



Deposited via The University of York.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/229683/>

Version: Published Version

Article:

Ezer, Daphne, Vong, Gina, Claydon, William et al. (2024) AraLeTA: An Arabidopsis leaf expression atlas across diurnal and developmental scales. *Plant Physiology*. kiae117.

ISSN: 0032-0889

<https://doi.org/10.1093/plphys/kiae117>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



AraLeTA: An Arabidopsis leaf expression atlas across diurnal and developmental scales

Gina Y.W. Vong [†], Kayla McCarthy [†], Will Claydon [†], Seth J. Davis , Ethan J. Redmond ,
Daphne Ezer ^{*}

Department of Biology, University of York, York YO10 5DD, UK

*Author for correspondence: daphne.ezer@york.ac.uk

[†]Joint first authors.

The author responsible for 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: Daphne Ezer (daphne.ezer@york.ac.uk).

Abstract

Mature plant leaves are a composite of distinct cell types, including epidermal, mesophyll, and vascular cells. Notably, the proportion of these cells and the relative transcript concentrations within different cell types may change over time. While gene expression data at a single-cell level can provide cell-type-specific expression values, it is often too expensive to obtain these data for high-resolution time series. Although bulk RNA-seq can be performed in a high-resolution time series, RNA-seq using whole leaves measures average gene expression values across all cell types in each sample. In this study, we combined single-cell RNA-seq data with time-series data from whole leaves to assemble an atlas of cell-type-specific changes in gene expression over time for Arabidopsis (*Arabidopsis thaliana*). We inferred how the relative transcript concentrations of different cell types vary across diurnal and developmental timescales. Importantly, this analysis revealed 3 subgroups of mesophyll cells with distinct temporal profiles of expression. Finally, we developed tissue-specific gene networks that form a community resource: an Arabidopsis Leaf Time-dependent Atlas (AraLeTa). This allows users to extract gene networks that are confirmed by transcription factor-binding data and specific to certain cell types at certain times of day and at certain developmental stages. AraLeTa is available at <https://regulatorynet.shinyapps.io/araleta/>.

Introduction

The coordination of spatial and temporal gene expression dynamics is fundamental to plant development and response to environmental stimuli. Organisms have distinct gene regulatory programs within different cell types, which regulate the changes in gene expression over diurnal (Yakir et al. 2011) and developmental timescales (Ma et al. 2005). Each of these regulatory programs coordinates the changes in gene expression over time to respond to both intrinsic (time of day and maturity) and extrinsic factors (environmental stimuli) (Sheen 1994). It is the coordination of transcriptional patterns in discrete cell types that provides determinate capacity for cell function in a context of its tissue and organ.

Multiple approaches have been used to measure gene expression over space and time, but each has its drawbacks. Through live image-based assays, it is possible to track gene expression of a small number of genes over both time and space (Shav-Tal et al. 2004). For instance, Gould et al. (2018) identified waves of circadian gene expression originating from the meristems. In contrast, RNA-seq enables researchers to measure gene expression of all mRNAs at once. Many researchers perform high temporal resolution RNA-seq time-series experiments to infer how gene expression changes over time (e.g. Krouk et al. 2010; Cortijo et al. 2017; Ezer and Keir 2019; Balcerowicz et al. 2021). However, these studies do not capture the cell-type-specific changes in gene expression. Additionally, over 80% of leaf expression in a bulk RNA-seq sample originates from mesophyll cells,

Received November 28, 2023. Accepted February 09, 2024. Advance access publication March 1, 2024.

© The Author(s) 2024. Published by Oxford University Press on behalf of American Society of Plant Biologists.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Open Access

which will mask gene expression patterns from other cell types, such as in the vasculature and epidermis tissues (Endo et al. 2014). Bulk RNA-seq also masks heterogeneity within mesophyll cell populations (Procko et al. 2022). Increasingly, single-cell and tissue-specific RNA-seq have been used to characterize cell-type-specific gene expression patterns in leaf (Liu et al. 2020; Kim et al. 2021; Lopez-Anido et al. 2021), root (Shahan et al. 2022), and meristem (Neumann et al. 2022) tissue, but it is prohibitively expensive to perform these in a high-resolution time series. Lee et al. (2023) have developed a developmental single-cell atlas, but it covers only 5 vegetative stages and is therefore not at a similar temporal resolution as existing bulk RNA-seq resources and would be too expensive to replicate under a wide range of experimental conditions. Recent efforts are underway to construct further Plant Cell Atlases based on single-cell analysis (Plant Cell Atlas Consortium et al. 2021), and it is important to find methods to best utilize these kinds of resources, especially when investigating processes that occur over time.

Our work here demonstrates the potential of integrating single-cell RNA-seq data and high-resolution time-series data to unravel cell-type-specific gene networks in *Arabidopsis* (*Arabidopsis thaliana*). Drawing inspiration from cancer cell dynamics research (Newman et al. 2019), we used CIBERSORTx, a powerful technique that combines single-cell RNA-seq and bulk RNA-seq data (Newman et al. 2019) and that outperforms other deconvolution methods in a large benchmarking study (Sutton et al. 2022). We inferred relative expression in various cell types across samples and estimated cell-type-specific gene expression values. By applying these techniques, we gained insights into the dynamics of expression within different cell types across varying temporal scales. Although Procko et al. (2022) identified 4 subpopulations of mesophyll with unclear distinguishing markers, our analysis revealed changes in their relative expression levels over diurnal and developmental scales, raising questions about cell-state changes and activity-level variations over time, as well as the relative light sensitivity of different mesophyll cell states. Moreover, differences in the relative expression of mesophyll subgroups between bolted and unbolted plants were observed. By incorporating transcription factor-binding data (O'Malley et al. 2016), we have provided a valuable resource for the *Arabidopsis* community: a leaf cell-type network available at <https://regulatorynet.shinyapps.io/araleta/>. We also highlight relevant portions of this network during diurnal and developmental timescales.

Results

Detection of cell-type transcriptional activity in bulk RNA-seq by utilizing single-cell RNA-seq data

First, we wish to confirm that we can accurately predict proportions of cell types in bulk RNA-seq samples using single-cell RNA-seq data from *A. thaliana* by training

CIBERSORTx on a training set of single-cell RNA-seq cells and then testing its accuracy on deconvolving simulated bulk RNA-seq samples constructed from subsets of the remaining cells (Fig. 1A). When we simulated bulk RNA samples containing a single-cell type, we correctly identified all tissue types, except for hydathodes, as these were often misclassified as mesophyll cells (Supplementary Fig. S1, A and B). We chose to use Procko et al. (2022) because it had a larger number of leaf cells than many other single-cell datasets, and therefore, we were able to include rarer cell types. Moreover, the authors did not artificially enrich their selection of specific cell types, and therefore the ratio of cell types would be roughly similar to that found in natural leaves. In this paper, we named the 3 mesophyll cell clusters (specifically, Clusters 1, 3, and 4) from Procko et al. (2022) as mesophyll Groups 1, 2, and 3, respectively. We also named the 3 clusters of unknown type (termed Clusters 10, 11, and 16) as unknown Groups 1, 2, and 3, respectively (Supplementary Fig. S1A).

Next, we simulated bulk RNA-seq samples with mixed cell types. We accurately predicted the relative abundance of simulated cell types between these samples (Supplementary Table S1), with Pearson's $R > 0.8$ for all cell types except for 2 unidentified cell types and sieve cells (Supplementary Fig. S2, Fig. 1B). However, we note that CIBERSORTx is best at predicting the relative amounts of cell types between samples, rather than the relative proportions of cell types within a sample—a known issue with gene expression deconvolution algorithms (Sutton et al. 2022). It consistently overestimated the proportion of stressed cells and underestimated the proportion of other cell types (Fig. 1C). These results confirm that CIBERSORTx, a technique initially developed for mammalian research (Newman et al. 2019), can be applied to plant systems.

CIBERSORTx was also successful at predicting the tissue composition of simulated bulk RNA-seq samples generated from other single-cell RNA-seq experiments (Fig. 1D, Supplementary Fig. S3). We chose to test our model on cells in Xia et al. (2022) as these were isolated using microdissection, so that we would be able to know the spatial localization of these cells. In contrast, in marker-based selection, there is always a risk that the marker is also expressed in an off-target cell type. Interestingly, microdissected spongy mesophyll cells from Xia et al. (2022) tended to be classified as mesophyll cells from Group 1, while the palisade cells were more evenly split among the 3 mesophyll groups that Procko et al. (2022) previously reported (Fig. 1D). This suggested that atlases of single-cell RNA-seq expression can serve as reusable resources in the community for estimating cell-type compositions of bulk samples.

Next, we explored the parameters of the signature matrix constructed from the training set to help with deconvolving the bulk RNA-seq samples. CIBERSORTx selected 4950 genes for use in predicting the cell-type proportion (Supplementary Fig. S4), and these have significant enrichment ($P < 0.01$) for Gene Ontology (GO) terms associated

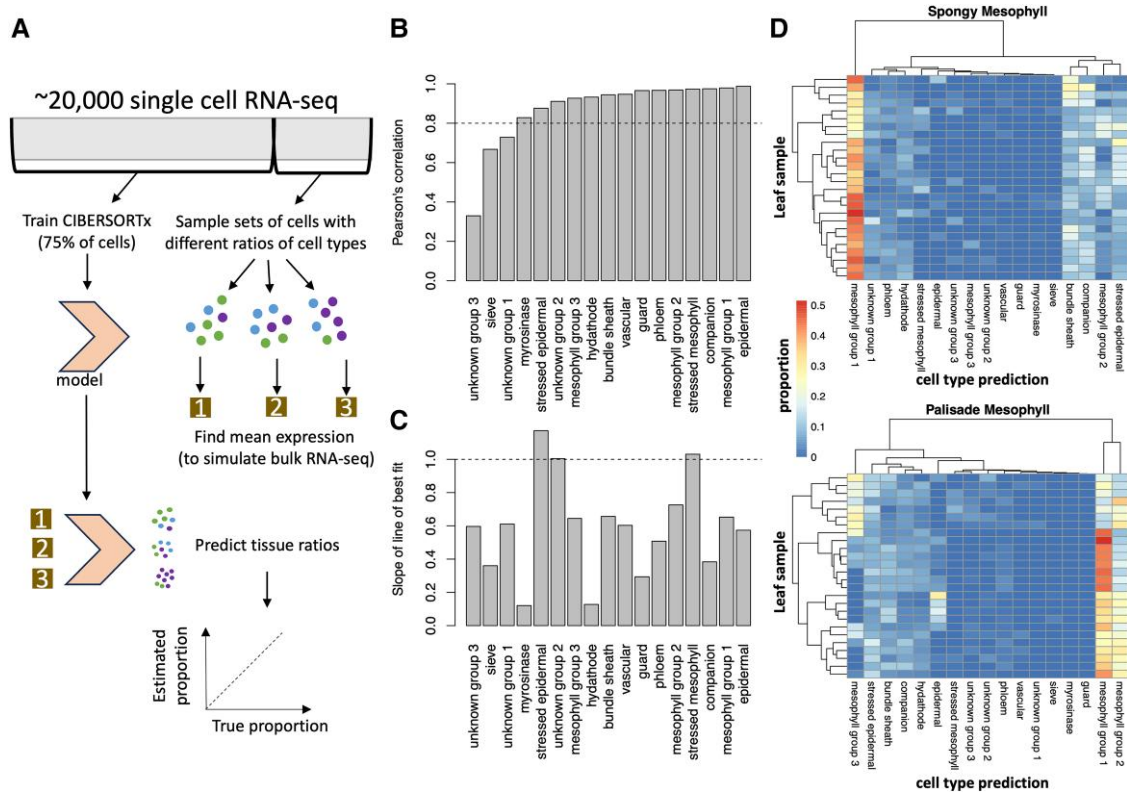


Figure 1. CIBERSORTx predictions for simulated bulk RNA-seq samples. **A)** Seventy-five percent of cells per cluster were used to train a CIBERSORTx model, while the remaining 25% of cells were repeatedly subsampled to generate 500 simulated bulk RNA-seq samples with mixed cell types (see the Materials and methods section). **B)** Pearson's correlation between the true cell-type proportions in the simulated bulk RNA-seq sample and the tissue proportions predicted by CIBERSORTx. All the correlations were statistically significant ($P < 0.05$), with Pearson's correlations >0.8 highlighted by the horizontal bar. **C)** The slope of the line of best fit for a comparison between the predicted and the true cell-type proportions (scatterplots in Supplementary Fig. S2). Values close to 1 (horizontal line) are most accurate. Taken together, these results suggest that CIBERSORTx can predict the relative proportions within the same cell type between RNA-seq samples, but that it consistently overpredicts or underpredicts certain cell types within an RNA-seq sample. **D)** Predicted cell-type proportions in microdissected leaf samples, with each row representing a leaf and each column representing a different predicted cell type. Note that phloem is short for phloem parenchyma.

with ion binding, catalytic activity, structural constituents of chromatin, transmembrane transporter activity, response to stimulus/stress, and amino acid metabolic processes (see Supplementary Table S2 [signature matrix] and Supplementary Table S3 [gProfiler outputs]). These cell-type-specific processes appear to be capable of distinguishing cell-type proportions.

Across developmental scales, transcriptional activity shifts from epidermal to vascular cell types

We next sought to identify how cell-type transcriptional activities changed over different temporal scales. For this, we applied the signature matrices that we previously inferred from single-cell RNA-seq data to interpret bulk RNA-seq time-series datasets. It is important to note that deconvolution algorithms like CIBERSORTx do not find the proportion of cells of each type but rather the proportion of transcripts that is attributable to different cell types. Changes in cell-type proportion over time could be attributable to changes

in the total number of cells at a certain time, increased activity of certain cell types, or cells changing their expression profile to mimic the expression of a different cell type. For conciseness, we will refer to the output of CIBERSORTx as the “activity” of specific cell types.

Next, we investigated whether the proportion of cell type expression levels varied across a developmental timescale (Fig. 2A, Supplementary Fig. S5A and Table S4), utilizing a leaf developmental time-series RNA-seq dataset (Woo et al. 2016). The reference single-cell RNA-seq experiment was performed on the first true leaves at 17 d postgermination, and therefore, it only captures cell-type-specific transcriptional activity at a single snapshot. During the growth-to-senescence transition, we detected a decrease in epidermal expression and an increase in vascular expression. Rarer cell types, like guard cells and hydathodes, seemed to have a higher proportion early in development, possibly because these cell types get diluted as the leaf expands (letsvaart et al. 2017) or because these rarer cell types have gene expression profiles that mimic other cell types

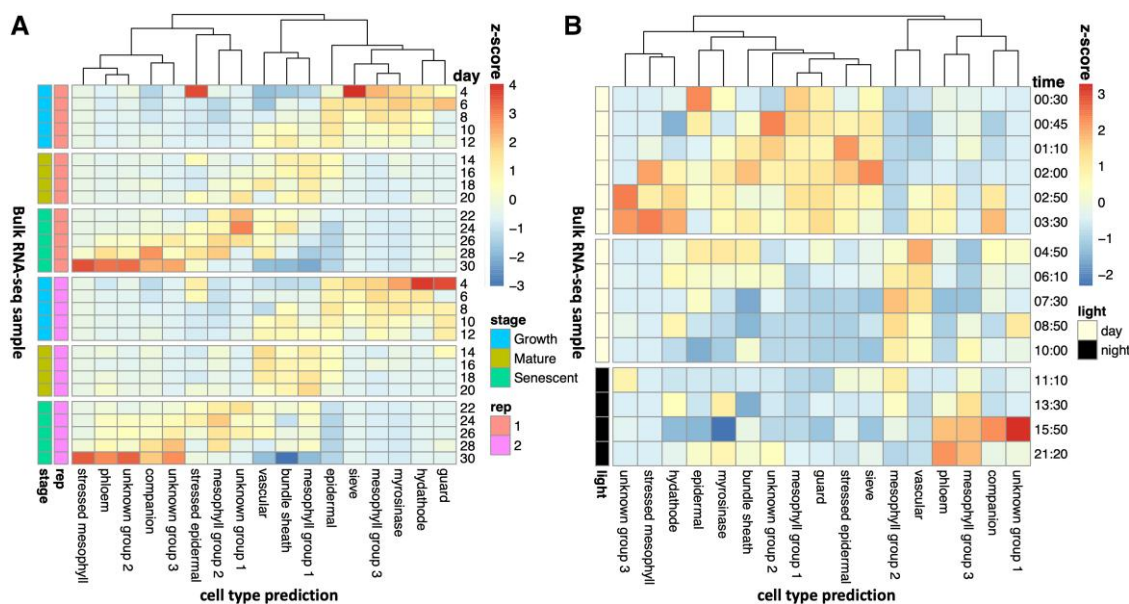


Figure 2. Shifts in cell-type proportions over developmental and diurnal time series. We predicted the proportion of cell types in **A**) developmental time series (Woo et al. 2016) and **B**) diurnal time series (Hickman et al. 2017), utilizing a reference leaf scRNA-seq dataset (Procko et al. 2022). In **A**), each row represents a different RNA-Seq sample, with the proportions normalized using z-scores. The developmental stage and the biological replicate are indicated by the colored bars. In **B**), the proportions of cell types were predicted independently for each of the 4 biological replicates, and then these were averaged for each row and normalized by z-scores. The times are relative to dawn.

when mature, as previously shown for guard cells (Adrian et al. 2015). Phloem parenchyma expression and stressed mesophyll expression were maximized in the senescing leaves compared with other developmental stages. This is consistent with phloem parenchyma cells contributing to nutrient redistribution in the senescing leaf and with the transformation of neighboring cell types into phloem cells during senescence (Hunziker et al. 2019). Although there is currently no leaf-specific developmental single-cell atlas to compare our results with, Lee et al. (2023) performed a whole rosette single-cell RNA-seq in 3 developmental timepoints. Both this dataset and CIBERSORTx show a decline in bundle sheath and vascular cells from Day 17 onward. Lee et al. (2023) showed that the proportion of epidermal cells stays stable, as the plant transitions to senescence, suggesting that our observed decline in epidermal cells may reflect a change in transcriptional activity, rather than a decline in cell count. Our results contradict Lee et al. (2023) in terms of phloem parenchyma and companion cell expression, which we predict increase and Lee et al. (2023) predict decline. These discrepancies may result from the activity of phloem parenchyma cells in nonleaf tissue in the Lee et al. (2023) dataset, as it includes whole rosettes. Moreover, Lee et al. (2023) appeared to have a larger proportion of senescing cell types in Day 17 rosettes than at older ages, and therefore, the sampled rosettes at this timepoint may have been stressed, making direct comparisons challenging.

Intriguingly, the 3 groups of nonstressed mesophyll cells also displayed distinct temporal patterns. Group 3 mesophylls had peak expression during the growth phase,

Group 1 mesophylls had peak expression at the mature leaf phase, and Group 2 mesophylls had peak expression during the mature-leaf-to-senescence transition period. Group 1 cells were found most frequently in the leaves, as in Procko et al. (2022), which is consistent with their sample collection timepoint of 17 d. This analysis suggests that mesophyll-specific expression profiles transition among the 3 different states identified by Procko et al. (2022) over developmental timescales.

Diurnal oscillations of mesophyll, epidermal, vascular, and phloem transcriptional activity

Additionally, we were interested in how the cell-type-specific gene expression patterns varied across diurnal timescales, as measured in a leaf diurnal time-series RNA-seq dataset (Hickman et al. 2017; see Fig. 2B, Supplementary Figs. S5B, S6, and Table S5). Procko et al. (2022) grew their leaves under continuous light to minimize the impacts of the clock, but plant clocks are synchronized by the initial seed imbibement and so they will still experience consistent daily oscillations (Zhong et al. 1998). Even if averages of transcripts appear “arrhythmic” after prolonged plant acclimation to constant conditions, the cells that pattern tissues are still robustly rhythmic, although asynchronous from each other (Yakir et al. 2011; Gould et al. 2018). Thus, the individual cells sampled by Procko et al. (2022) will each be in an unknown phase of the circadian clock.

Epidermal expression is maximized in the morning. Meanwhile, vascular expression peaks in the afternoon. We

hypothesize that this is when the plant is most water-stressed, due to the tendency for this time of day to have a higher temperature. Phloem parenchyma expression peaks at the end of the night when passive loading of sugars through plasmodesmata is replaced with active apoplasmic loading (Wei et al. 2021). Interestingly, epidermal stress cell expression peaks during the ZT1 dawn burst of expression (Balcerowicz et al. 2021), while mesophyll stress expression peaks a few hours afterward (ZT2-4). One hypothesis is that epidermal cells become stressed by the lights suddenly turning on within the growth cabinet, while mesophyll cells become stressed because of reactive oxygen species accumulation as a result of photosynthesis.

The 3 nonstressed mesophyll groups primarily have peak expression at different times of the day. The mesophyll cells that are predicted to be active during early development (Group 3) are also predicted to be active late at night. The mesophyll cells that were active in mature leaves (Group 1) were also active in the morning. A comparison with Xia et al. (2022) suggests that this mesophyll group may be enriched for spongy mesophyll cells. The mesophyll group that was most active during the transition to senescence (Group 2) was also active during the afternoon and early evening. The mesophyll subgroups each appear to have distinct temporal transcriptional profiles, both on a diurnal scale and on developmental scale.

Expression of time-of-day–dependent light-sensitive genes across cell types

Next, we evaluated whether the cell types had different levels of expression of light-responsive genes. We hypothesized that Group 1 mesophyll cells may include more of the genes that are induced by light exposure at the end of the night, as these cells are most active in the morning. To evaluate this hypothesis, we analyzed the expression of light-induced genes (Rugnone et al. 2013) in the Procko et al. (2022) single-cell dataset. Consistent with our hypothesis, genes that are induced by light exposure at night (Fig. 3A) tended to be found in mesophyll Group 1 (morning/mature phase mesophyll group). This set of mesophyll Group 1 genes included genes encoding NIGHT LIGHT–INDUCIBLE AND CLOCK–REGULATED1,3 (LNK1, LNK3), which help entrain the circadian clock in response to light in the morning (Xie et al. 2014), SALT TOLERANCE (STO), which is a BBX family protein involved in the morning dawn burst of expression (Balcerowicz et al. 2021), and members of the light-harvesting complex in photosynthesis (LHCB2.4, LHCA4). Thus, mesophyll clusters indeed harbored notable light-induced genes.

A different set of genes that are light-sensitive at night were more highly expressed in stressed mesophyll cells. These included genes encoding DNAJ, a heat shock protein (Pulido and Leister 2017), SERINE/ARGININE RICH-LIKE PROTEIN 45A (SR45a), a stress-induced splicing factor that regulates anthocyanin accumulation (Gulledge et al. 2012;

Albaqami 2023), and MULTIPROTEIN BRIDGING FACTOR 1C (MBF1C), whose expression is elevated in response to a wide range of stresses (Lee and Bailey-Serres 2019). These results suggest that light exposure earlier than anticipated may induce a light stress response, in addition to activating morning mesophyll expression, and that these 2 responses impact the activity of 2 different sets of genes in 2 different cell types. In contrast, genes that are induced by light exposure after an extended night (Fig. 3B) are primarily expressed either in the mesophyll or in stressed epidermal cells.

Despite the plants in Procko et al. (2022) being exposed to continuous light, some cell types continue to express genes that are repressed under light exposure. Specifically, unknown Groups 2 and 3 contain genes that have decreased expression under light exposure at night like *SENESCENCE1* (*SEN1*), *DARK INDUCIBLE 10* (*DIN10*), and *GLUTAMINE-DEPENDENT ASPARAGINE SYNTHASE 1* (*ASN1*), which are all also induced by senescence (Fujiki et al. 2001). Coupled with our previous observations that their activity peaks during late senescence (Fig. 2A), we propose that these cell types are associated with senescence. Stressed epidermal cells also contain high levels of light-repressed genes. Genes that have reduced expression under light exposure are not expressed highly in the cell types that are predominantly found enriched during the night (phloem, mesophyll Group 3, and companion cells, Fig. 2B), but this may be because the scRNA-seq was performed in plants under continuous light (Fig. 3, C and D). This suggests that it may be wise to perform scRNA-seq on more realistic diurnal conditions, sampling multiple times a day, to adequately capture the cell-specific transcription over the course of a day.

Cell-type-specific regulatory program during bolting

We have observed that there are differences in the transcriptional activity of cell types across a developmental and a diurnal timescale. Next, we decided to focus on the changes that happen during a rapid developmental transition, specifically bolting, which coincides with the start of senescence (Redmond et al. 2023). Mimicking the pattern observed during the developmental time series (Fig. 2A), the relative expression of mesophyll Group 3 (the growth-phase mesophyll) went down, while the expression of mesophyll Groups 1 and 2 (the mature phase and senescence mesophyll) went up over pseudotime (Fig. 4A, Supplementary Fig. S7, A and B and Table S6), where pseudotime refers to the predicted ordering of individual plants in Redmond et al. (2023) over a developmental trajectory on the basis of bulk RNA-seq data (Fig. 4B, Supplementary Table S6). Also consistent with the developmental time series, we observe a decrease in epidermal and sieve cell activity. The consistency between these cell-type activity predictions between the developmental atlas (Woo et al. 2016) and the bolting pseudotime (Redmond et al. 2023) suggests that these developmental shifts in cell activity are robust and confirms that the pseudotime is effectively ordered by a developmental trajectory.

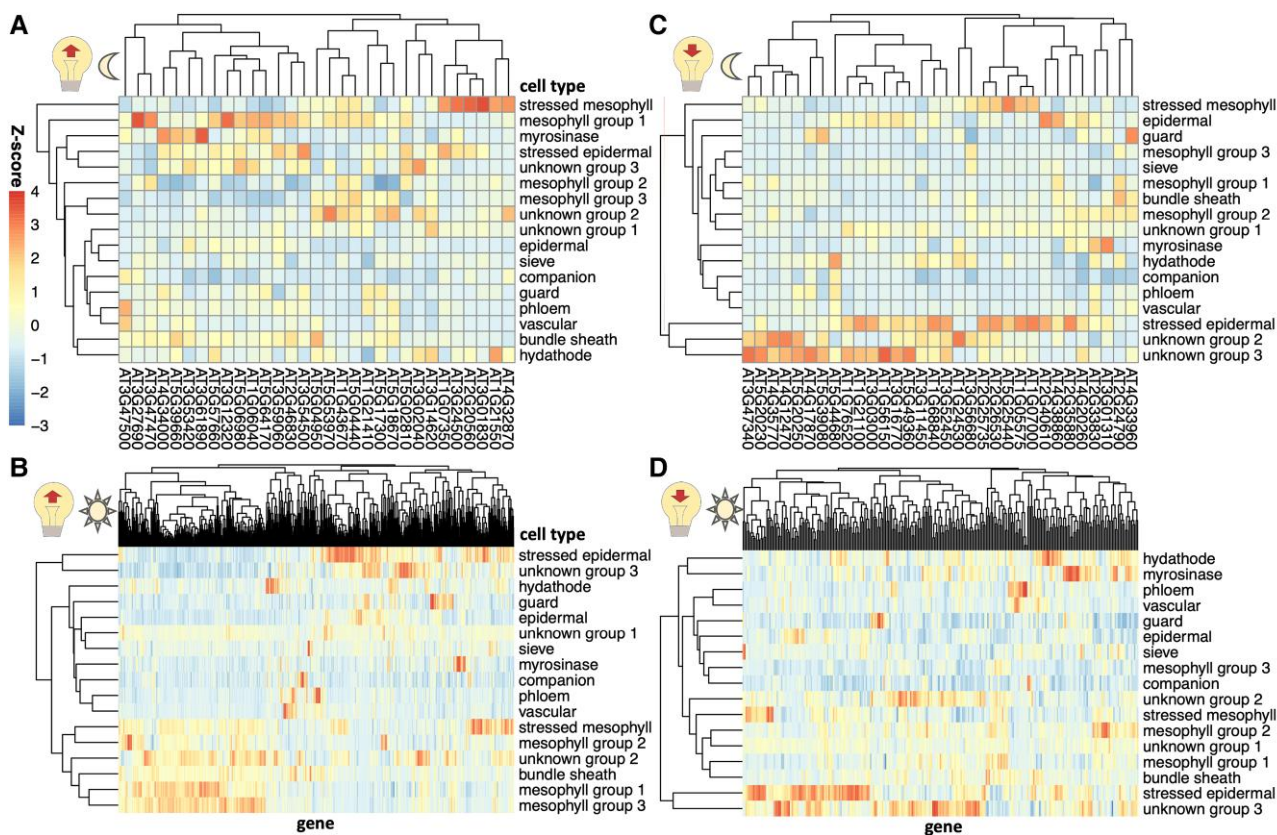


Figure 3. The expression of light-sensitive genes in single-cell leaf data. Sets of genes were selected that are light-induced **A**) and **B**) or light-repressed **C**) and **D**) during either a nocturnal light treatment **A**) and **C**) or a light treatment after an extended night **B**) and **D**), based on [Rugnone et al. \(2013\)](#). The mean expression of these genes in cells in each cluster in [Procko et al. \(2022\)](#) was calculated, and z-scores were calculated for each column.

We were curious whether the changes in cell types over time were associated more with the plant's development or with its other traits such as biomass or leaf size. For each cell type, the Spearman's correlation coefficient was calculated between the proportion of that cell type in each plant and that plant's pseudotime, wet biomass, or leaf area ([Fig. 4C](#), [Supplementary Fig. S8](#)). The proportion of vascular cells, bundle sheath, and sieve cells were more strongly correlated with biomass than with pseudotime, suggesting that the size of the plant may be associated with water and sugar transport. On the other hand, mesophyll, epidermal, and companion cells were more associated with the pseudotime and therefore may be more closely associated with development.

Next, we investigated the cell-type-specific processes taking place during bolting, utilizing the capacity of CIBERSORTx to impute cell-type-specific expression from bulk RNA-seq data. The cell-type-specific genes increased or decreased their expression at different points in the pseudotime using the bulk RNA-seq data from [Redmond et al. \(2023\)](#) ([Fig. 5A](#), [Supplementary Fig. S9A](#)), suggesting that at least part of the differences in the timings of expression of genes across pseudotime could be a result of different activity levels of different cell types. Most of

the genes were identified as being associated with mesophyll Group 2 ([Fig. 5B](#), [Supplementary Fig. S9B](#)). There are some consistencies with the mean expression levels of these genes in the scRNA-seq dataset ([Procko et al. 2022](#)), specifically among mesophyll Group 2 cells and companion cells for genes that increase with pseudotime and with mesophyll Group 2/3 and sieve cells for genes that decrease with pseudotime ([Fig. 5C](#), [Supplementary Fig. S9C](#)). However, there are also some inconsistencies, such as being unable to reconstitute the epidermal expression pattern. Over all genes, the imputed expressions are relatively consistent with the known gene expression values ([Supplementary Fig. S10](#)).

First, we analyzed the GO terms of cell-type-specific genes whose expression increased over pseudotime ([Supplementary Table S7](#)). Those associated with mesophyll Group 2 tended to be found in membranes and involved in adenosine triphosphate, guanine triphosphate, and nicotinamide adenine dinucleotide+ binding (GO:0005524, GO:0005096, and GO:0003953) and the regulation of endocytosis (KEGG:04144). Genes found in the epidermis were enriched via vesicle-mediated transport (GO:0016192). Vascular genes were associated with sulfur metabolism (KEGG:00920) and response to nutrient levels (GO:0031669).

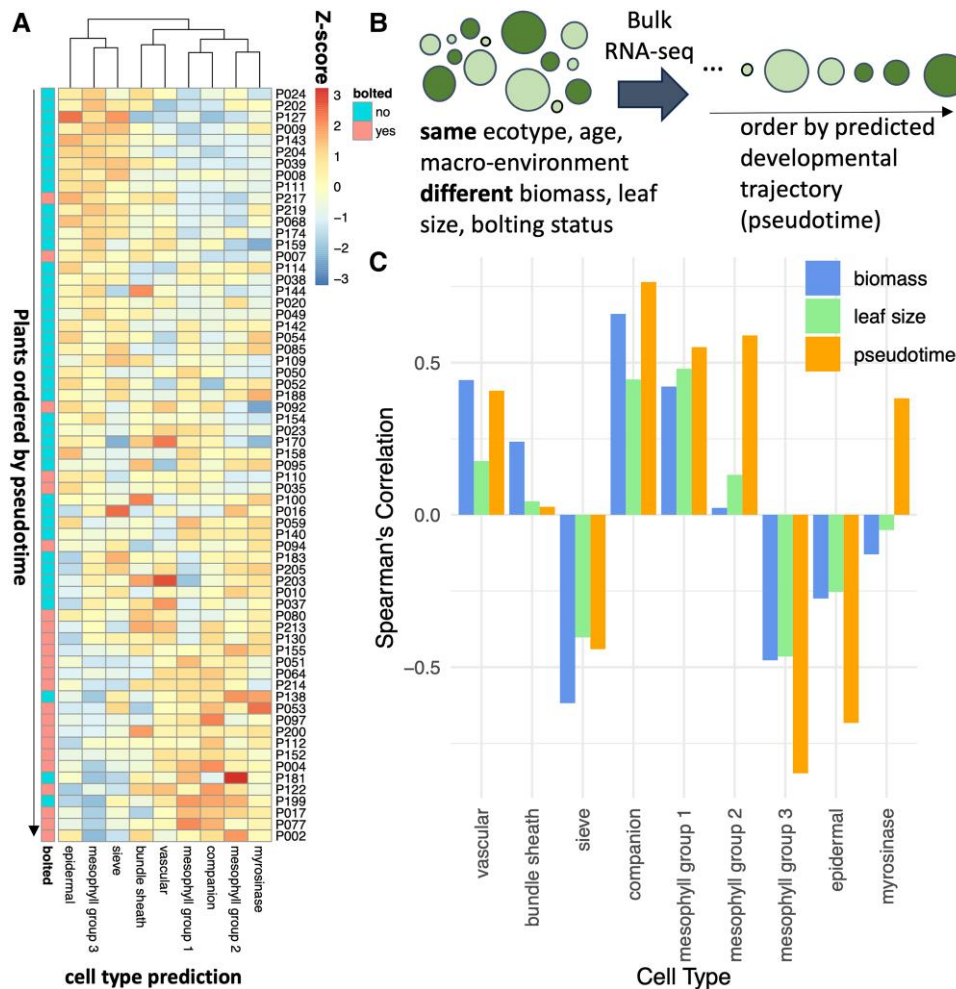


Figure 4. Cell activity changes during bolting. **A**) We predicted the proportion of cell types in plants immediately before and after bolting (Redmond et al. 2023), utilizing a reference leaf scRNA-seq dataset (Procko et al. 2022). Each row represents a different RNA-seq sample (representing a single plant), with the proportions normalized using z-scores. **B**) The plants in **A**) are ordered by pseudotime, which was calculated by Redmond et al. (2023) as an arrangement of 65 individual plants sampled along a developmental trajectory. **C**) For each cell type, the Spearman ranked correlation was calculated between the proportion of that cell type in each plant versus another trait of the plant (biomass, leaf size, or pseudotime).

Among genes that reduced their expression over pseudotime (Supplementary Table S8), there was enrichment in chloroplast and photosynthesis-related GO terms in epidermal and mesophyll cell types. Epidermal cells were also enriched in porphyrin (a pigment) metabolism (KEGG:00860) and amino acid metabolism-related processes (KEGG:00300). Mesophyll Group 3 was associated with the cell cycle (GO:0007049) and DNA replication initiation (GO:0006270), while both Groups 2 and 3 were enriched for components of the ribosome (GO:0005840) compared with other cell types. We noted that mesophyll Group 3 was also expressed in early development and at night, and therefore, potentially this cell type was in a dividing and growing state.

AraLeTA: an Arabidopsis Leaf Time-dependent Atlas

Our previous results have demonstrated that there are different regulatory programs that are active at different times

of day, different developmental stages, and different cell types. As a resource for the community, we have developed AraLeTA (Arabidopsis Leaf Time-dependent Atlas; <https://regulatorynet.shinyapps.io/araleta/>), which can be used to identify the portions of the Arabidopsis gene regulatory network that are active in different contexts, by filtering based on expression values (Fig. 6). As its basis, it utilizes the DNA affinity purification sequencing (DAP-seq) network developed by O'Malley et al. (2016), but it enables the user to filter the network by the age, time of day, and cell types of interest, highlighting edges in which both the source and the target are expressed above a threshold value under the relevant conditions. This kind of thresholding approach uses similar criteria for filtering the gene regulatory network as what was used by Ferrari et al. (2022) and has been shown to enrich the gene network for true edges. The AraLeTA network can be downloaded or visualized as a heatmap or graph.

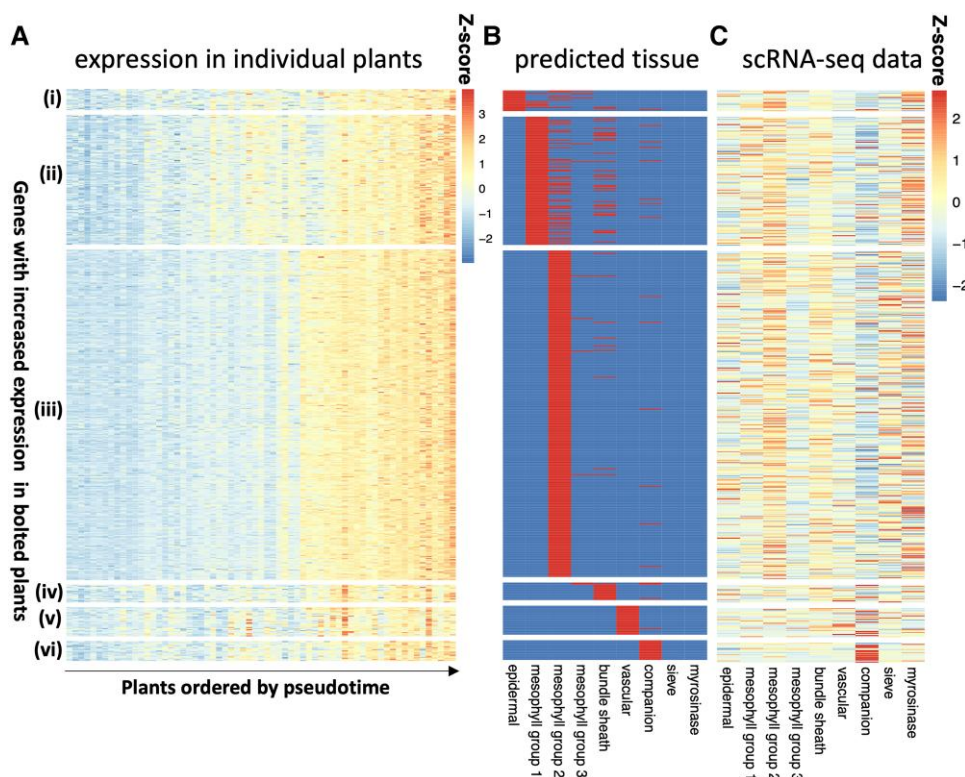


Figure 5. Cell-type-specific gene expression of bolting-related genes. Using the high-resolution imputing function in CIBERSORTx, we predicted cell-type-specific expression of genes that were differentially expressed in bolted/unbolted plants. In this study, we show the results for the genes that have higher expression levels in bolted plants, with the inverse gene set shown in [Supplementary Fig. S9](#). **A**) The z-score of the expression of these genes in the bulk RNA-seq experiment ([Redmond et al. 2023](#)), grouped by the cell type in which they were predicted to be expressed and ordered by pseudotime. **B**) The cell-type assignment, with red indicating that a gene is expected to be found in that cell type. **C**) The z-score of the mean expression of these genes in the scRNA-seq dataset ([Procko et al. 2022](#)).

As it is often difficult to make sense of large “hairball” networks, we have incorporated PAFway additionally, a program for identifying the relationships between functional terms within the topology of the network ([Mahjoub and Ezer 2020](#)). This allows us to find patterns in the rewiring of the network over time, within specific cell types.

To highlight the utility of AraLeTA, we focused on 4 conditions: (i) young (4 to 14 d postgermination) mesophyll (Groups 1, 2, and 3) cells, (ii) old (20 to 30 d postgermination) mesophyll cells, (iii) young vascular (bundle sheath, phloem, vascular, companion, guard, and sieve) cells, and (iv) old vascular cells ([Fig. 6A](#)). To select thresholds for our networks, we varied the single-cell and bulk RNA-seq threshold parameters and evaluated their impact on the number of significant functional edges ($P < 0.05$) selected by the PAFway network, choosing thresholds that generated either local maxima or inflection points in the size of the PAFway network ([Supplementary Fig. S11](#)). As expected, young mesophyll cells included many associations with light response ([Fig. 6B](#)). Interestingly, there were more defense-, jasmonic acid-, and salicylic acid (SA)-related edges in vascular cells, reflecting the transport mechanism of these plant hormones ([Supplementary Fig. S12, A and B](#)). The old plant networks were centered on SA, a senescence-associated plant

hormone, and defense ([Fig. 6C](#)). Many of the genes with GO terms related to defense are also involved in senescence, especially in response to necrotrophic pathogens ([Woo et al. 2016](#)). [Redmond et al. \(2023\)](#) also demonstrated that processes such as ubiquitination, endocytosis, cell cycle, translation, and response to redox are all perturbed at the onset of senescence, and these terms all appear in the network associated with older plants. These general trends confirm that the filtering criteria are adequately selecting relevant portions of the DAP-seq network.

Discussion

Cell-type-specific transcriptional activity: what does it mean?

In this study, we illustrated how we used CIBERSORTx to infer the transcriptional activity of different cell types within a bulk RNA-seq sample. One must be nuanced in our interpretation of what a change in transcriptional activity means from a biological standpoint ([Supplementary Fig. S13](#)). When one cell type is predicted to have a higher proportion than another cell type, there are many alternative explanations: (i) there may be a higher proportion of one cell type in the sample relative to the other ([Supplementary Fig. S13A](#)). (ii)

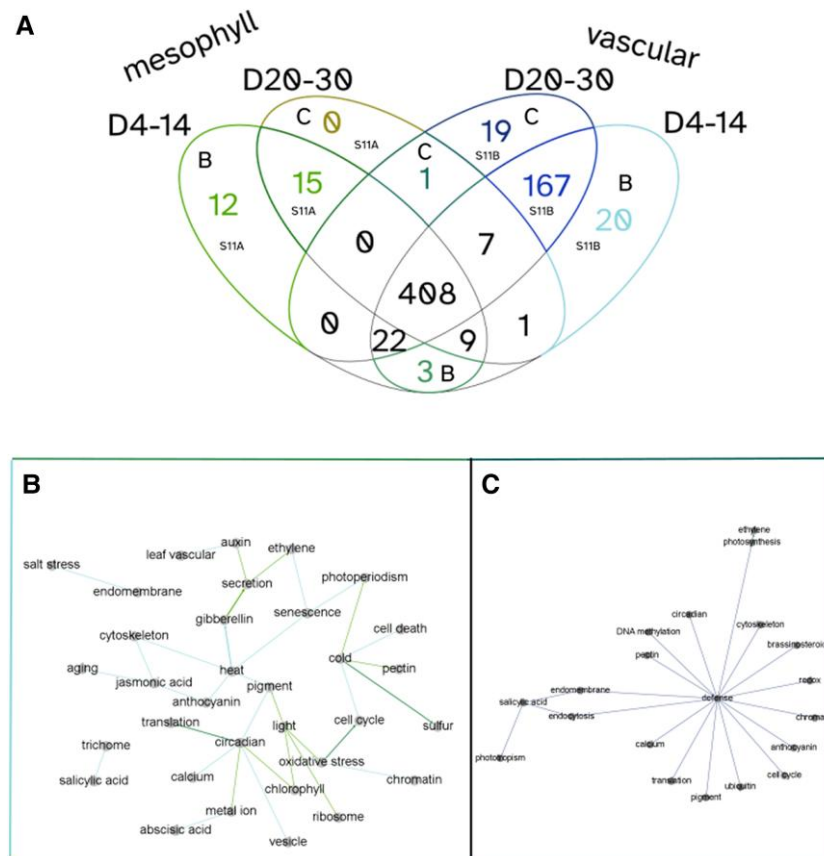


Figure 6. PAFway networks across different tissue types at different ages. Using PAFway, we generated networks of functional terms to assess changes in function between tissue types as they age. **A**) Mesophyll (Groups 1, 2, and 3) and vascular-associated (bundle sheath, phloem, vascular, companion, guard, and sieve) cells showed a large overlap in functional edges across age groups, with some network edges that are unique to specific tissues and ages. **B**) A network of functional terms that are associated with young mesophyll and vascular cells. **C**) A network of functional terms associated with old mesophyll and vascular cells. Together, these show the changing network over time, suggesting that these cells perform different physiological functions as they age. The colors of the edges in **B**) and **C**) correspond to the associated colored sections of the Venn diagram in **A**).

One cell type may have a higher overall transcription rate than the other (Supplementary Fig. S13B). (iii) There may be only one cell type that fluctuates between transcriptional states. For instance, it is unclear whether it is possible for cells that belong to a particular mesophyll group to transition into a cell from a different mesophyll group (Supplementary Fig. S13C); (iv) cells may express a transcriptional pattern that is similar to other cell types at certain timepoints, and the algorithm may be misassigning the transcriptional activity to these cell-type categories (Supplementary Fig. S13D). To make this latter point more concrete: while most photosynthesis in *Arabidopsis* takes place in mesophyll cells, other cell types also contain some low density of chloroplasts (Ishikawa et al. 2020). The increased transcriptional activity of photosynthetic processes in these other cell types at certain times of day may lead to an overestimation of mesophyll transcriptional activity by CIBERSORTx, as photosynthesis-related processes are used as markers for mesophyll transcriptional activity. We choose to use the term “cell-type-specific transcriptional activity” to highlight the fact that we are analyzing the relative frequency of

cell-type-specific diurnal patterns of transcription. Despite the multiple different underlying processes that can lead to changes in cell-type-specific transcriptional activity, as we define here, this concept is still very useful, because it provides us with a summary of the distinct transcriptional programs that are occurring in samples that contain mixes of different cell types.

Shifts in cell transcriptional activity over different timescales

We show that the transcriptional activity of several different cell types varies both diurnally and developmentally, as summarized in Fig. 7, A and B. This suggests that waves of expression in bulk RNA-seq time series may represent waves of transcriptional activities of different cell types, rather than waves of regulatory activity—an assumption that is often held when inferring gene regulatory networks based on bulk RNA-seq data (Huynh-Thu and Geurts 2018). Time-series single-nucleus RNA-seq (Lee et al. 2023) and spatial transcriptomics (Giacomello 2021) in plants promise to better distinguish temporal and spatial waves.

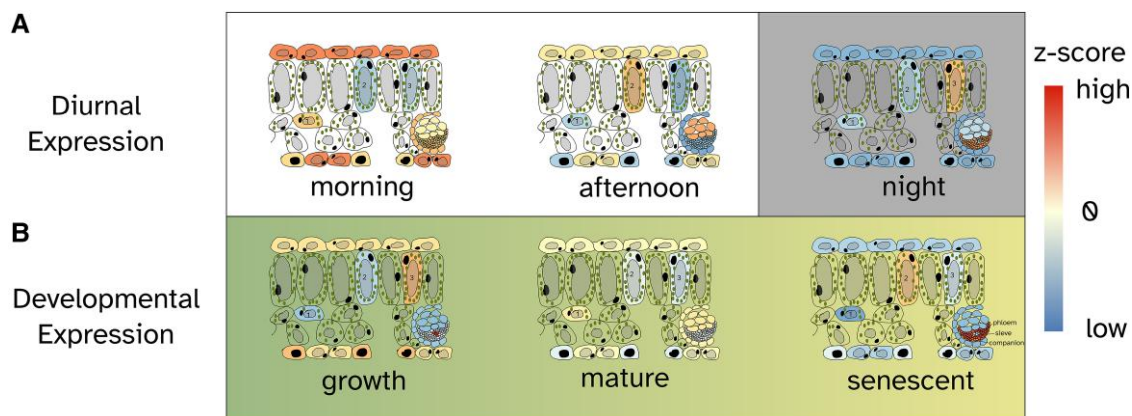


Figure 7. A graphical abstract showing the cell-type-specific expression pattern. This refers to the relative changes in gene expression values across diurnal **A**) and developmental **B**) time series. The colors correspond to the color scale of the heatmaps in Fig. 2.

Patterns of mesophyll transcriptional states

Single-cell RNA-seq and single-nucleus RNA-seq have consistently identified subclustering of mesophyll cells, both in *Arabidopsis* and in rice (Kim et al. 2021; Wang et al. 2021; Procko et al. 2022; Tenorio Berrío et al. 2022; Xia et al. 2022). These studies have enabled the identification of cell-type-specific markers for different subgroups of mesophyll; however, the main roles of these subclusters have not been fully characterized, and the mesophyll groups do not have obvious physical differences. Our analysis augments Procko et al.'s (2022) analysis of the 3 mesophyll groups (as summarized in Table 1), by revealing the time in which each group is most likely transcriptionally active, both on a diurnal scale and on a developmental scale.

It is important to remember that different parts of the same individual plant may perceive time in different ways, a form of intraorganismal heterochrony. Gould et al. (2018) found that there are waves of circadian expression that spread spatially from the meristem and root tips to the remainder of the plant, which could result in circadian asynchrony between cells in the leaf under free-running conditions. Asynchrony of the clock may also arise between cells in the leaf due to the entrainment of the clock in the morning through exogenous sugars (Haydon et al. 2013). There is also intraorganismal heterochrony in relation to biological age. Different leaves will be at different stages of development at the same chronological time (Efroni et al. 2008; Redmond et al. 2023). It may even be possible for different parts of the same leaf to have different biological ages at the same chronological age, due to localized stresses. The individual transcriptional profiles sampled using a single-cell or single-nucleus RNA-seq method may represent not only heterogeneous cell types but also heterogeneous biological times. By combining single-cell RNA-seq data with high temporal resolution bulk RNA-seq, we may be able to begin to dissect this kind of temporal heterogeneity too.

CIBERSORTx enables greater exploitation of RNA-seq datasets

While single-cell RNA-seq is becoming increasingly common in plants, it is still too expensive and cumbersome to perform these experiments over high-resolution time series, under a wide range of environmental conditions and under a wide range of genotypes. In addition, there are tens of thousands of existing bulk RNA-seq datasets available that could vary in their cell-type proportion (Zhang et al. 2020). In this study, we show that plant scRNA-seq datasets can be used to train a CIBERSORTx model that can be used to algorithmically dissect the bulk RNA-seq samples by their cell types. This will confer *Arabidopsis* researchers who do not have the capacity to do scRNA-seq the ability to exploit these data for enhancing their research. Additionally, this kind of analysis could be used to help decide which timepoints would be most informative for performing single-cell RNA-seq, using a similar computational approach as in Ezer and Keir (2019). Moreover, AraLeTA is a useful platform for utilizing existing bulk RNA-seq time-series data, single-cell RNA-seq data, and transcription factor-binding data to isolate spatial and temporal regulatory processes of interest. Together, CIBERSORTx and AraLeTA provide us with an atlas of leaf expression in *Arabidopsis* over cell type and over time.

Materials and methods

Running CIBERSORTx

Seurat was used to visualize and process the single-cell RNA-seq data (Hao et al. 2021). In the matrix provided to CIBERSORTx (Newman et al. 2019) to generate the signature matrix, genes that were very lowly or very highly expressed were filtered out, with $\ln(\text{total read count})$ between 4 and 10. For the purposes of testing CIBERSORTx on the simulated bulk RNA-seq data, a random sample of 75% of the cells was used to generate the signature matrix and a Pearson's correlation, and correlation tests were deployed on the remaining 25% testing set. For the rest of the paper, all cells were used to

Table 1. A summary of the properties of mesophyll cell types

Mesophyll	Period of development	Time of day	Light-sensitive genes	Impact of bolting on relative expression
1	Mature	Morning	<i>LNK1, LNK3, STO, LHCB2, LHCA4</i>	Increase
2	Senescence	Afternoon to early evening		Increase
3	Growth	Night		Decrease
Stressed	Late senescence	2 to 4 h after dawn	<i>DNAJ, SR45a, MBF1C</i>	N/A

generate the signature matrix. The cluster designations used in Procko et al. (2022) were used as the phenotype classes. In all cases, we disabled quantile normalization (which is recommended for RNA-seq data) and used 100 runs for the permutation tests. For imputing the gene expressions in each sample in the bolted plants, we used only the 9 most abundant cell types and “other” and focused only on genes that were identified as upregulated or downregulated in bolted/not bolted plants, according to Redmond et al. (2023).

Generating simulated bulk RNA-seq samples

To simulate bulk RNA-seq samples composed on a single cell type, the remaining cells from Procko et al. (2022) that were not used to generate the model were randomly partitioned into 2 equal groups for each cell type and summed together, forming Replicates 1 and 2 for each cell type. We also simulated bulk RNA-seq samples from the single-cell microdissected samples from Xia et al. (2022). In this case, we summed over the single-cell RNA-seq samples from each cell type associated with a specific leaf.

To simulate the mixed bulk RNA-seq samples, we not only wanted to have somewhat realistic cell-type proportions but also have a wide variation in cell types between samples. For each cell type, we first calculated the proportion of cells of that type in the single-cell RNA-seq dataset. Then, per cell type, we simulated the proportion of cells within the bulk RNA-seq sample according to a normal distribution pattern, with a mean and *sd* equal to the cell type proportion within the single-cell RNA-seq dataset. To calculate the final number of cells per cell type, we rounded up any negative values to 0 and multiplied this by the average number of cells we wanted per sample (600). We randomly selected the specified number of cells from each cell type and then calculated the sum of transcript counts per gene for the associated set of cells and considered that to be our simulated bulk RNA-seq sample.

Bioinformatics analysis

All clustering was performed using the default hierarchical clustering parameters of the heatmap package (Kolde 2019). GO enrichment analysis was performed using gProfiler (Kolberg et al. 2023). Networks were visualized using Gephi (Bastian et al. 2009) and iGraph (Csárdi et al. 2023). PAFway (Mahjoub and Ezer 2020) was used to generate a graph of the related functional annotations of the basis of the text analysis of the GO Slim annotations provided by Berardini et al. (2015) and the topology of the DAP-seq gene network (O'Malley et al. 2016). AraLeTA was developed as a Shiny

App (Chang et al. 2023). The PAFway networks were visualized using Gephi (Bastian et al. 2009). All codes for generating the figures are available at <https://github.com/stressedplants/AraletaScripting/>. All codes for the Shiny app are available at <https://github.com/stressedplants/AraLeTA/>.

Accession numbers

The data underlying this article are available in the NCBI's Gene Expression Omnibus under data libraries under accession numbers GSE43616, GSE182414, and GSE184511 and in the NCBI's Sequence Read Archive under PRJNA668247, PRJNA224133, and PRJNA395645. All TAIR IDs are found in Supplementary Tables S2, S4 to S6.

Acknowledgments

The Viking cluster was used during this project, which is a high-performance compute facility provided by the University of York. We are grateful for computational support from the IT Services and the Research IT Team, University of York.

Author contributions

D.E. designed the research. All authors performed the research. D.E. and G.Y.W.V. contributed new analytic/computational tools. D.E., G.Y.W.V., K.M., and W.C. analyzed the data. All authors wrote the paper.

Supplementary data

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

Supplementary Figure S1. The ability of CIBERSORTx to detect “mock” bulk RNA-seq samples, consisting of a single cell type.

Supplementary Figure S2. The ability of CIBERSORTx to predict the relative composition of “mock” bulk RNA-seq samples with mixed tissues.

Supplementary Figure S3. The assignment of cell type to microdissected tissues.

Supplementary Figure S4. The analysis of a signature matrix.

Supplementary Figure S5. Unscaled heatmaps from Fig. 2.

Supplementary Figure S6. The impact of methyl jasmonate (MeJA) on cell-type proportions.

Supplementary Figure S7. Unscaled heatmaps from Fig. 4.

Supplementary Figure S8. Scatterplots underlying Fig. 4C.

Supplementary Figure S9. Cell-type-specific gene expression of genes that have low expression in leaves of bolted plants.

Supplementary Figure S10. A comparison of predicted gene expression values and imputed values.

Supplementary Figure S11. The identification of scRNA and bulk RNA thresholding values.

Supplementary Figure S12. A comparison of predicted gene expression values and imputed values.

Supplementary Figure S13. Factors influencing transcriptional activity.

Supplementary Table S1. Simulated bulk RNA-seq samples.

Supplementary Table S2. A signature matrix learned by CIBERSORTx.

Supplementary Table S3. A gProfiler GO summary of signature genes.

Supplementary Table S4. Predicted transcriptional activities of cell types across developmental scales.

Supplementary Table S5. Predicted transcriptional activities of cell types across diurnal scales and with MeJA exposure.

Supplementary Table S6. Predicted transcriptional activities of cell types in bolted and unbolted plants.

Supplementary Table S7. A gProfiler GO summary of cell-type-specific genes (more expressed in bolting cells).

Supplementary Table S8. A gProfiler GO summary of cell-type-specific genes (less expressed in bolting cells).

Funding

We would like to acknowledge the following funding sources: the Royal Society (RGS\R2\212345: D.E.), Biotechnology and Biological Sciences Research Council (Impact Acceleration Account) IAA (BB/S506795/1: D.E.), Biotechnology and Biological Sciences Research Council (Responsive Mode) (BB/V006665/1: D.E. and S.J.D.), the Biotechnology and Biological Sciences Research Council (White Rose Doctoral Training Partnership) (BB/T007222/1: G.Y.W.V. and E.J.R.), and GenerationResearch (W.C.).

Conflict of interest statement. None declared.

References

- Adrian J, Chang J, Ballenger CE, Bargmann BOR, Alassimone J, Davies KA, Lau OS, Matos JL, Hachez C, Lanctot A, et al.** Transcriptome dynamics of the stomatal lineage: birth, amplification, and termination of a self-renewing population. *Dev Cell*. 2015;**33**(1): 107–118. <https://doi.org/10.1016/j.devcel.2015.01.025>
- Albaqami M.** The splicing factor SR45 negatively regulates anthocyanin accumulation under high-light stress in *Arabidopsis thaliana*. *Life*. 2023;**13**(6):1386. <https://doi.org/10.3390/life13061386>
- Balcerowicz M, Mahjoub M, Nguyen D, Lan H, Stoeckle D, Conde S, Jaeger KE, Wigge PA, Ezer D.** An early-morning gene network controlled by phytochromes and cryptochromes regulates photomorphogenesis pathways in *Arabidopsis*. *Mol Plant*. 2021;**14**(6): 983–996. <https://doi.org/10.1016/j.molp.2021.03.019>

- Bastian M, Heymann S, Jacomy M.** 2009. Gephi: an open source software for exploring and manipulating networks. In: Proceedings of the International AAAI Conference on Web and Social Media, California, USA, vol. 3, p. 361–362.
- Berardini TZ, Reiser L, Li D, Mezheritsky Y, Muller R, Strait E, Huala E.** The Arabidopsis information resource: making and mining the “gold standard” annotated reference plant genome. *Genesis*. 2015;**53**(8):474–485. <https://doi.org/10.1002/dvg.22877>
- Chang W, Cheng J, Allaire J, Sievert C, Schloerke B, Xie Y, Allen J, McPherson J, Dipert A, Borges B.** shiny: web application framework for R. R package version 1.7.4.1. 2023.
- Cortijo S, Charoensawan V, Brestovitsky A, Buning R, Ravarani C, Rhodes D, van Noort J, Jaeger KE, Wigge PA.** Transcriptional regulation of the ambient temperature response by H2A.Z nucleosomes and HSF1 transcription factors in *Arabidopsis*. *Mol Plant*. 2017;**10**(10):1258–1273. <https://doi.org/10.1016/j.molp.2017.08.014>
- Csárdi G, Nepusz T, Müller K, Horvát S, Traag V, Zanini F, Noom D.** 2023. igraph for R: R interface of the igraph library for graph theory and network analysis. [accessed 2023 Nov 1]. Zenodo. <https://zenodo.org/record/7682609>.
- Efroni I, Blum E, Goldshmidt A, Eshed Y.** A protracted and dynamic maturation schedule underlies *Arabidopsis* leaf development. *Plant Cell*. 2008;**20**(9):2293–2306. <https://doi.org/10.1105/tpc.107.057521>
- Endo M, Shimizu H, Nohales MA, Araki T, Kay SA.** Tissue-specific clocks in *Arabidopsis* show asymmetric coupling. *Nature*. 2014;**515**(7527):419–422. <https://doi.org/10.1038/nature13919>
- Ezer D, Keir J.** NITPicker: selecting time points for follow-up experiments. *BMC Bioinformatics*. 2019;**20**(1):166. <https://doi.org/10.1186/s12859-019-2717-5>
- Ferrari C, Manosalva Pérez N, Vandepoele K.** MINI-EX: integrative inference of single-cell gene regulatory networks in plants. *Mol Plant*. 2022;**15**(11):1807–1824. <https://doi.org/10.1016/j.molp.2022.10.016>
- Fujiki Y, Yoshikawa Y, Sato T, Inada N, Ito M, Nishida I, Watanabe A.** Dark-inducible genes from *Arabidopsis thaliana* are associated with leaf senescence and repressed by sugars. *Physiol Plant*. 2001;**111**(3): 345–352. <https://doi.org/10.1034/j.1399-3054.2001.1110312.x>
- Giacomello S.** A new era for plant science: spatial single-cell transcriptomics. *Curr Opin Plant Biol*. 2021;**60**:102041. <https://doi.org/10.1016/j.pbi.2021.102041>
- Gould PD, Domijan M, Greenwood M, Tokuda IT, Rees H, Kozma-Bognar L, Hall AJ, Locke JC.** Coordination of robust single cell rhythms in the *Arabidopsis* circadian clock via spatial waves of gene expression. *Elife*. 2018;**7**:e31700. <https://doi.org/10.7554/eLife.31700>
- Gulledge AA, Roberts AD, Vora H, Patel K, Loraine AE.** Mining *Arabidopsis thaliana* RNA-seq data with integrated genome browser reveals stress-induced alternative splicing of the putative splicing regulator SR45a. *Am J Bot*. 2012;**99**(2):219–231. <https://doi.org/10.3732/ajb.1100355>
- Hao Y, Hao S, Andersen-Nissen E, Mauck WM III, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zager M, et al.** Integrated analysis of multimodal single-cell data. *Cell*. 2021;**184**(13):3573–3587.e29. <https://doi.org/10.1016/j.cell.2021.04.048>
- Haydon MJ, Mielczarek O, Robertson FC, Hubbard KE, Webb AAR.** Photosynthetic entrainment of the *Arabidopsis thaliana* circadian clock. *Nature*. 2013;**502**(7473):689–692. <https://doi.org/10.1038/nature12603>
- Hickman R, Van Verk MC, Van Dijken AJH, Mendes MP, Vroegop-Vos IA, Caarls L, Steenbergen M, Van der Nagel I, Wesselink GJ, Jironkin A, et al.** Architecture and dynamics of the jasmonic acid gene regulatory network. *Plant Cell*. 2017;**29**(9): 2086–2105. <https://doi.org/10.1105/tpc.16.00958>
- Hunziker P, Halkier BA, Schulz A.** *Arabidopsis* glucosinolate storage cells transform into phloem fibres at late stages of development. *J Exp Bot*. 2019;**70**(16):4305–4317. <https://doi.org/10.1093/jxb/erz176>
- Huynh-Thu VA, Geurts P.** dynGENIE3: dynamical GENIE3 for the inference of gene networks from time series expression data. *Sci Rep*. 2018;**8**(1):3384. <https://doi.org/10.1038/s41598-018-21715-0>

- Ietswaart R, Rosa S, Wu Z, Dean C, Howard M.** Cell-size-dependent transcription of FLC and its antisense long non-coding RNA COOLAIR explain cell-to-cell expression variation. *Cell Syst.* 2017;4(6):622–635.e9. <https://doi.org/10.1016/j.cels.2017.05.010>
- Ishikawa H, Yasuzawa M, Koike N, Sanjaya A, Moriyama S, Nishizawa A, Matsuoka K, Sasaki S, Kazama Y, Hayashi Y, et al.** Arabidopsis PARC6 is critical for plastid morphogenesis in pavement, trichome, and guard cells in leaf epidermis. *Front Plant Sci.* 2020;10:1665. <https://doi.org/10.3389/fpls.2019.01665>
- Kim J-Y, Symeonidi E, Pang TY, Denyer T, Weidauer D, Bezdruczyk M, Miras M, Zöllner N, Hartwig T, Wudick MM, et al.** Distinct identities of leaf phloem cells revealed by single cell transcriptomics. *Plant Cell.* 2021;33(3):511–530. <https://doi.org/10.1093/plcell/koaa060>
- Kolberg L, Raudvere U, Kuzmin I, Adler P, Vilo J, Peterson H. g.** Profiler—interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Res.* 2023;51(W1):W207–W212. <https://doi.org/10.1093/nar/gkad347>
- Kolde R.** pheatmap: Pretty heatmaps. R package version 1.0.12. 2019.
- Krouk G, Mirowski P, LeCun Y, Shasha DE, Coruzzi GM.** Predictive network modeling of the high-resolution dynamic plant transcriptome in response to nitrate. *Genome Biol.* 2010;11(12):R123. <https://doi.org/10.1186/gb-2010-11-12-r123>
- Lee TA, Bailey-Serres J.** Integrative analysis from the epigenome to transcriptome uncovers patterns of dominant nuclear regulation during transient stress. *Plant Cell.* 2019;31(11):2573–2595. <https://doi.org/10.1105/tpc.19.00463>
- Lee TA, Nobori T, Illouz-Eliaz N, Xu J, Jow B, Nery JR, Ecker JR.** A single-nucleus atlas of seed-to-seed development in Arabidopsis. *bioRxiv* 533992. <https://doi.org/10.1101/2023.03.23.533992>, 24 March 2023, preprint: not peer reviewed.
- Liu Z, Zhou Y, Guo J, Li J, Tian Z, Zhu Z, Wang J, Wu R, Zhang B, Hu Y, et al.** Global dynamic molecular profiling of stomatal lineage cell development by single-cell RNA sequencing. *Mol Plant.* 2020;13(8):1178–1193. <https://doi.org/10.1016/j.molp.2020.06.010>
- Lopez-Anido CB, Vatén A, Smoot NK, Sharma N, Guo V, Gong Y, Anleu Gil MX, Weimer AK, Bergmann DC.** Single-cell resolution of lineage trajectories in the Arabidopsis stomatal lineage and developing leaf. *Dev Cell.* 2021;56(7):1043–1055.e4. <https://doi.org/10.1016/j.devcel.2021.03.014>
- Ma L, Sun N, Liu X, Jiao Y, Zhao H, Deng XW.** Organ-specific expression of Arabidopsis genome during development. *Plant Physiol.* 2005;138(1):80–91. <https://doi.org/10.1104/pp.104.054783>
- Mahjoub M, Ezer D.** PAFway: pairwise associations between functional annotations in biological networks and pathways. *Bioinformatics.* 2020;36(19):4963–4964. <https://doi.org/10.1093/bioinformatics/btaa639>
- Neumann M, Xu X, Smaczniak C, Schumacher J, Yan W, Blüthgen N, Greb T, Jönsson H, Traas J, Kaufmann K, et al.** A 3D gene expression atlas of the floral meristem based on spatial reconstruction of single nucleus RNA sequencing data. *Nat Commun.* 2022;13(1):2838. <https://doi.org/10.1038/s41467-022-30177-y>
- Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, Khodadoust MS, Esfahani MS, Luca BA, Steiner D, et al.** Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol.* 2019;37(7):773–782. <https://doi.org/10.1038/s41587-019-0114-2>
- O'Malley RC, Huang SC, Song L, Lewsey MG, Bartlett A, Nery JR, Galli M, Gallavotti A, Ecker JR.** Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell.* 2016;165(5):1280–1292. <https://doi.org/10.1016/j.cell.2016.04.038>
- Plant Cell Atlas Consortium;** Jha SG, Borowsky AT, Cole BJ, Fahlgren N, Farmer A, Huang SC, Karia P, Libault M, Provart NJ, et al. Vision, challenges and opportunities for a plant cell atlas. *Elife.* 2021;10:e66877. <https://doi.org/10.7554/eLife.66877>
- Procko C, Lee T, Borsuk A, Bargmann BOR, Dabi T, Nery JR, Estelle M, Baird L, O'Connor C, Brodersen C, et al.** Leaf cell-specific and single-cell transcriptional profiling reveals a role for the palisade layer in UV light protection. *Plant Cell.* 2022;34(9):3261–3279. <https://doi.org/10.1093/plcell/koac167>
- Pulido P, Leister D.** Novel DNAJ-related proteins in *Arabidopsis thaliana*. *New Phytol.* 2017;217(2):480–490. <https://doi.org/10.1111/nph.14827>
- Redmond EJ, Ronald J, Davis SJ, Ezer D.** Single-plant-omics reveals the cascade of transcriptional changes during the vegetative-to-reproductive transition. *bioRxiv* 557157. <https://doi.org/10.1101/2023.09.11.557157>, 12 September 2023, preprint: not peer reviewed.
- Rugnone ML, Faigón Soverna A, Sanchez SE, Schlaen RG, Hernando CE, Seymour DK, Mancini E, Chernomoretz A, Weigel D, Más P, et al.** LNK genes integrate light and clock signaling networks at the core of the Arabidopsis oscillator. *Proc Natl Acad Sci U S A.* 2013;110(29):12120–12125. <https://doi.org/10.1073/pnas.1302170110>
- Shahan R, Hsu C-W, Nolan TM, Cole BJ, Taylor IW, Greenstreet L, Zhang S, Afanassiev A, Vlot AHC, Schiebinger G, et al.** A single-cell Arabidopsis root atlas reveals developmental trajectories in wild-type and cell identity mutants. *Dev Cell.* 2022;57(4):543–560.e9. <https://doi.org/10.1016/j.devcel.2022.01.008>
- Shav-Tal Y, Singer RH, Darzacq X.** Imaging gene expression in single living cells. *Nat Rev Mol Cell Biol.* 2004;5(10):855–862. <https://doi.org/10.1038/nrm1494>
- Sheen J.** Feedback control of gene expression. *Photosynth Res.* 1994;39(3):427–438. <https://doi.org/10.1007/BF00014596>
- Sutton GJ, Poppe D, Simmons RK, Walsh K, Nawaz U, Lister R, Gagnon-Bartsch JA, Voineagu I.** Comprehensive evaluation of deconvolution methods for human brain gene expression. *Nat Commun.* 2022;13(1):1358. <https://doi.org/10.1038/s41467-022-28655-4>
- Tenorio Berrío R, Verstaen K, Vandamme N, Pevernagie J, Achon I, Van Duyse J, Van Isterdael G, Saeys Y, De Veylder L, Inzé D, et al.** Single-cell transcriptomics sheds light on the identity and metabolism of developing leaf cells. *Plant Physiol.* 2022;188(2):898–918. <https://doi.org/10.1093/plphys/kiab489>
- Wang Y, Huan Q, Li K, Qian W.** Single-cell transcriptome atlas of the leaf and root of rice seedlings. *J Genet Genomics.* 2021;48(10):881–898. <https://doi.org/10.1016/j.jjgg.2021.06.001>
- Wei X-Y, Collings DA, McCurdy DW.** Review: more than sweet: new insights into the biology of phloem parenchyma transfer cells in Arabidopsis. *Plant Sci.* 2021;310:110990. <https://doi.org/10.1016/j.plantsci.2021.110990>
- Woo HR, Koo HJ, Kim J, Jeong H, Yang JO, Lee IH, Jun JH, Choi SH, Park SJ, Kang B, et al.** Programming of plant leaf senescence with temporal and inter-organellar coordination of transcriptome in Arabidopsis. *Plant Physiol.* 2016;171(1):452–467. <https://doi.org/10.1104/pp.15.01929>
- Xia K, Sun HX, Li J, Li J, Zhao Y, Chen L, Qin C, Chen R, Chen Z, Liu G, et al.** The single-cell stereo-seq reveals region-specific cell subtypes and transcriptome profiling in Arabidopsis leaves. *Dev Cell.* 2022;57(10):1299–1310.e4. <https://doi.org/10.1016/j.devcel.2022.04.011>
- Xie Q, Wang P, Liu X, Yuan L, Wang L, Zhang C, Li Y, Xing H, Zhi L, Yue Z, et al.** LNK1 and LNK2 are transcriptional coactivators in the Arabidopsis circadian oscillator. *Plant Cell.* 2014;26(7):2843–2857. <https://doi.org/10.1105/tpc.114.126573>
- Yakir E, Hassidim M, Melamed-Book N, Hilman D, Kron I, Green RM.** Cell autonomous and cell-type specific circadian rhythms in Arabidopsis. *Plant J.* 2011;68(3):520–531. <https://doi.org/10.1111/j.1365-3113X.2011.04707.x>
- Zhang H, Zhang F, Yu Y, Feng L, Jia J, Liu B, Li B, Guo H, Zhai J.** A comprehensive online database for exploring ~20,000 public Arabidopsis RNA-seq libraries. *Mol Plant.* 2020;13(9):1231–1233. <https://doi.org/10.1016/j.molp.2020.08.001>
- Zhong HH, Painter JE, Salomé PA, Straume M, McClung CR.** Imbibition, but not release from stratification, sets the circadian clock in Arabidopsis seedlings. *Plant Cell.* 1998;10(12):2005–2017. <https://doi.org/10.1105/tpc.10.12.2005>