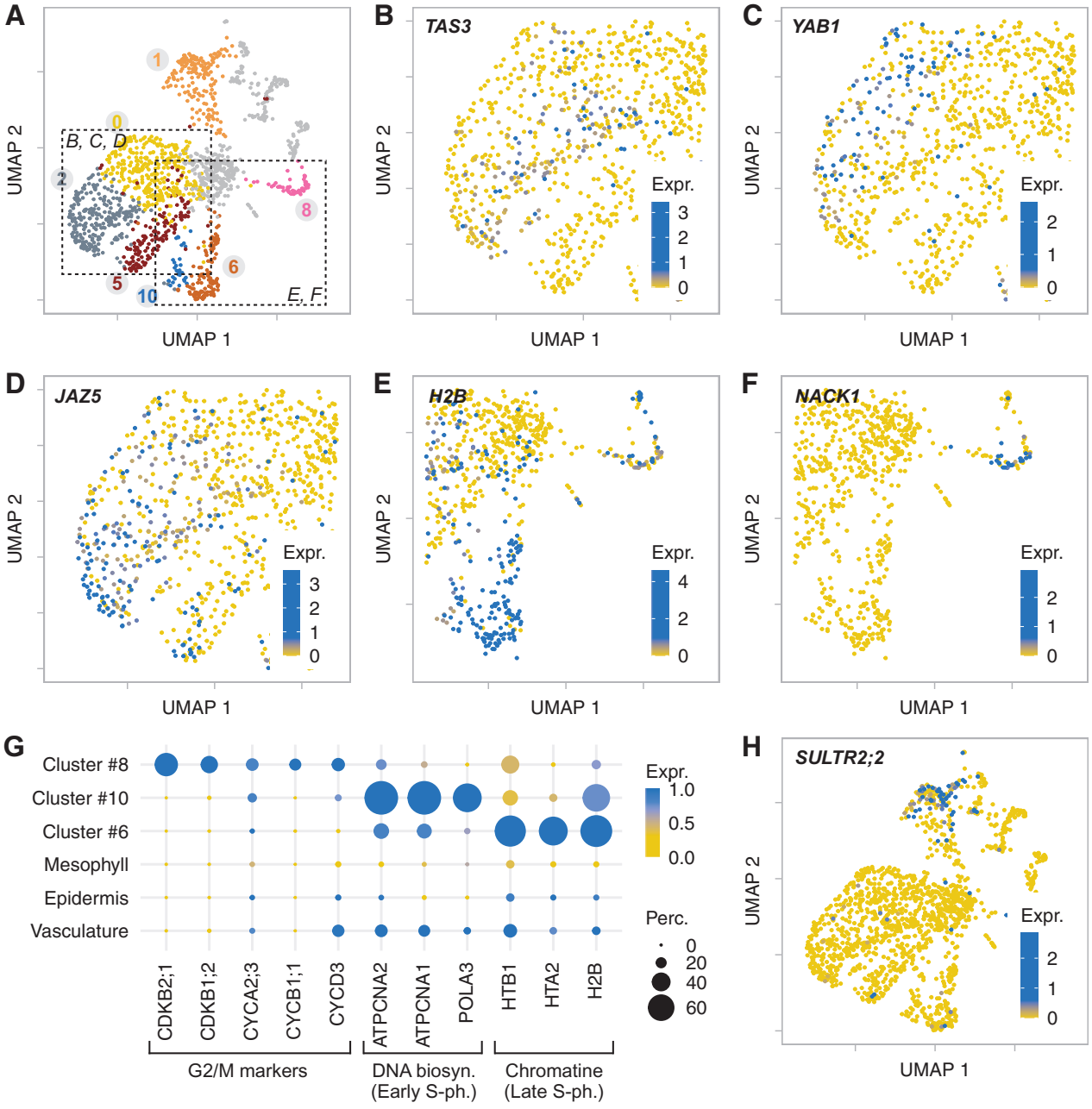

Figure 4 Identification of the mesophyll cell populations. A, Highlights of the mesophyll clusters in the UMAP plot. The dashed-line rectangles represent the zones on which the indicated panels are zoomed in. B, Close-up view (see A) of the UMAP plot with the normalized expression of the adaxial spongy mesophyll marker gene *TAS3*. C, Close-up view (see A) of the UMAP plot with the normalized expression of the abaxial palisade mesophyll marker gene *YAB1*. D, Close-up view (see A) of the UMAP plot with the normalized expression of stress-related gene *JASMONATE-ZIM DOMAIN 5*. E, Close-up view (see A) of the UMAP plot with the normalized expression of the histone gene *H2B*. F, Close-up view (see A) of the UMAP plot with the normalized expression of the cell division marker gene *NACK1*. G, Dotplot with marker genes of the G2/M cell cycle phase transition, or specific subphases of S-phase (DNA biosynthesis in early S-phase or chromatine compacting in late S-phase). The color of the dots represents the average expression in the cells expressing the gene, whereas the size of the dots represents the percentage of cells expressing the gene. H, UMAP plot centered on the mesophyll clusters with the normalized expression of the BS marker gene *SULTR2;2*. B–F, Full UMAP plots can be found in Supplemental Figure S28. *NACK1 = NPK1- ACTIVATING KINESIN1, CDKB2;1/1;2 = CYCLIN-DEPENDENT KINASE B2;1/1;2, ATPCNA1/2 = PROLIFERATING CELLULAR NUCLEAR ANTIGEN1/2, POLA3 = DNA PRIMASE POLA, HTB1 = HISTONE B2, HTA2 = HISTONEA2.*

visualized the expression of the BS (and vasculature) marker *SULPHATE TRANSPORTER 2;2* (*SULTR2;2*, AT1G77990) (Kirschner et al., 2018). Here, we found *SULTR2.2* to be specifically expressed in the mesophyll part of cluster #1, thus this subpopulation was defined as BSs (Figure 4H). Finally, as mentioned earlier, cluster #5 was characterized by the expression of genes related to cell wall modifications and cytokinesis (Supplemental Dataset S4). Whether this

population is the result of cell-wall digestion by protoplasting, or a group of cells undergoing cell wall modifications in the leaf, remains unclear. In conclusion, scRNA-seq allowed not only to distinguish two mesophyll clusters along a different stress and polarity gradient, but also to discriminate mesophyll cells undergoing three distinct stages of the cell cycle.

Diverse cell populations form the main vasculature tissues and provide spatial separation during glucosinolate metabolism

Embedded in the mesophyll, the vasculature is responsible for bidirectional transport and distribution of molecules to the destination cells. The vasculature is a very heterogenic tissue and, hence, we expected to identify multiple different clusters, including xylem parenchyma (XP), phloem parenchyma (PP), phloem CCs, with sieve elements (SEs) and TEs, and xylem parenchyma (XP) (Figure 5 and Table 1).

First, we aimed at delineating the identity of the PP and the conducting cells of the phloem. For the PP, we visualized the expression of the genes *SWEET11* (AT3G48740) and *SWEET12* (AT5G23660), encoding sucrose efflux transporters (Chen et al., 2012), and defined a lower group of cells in the center of the vasculature population as the PP (Figure 5A). The CCs associate with SEs, two cell types that express common markers during early development, such as $H^{(+)}$-*ATPASE 3* (*AHA3*, AT5G57350) and *ALTERED PHLOEM DEVELOPMENT* (*APL*, AT1G79430) (DeWitt and Sussman, 1995; Bonke et al., 2003; Otero and Helariutta, 2016). Visualizing the expression of these genes highlighted a group of cells belonging to cluster #8 (Figure 5A). Another group of cells expressed *SODIUM POTASSIUM ROOT DEFECTIVE 1* (*NaKR1*, AT5G02600), *SUCROSE-PROTON SYMPORTER* (*SUC2*, AT1G22710), and *PHLOEM PROTEIN 2-A1* (AT4G19840), typical CC markers; this part of the cluster could be considered as the most differentiated CC population (Truernit and Sauer, 1995; Tian et al., 2010; Cayla et al., 2015; Figure 5A). Interestingly, a small group of cells in the CC cluster expressed the typical SE marker genes *SIEVE ELEMENT OCCLUSION-RELATED 1* (*SEOR1*, AT3G01680) and *2* (*SEOR2*, AT3G01670) (Anstead et al., 2012) and were therefore considered as differentiating SE cells (Figure 5A). Finally, within cluster #8 and included in the CC population, we found very few cells expressing the typical TE differentiation markers involved in secondary cell wall biosynthesis, like *CELLULOSE SYNTHASE 8* (AT4G18780) (Schuetz et al., 2013) or *COBRA-LIKE4* (AT5G15630) (McNair, 2015; Figure 5A).

Next, the xylem was delineated by the expression pattern of the xylem marker genes *FASCICLIN-LIKE ARABINOGALACTAN-PROTEIN 12* (AT5G60490), *HOMEOBOX GENE 8* (*ATHB-8*, AT4G32880), and *ACAULIS 5* (AT5G19530) (Baima et al., 2001; Muñiz et al., 2008; Endo et al., 2018; Figure 5A). This highlighted a group of cells of cluster #1, that we defined as the XP-1. Lastly, although in Arabidopsis the production of secondary vascular tissue in leaves is very rudimentary (Yu et al., 2010), we used two cambium markers (*PHLOEM INTERCALATED WITH XYLEM* (AT5G61480) and *WUSCHEL-RELATED HOMEOBOX4* (AT1G46480)) to confirm that the cells located between XP and PP were likely containing the cambium population (Hirakawa et al., 2010; Agusti et al., 2011; Figure 5A).

Once the main vasculature populations were identified using the known marker genes, a population of vasculature cells close to the BS remained unexplored. We performed a differential gene expression analysis on this population, which revealed a plethora of genes related to glucosinolate (GSL) metabolism. Two major types of GSL, aliphatic and indole GSL, serve different biological functions and are synthetized in spatially separated cell populations of PP (Figure 5B; Nintemann et al., 2018). We found expression of the cytochrome *CYP83B1* (AT4G31500), involved in indole GSL biosynthesis, along with the PP, as well as in the BS. Interestingly, transcripts of the *CYP* gene involved in aliphatic GSL biosynthesis, *CYP83A1* (AT4G13770), highlighted another PP population that was separated from the conducting phloem vessels (Figure 5B; Nintemann et al., 2018). We, therefore, labeled this as the PP-2 population (and renamed the previous one "PP-1"; Supplemental Figure S10A). Genes expressing GSL hydrolytic enzymes, the myrosinases, are expressed in GCs and specialized cells, referred to as myrosin idioblasts (MIs): upon wounding, GSLs, and myrosinases are mixed (Xue et al., 1995; Koroleva et al., 2000; Andréasson et al., 2001; Li and Sack, 2014). We used the expression pattern of two genes encoding myrosinases, *THIOGLUCOSIDE GLUCOHYDROLASE 1* and *2* (*TGG1*, AT5G26000; *TGG2*, AT5G25980) to identify the vascular part of cluster #9 (Figure 5A) as the MIs (Barth and Jander, 2006; Figure 5B). *TGG1* was expressed in GCs and MIs, while *TGG2/BGLU37* was found exclusively in MIs (Figure 5B; Supplemental Figure S10A).

Next, we studied the identity of a small population of cells located close to the PP-2 cluster (Figure 5A). A DE analysis (Supplemental Dataset S4) revealed genes involved in sulfate and GSL transport, particularly *SULTR2;1* (AT5G10180) and *GSL TRANSPORTER 2* (*GTR2*, AT5G62680), which were previously described to be expressed in leaf veins. Additionally, this population was enriched in genes involved in lignin biosynthesis, such as several members of the *4-COUMARATE: COA LIGASE* (*4CL*) gene family or *TRANSALDOLASE 2* (AT5G13420) (Vanholme et al., 2013; Supplemental Figure S10A). Interestingly, XPs were reported to provide monolignols to adjacent TEs (Smith et al., 2017). We, therefore, speculated that this population contains XPs, without typical xylem identity, involved in TE lignification; we called this population XP-2. Altogether, we identified nine putative cell populations within the vasculature tissue (Supplemental Figure S10B). The hypothetical spatial organization of vasculature clusters, which were mainly defined by the expression of genes involved in GSL biosynthesis, storage, or breakdown, is depicted in Supplemental Figure S10C.

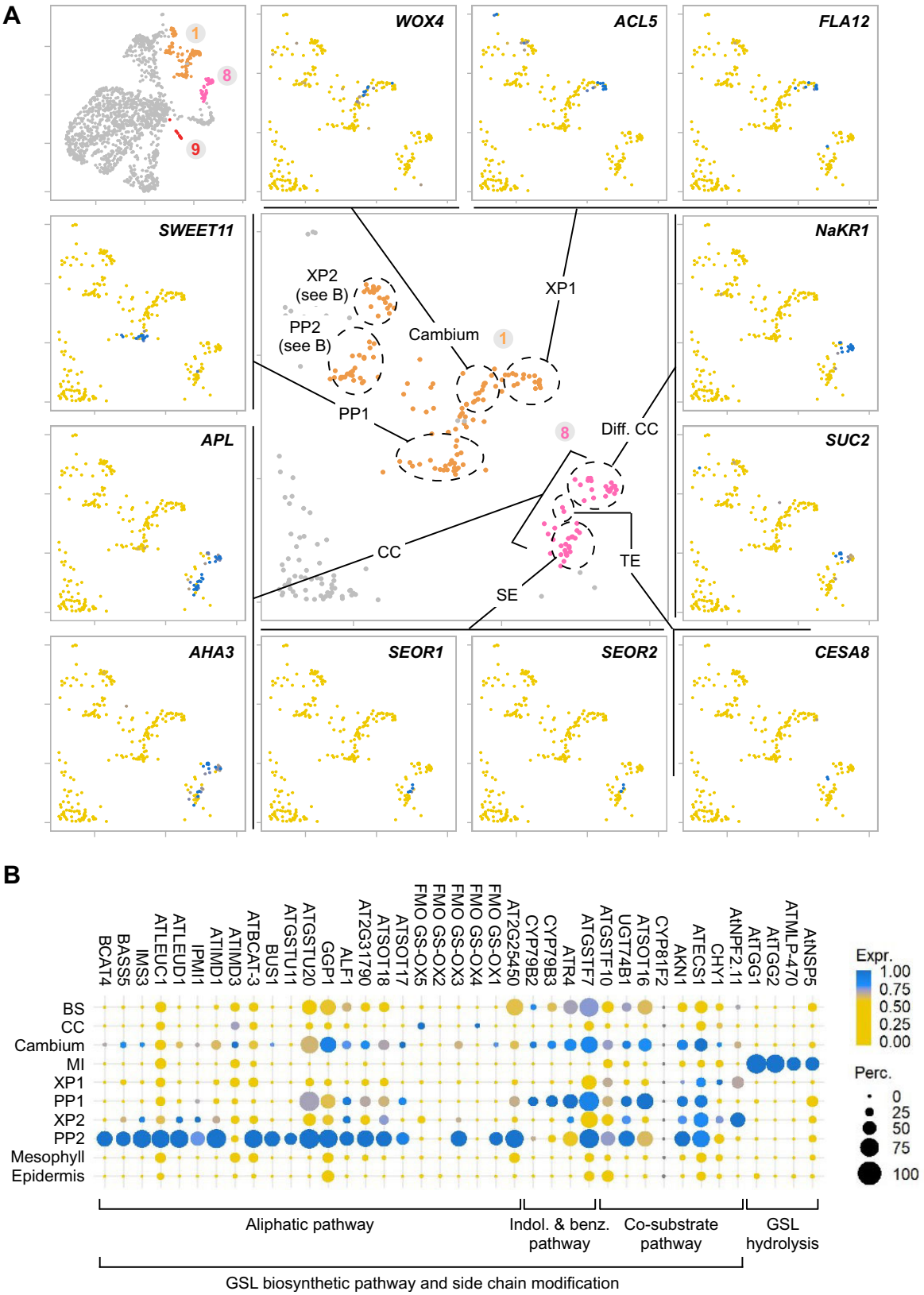

Figure 5 Identification of vascular cell populations and their role in the glucosinolate biosynthesis pathway. A, On the top-left panel, the clusters classified as the vascular tissue are highlighted. The middle panel represents a close-up view of clusters #1 and #8 with indication of the subpopulations of the vascular tissue. These were identified based on the expression of the marker genes depicted in the surrounding panels or (for XP-2 and PP-2) based on the expression patterns shown in (B). Different UMAP plots (close-up on clusters #1 and #8) show normalized expression of the genes (blue = high, yellow = low). Full UMAP plots can be found in Supplemental Figure S28. *X*- and *Y*-axis depict UMAP 1 and 2, respectively. B, Dot plot showing the normalized expression of genes involved in glucosinolate hydrolysis and biosynthesis. The size of the dots represents the percentage of cells in the clusters that express the gene. The color represents the relative expression of the gene (yellow = low, blue = maximal expression of the gene). *ACL5* = *ACAULIS5*, *FLA12* = *FASCICLIN-LIKE ARABINOGALACTAN-PROTEIN12*.

The mesophyll transcriptome is most sensitive to low water potential

As we built the presented leaf atlas on plants exposed to normal and lower water potential (i.e. mild drought), we aimed to explore how these two different growth conditions could affect the leaf single-cell transcriptome. As mentioned before, the low water potential did not have dramatic effects on the transcriptome, as none of the clusters was specifically present in either the WW, or the mild drought sample (Supplemental Figure S3D). Notably, the "stress-related" mesophyll cluster #2 is not predominantly composed of drought-stressed cells, and the genes that are enriched in that cluster (Supplemental Dataset S5) are not the ones DE under drought (Supplemental Dataset S1), indicating that the stress-related transcriptome of cluster #2 was not caused by the mild drought. To investigate how drought affected each main leaf tissue, we first identified, for each tissue, the genes that show at least 1.5-fold expression change between both conditions (Supplemental Dataset S8). Interestingly, very little overlap was found in these mild drought-responsive genes between the different tissues, with exception of *CSD2* (Figure 6, A–C). Instead, per main tissue, at least 1/3 of the response to low water potential was tissue-specific. Among the tissue-specific genes, notable examples included *CALEOSIN 3* (AT2G33380) and *COLD REGULATED 15B* (AT2G42530), induced by low water potential only in the mesophyll tissue, and the cellular wax biosynthesis gene *ECERIFERUM* (*CER1*, AT1G02205), specifically induced in the epidermis cells (Supplemental Figure S11A).

Second, as every cluster is composed of cells coming from both growth conditions, we aimed to calculate which main tissue was most sensitive to the drought. For this, we calculated the effect of drought on each of the molecular dimensions that contributed to the overall variance in this dataset and observed that drought mainly affected Dim.1, Dim.2, Dim.3, and Dim.6. As mentioned earlier, Dim.1 and Dim.2 were explained by genes involved in photosynthesis and differentiation, respectively (Figure 2E), while the genes that contribute to Dim.3 and Dim.6 appeared to be related to stress response (Supplemental Dataset S8). We plotted the cells of each main cluster along the four principal components that were most different between both conditions and calculated to what extent the two conditions shifted the position of the cells (Supplemental Figure S11B). Overall, the effect of drought on Dim.1 and Dim.2 was largely negative, while drought had a positive effect on Dim.3 and Dim.6, as could be expected. Under low water potential, mesophyll cells showed a clearly lower Dim.1 and Dim.2, and an increase in Dim.3 and Dim.6 (Supplemental Figure 11B). In contrast, the vasculature cells and initials were only mildly affected by drought. Epidermal cells showed an upwards shift in Dim.3 and a moderate decrease in Dim.1, which was mainly triggered by the cells with a low differentiation level (Dim.2) (Supplemental Figure S11B). Overall, summarizing the effects of drought on each of these main dimensions in the different tissues suggests that the mesophyll cells are most sensitive to the changing soil water potential (Figure 6D).

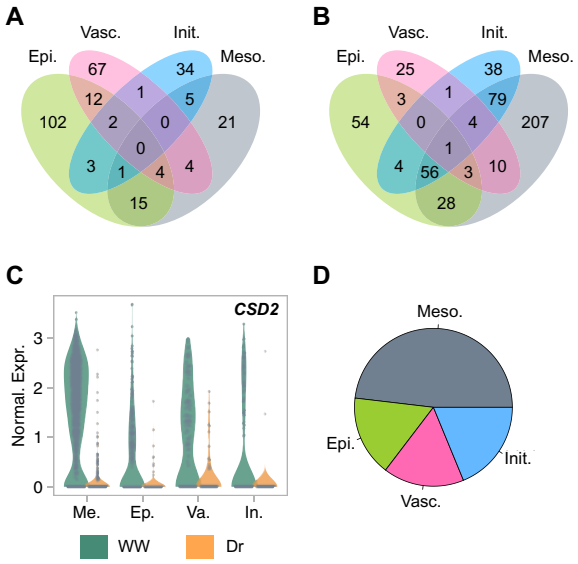

Figure 6 Transcriptome responses to lower water potential in the main leaf tissues. A, Venn diagram of genes upregulated under low water potential (compared to WW conditions). B, Venn diagram of the genes upregulated under normal water potential (compared to low water potential). C, Normalized expression of the COPPER SUPEROXIDE DISMUTASE2 (CSD2) gene, the gene in the overlap of the Venn diagram shown in (B). D, Effect of the low water potential treatment on the four main tissues represented by the sum of the absolute difference in median of each major dimension under low vs. normal watering conditions, as is depicted in Supplemental Figure S11.

Marker gene identification and a browsable dataset to explore the leaf single-cell atlas

In the previous parts, we have identified 14 different cell-type populations with their unique transcriptomic landscape. Here, we aimed to attribute the major biological function to each of the tissues, as well as to identify new marker genes (Figure 7). For the first purpose, we performed an in silico functional analysis of the top 100 (ranked by adjusted p-value) DE genes in each population (Figure 7A). Then, the top 10 GO terms in each population were visualized as a circular gene-concept network illustrating the connections between DE genes and GO terms (Supplemental Figures S12–26), of which an example is depicted for the GC population (Figure 7B). The five most relevant biological processes (BP) were used to functionally annotate each cluster (Table 2). Next, we selected marker genes for each of the identified populations based on the conditions for tissue-specificity previously described. Additionally, here, we allowed the selection of highly specific genes in small populations: mildly induced genes in the target population (average $\log_2$FC > 0.25) that were nearly absent in the background population (= in $< 1\%$ of the cells) but expressed in at least 30% more cells in the target population were included. The specificity of the top 10 tissue-specific

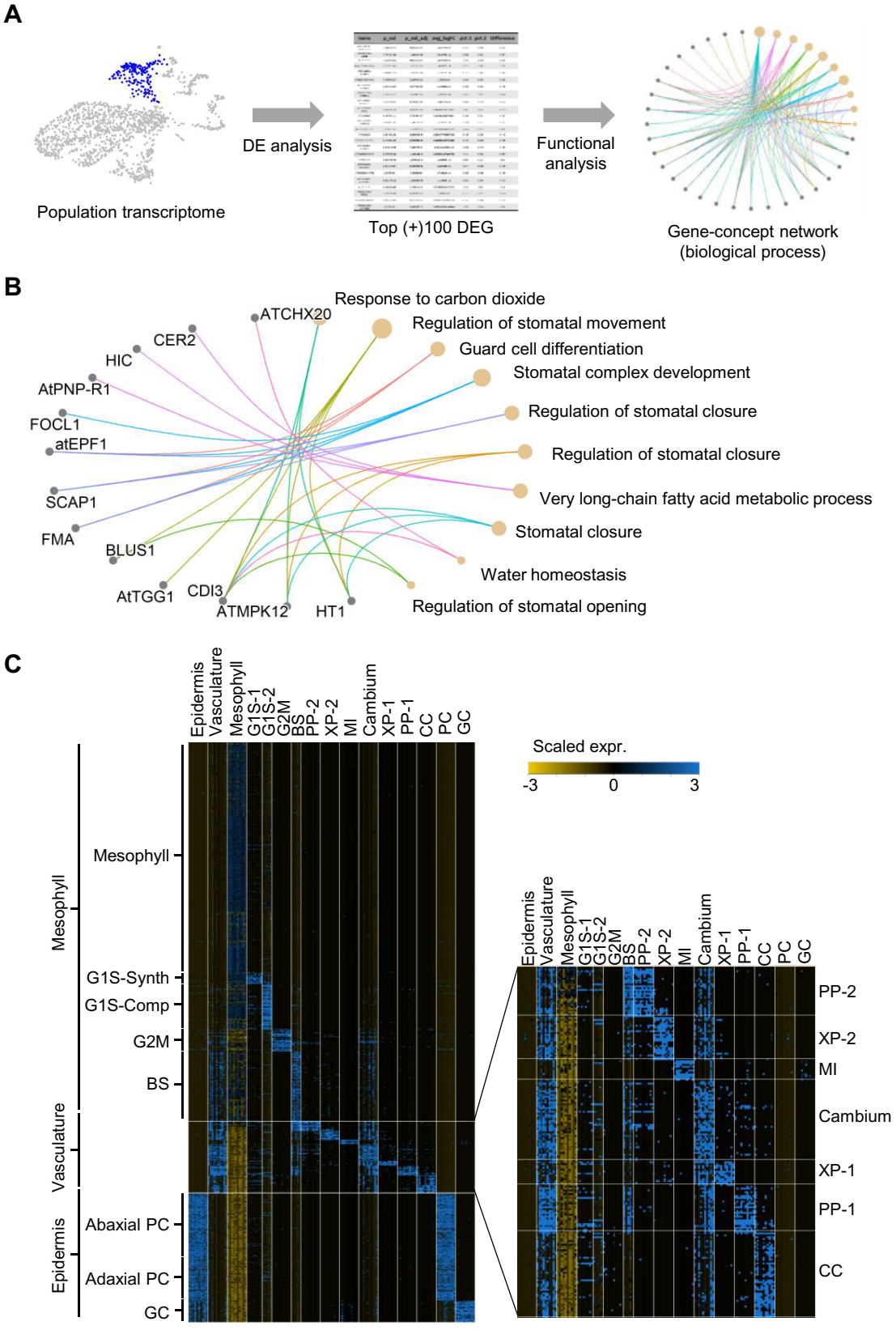

Figure 7 Functional analysis and marker gene identification in fourteen leaf cell populations. A, Schematic view of the functional analysis workflow performed for each population. Top 10 BP gene-concept networks for each population are represented in Supplemental Figures S12–26. B, Example of a gene-concept network circular plot of the GC population showing the top 10 BP based on the expression of the top 100 upregulated DE genes captured in this population. C, Heat map illustrating the scaled expression and specificity of specific marker genes identified in every leaf-cell population with a close-up view of the vasculature population at the right. The bar at the left indicates the cluster annotated. The bar at the top indicates the top 10 tissue-specific markers, based on the adjusted *P*-value, selected from this dataset. Colors represent the scaled expression of the gene (yellow = low, blue = high).

Table 2 Top-5 BP from GO terms obtained from the top-100 positively differentially expressed genes of 14 leaf populations

Cell population	Go enrichment BP
Mesophyll	Photosynthesis, photosynthesis light reaction, generation of precursor metabolites and energy, electron transport chain, response to light intensity
Cell cycle G1S (DNA biosynthesis)	DNA replication, DNA-dependent DNA replication, DNA repair, cellular response to DNA damage stimulus, mitotic cell cycle
Cell cycle G1S (Chromatin organization)	Chromatin organization, photosynthesis, photosynthesis, light reaction, protein–DNA complex subunit organization, negative regulation of RNA metabolic process
Cell cycle G2M	Mitotic cell cycle process, mitotic cell cycle, regulation of cell cycle, microtubule cytoskeleton organization, cytoskeleton organization
BS	Sulfur compound biosynthetic process, sulfur compound metabolic process, S-glycoside metabolic process, glycosinolate metabolic process, glucosinolate metabolic process
Companion cell phloem	Regionalization, phloem or xylem histogenesis, meristem development, cation homeostasis, photoperiodism
Cambium	Response to auxin, meristem development, phloem or xylem histogenesis, alpha-amino acid metabolic process, S-glycoside metabolic process
XP-1	Response to auxin, phloem, or xylem histogenesis, cellular response to auxin stimulus, auxin transport, regionalization
PP-1	Secondary metabolite biosynthetic process, sulfur compound biosynthetic process, S-glycoside metabolic process, glycosinolate metabolic process, glucosinolate metabolic process
XP-2	Phenylpropanoid metabolic process, phenylpropanoid biosynthetic process, secondary metabolite biosynthetic process, response to wounding, response to jasmonic acid
PP-2	Sulfur compound biosynthetic process, S-glycoside biosynthetic process, glucosinolate biosynthetic process, glucosinolate biosynthetic process, glycosyl compound biosynthetic process
MIs	S-glycoside metabolic process, glycosinolate metabolic process, glucosinolate metabolic process, cellular carbohydrate metabolic process, glycosyl compound metabolic process
PC	Fatty acid metabolic process, fatty acid biosynthetic process, monocarboxylic acid biosynthetic process, lipid transport, lipid localization
GC	External encapsulating structure organization, plant-type cell wall organization, anion transport, carbohydrate biosynthetic process, organic acid transport

genes is represented in Figure 7C (complete list in Supplemental Dataset S9).

In order to make the entire dataset publicly available to the scientific community, we integrated the data in the Plant Single Cell Atlas on-line browser tool (Wendrich et al., 2020), available at https://bioit3.irc.ugent.be/plant-sc-atlas/. The leaf single-cell tool home page presents an overview of the 14 annotated clusters (Figure 1B), accompanied by a brief introduction of the biological functions of each of them. Besides this, the tool can be used to easily visualize the expression of any gene of interest that was captured in the leaf single-cell dataset. More advanced analyses, including the identification of genes that correlate with a gene of interest in the single-cell transcriptome, or the identification of clusters in which multiple genes co-expressed, are also possible in the current version of this tool.

Discussion

The data presented here offer a single-cell level insight into the transcriptome of young leaves, under two distinct environmental growth conditions. After sequencing, the number of high-quality cells obtained from the drought sample was significantly lower than from the WW sample. However, the proportion of cells coming from the two conditions was stable across the different cell populations. Although we have no clear explanation for the lower number of high-quality drought stress protoplasts, we can speculate that the different leaf or cell structure, with an increased wax production and a thicker cell wall, respectively, might have hindered efficient protoplasting. Alternatively, it is plausible that the lower water potential within the drought-stressed cells increased their fragility in the protoplasting buffer, which is iso-osmotic with the content of WW protoplasts. Despite these issues, this analysis allowed for the identification of 14 unique cell populations. However, not all the known leaf cell types, such as the trichomes, were identified. Trichomes are very large and highly modified cells with a rigid cell wall, thus being highly resistant to protoplasting (Werker, 2000). Additionally, we did not identify any S-cell, lignified cells storing GSL. However, as there are currently no known S-cell marker genes, it remains unclear whether S-cells were present—but unidentifiable—among the XP-2 cells of the dataset, or whether these cells were not captured by the experimental conditions used in this study. The latter scenario could be explained by the programmed cell death that these cells undergo in an early stage of differentiation (Hunziker et al., 2019). Each of the identified cell populations showed a similar gene expression level except for the initials and, to a lower extent, the cell wall-related cluster. These clusters, for which it is unclear whether they have a biological meaning or result from an artifact, typically showed low expression of genes that were ubiquitously expressed in the mesophyll, epidermis, and vasculature.

By single-cell profiling of young leaves, we aimed to find both cells that undergo division and that are expanding/differentiating. Interestingly, the way these young, dividing cells integrated in the map of each tissue was very different depending on the cell type. Among the epidermal PCs, a tissue that loses photosynthetic activity during differentiation (Charuvi et al., 2012), we identified young cells expressing photosynthesis-related genes at one side of the clusters, hinting toward a developmental gradient. However, such continuous developmental gradient was dominated by a discrete polarity gradient that defined PC clustering. Between the abaxial and adaxial PC population, a small group of cells, that represent meristemoid mother cells (MMCs), showed a high meristematic activity. MMCs divide into pavement and GCs and co-express *ATML1*, *MUTE*, and *SPCH* (Liu et al., 2020), which was observed in this small population of cells. In contrast, in the mesophyll population, the transcriptomic identity related to cell cycle activity was very pronounced and dominated over the spongy or palisade parenchyma identity. This is very reminiscent to observations made in roots where the cluster of dividing cells overruled tissue identity in root tips (Wendrich et al., 2020). Three mesophyll cell populations were identified based on their cell cycle stage. The cells of the G2/M cluster have the highest meristematic activity and are dividing, while the cells of the DNA synthesis and chromatin compacting clusters are likely in the early and late S-phase, respectively, either in a mitotic or an endoreduplication cycle. Finally, in the center of the vasculature clusters, we found a population of young cells with meristematic activity, corresponding to the procambial population. In line with the study of Kim et al. (2021), this population was connected to the XP and PP population, but not to the CC cluster. The CC cluster also contained SEs, as both originate from the same mother cells (Bonke et al., 2003; Kim et al., 2021). Overall, the developmental stage that was harvested in this study was suitable to identify young or meristematic cells in each main tissue.

In previous root scRNA-seq studies, cluster annotation was performed by overlaying the single-cell transcriptome with tissue-specific expression datasets generated by microarrays (Brady et al., 2007; Cartwright et al., 2009). For leaves, besides the study of Svozil et al. (2015) that generated mass spectrometry data per main leaf tissue, this type of dataset was not available. Therefore, we annotated each tissue manually based on individual studies reporting tissue markers. To allow easier cluster annotation in future leaf scRNA-seq studies, we identified additional marker genes for each of the tissues. The identification of leaf tissues based on multiple marker genes will be important in future studies, particularly if scRNA-seq analyses are performed under different environmental conditions or using various genetic perturbations, of which some might affect the expression of the classical marker genes. For example, the mesophyll is largely defined by the expression of photosynthesis-related genes, however, abiotic stresses like oxidative stress dramatically affect leaf greening. In this context, marker genes unrelated to photosynthesis, like *ELP* and *GER3* found here as good mesophyll markers, could contribute to unbiased tissue identification. Moreover, single marker genes often fail to highlight a whole tissue, as was observed with the heterogeneous vasculature, or with the epidermis, where the typical polarity markers highlighted only the most differentiated part of the epidermis. We identified additional polarity markers that appeared at different positions on the differentiation gradient of the PCs. For example, *CHS* and *YAB3* or AT4G21910 marked the adaxial and abaxial PC population, respectively, along the whole developmental gradient. In contrast, *AUXIN-RESISTANT3 (AXR3)* or AT5G62210, and *YAB5* only appeared at a later developmental stage in the adaxial and abaxial PC population, respectively. Notably, no genes were found that specifically mark one epidermal cell layer only in the less differentiated cells. This suggests that genes that contribute to the polarity-identity of the PCs in young leaves remain active during later developmental stages.

This study also highlighted the potential of uncovering spatial separation between different branches of metabolic pathways, as illustrated for the GSLs. GSLs are specialized metabolites present in many crucifer plants and contribute to the defense against microbes and insects. On the one hand, indole GSLs were mainly enriched in BS and PP-1, the PP population directly in contact with the SE (Nintemann et al., 2018), which was also enriched in genes involved in sugar loading to the phloem, like *SWEET11/12*. Aliphatic GSLs, on the other hand, are specifically enriched in the PP-2 population not neighboring the SE cells (Nintemann et al., 2018). These data suggest that only indole GSLs are synthesized near the SEs, allegedly in order to be loaded in the phloem to be transported over long distances. In contrast, the aliphatic GSLs were suggested to be transported via the apoplast to surrounding S-cells for storage (Nintemann et al., 2018). Interestingly, whereas GSL transporters are still to be identified, in these clusters, we observed high expression of numerous genes with unknown functions; they could lay the basis of future research to elucidate GSL metabolism and transport. Similarly, lignin metabolism and transport genes allowed us to understand the two populations forming the xylem. Both XP-1 and XP-2 were enriched in lignification genes, but we hypothesize that XP-2 cells were parenchymal cells exporting monolignols to the adjacent SE cells, according to the "good neighbor hypothesis" (Smith et al., 2017), whereas XP-1 cells would not export lignin but undergo lignification themselves. In this case, the XP-1 population would actually consist of young developing TE cells, or possibly also contain S cells undergoing lignification (Hunziker et al., 2019).

Finally, the presented scRNA-seq dataset offered a glimpse on the tissue-level responses to mild drought stress in leaves. Whether abiotic stress affects the transcriptome uniformly across all cell types is an intriguing question. Here, we showed that the response to mild drought was highly dependent on the leaf tissue. This was not surprising, given that the drought response is extremely complex and fine-

tuned depending on the organ, stress level, time of day, stress duration and, thus, also the cell identity (Verslues, 2017). Similar tissue-specific responses were also observed in the study of Arabidopsis roots exposed to a heat shock (Jean-Baptiste et al., 2019). Overall, when profiling drought-triggered transcriptomic changes at organ level, one observes a downregulation of growth-related, energy-consuming processes (e.g. protein synthesis) along with an upregulation of stress defense genes, a balance established to ensure survival in case the stress becomes severe (Claeys and Inzé, 2013). The same "dichotomy" in gene expression was reported under heat shock, and another study demonstrated that it holds true within a single cell (Teves and Henikoff, 2011; Jean-Baptiste et al., 2019). Interestingly, our data rather suggest that leaves exposed to mild drought stress spatially divide these opposite responses: whereas the mesophyll showed a dramatic downregulation of genes, the epidermis was the tissue in which most genes were induced. This makes biological sense, as the mesophyll is the energy source of the leaf, while the epidermis plays a prominent role in leaf protection. Overall, plant responses to stress are so complex that the key to understand them relies on the study of the networks at single-cell level: by spatially separating, and thus simplifying the responses, tissue-specific pathways that contribute to better stress defense might be pinpointed, and could ultimately be engineered specifically in the tissue where they do matter, without causing unwanted side effects in other tissues or organs.

Materials and methods

Plant material and growth conditions

Arabidopsis (*A. thaliana*) Col-0 was grown under a long-day regime (16-h light and -h dark, intensity 110–120 µmol $m^{-2}s^{-1}$) at 21°C on the weighing, imaging and watering automated machine (WIWAM) platform (http://www.wiwam.be). Five seedlings were grown per pot in order to obtain 40 seedlings per sample (one replicate). The seeds were sown in 85 ± 1 g of Saniflor compost (Van Isreal N.V., Geraardsbergen, Belgium) with an average absolute water content of 70%. Plants were watered daily until 8 days after stratification (DAS) with a well-watered regime of soil relative water content 69% (2.2 g_{water}/g_{soil}). At 9 DAS, half of the pots (randomized) were not watered until harvest (14 DAS), while the other half kept the WW regime.

Sample preparation

Protoplasting protocol was adapted from Ryu et al. (2019). Briefly, the third leaf of 40 plants per sample was harvested, chopped, and incubated in 2 mL enzyme solution (0.4 M mannitol, 10 mM $CaCl_2$, 20 mM MES, 20 mM KCl, 2% (wt/vol) cellulase, and 0.5% (wt/vol) macero enzyme) constantly shaking for 30 min at room temperature. After a 7-min centrifugation at 100*g*, the pellet was resuspended in 1 mL washing solution (0.4 M mannitol, 10 mM $CaCl_2$, 20 mM MES, and 20 mM KCl). Resuspended cells were filtered through a 50-μm cell strainer, which was washed with 0.5 mL of washing solution. Cells were sorted on the Melody FACS, where gates to sort out debris of dead cells were predefined via a test sample that was 4′,6-diamidino-2-fenylindole (DAPI)-stained (viable cell ratio 31%, prior to sorting). Subsequently, 100,000 cells per sample were sorted with the Melody FACS in $<$10 min. Cells were collected in wash buffer (0.4 M mannitol, 10 mM $CaCl_2$, 20 mM MES, and 20 mM KCl) in a well of a 24-well plate. After centrifugation and buffer exchange, the protoplasts were loaded on the 10X instrument.

Single-cell RNA-seq library sample preparation, sequencing, and data processing

Single-cell suspensions were loaded on a GemCode Chromium Instrument (10X Genomics) to generate single-cell gel beads-in-emulsion (GEMs) using the Chromium Single Cell 3′ Gel Bead and Library Kit, version 3 (10X Genomics, Cat. 1000075) and the Chromium i7 Multiplex Kit (10X Genomics, Cat. 120262) according to the manufacturer's instructions. In brief, reverse-transcription was performed in a 96-deep-well reaction module at 53°C for 45 min, 85°C for 5 min and ending at 4°C. Next, GEMs were broken and cDNA was cleaned up with DynaBeads MyOne Silane Beads (Thermo Fisher Scientific, Waltham, MA, Cat. 37002D) and SPRIselect Reagent Kit (Beckman Coulter, Brea, CA, USA, No. B23318). Full-length, barcoded cDNA was amplified at 98°C for 3 min, 12 cycles at 98°C for 15 s, 67°C for 20 s and 72°C for 1 min, 1 cycle at 72°C for 1 min and ending at 4°C. Illumina-ready sequencing libraries were prepared after clean-up with the SPRIselect Reagent Kit and enzymatic fragmentation via end-repair, A-tailing, adapter ligation, post-ligation SPRIselect clean-up, and sample index PCR using the Chromium Single Cell 3′ Library Kit (10X Genomics, Cat 1000078). The cDNA content of pre-fragmentation and post-sample index polymerase chain reaction (PCR) samples was analyzed using the 2100 BioAnalyzer (Agilent, Santa Clara, CA, USA). Sequencing libraries were loaded on an Illumina NovaSeq flow cell performed at the VIB Nucleomics Core (VIB, Leuven, Belgium) with sequencing settings according to the recommendations of 10x Genomics for 3′-gene expression. The read lengths were the 10X recommendations for v3 chemistry: I1 = 8; *R*1 = 28; *R*2 = 91.

The Cell Ranger suite (10X Genomics, version 3.0.2) was used to perform sample demultiplexing and to generate FASTQ files for read 1, read 2, and the i7 sample index for the libraries. Read 2 of the gene expression libraries was mapped to the *A. thaliana* reference genome (TAIR10.40) using STAR. Subsequent barcode processing and gene counting was performed using the Cell Ranger.

Important metrics for the WW and drought sample are the following: Mean reads per cell: 110,770 (WW) and 298,794 (drought), median genes per cell: 1,514 (WW) and 1,309 (drought), fraction reads in cells: 68.7% (WW) and 55.3% (drought), median unique molecular identifyer (UMI)

counts per cell: 3,932 (WW) and 3,070 (drought), number of reads: 178,451,134 (WW), and 176,886,079 (drought), sequencing saturation: 66.0% (WW) and 81.9% (drought). Cell calling and removal of empty droplets was performed by applying the CB2 algorithm (Ni et al., 2020) to the raw feature cell barcode matrices generated by Cellranger count software. This resulted in the filtered digital gene expression matrices that were subsequently processed and filtered using Seurat (version 3.2.3 with R version 3.6.0; Stuart et al., 2019). Low-abundance genes were removed by removing all genes that were not expressed in at least three cells and only protoplasts in which at least 200 genes were detected were retained. Finally, cells with a UMI count lower than 500 counts were removed (Supplemental Figure S27). Normalization, detection of high variable genes, scaling, clustering, and dimensionality reduction were performed using Seurat (top 30 principal components). The absence of a strong doublet-effect on the clustering was verified. After validating the biological meaning of the clusters (i.e. marker genes expression did not indicate any overclustering), a final cluster resolution to assess the high diversity of cell types was set at 0.9. The vascular tissue, composed of multiple small populations, required subsequent manual subdivision based on marker gene expression.

Figure generation

All figures were made in R version 4.0.3 using the normalized gene expression, except for heatmaps where the scaled data were used. Used packages include "ggplot 2" (scatterplots and dot plots), "gplots" (heatmap.2), and "VennDiagram". For each main figure where an inset is shown, a full visualization is depicted in Supplemental Figure S28. Dot plots only take into account cells in which the depicted gene is expressed in at least 5% of the cells of a cluster. The dot size represents the ratio of cells expressing a gene over the total number of cells in that cluster. For the analyses involving dimensional analysis, the "prcomp" function was used. For the genes shown in Figure 6, we filtered out genes that were low-expressed in both conditions (normalized expression <0.5).

For the functional analysis, we used the cnetplot and dot plot functions from the Rpackage "ClusterProfiler" (Yu et al., 2012). A GO BP concept network plot (cnetplot) depicting the linkages between DE genes and biological concept term for each population was constructed according to the library "org.At.tair.db", a genome-wide annotation for Arabidopsis, primarily based on mapping using TAIR identifiers (Carlson, 2019). For each population, significant enrichments from top 100 positively DE genes with a P.adj < 0.01 were obtained. Afterward, a cnetplot and dot plot displaying the top 10 enriched Biological Process terms and gene linkages were generated.

Identification of DE genes per cluster or condition and cluster-specific marker genes

Differential gene expression analysis in clusters was performed by comparing the expression profile of each cluster with that of the rest of the dataset. When mentioned specifically that two clusters were compared with each other (e.g. abaxial and adaxial epidermis, or cluster #0 and cluster #2), then the DE analysis was performed by comparing only these two clusters with each other. DE analysis between conditions (WW conditions versus low water potential) was performed by comparing the expression profile of all cells from the WW sample with that of all cells of the drought stress sample. For all DE analyses, a Wilcoxon Rank Sum test was done through the Seurat function "FindMarkers." *P*-value adjustment was performed using Bonferroni correction. Clusters with the same cell annotation based on gene expression analysis were combined to generate a more comprehensible dataset. Merging the data corresponding with both conditions was done using the built-in merge function of Seurat. Enriched genes were defined as those genes significantly induced (Avg logFC > 0.25).

Cluster-specific marker genes for main clusters are the DE genes (P.adj < 0.05) fulfilling one of these conditions: (1) Avg logFC > 2 and expressed in more than 30% of the cells in the population compared to the background population (difference > 0.3); (2) Avg logFC > 0.25) in the selected population but difference > 0.5 and expressed in $<5\%$ of the background population (pct.2 < 0.05). Additional condition for markers individual clusters: (3) Avg logFC > 0.25 and pct2 < 0.01 and difference > 0.3.

Web tool availability for data visualization

All single-cell datasets presented in this manuscript can be accessed via our online leaf protoplast atlas https://bioit3.irc.ugent.be/plant-sc-atlas/leaf (for a detailed analysis of this dataset, please visit: https://www.psb.ugent.be/sc-leaf-yield lab). This tool presents the normalized data. For violin plots, no filtering was applied (all cells not expressing a gene are also depicted). Pearson correlation coefficient analysis displays the top 1,000 correlating genes with *P*-value (no *P*-value threshold). For the visualization of co-expressed genes, the level of co-expression is depicted as a bivariate gradient between the distance and the summation of the normalized gene expression of both genes.

Accession numbers

The single-cell RNA-seq data are available in the ArrayExpress database (https://www.ebi.ac.uk/arrayexpress) under the accession number E-MTAB-11006.

Supplemental data

The following materials are available in the online version of this article.

Supplemental Figure S1. Schematic cross-section of a mature leaf.

Supplemental Figure S2. t-distributed stochastic neighbor-embedding representation of the Arabidopsis leaf transcriptomic landscape profiled by scRNA-sequencing.

Supplemental Figure S3. Effect of the reduced water potential on the transcriptome and clustering.

Supplemental Figure S4. Expression pattern of additional main cluster markers.

Supplemental Figure S5. Expression pattern of photosynthesis-related and chloroplast genes.

Supplemental Figure S6. Heat map illustrating the scaled expression level of transcripts of tissue-specific proteins present in the sc-RNAseq dataset.

Supplemental Figure S7. Developmental gradient in the epidermal cell population.

Supplemental Figure S8. Identification of GCs and stomatal development.

Supplemental Figure S9. Expression pattern of additional cell cycle markers.

Supplemental Figure S10. Hypothetical spatial representation of vasculature tissues identified from scRNA-seq.

Supplemental Figure S11. Effect of lower water potential on transcript levels and main principal components (dimensions).

Supplemental Figure S12. Functional analysis of the mesophyll.

Supplemental Figure S13. Functional analysis of the cells in G1/S compaction.

Supplemental Figure S14. Functional analysis of cells in G1/S DNA synthesis.

Supplemental Figure S15. Functional analysis of cells in G2/M cell cycle phase.

Supplemental Figure S16. Functional analysis of the BS.

Supplemental Figure S17. Functional analysis of the CCs.

Supplemental Figure S18. Functional analysis of the cambium.

Supplemental Figure S19. Functional analysis of the XP-1.

Supplemental Figure S20. Functional analysis of the XP-2.

Supplemental Figure S21. Functional analysis of the PP-1.

Supplemental Figure S22. Functional analysis of the PP-2.

Supplemental Figure S23. Functional analysis of the MIs.

Supplemental Figure S24. Functional analysis of the PCs.

Supplemental Figure S25. Functional analysis of the GCs.

Supplemental Figure S26. Functional analysis of the cell wall-related cluster.

Supplemental Figure S27. Protoplast number in function of UMI threshold.

Supplemental Figure S28. Full visualization of the UMAP plots insets shown in Figures 3–5.

Supplemental Dataset S1. Drought-responsive differential gene expression analysis.

Supplemental Dataset S2. Differential gene expression analysis of the main tissue populations.

Supplemental Dataset S3. Marker gene identification for the main tissue populations.

Supplemental Dataset S4. Differential gene expression analysis of 15 leaf populations.

Supplemental Dataset S5. Differential gene expression analysis between the adaxial and abaxial PC populations.

Supplemental Dataset S6. Marker gene identification for the adaxial and abaxial PC populations.

Supplemental Dataset S7. Differential gene expression analysis between the cluster #0 and #2 of the mesophyll population.

Supplemental Dataset S8. Drought-responsive genes per main tissue.

Supplemental Dataset S9. Marker gene identification of 14 leaf cell populations.

Acknowledgments

We thank the members of the Systems Biology of Yield team for the nice working atmosphere, particularly Hannes and Ying for the constructive discussions. We are especially grateful to Dr Annick Bleys, Dr Hannes Vanhaeren, and Dr Tom Vanhauteghem for the critical reading of the manuscript and the suggestions to improve it. We would also like to thank Dr Bert de Rybel and the Plant Single Cell platform at PSB for their guidance in this project, as well as Dr Lieven Sterck for uploading the data in the public depository. Finally, this work would not have been possible without the help of the VIB Flow Core, VIB Nucleomics Core, and the Singularity single-cell platform.

Funding

This work was supported by Ghent University (Bijzonder Onderzoeksfonds Methusalem Project BOF08/01M00408) and by the Research Foundation Flanders (G010820N). Marieke Dubois is a postdoctoral fellow of Flanders Research Foundation (FWO no. 12Q7919N). Ignacio Achon was benefited from a doctoral fellowship "Don Carlos Antonio Lopez" by El Programa Nacional de Becas from Paraguay (BECAL #164/2017).

Conflict of interest statement. The authors declare no conflict of interest.

References

Agusti J, Lichtenberger R, Schwarz M, Nehlin L, Greb T (2011) Characterization of transcriptome remodeling during cambium formation identifies *MOL1* and *RUL1* as opposing regulators of secondary growth. PLoS Genet **7:** e1001312

Amir EAD, Davis KL, Tadmor MD, Simonds EF, Levine JH, Bendall SC, Shenfeld DK, Krishnaswamy S, Nolan GP, Pe'er D (2013) viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. Nat Biotechnol **31:** 545–552

Andréasson E, Jørgensen LB, Höglund A-S, Rask L, Meijer J (2001) Different myrosinase and idioblast distribution in Arabidopsis and Brassica napus. Plant Physiol **127:** 1750–1763

Andriankaja M, Dhondt S, De Bodt S, Vanhaeren H, Coppens F, De Milde L, Mühlenbock P, Skirycz A, Gonzalez N, Beemster GTS, et al. (2012) Exit from proliferation during leaf development in *Arabidopsis thaliana*: a not-so-gradual process. Dev Cell **22:** 64–78

Anstead JA, Froelich DR, Knoblauch M, Thompson GA (2012) Arabidopsis P-protein filament formation requires both *AtSEOR1* and *AtSEOR2*. Plant Cell Physiol **53:** 1033–1042

Baima S, Possenti M, Matteucci A, Wisman E, Altamura MM, Ruberti I, Morelli G (2001) The Arabidopsis ATHB-8 HD-zip

protein acts as a differentiation-promoting transcription factor of the vascular meristems. Plant Physiol **126:** 643–655

Barth C, Jander G (2006) Arabidopsis myrosinases *TGG1* and *TGG2* have redundant function in glucosinolate breakdown and insect defense. Plant J **46:** 549–562

Bezrutczyk M, Zöllner NR, Kruse CP, Hartwig T, Lautwein T, Köhrer K, Frommer WB, Kim JY (2021) Evidence for phloem loading via the abaxial bundle sheath cells in maize leaves. Plant Cell **3**: 531–547

Biedroń M, Banasiak A (2018) Auxin-mediated regulation of vascular patterning in *Arabidopsis thaliana* leaves. Plant Cell Rep **37:** 1215–1229

Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Stat Mech **2008:** P10008

Bonke M, Thitamadee S, Mähönen AP, Hauser MT, Helariutta Y (2003) APL regulates vascular tissue identity in *Arabidopsis*. Nature **426:** 181–186

Brady SM, Orlando DA, Lee J-Y, Wang JY, Koch J, Dinneny JR, Mace D, Ohler U, Benfey PN (2007) A high-resolution root spatiotemporal map reveals dominant expression patterns. Science **318:** 801–806

Burow M, Rice M, Hause B, Gershenzon J, Wittstock U (2007) Cell-and tissue-specific localization and regulation of the epithiospecifier protein in *Arabidopsis thaliana*. Plant Mol Biol **64:** 173–185

Carlson M (2019) org.Hs.eg.db: Genome wide annotation for Human. R Package version 832, https://bioconductor.org/packages/release/data/annotation/html/org.Hs.eg.db.html

Cartwright DA, Brady SM, Orlando DA, Sturmfels B, Benfey PN (2009) Reconstructing spatiotemporal gene expression data from partial observations. Bioinformatics **25:** 2581–2587

Cayla T, Batailler B, Le Hir R, Revers F, Anstead JA, Thompson GA, Grandjean O, Dinant S (2015) Live imaging of companion cells and sieve elements in Arabidopsis leaves. PLoS One **10:** e0118122

Charuvi D, Kiss V, Nevo R, Shimoni E, Adam Z, Reich Z (2012) Gain and loss of photosynthetic membranes during plastid differentiation in the shoot apex of *Arabidopsis*. Plant Cell **24:** 1143–1157

Chen LQ, Qu XQ, Hou BH, Sosso D, Osorio S, Fernie AR, Frommer WB (2012) Sucrose efflux mediated by SWEET proteins as a key step for phloem transport. Science **335:** 207–211

Chen Y, Dubois M, Vermeersch M, Vanhaeren H, Inzé D (2021) Distinct cellular strategies determine sensitivity to mild drought of Arabidopsis natural accessions. Plant Physiol **186**: 1171–1185

Chitwood DH, Nogueira FT, Howell MD, Montgomery TA, Carrington JC, Timmermans MC (2009) Pattern formation via small RNA mobility. Genes Dev **23:** 549–554

Claeys H, Inzé D (2013) The agony of choice: how plants balance growth and survival under water-limiting conditions. Plant Physiol **162:** 1768–1779

Cools T, Iantcheva A, Maes S, Van den Daele H, De Veylder L (2010) A replication stress-induced synchronization method for *Arabidopsis thaliana* root meristems. Plant J **64:** 705–714

Denyer T, Ma X, Klesen S, Scacchi E, Nieselt K, Timmermans MC (2019) Spatiotemporal developmental trajectories in the Arabidopsis root revealed using high-throughput single-cell RNA sequencing. Dev Cell **48:** 840–852. e845

DeWitt ND, Sussman MR (1995) Immunocytological localization of an epitope-tagged plasma membrane proton pump (H(+)-ATPase) in phloem companion cells. Plant Cell **7:** 2053–2067

Dolan L, Janmaat K, Willemsen V, Linstead P, Poethig S, Roberts K, Scheres B (1993) Cellular organisation of the *Arabidopsis thaliana* root. Development **119:** 71–84

Dubois M, Claeys H, Van den Broeck L, Inzé D (2017) Time of day determines Arabidopsis transcriptome and growth dynamics under mild drought. Plant Cell Environ **40:** 180–189

Endo S, Iwamoto K, Fukuda H (2018) Overexpression and cosuppression of xylem-related genes in an early xylem differentiation stage-specific manner by the *AtTED4* promoter. Plant Biotechnol J **16:** 451–458

Facette MR, Smith LG (2012) Division polarity in developing stomata. Curr Opin Plant Biol **15:** 585–592

Farmer A, Thibivilliers S, Ryu KH, Schiefelbein J, Libault M (2021) Single-nucleus RNA and ATAC sequencing reveals the impact of chromatin accessibility on gene expression in Arabidopsis roots at the single-cell level. Mol Plant **14**: 372–383

Fett JP, Coleman JR (1994) Characterization and expression of two cDNAs encoding carbonic anhydrase in *Arabidopsis thaliana*. Plant Physiol **105:** 707–713

Fowler TJ, Bernhardt C, Tierney ML (1999) Characterization and expression of four proline-rich cell wall protein genes in Arabidopsis encoding two distinct subsets of multiple domain proteins. Plant Physiol **121:** 1081–1091

Glover B, Martini C (2000) Specification of epidermal cell morphology. Adv Bot Res **31**: 193–217

Gray JE, Holroyd GH, Van Der Lee FM, Bahrami AR, Sijmons PC, Woodward FI, Schuch W, Hetherington AM (2000) The HIC signalling pathway links CO 2 perception to stomatal development. Nature **408:** 713–716

Hirakawa Y, Kondo Y, Fukuda H (2010) TDIF peptide signaling regulates vascular stem cell proliferation via the *WOX4* homeobox gene in Arabidopsis. Plant Cell **22:** 2618–2629

Houbaert A, Zhang C, Tiwari M, Wang K, de Marcos Serrano A, Savatin DV, Urs MJ, Zhiponova MK, Gudesblat GE, Vanhoutte I, et al. (2018) POLAR-guided signalling complex assembly and localization drive asymmetric cell division. Nature **563:** 574–578

Hunziker P, Halkier BA, Schulz A (2019) Arabidopsis glucosinolate storage cells transform into phloem fibres at late stages of development. J Exp Bot **70:** 4305–4317

Hwang JE, Hong JK, Lim CJ, Chen H, Je J, Yang KA, Kim DY, Choi YJ, Lee SY, Lim CO (2010) Distinct expression patterns of two Arabidopsis phytocystatin genes, AtCYS1 and AtCYS2, during development and abiotic stresses. Plant Cell Rep **29:** 905–915

Inada N, Wildermuth MC (2005) Novel tissue preparation method and cell-specific marker for laser microdissection of *Arabidopsis* mature leaf. Planta **221:** 9–16

Jean-Baptiste K, McFaline-Figueroa JL, Alexandre CM, Dorrity MW, Saunders L, Bubb KL, Trapnell C, Fields S, Queitsch C, Cuperus JT (2019) Dynamics of gene expression in single root cells of *Arabidopsis thaliana*. Plant Cell **31:** 993–1011

Jin X, Wang RS, Zhu M, Jeon BW, Albert R, Chen S, Assmann SM (2013) Abscisic acid–responsive guard cell metabolomes of Arabidopsis wild-type and gpa1 G-protein mutants. Plant Cell **25:** 4789–4811

Kalve S, De Vos D, Beemster GTS (2014) Leaf development: a cellular perspective. Front Plant Sci **5:** 362

Kanaoka MM, Pillitteri LJ, Fujii H, Yoshida Y, Bogenschutz NL, Takabayashi J, Zhu JK, Torii KU (2008) *SCREAM/ICE1* and *SCREAM2* specify three cell-state transitional steps leading to *Arabidopsis* stomatal differentiation. Plant Cell **20:** 1775–1785

Kazama T, Ichihashi Y, Murata S, Tsukaya H (2010) The mechanism of cell cycle arrest front progression explained by a *KLUH/CYP78A5*-dependent mobile growth factor in developing leaves of *Arabidopsis thaliana*. Plant Cell Physiol **51:** 1046–1054

Keerthisinghe S, Nadeau JA, Lucas JR, Nakagawa T, Sack FD (2015) The Arabidopsis leucine-rich repeat receptor-like kinase MUSTACHES enforces stomatal bilateral symmetry in A rabidopsis. Plant J **81:** 684–694

Kidner CA, Timmermans MC (2010) Signaling sides: adaxial–abaxial patterning in leaves. Curr Top Dev Biol **91:** 141–168

Kim JY, Symeonidi E, Pang TY, Denyer T, Weidauer D, Bezrutczyk M, Miras M, Zöllner N, Hartwig T, Wudick MM (2021) Distinct identities of leaf phloem cells revealed by single cell transcriptomics. Plant Cell **33**: 511–530

Kim J, Jung JH, Lee SB, Go YS, Kim HJ, Cahoon R, Markham JE, Cahoon EB, Suh MC (2013) Arabidopsis 3-ketoacyl-coenzyme a synthase9 is involved in the synthesis of tetracosanoic acids as precursors of cuticular waxes, suberins, sphingolipids, and phospholipids. Plant Physiol **162:** 567–580

Kirschner S, Woodfield H, Prusko K, Koczor M, Gowik U, Hibberd JM, Westhoff P (2018) Expression of *SULTR2;* 2, encoding a low-affinity sulphur transporter, in the Arabidopsis bundle sheath and vein cells is mediated by a positive regulator. J Exp Bot **69:** 4897–4906

Koroleva OA, Davies A, Deeken R, Thorpe MR, Tomos AD, Hedrich R (2000) Identification of a new glucosinolate-rich cell type in Arabidopsis flower stalk. Plant Physiol **124:** 599–608

Kunst L, Samuels AL (2003) Biosynthesis and secretion of plant cuticular wax. Prog Lipid Res **42:** 51–80

Li M, Sack FD (2014) Myrosin idioblast cell fate and development are regulated by the Arabidopsis transcription factor FAMA, the auxin pathway, and vesicular trafficking. Plant Cell **26:** 4053–4066

Lieber D, Lora J, Schrempp S, Lenhard M, Laux T (2011) *Arabidopsis WIH1* and *WIH2* genes act in the transition from somatic to reproductive cell fate. Curr Biol **21:** 1009–1017

Liu Q, Liang Z, Feng D, Jiang S, Wang Y, Du Z, Li R, Hu G, Zhang P, Ma Y (2021) Transcriptional landscape of rice roots at the single-cell resolution. Mol Plant **14:** 384–394

Liu Z, Zhou Y, Guo J, Li J, Tian Z, Zhu Z, Wang J, Wu R, Zhang B, Hu Y (2020) Global dynamic molecular profiling of stomatal lineage cell development by single-cell RNA sequencing. Mol Plant **13:** 1178–1193

Long Y, Liu Z, Jia J, Mo W, Fang L, Lu D, Liu B, Zhang H, Chen W, Zhai J (2021) FlsnRNA-seq: protoplasting-free full-length single-nucleus RNA profiling in plants. Genome Biol **22:** 1–14

Mahroug S, Courdavault V, Thiersault M, St-Pierre B, Burlat V (2006) Epidermis is a pivotal site of at least four secondary metabolic pathways in *Catharanthus roseus* aerial organs. Planta **223:** 1191–1200

Marks MD, Wenger JP, Gilding E, Jilk R, Dixon RA (2009) Transcriptome analysis of *Arabidopsis* wild-type and *gl3-sst sim* trichomes identifies four additional genes required for trichome development. Mol Plant **2:** 803–822

McInnes L, Healy J, Melville J (2018) Umap: Uniform manifold approximation and projection for dimension reduction. arXiv, http://arxivorg/abs/180203426, Published online December 6, 2018.

McNair G (2015) COBRA-LIKE4: a GPI-anchored protein functioning as a mediator of cellulose ultrastructure in herbaceous and woody plants. UBC Theses and Dissertations. University of British Columbia, Vancouver, Canada

Membré N, Bernier F, Staiger D, Berna A (2000) *Arabidopsis thaliana* germin-like proteins: common and specific features point to a variety of functions. Planta **211:** 345–354

Moroney J, Bartlett S, Samuelsson G (2001) Carbonic anhydrases in plants and algae. Plant Cell Environ **24:** 141–153

Muñiz L, Minguet EG, Singh SK, Pesquet E, Vera-Sirera F, Moreau-Courtois CL, Carbonell J, Blázquez MA, Tuominen H (2008) ACAULIS5 controls *Arabidopsis* xylem specification through the prevention of premature cell death. Development **135:** 2573–2582

Mustroph A, Zanetti ME, Girke T, Bailey-Serres J (2013) Isolation and analysis of mRNAs from specific cell types of plants by ribosome immunopurification. Methods Mol Biol **959**: 277–302

Ni Z, Chen S, Brown J, Kendziorski C (2020) CB2 improves power of cell detection in droplet-based single-cell RNA sequencing data. Genome Biol **21:** 1–10

Nintemann SJ, Hunziker P, Andersen TG, Schulz A, Burow M, Halkier BA (2018) Localization of the glucosinolate biosynthetic enzymes reveals distinct spatial patterns for the biosynthesis of indole and aliphatic glucosinolates. Physiol Plant **163:** 138–154

Otero S, Helariutta Y (2016) Companion cells: a diamond in the rough. J Exp Bot **68**: 71–78

Pillitteri LJ, Dong J (2013) Stomatal development in Arabidopsis. Arabidopsis Book **11:** e0162

Pillitteri LJ, Sloan DB, Bogenschutz NL, Torii KU (2007) Termination of asymmetric cell division and differentiation of stomata. Nature **445:** 501–505

Potter SS (2018) Single-cell RNA sequencing for the study of development, physiology and disease. Nat Rev Nephrol **14:** 479–492

Reynoso MA, Juntawong P, Lancia M, Blanco FA, Bailey-Serres J, Zanetti ME (2015) Translating Ribosome Affinity Purification (TRAP) followed by RNA sequencing technology (TRAP-SEQ) for quantitative assessment of plant translatomes. Methods Mol Biol **1284**: 185–207

Rich-Griffin C, Stechemesser A, Finch J, Lucas E, Ott S, Schäfer P (2020) Single-cell transcriptomics: a high-resolution avenue for plant functional genomics. Trend Plant Sci **25:** 186–197

Rusconi F, Simeoni F, Francia P, Cominelli E, Conti L, Riboni M, Simoni L, Martin CR, Tonelli C, Galbiati M (2013) The *Arabidopsis thaliana MYB60* promoter provides a tool for the spatio-temporal control of gene expression in stomatal guard cells. J Exp Bot **64:** 3361–3371

Ryu KH, Huang L, Kang HM, Schiefelbein J (2019) Single-cell RNA sequencing resolves molecular relationships among individual plant cells. Plant Physiol **179:** 1444–1456

Satterlee JW, Strable J, Scanlon MJ (2020) Plant stem-cell organization and differentiation at single-cell resolution. Proc Natl Acad Sci USA **117:** 33689–33699

Schuetz M, Smith R, Ellis B (2013) Xylem tissue specification, patterning, and differentiation mechanisms. J Exp Bot **64:** 11–31

Seyfferth C, Renema J, Wendrich JR, Eekhout T, Seurinck R, Blob B, Saeys Y, Helariutta Y, Birnbaum KD, De Rybel B (2021) Advances and opportunities of single-cell transcriptomics for plant research. Ann Rev Plant Biol **72**: 847–866

Shitov AV, Terentyev VV, Zharmukhamedov SK, Rodionova MV, Karacan M, Karacan N, Klimov VV, Allakhverdiev SI (2018) Is carbonic anhydrase activity of photosystem II required for its maximum electron transport rate? Biochim Biophys Acta (BBA)-Bioenergetics **1859:** 292–299

Shulse CN, Cole BJ, Ciobanu D, Lin J, Yoshinaga Y, Gouran M, Turco GM, Zhu Y, O'Malley RC, Brady SM (2019) High-throughput single-cell transcriptome profiling of plant cell types. Cell Rep **27:** 2241–2247. e2244

Skirycz A, Vandenbroucke K, Clauw P, Maleux K, De Meyer B, Dhondt S, Pucci A, Gonzalez N, Hoeberichts F, Tognetti VB, et al. (2011) Survival and growth of *Arabidopsis* plants given limited water are not equal. Nat Biotechnol **29:** 212–214

Slane D, Bayer M (2017) Cell type-specific gene expression profiling using fluorescence-activated nuclear sorting. Methods Mol Biol **1629**: 27–35

Smith RA, Schuetz M, Karlen SD, Bird D, Tokunaga N, Sato Y, Mansfield SD, Ralph J, Samuels AL (2017) Defining the diverse cell populations contributing to lignification in Arabidopsis stems. Plant Physiol **174:** 1028–1036

Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck III WM, Hao Y, Stoeckius M, Smibert P, Satija R (2019) Comprehensive integration of single-cell data. Cell **177:** 1888–1902. e1821

Sugano SS, Shimada T, Imai Y, Okawa K, Tamai A, Mori M, Hara-Nishimura I (2010) Stomagen positively regulates stomatal density in Arabidopsis. Nature **463:** 241–244

Svozil J, Gruissem W, Baerenfaller K (2015) Proteasome targeting of proteins in Arabidopsis leaf mesophyll, epidermal and vascular tissues. Front Plant Sci **6:** 376

Tameshige T, Fujita H, Watanabe K, Toyokura K, Kondo M, Tatematsu K, Matsumoto N, Tsugeki R, Kawaguchi M, Nishimura M (2013) Pattern dynamics in adaxial-abaxial specific gene expression are modulated by a plastid retrograde signal during *Arabidopsis thaliana* leaf development. PLoS Genet **9:** e1003655

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

Thellmann M, Andersen TG, Vermeer JE (2020) Translating ribosome affinity purification (trap) to investigate *Arabidopsis thaliana* root development at a cell type-specific scale. J Vis Exp (Jove) **159:** e60919

Tian H, Baxter IR, Lahner B, Reinders A, Salt DE, Ward JM (2010) Arabidopsis NPCC6/NaKR1 is a phloem mobile metal binding protein necessary for phloem function and root meristem maintenance. Plant Cell **22:** 3963–3979

Truernit E, Sauer N (1995) The promoter of the *Arabidopsis thaliana SUC2* sucrose-H + symporter gene directs expression of β-glucuronidase to the phloem: evidence for phloem loading and unloading by *SUC2*. Planta **196:** 564–570

Tsukaya H (2002) Leaf development. Arabidopsis Book **1**: e0072

Tsukaya H (2013) Leaf development. Arabidopsis Book **11:** e0163

Turner S, Sieburth LE (2003) Vascular patterning. Arabidopsis Book **2:** e0073

Uemoto K, Araki T, Endo M (2018) Isolation of Arabidopsis palisade and spongy mesophyll cells. Methods Mol Biol **1830**: 141–148

Vanholme R, Cesarino I, Rataj K, Xiao Y, Sundin L, Goeminne G, Kim H, Cross J, Morreel K, Araujo P, et al. (2013) Caffeoyl shikimate esterase (CSE) is an enzyme in the lignin biosynthetic pathway in *Arabidopsis*. Science **341:** 1103–1106

Verslues PE (2017) Time to grow: factors that control plant growth during mild to moderate drought stress. Plant Cell Environ **40:** 177–179

Wang Y, Jiao Y (2014) Translating ribosome affinity purification (TRAP) for cell-specific translation profiling in developing flowers. Methods Mol Biol **1110**: 323–328

Weimer AK, Matos JL, Sharma N, Patell F, Murray JA, Dewitte W, Bergmann DC (2018) Lineage-and stage-specific expressed *CYCD7; 1* coordinates the single symmetric division that creates stomatal guard cells. Development **145**: dev160671

Wendrich JR, Yang B, Vandamme N, Verstaen K, Smet W, Van de Velde C, Minne M, Wybouw B, Mor E, Arents HE (2020) Vascular transcription factors guide plant epidermal responses to limiting phosphate conditions. Science **370**: eaay4970

Werker E (2000) Trichome diversity and development. Adv Bot Res **31**: 1–35

Xue J, Jørgensen M, Pihlgren U, Rask L (1995) The myrosinase gene family in *Arabidopsis thaliana*: gene organization, expression and evolution. Plant Mol Biol **27:** 911–922

Yang C, Ye Z (2013) Trichomes as models for studying plant cell differentiation. Cell Mol Life Sci **70:** 1937–1948

Yu G, Wang LG, Han Y, He QY (2012) Cluster profiler: an R package for comparing biological themes among gene clusters. Omics: J Integr Biol **16:** 284–287

Yu Y, Feng Z, Wang G, Li F, Du X, Zhu J (2010) Initiation of dedifferentiation and structural changes in in vitro cultured petiole of *Arabidopsis thaliana*. Protoplasma **241:** 75–81

Zhang TQ, Xu ZG, Shang GD, Wang JW (2019) A single-cell RNA sequencing profiles the developmental landscape of Arabidopsis root. Mol Plant **12:** 648–660