Root Development
&
Acquired Thermotolerance Regulatory Targets
in *Arabidopsis thaliana*

Ken Jean-Baptiste

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2020

Reading Committee:

Christine Queitsch, Chair

Phil Green

Takato Imaizumi

Program Authorized to Offer Degree:

Genome Science

University of Washington

Abstract

Root Development and Acquired Thermotolerance Regulatory Targets in *Arabidopsis thaliana*

Ken Jean-Baptiste

Chair of the Supervisory Committee:
Associate Professor Christine Queitsch
Department of Genome Sciences

In this thesis, I will describe my work in two research topics using *Arabidopsis thaliana*: root development and acquired thermotolerance.

1. Arabidopsis roots are an excellent model for studying cell development because the root tissue has a mix of undifferentiated cells, developing cells, and mature cells. The roots are structured such that undifferentiated cells reside at the root tip, and mature cells are closer to the plant shoot. Much of the knowledge gained from researching plant roots was gained from microscopy based approaches and transgenic plant lines. These approaches are quite labor intensive and have not been applied in many plant/ crop species. Here I present research profiling the transcriptome of the root using single cell RNA-seq (scRNA-seq). In just one experiment, the scRNA-seq data was able to capture all the major cell types within the root, to recapitulate known developmental changes in transcription associated with each cell type, and to identify novel genes that change over the course of cell development. Next I describe a collaboration project studying the process of lateral root formation, a process in which the developmental state of a mature cell resets to form a population of cells capable of developing into any other cell type in the root. Using scRNA-seq again, genes known to play major roles in lateral root were identified as well as genes that previously had no association with lateral root development.

2. Acquired thermotolerance is a conserved mechanism that allows an organism to survive lethal heat stress if they were recently exposed to a milder heat stress. Acquired thermotolerance is a form of transcriptional memory that primes an organism to respond more effectively to heat stress. This is achieved through changes in how protein coding genes and other elements of the genome are regulated. The full extent of loci that are regulated in order to achieve acquired thermotolerance is not known. Here I present work assaying chromatin accessibility in Arabidopsis in various stages of acquired thermotolerance to identify loci whose accessibility changes in response to heat stress. Loci that become more accessible during heat stress are able to interact with transcription factors and other DNA binding molecules. Specifically, I focus on a previously identified phenomenon of heat induced gene body accessibility seen in plants without acquired thermotolerance responding to lethal heat stress. Here I show that gene body accessibility does occur during a mild heat stress, but most heat induced accessibility changes are temporary. I also identify genes that are heat accessible only in plants with acquired thermotolerance. Next, the role the histone variant H2A.Z may play in acquired thermotolerance was explored using an H2A.Z incorporation mutant, *arp6*. It was found that *arp6* had a defect in its memory of prior heat stress and could not maintain acquired thermotolerance as long as a wild type plant could. Lastly, the possibility that acquired thermotolerance changed chromosome conformation was explored. Although heat stress did not produce large changes in interactions between regions of the nuclear genome, it was found that the number of interactions between the nuclear genome and the chloroplast genome increased drastically during lethal heat stress. This increase was prevented in plants with acquired thermotolerance.

Table of Contents

List of Figures

Chapter 2

List of Tables

Acknowledgements

There are many people to thank for how far I've gotten including my parents, my advisor Christine Queitsch, my project collaborators Hardik, and Amy, my thesis committee, the Queitsch lab, particularly my mentors Kerry Bubb and Cristina Alexandre, and the Genome Sciences community. Thanks everybody.

Chapter 1. Introduction

Thesis Overview

In broad terms, my dissertation work focused on analyzing genomic data in *Arabidopsis thaliana* to better understand gene regulation. This took the form of three computational projects, two of which involved studying Arabidopsis root development using single cell RNA-sequencing (scRNA-seq) and the third involved studying heat stress response with the Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq). Here I will provide background for these projects.

The technology used to study the transcriptome has improved rapidly in the past 30 years.

Since the 1990s, researchers have been assaying transcriptomes using microarray technologies, however with the reduced cost of next generation sequencing (NGS) in the early 2000s, the preferred method of studying the transcriptome became RNA-seq (Shendure et al. 2017). One major advantage RNA-seq has over microarray technologies is that RNA-seq allowed for the quantification of RNA sequences, whereas microarray technologies relied on the hybridization of RNA sequences with complementary DNA sequences on a chip to create a fluorescent signal which correlated with the number of hybridizing RNA sequences from the sample of interest. RNA-seq made the statistics surrounding transcriptomics analysis much simpler and allowed for the quantification of previously unknown RNA transcripts.

One limitation of both RNA-seq and microarray technologies is cell resolution. For both methods, RNA is typically extracted from bulk populations of cells and the resulting transcriptome may be a mix of cell types. To obtain tissue/ cell type specific transcriptomes, tissue dissection and cell sorting approaches are often deployed to isolate one tissue/ cell type. Single cell resolution transcriptomics was still desirable, however, since it was difficult to isolate rare cell types with dissection methods and cell sorting relied on prior knowledge of cell type specific markers. In 2009, the first scRNA-seq paper was published describing the

transcriptome of quite literally a single cell, a mouse blastomere (Tang et al. 2009). Over the past 10 years, scRNA-seq technologies have evolved to allow for the sequencing of thousands of cells with just one experiment (Chen et al. 2019). For my dissertation, the scRNA-seq platform, 10X, was used to study Arabidopsis roots (Zheng et al. 2017).

Arabidopsis roots are a great model for studying cell development and scRNA-seq can add to this large body of work

Single cell technologies, particularly scRNA-seq, can allow for great strides in the field of cell development. A prominent example of the use of scRNA-seq was the construction of a gene expression atlas was done in *C. elegans* (Packer et al. 2019). In this study, Packer *et al.* 2019, used scRNA-seq to generate a gene expression profile of every *C. elegans* cell. Taking advantage of decades of cell development research in *C. elegans*, they were able to label each cell in their dataset and to detail the subtle transcriptional changes that occur throughout *C. elegans* development. In order to do this work in other organisms, extensive prior knowledge of cell type specific marker genes must be available. Fortunately, in Arabidopsis, extensive work has been done to generate cell types specific markers for the different cell types in the Arabidopsis root.

The biology of the Arabidopsis root makes it a great model for studying cell development. The root consists of three developmental zones: the meristematic zone, the elongation zone, and the maturation zone. Less developed cells originate from the meristematic zone, grow and elongate in the elongation zone, and become fully developed in the maturation zone (Ishikawa and Evans 1995; Verbelen et al. 2006). Because of the developmental layout of the root, it allows one to study cell development by being able to capture many different cell types and various stages of their development.

The root meristem gives rise to the major cell layers: endodermis (outermost), cortex, endodermis, and stele (innermost), as well as the root specific cells columella

The meristematic zone consists of cells at the root tip and is where cell division and cell fate determination occurs and where cell polarity is determined (Dolan et al. 1993). The root tip contains a small population of cells known as the quiescent center which have stem cell-like properties. Quiescent center cells are a population of slowly dividing cells that have the potential to develop into any other cell type in the root (Clowes 1958). Within a few cell divisions, the dividing quiescent center cells determine their fate, which is influenced by their position relative to other cells in the meristem. This gives rise to a set of cells, referred to as initials, fated to form the major tissue types/ cell layers of the root. There are four initial cell populations in the roots: the columella initials, the epidermis/lateral root cap initials, the cortex/ endodermis initials, and the stele initials (Dolan et al. 1993).

The columella initials give rise to the columella cells. Together with the lateral root cap cells which emerge from the epidermis/lateral root cap initials, the columella serves as the root cap for the root. As the name implies, root cap cells are the most exterior cells at the root tip functioning as a protective dome for the meristematic zone. The columella cells serve as the spearhead of the root tip and the lateral root cap cells are positioned to the side of the columella cells. Root cap cells not only protect the root, they also facilitate the growth propelled movement of the root through soil and are involved in graviperception (McCully 1999; Hawes et al. 2000; Iijima et al. 2008).

As mentioned previously, the epidermis/lateral root cap initials give rise to lateral root cap cells, but they also can develop into the epidermis/epidermal cells. Epidermal cells are the most exterior cell layer in the Arabidopsis root and line the majority of the root with the exception of the root tip. There are two distinct epidermal cell types: hair and non-hair (also referred to as trichoblast and atrichoblast respectively in the literature). Increasing the surface area of the root, hair cells extend out from the epidermis to help gather water and nutrients from the surrounding

soil (Schiefelbein et al. 1992). Nonhair cells do not extend out from the root and simply serve as a protective layer.

The cortex/endodermis initials can develop into either cortex cells or endodermis cells. The cortex cell layer resides between the epidermis layer and the endodermis cell layer. The cortex cells send an intercellular signal to adjacent epidermis cells which influences which epidermis cells develop into hair or nonhair cells (Lee and Schiefelbein 2002). Epidermis cells touching one cortex cell become nonhair cells, while epidermis cells touching two cortex cells become hair cells. Endodermis cells have a number of functions including regulating the flow of water and hormones between cell layers and sending intercellular signals to influence tissue patterning in the innermost cell layer, the stele cell layer (Carlsbecker et al. 2010; Miyashima et al. 2011; Miyashima and Nakajima 2011).

The stele initials are the progenitor cells for the stele/ vascular layer of the root. There are several types of stele cells, but the two most well known stele cells are the xylem and the phloem. As vascular tissues, the xylem cells transport water and nutrients from the roots to the rest of the plant while the phloem transports sugars and other metabolites from the main body to the roots. Surrounding the xylem and the phloem cells and directly adjacent to the endodermis cell layer are the pericycle cells. Pericycle cells are not vascular tissue, but they play roles in loading the xylem cells with nutrients and unloading the phloem cells of metabolites (Oparka 1990; Dannel et al. 1998; Takano et al. 2002; Ross-Elliott et al. 2017). The two types of pericycle cells, the xylem pole pericycle and the phloem pole pericycle, are named after their proximity to either the xylem or the phloem and whether or not they help in xylem loading or phloem unloading. The xylem pole pericycle cells have a secondary role in lateral root formation.

Lateral roots originate from the xylem pole pericycle

Aside from the developmental structure of the plant root, the lateral root formation is another reason why the root system is a great model for plant development. Lateral root formation is a developmental process by which the main plant root grows additional side roots that emerge at various points from the maturation zone. These lateral roots have all the same cell layers and the developmental structures as the main root; these roots serve to increase the plant's capability to intake water and nutrients from the environment (Laskowski et al. 1995). Lateral root formation is dependent on auxin signaling and the xylem pole pericycle cells. Xylem pole pericycle cells, despite being fully differentiated, can reset their development back to a stem-cell like state to form a population of cells capable of forming a new lateral root (Parizot et al. 2008). This population of cells fated to form a lateral root is referred to as the lateral root primordia. The precise molecular mechanisms that cause a xylem pole pericycle cell to become a lateral root primordium is a topic that is actively being studied, but this transition is dependent on the transport and accumulation of auxin, a naturally occurring plant hormone. An auxin signal that originates from the root tip accumulates in some xylem pole pericycle cells just before these cells enter the elongation zones (Casimiro et al. 2001; De Smet et al. 2007; Parizot et al. 2008; Parizot et al. 2012). These cells are the ones with the capability to form lateral root primordia.

Root development research has been driven by microscopy and transgenic plant lines

Prior to my dissertation, much of the knowledge root developmental knowledge was gained through microscopy studies, and the generation of mutant and transgenic plant lines. No scRNA-seq data had been generated in plant roots beforehand. The most relevant Arabidopsis resource for the scRNA-seq data analysis discussed in this paper is a set of cell type specific GFP marker lines (Birnbaum et al. 2003; Laplaze et al. 2005; Nawy et al. 2005; Lee et al. 2006; Brady et al. 2007). These GFP marker lines were enhancer trap lines with some lines using the promoters of known genes and others using sequences whose function is unknown. Using

microdissections, the roots of each GFP marker line were cut into 13 sections longitudinally in an attempt to capture different stages of a cell's development. Microarrays were then used to measure the transcriptomes of these developmental slices from each GFP marker line (Brady et al. 2007). Later on, an RNA-seq data set was generated for each GFP marker line (Li et al. 2016).

These resources required years of research effort to generate and was quite labor intensive. In order to make the GFP lines, many transgenic plant lines had to be created and screened for cell type specific expression of GFP. The emerging technology of scRNA-seq can allow researchers to get single cell resolution of transcriptomes without the need to perform microdissections or to rely on GFP lines. A major motivation of part of my dissertation work was to determine if cell type specific signals could be resolved by performing scRNA-seq on Arabidopsis roots. Cell type specific resources were helpful to me in my dissertation because one of the challenges of scRNA-seq analysis is the classification of cells into the correct cell type. Using the transcriptome data generated from these past experiments, I was able to infer the cell type of each cell in the root scRNA-seq data. From there I was able to identify previously known genes involved in root development as well as genes not annotated for their developmental roles. Afterwards, I applied these expertise to study lateral root formation in roots.

Acquired Thermotolerance is a great example of transcriptional memory

The second topic I studied related to acquired thermotolerance and heat stress. Acquired thermotolerance is a phenomenon by which organisms survive lethal temperature treatments by experiencing a pretreatment at a mild heat stress (Lin et al. 1984; Lindquist and Craig 1988). This phenomenon is observed across all domains of life, although the underlying molecular mechanisms differ between eukaryotes, prokaryotes, and archaea (Trent et al. 1994). As with any heat stress response, acquired thermotolerance requires the rapid upregulation of

HEAT SHOCK PROTEINS (HSPs) to serve as chaperones to protect other proteins from denaturing and the rapid downregulation of other proteins and genes (Sanchez and Lindquist 1990; Parsell et al. 1994; Becker and Craig 1994). Acquired thermotolerance also requires thermomemory, a molecular mechanism that "remembers" prior heat stress. The mechanisms that produce thermomemory are not entirely understood.

In Arabidopsis, HSP101 and HEAT SHOCK TRANSCRIPTION FACTOR A2 (HSFA2) play crucial roles in acquired thermotolerance. HSP101 is required to obtain acquired thermotolerance and HSFA2 is required to maintain acquired thermotolerance via thermomemory (Queitsch et al. 2000; Charng et al. 2007). Constitutive expression of either the *HSP101* gene and the *HSFA2* gene do enhance thermotolerance, but the transcript and protein levels of these constitutive expression lines were significantly less than the levels obtained by pretreated plants (Queitsch et al. 2000; Charng et al. 2007). This implied that these genes, and by extension acquired thermotolerance, had intricate regulation.

In order to identify potential targets for regulation, chromatin accessibility, is typically assayed, i.e. how accessible a given region of the genome is for binding. Previous work in the Queitsch lab assayed chromatin accessibility during a lethal heat stress at 45 ℃ and found that most sites remained unchanged in terms of chromatin accessibility (Sullivan et al. 2014). A small number of loci became inaccessible or less accessible during a heat stress, and another set of loci were only heat accessible. The most striking feature of these heat accessible loci were that entire gene bodies would become accessible. Examples of these heat accessible loci were the gene bodies of *HSP101* and *HSFA2*. To determine if this heat accessible gene body phenomenon occurs during non lethal heat stresses and to identify targets for regulation during acquired thermotolerance, ATAC-seq was performed at various stages of acquired thermotolerance.

We were also interested in some mechanisms by which plants maintained thermomemory. One hypothesis we had was that thermomemory is achieved through a histone

variant known as H2A.Z. H2A.Z is evicted from stress responsive genes upon induction, but is gradually incorporated back after the initial induction (Zhang et al. 2005). To explore this hypothesis, we used H2A.Z incorporation mutant *ACTIN RELATED PROTEIN 6 (arp6)*. We evaluated the thermomemory of this mutant as well as its chromatin accessibility to determine what role, if any, H2A.Z incorporation may play in heat induced gene body accessibility. Another hypothesis we explored was the possibility that heat stress caused changes in chromosome conformation and that these conformational changes were a part of thermomemory. In yeast, studies of transcriptional memory showed that recently expressed genes, whose expression was induced by a stimulus, localized to different areas of the nucleus cause these recently expressed genes to be expressed more quickly in response to a second stimulus (Brickner et al. 2007). This relocalization was dependent on the histone variant H2A.Z. To explore this chromosome conformation hypothesis, we performed Hi-C, a high throughput sequencing method to measure the interaction frequencies between loci, in wild type plants and in *arp6*.

In the next three chapters I will present the following work:

- In Chapter 2, I will discuss my work analyzing one of the first scRNA-seq datasets in Arabidopsis roots, which included finding previously known transcriptional changes associated with development and describing novel transcriptional changes during root development.
- In Chapter 3, I will apply the knowledge and skills discussed in Chapter 2 to analyze lateral root development. This dataset improved upon the previous one in that it had multiple time points which proved helpful in not only identifying cell types, but describing developmental relationships and it utilized microdissections to enrichment for cell types that were not well represented when performing scRNA-seq on the entire root.
- In Chapter 4, I discuss my work analyzing the ATAC-seq data from multiple stages of acquired thermotolerance. Here I determine whether or not the heat induced gene body

accessibility phenomenon occurs during non-lethal heat stresses, explore the role efficient H2A.Z incorporation has in thermomemory and the heat induced gene body accessibility, and whether or not heat stress causes changes in chromosome conformation.

Chapter contributions and public availability

Before moving on, I would like to quickly detail the contributions of every author involved in the work I will discuss. In the work discussed in Chapters 2-4, I primarily was responsible for processing and analyzing sequence data (scRNA-seq for Chapter 2 and Chapter 3; ATAC-seq and Hi-C for Chapter 4). In order for me to analyze data, data must be collected in well designed and well conducted experiments! Below I shall list these individuals who made my analysis possible. I would also like to acknowledge others who assisted in my analysis. Lastly, I and several other co-authors contributed to the writing of Chapter 2 and Chapter 3 while I was the main writer of Chapter 4.

Chapter 2

Chapter 2 is available publicly as:

Jean-Baptiste K, McFaline-Figueroa JL, Alexandre CM, Dorrity MW, Saunders L, Bubb KL, Trapnell C, Fields S, Queitsch C, Cuperus JT. Dynamics of Gene Expression in Single Root Cells of *Arabidopsis thaliana*. Plant Cell. 2019 May;31(5):993-1011. doi: 10.1105/tpc.18.00785. Epub 2019 Mar 28. PMID: 30923229.

J.L.MF and K.L.B. assisted in analyzing a portion of the scRNA-seq pertaining to a heat stress experiment. C.M.A. performed much of the benchwork and experiments which includes growing, and harvesting plant tissue, conducting heat stress experiments, and protoplasting plant tissue for scRNA-seq. M.W.D. assisted in a transcription factor analysis. L.S. assisted in performing a RNA velocity analysis. C.T. and S.F. provided guidance for the project as well as providing feedback for the manuscript. C.Q. and J.T.C. also provided guidance for the project as well as

contributing to the writing of the manuscript and figure design. J.T.C. also performed scRNA-seq with the aid of C.M.A.

Chapter 3

Chapter 3 is currently available as a pre-print on bioRxiv at:

Gala HP, Lanctot A, Jean-Baptiste K, Guiziou S, Chu JC, Zemke JE, George W, Queitsch C, Cuperus JT, Nemhauser JL. A single cell view of the transcriptome during lateral root initiation in *Arabidopsis thaliana*. bioRxiv 2020.10.02.324327; doi: https://www.biorxiv.org/content/10.1101/2020.10.02.324327v2

H.P.G., A.L., S.G., J.C.C., J.E.Z., W.G., and J.T.C performed much of the benchwork for this project which included but is not limited to growing and harvesting plant tissue, constructing transgenic lines for gene validation experiments, and performing scRNA-seq. H.P.G and A.L. also contributed to the writing of the manuscript and figure design as well as assisting in comparing our findings to prior research in lateral root formation. C.Q., J.T.C., and J.L.N. provided guidance for the project and provided feedback on the manuscript.

Chapter 4

Chapter 4 is currently not publicly available, but shall be made available on bioRxiv in the near future. Below are the list of authors/ contributors to this work as well as it's tentative title.

Jean-Baptiste K, Mason GA, Alexandre CM, Bubb KL, Dorrity MW, Rival P, Sullivan AO, Cuperus JT, and Queitsch C. The Role of Chromatin Accessibility, Chromosome Conformation, and Efficient H2A.Z Incorporation in Acquired Thermotolerance in *Arabidopsis thaliana.*

GAM, CMA, MWD, PR, AOS, and JTC performed much of the benchwork for this project which included but is not limited to growing and harvesting plant tissue, heat stressing plants, performing the ATAC-seq and Hi-C protocols, and sequencing. Hi-C,in particular, was performed with the aid of Phase Genomics (https://phasegenomics.com/). CMA, KLB, JTC, and CQ provided feedback on the manuscript. KLB and CQ guided the project.

Chapter 2. Dynamics of gene expression in single root cells of *A. thaliana*

2.1 Abstract

Single-cell RNA-seq can yield high-resolution cell-type-specific expression signatures that reveal new cell types and the developmental trajectories of cell lineages. Here, we apply this approach to *A. thaliana* root cells to capture gene expression in 3,121 root cells. We analyze these data with Monocle 3, which orders single cell transcriptomes in an unsupervised manner and uses machine learning to reconstruct single-cell developmental trajectories along pseudotime. We identify hundreds of genes with cell-type-specific expression, with pseudotime analysis of several cell lineages revealing both known and novel genes that are expressed along a developmental trajectory. We identify transcription factor motifs that are enriched in early and late cells, together with the corresponding candidate transcription factors that likely drive the observed expression patterns. We assess and interpret changes in total RNA expression along developmental trajectories and show that trajectory branch points mark developmental decisions. Finally, by applying heat stress to whole seedlings, we address the longstanding question of possible heterogeneity among cell types in the response to an abiotic stress. Although the response of canonical heat shock genes dominates expression across cell types, subtle but significant differences in other genes can be detected among cell types. Taken together, our results demonstrate that single-cell transcriptomics holds promise for studying plant development and plant physiology with unprecedented resolution.

2.2 Introduction

Many features of plant organs such as roots are traceable to specialized cell lineages and their progenitors (Irish, 1991; Petricka et al., 2012). The developmental trajectories of these lineages have been based on tissue-specific and cell-type-specific expression data derived from tissue dissection and reporter gene-enabled cell sorting (Birnbaum et al., 2003; Brady et al.,

2007; Li et al., 2016). However, tissue dissection is labor-intensive and imprecise, and cell sorting requires prior knowledge of cell-type-specific promoters and genetic manipulation to generate reporter lines. Few such lines are available for plants other than the reference plant *Arabidopsis thaliana* (Rogers and Benfey, 2015). Advances in single-cell transcriptomics can replace these labor-intensive approaches. Single-cell RNA-seq has been applied to heterogeneous samples of human, worm, and virus origin, among others, yielding an unprecedented depth of cell-type-specific information (Cao et al., 2017; Irish, 1991; Packer and Trapnell, 2018; Russell et al., 2018; Trapnell, 2015; Trapnell et al., 2014).

While several examples of single cell RNA-seq have been carried out in *Arabidopsis* (Efroni et al., 2016, 2015; Brennecke et al., 2013), they were restricted to only a few cells or cell types. No whole organ single-cell RNA-seq has been attempted in any plant species. The *Arabidopsis* examples focused on root tips, finely dissecting the dynamics of regeneration or assaying technical noise across single cells in a single cell type. Thus, a need exists for larger scale technology that allows a more complete characterization of the dynamics of development across many cell types in an unbiased way. Such technology would increase our ability to assay cell types without reporter gene-enabled cell sorting, identify developmental trajectories, and provide a comparison of how different cell types respond to stresses or drugs. Several high-throughput methods have been described for sequencing of RNA at a high throughput of single cells. Most of these, including most droplet-based methods, rely on the 3’ end capture of RNAs. However, unlike with bulk RNA-seq, the data from single cell methods can be sparse, such that genes with low expression can be more difficult to study. Here, we take advantage of expression data from root-specific reporter lines in *A. thaliana* (Birnbaum et al., 2003; Brady et al., 2007; Cartwright et al., 2009; Li et al., 2016) to explore the potential of single-cell RNA-seq to capture the expression of known cell-type-specific genes and to identify new ones. We focus on roots of mature seedlings and probe the developmental trajectories of several cell lineages.

2.3 Results

Single cell RNA-seq of whole A. thaliana roots reveals distinct populations of cortex, endodermis, hair, non-hair, and stele cells

We used whole *A. thaliana* roots from seven-day-old seedlings to generate protoplasts for transcriptome analysis using the 10x Genomics platform (**Supplemental Figure 2.1A**). We captured 3,121 root cells to obtain a median of 6,152 unique molecular identifiers (UMIs) per cell. UMIs here are 10 base random tags added to the cDNA molecules that allow us to differentiate unique cDNAs from PCR duplicates. These UMIs corresponded to the expression of a median of 2,445 genes per cell and a total of 22,419 genes, close to the gene content of *A. thaliana*. Quality measures for sequencing and read mapping were high. Of the approximately 79,483,000 reads, 73.5% mapped to the TAIR10 *A. thaliana* genome assembly, with 67% of these annotated transcripts. These values are well within the range reported for droplet-based single-cell RNA-seq in animals and humans.

For data analysis, we applied Monocle 3, which orders transcriptome profiles of single cells in an unsupervised manner without *a priori* knowledge of marker genes (Qiu et al., 2017a; Qiu et al., 2017b; Trapnell et al., 2014). We used the 1500 genes in the data set (**Supplemental Data Set 2.1**) that showed the highest variation in expression (**Supplemental Figure 2.1B**). For unsupervised clustering, we used 25 principal components (PC). These 25 PCs accounted for 72.5% of the variance explained by the first 100 PCs, with the first PC explaining 11% and the 25th PC explaining 0.9% (**Supplemental Figure 2.1C**). Cells were projected onto two dimensions using the uniform manifold approximation and projection (UMAP) method (McInnes and Healy, 2018) and clustered, resulting in 11 clusters (**Figure 2.1A**) (Blondel et al., 2008). Most clusters showed similar levels of total nuclear mRNA, although clusters 9 and 11 were exceptions with higher levels (**Supplemental Figure 2.1D**). Because some of the UMAP

clusters, specifically clusters 9 and 11, consisted of cells that had higher than average amounts of nuclear mRNA, we were concerned that these clusters consisted merely of cells that were doublets, *i.e.* two (or more) cells that received the same barcode and that resulted in a hybrid transcriptome. As cells were physically separated by digestion, it was possible that two cells remained partially attached. In order to identify potential doublets in our data, we performed a doublet analysis using Scrublet (Wolock et al., 2018), which uses barcode and UMI information to calculate the probability that a cell is a doublet. This analysis identified only 6 cells, of 3,021 cells analyzed, as doublets, spread across multiple UMAP clusters and multiple cell types (**Supplemental Figure 2.1E**). Overall, given the low number of doublets, we did not attempt to remove these cells.

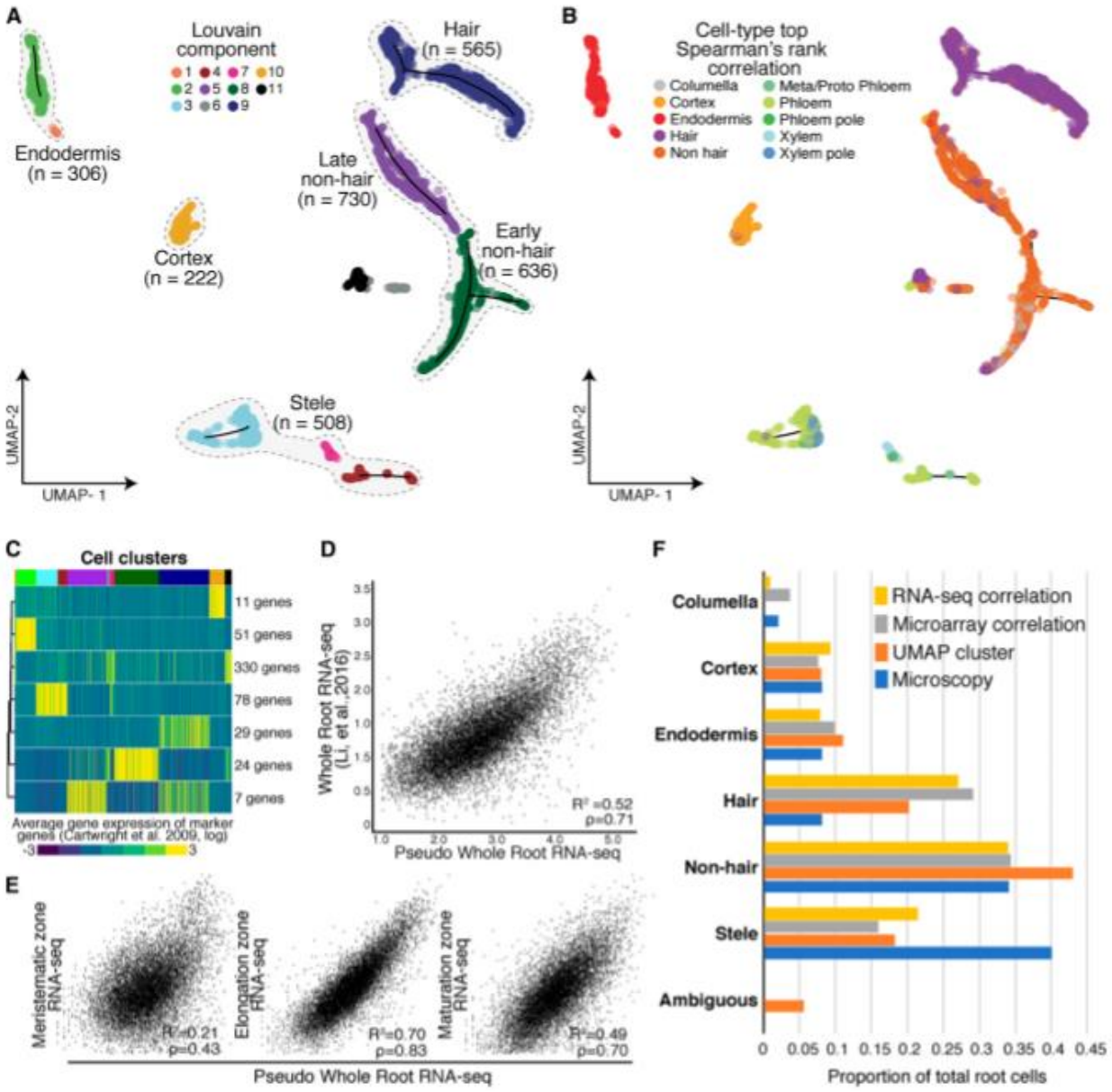


Figure 2.1. **Annotation of cell and tissue types for single-cell RNA-seq of whole *A. thaliana* roots.**

(A) Root cells were clustered and projected onto two-dimensional space with UMAP (McInnes and Healy, 2018). Filled circles represent individual cells; colors represent their respective Louvain component. Monocle 3 trajectories (black lines) are shown for clusters in which a trajectory could be identified. **(B)** Filled circles represent individual cells; colors indicate cell and tissue type based on highest Spearman's rank correlation with sorted tissue-specific bulk expression data (Brady et al., 2007; Cartwright et al., 2009). **(C)** Known marker genes (Brady et

al., 2007; Cartwright et al., 2009) were used to cluster single-cell gene expression profiles based on similarity. The expression of 530 known marker genes was grouped into 7 clusters, using k-means clustering. Mean expression for each cluster (rows) is presented for each cell (columns). Cells were ordered by their respective Louvain component indicated above by color (see **A,** starting at component 1 at left). Number of genes in each cluster is denoted at right. **(D)** Single-cell RNA-seq pseudo-bulked expression data are compared to bulk expression data of whole roots (Li et al., 2016). **(E)** Single-cell pseudo-bulk expression data are compared to bulk-expression data of the three developmental regions of the *A. thaliana* root (Li et al., 2016). **(F)** Proportions of cells as annotated by either UMAP (in **A**), Spearman's rank correlation (in **B**), or Pearson's rank in **Supplemental Figure 2.2**, are compared to proportions determined by microscopy (Brady et al., 2007; Cartwright et al., 2009).

To assign these clusters to cell types, we performed three complementary analyses relying on two expression data sets from tissue-specific and cell-type-specific reporter lines: an earlier one generated with microarrays (Brady et al., 2007; Cartwright et al., 2009) and a more recent one generated with RNA-seq and a greater number of lines (Li et al., 2016). We first compared the microarray expression data for each reporter line to the gene expression values in each single cell, using Spearman's rank correlations to assign each cell a cell type identity based on highest correlation of gene expression (**Figure 2.1B, Supplemental Data Set 2.2**) (Brady et al., 2007; Cartwright et al., 2009). Second, we compared the RNA-seq expression data to the gene expression values in each single cell by Pearson's correlation (Li et al., 2016, **Supplemental Figure 2.2A**). Third, we examined the expression of 530 cell-type-specific marker genes (Brady et al., 2007) by defining seven marker gene clusters with k-means clustering and calculating their average expression for each cell. We then compared each cell's UMAP Louvain component cluster assignment (**Figure 2.1A**) with its marker-gene-based assignment. Louvain components refer to the Louvain method for community detection (Blondel et al., 2008) which is implemented in Monocle 3. Unlike k-means clustering for which the user provides the desired number of clusters to partition a dataset, Louvain clustering optimizes modularity (*i.e.* the separation of clusters based on similarity within a cluster and among

clusters), aiming for high density of cells within a cluster compared to sparse density for cells belonging to different clusters. The 11 clusters presented in **Figure 2.1A** optimized the modularity of the generated expression data and were not defined by us.

In general, the UMAP clusters showed high and cluster-specific expression of marker genes. For example, cells in cluster 10 showed high and specific mean expression of cortex marker genes (**Figure 2.1C, Supplemental Figure 2.3, Supplemental Data Set 2.3**). Both expression correlations and marker gene expression allowed us to assign the Louvain components to five major groups: root hair cells, non-hair cells (containing both an early and late cluster), cortex cells, endodermis cells and stele cells (containing both xylem and phloem cells) (**Figure 2.1A**). Although some cells were most highly correlated in expression with the cell type columella in Spearman's rank tests and RNA-seq Pearson's correlation, these cells co-clustered with non-hair cells (**Figure 2.1B, Supplemental Figure 2.2**). This finding is consistent with bulk RNA-seq data of sorted cells (Li et al., 2016). Specifically, the *PET11* (columella) - sorted bulk RNA-seq data are most similar to bulk RNA-seq data sorted for *GL2* and *WER* (Li et al., 2016), both of which mark non-hair cells (Petricka et al., 2012). Therefore, these cells were grouped as early non-hair cells with other non-hair cells in Louvain component 8. As their expression values were best correlated with RNA-seq data for *WER*-sorted cells, they likely represent a mix of early non-hair and lateral root cap cells, which have very similar expression profiles (**Supplemental Figure 2.2**).

We assessed the extent to which combined single-cell root expression data resembled bulk whole root expression data (Li et al., 2016) (**Figure 2.1D, E**). We observed strong correlations between these two data sets (Pearson's correlation coefficient [R^2]=0.52, Spearman's r=0.71). We also compared the combined single-cell expression data to three bulk expression data sets representing the major developmental zones in the *A. thaliana* root: the meristematic zone, the elongation zone, and the maturation zone (**Figure 2.1E**). We observed

the highest correlation of single-cell and bulk expression in the elongation zone (R^2=0.70, r=0.83) and a lower correlation in the maturation zone (R^2=0.58, r=0.70). This observation is surprising given the more mature developmental stage of the harvested roots (**Supplemental Figure 2.1A**), and likely reflects that younger cells are more easily digested during protoplasting and contribute in greater numbers to the gene expression data. As expected, single-cell and bulk expression were poorly correlated in the meristematic zone (R^2=0.11, r=0.43), as meristematic tissue accounts for only a small proportion of mature roots. Furthermore, we compared tissue-specific expression (Li et al., 2016) to expression both in the annotated cell clusters and in cells expressing appropriate marker genes. In general, we found strong correlations among these data sets, suggesting that the clusters are annotated correctly (**Supplemental Table 2.1**).

We also compared the relative representation of root cell types between our data set and estimates based on microscopy studies (**Figure 2.1F**) (Brady et al., 2007; Cartwright et al., 2009). Independent of annotation method, we observed the expected numbers of cortex (222 Spearman's/ 233 UMAP), endodermis (306/ 304), non-hair cells (1201/ 1061) and columella cells (111/ no UMAP cluster). Hair cells (565/ 898) were overrepresented whereas stele cells (508/ 490) were underrepresented, possibly reflecting a bias in the protoplast preparation of whole root tissue.

Protoplasting, the removal of the plant cell wall, alters the expression of 346 genes (Birnbaum et al., 2003); 76 of these genes were included in the 1500 genes with the highest variation in expression (**Supplemental Data Set 2.1, Supplemental Figure 2.1B**) that we used for clustering. Some of the 76 genes showed cell-type-specific expression. To exclude the possibility that the expression pattern of these genes produced artefactual clusters and cell-type annotations, we removed them from the analysis and re-clustered, which resulted in a similar UMAP visualization, with similar numbers of Louvain components and cell types.

Single-cell RNA-seq of identifies novel genes with cell-type and tissue-type-specific expression

Some marker genes are not expressed exclusively in a single cell type, making it desirable to identify additional genes with cell-type-specific expression. We first confirmed the high and cluster-specific expression of well-known marker genes (**Figure 2.2A, Supplemental Figure 2.4**) (Li et al., 2016) such as the root-hair-specific *COBL9*, the endodermis-specific *SCR* and the three stele-specific genes *MYB46* (xylem-specific), *APL* (phloem-specific), and *SUC2* (phloem-specific). The nonspecific expression of the quiescent center cell marker genes *WOX5* and *AGL42* is likely due to the failure to capture sufficient numbers of these rare cells. The nonspecific expression of *WOL* and the more heterogeneous pattern of both *WER* and *GL2* expression have been previously observed (Brady et al., 2007; Winter et al., 2007).

Second, to find novel marker genes, we identified genes with significantly different expression within and among Louvain component clusters by applying the Moran's I test implemented in Monocle 3. We found 317 genes with cluster-specific expression, 164 of which were novel, including at least one in each cluster (**Figure 2.2A, Supplemental Data Set 4)**. Using cell-type annotations rather than Louvain clusters, we identified 510 genes with cell-type-specific expression, of which 317 overlapped with the Louvain component cluster-specific expression genes, as well as an additional 125 novel genes, some of which have been implicated in the development of a cell lineage in targeted molecular genetics studies.

For example, the stele-specific AT1G8810 (*ABS5, T5L1*) gene (cluster 7, **Figure 2.2A**) encodes a bHLH protein that promotes vascular cell division and differentiation as part of a heterodimer with second bHLH protein, LHW (Katayama et al., 2015; Ohashi-Ito et al., 2014). Another stele-specific gene, AT4G36160 (*ANAC076*, *VND2*) (cluster 7), encodes a ClassIIB NAC-domain transcription factor that contributes to xylem vessel element differentiation by

promoting secondary cell wall formation and programmed cell death (Tan et al., 2018). In tissue-specific bulk data (Brady et al., 2007; Winter et al., 2007), both genes show xylem-specific expression consistent with their biological functions; *T5L1* expression is high only in the meristematic and elongation zones, while *VND2* expression starts in the elongation zone and persists throughout the maturation zone. Other genes, not previously implicated in root development, show tissue-specific bulk expression patterns consistent with the single-cell data. For example, AT1G54940 (*GUX4*, *PGSIP4*), which encodes a xylan glucuronosyltransferase (Lee et al., 2012; Mortimer et al., 2010), was specifically expressed in hair cells (cluster 9) and is most highly expressed in cells destined to become hair cells in the elongation zone and in differentiated hair cells in the maturation zone (Brady et al., 2007; Cartwright et al., 2009).

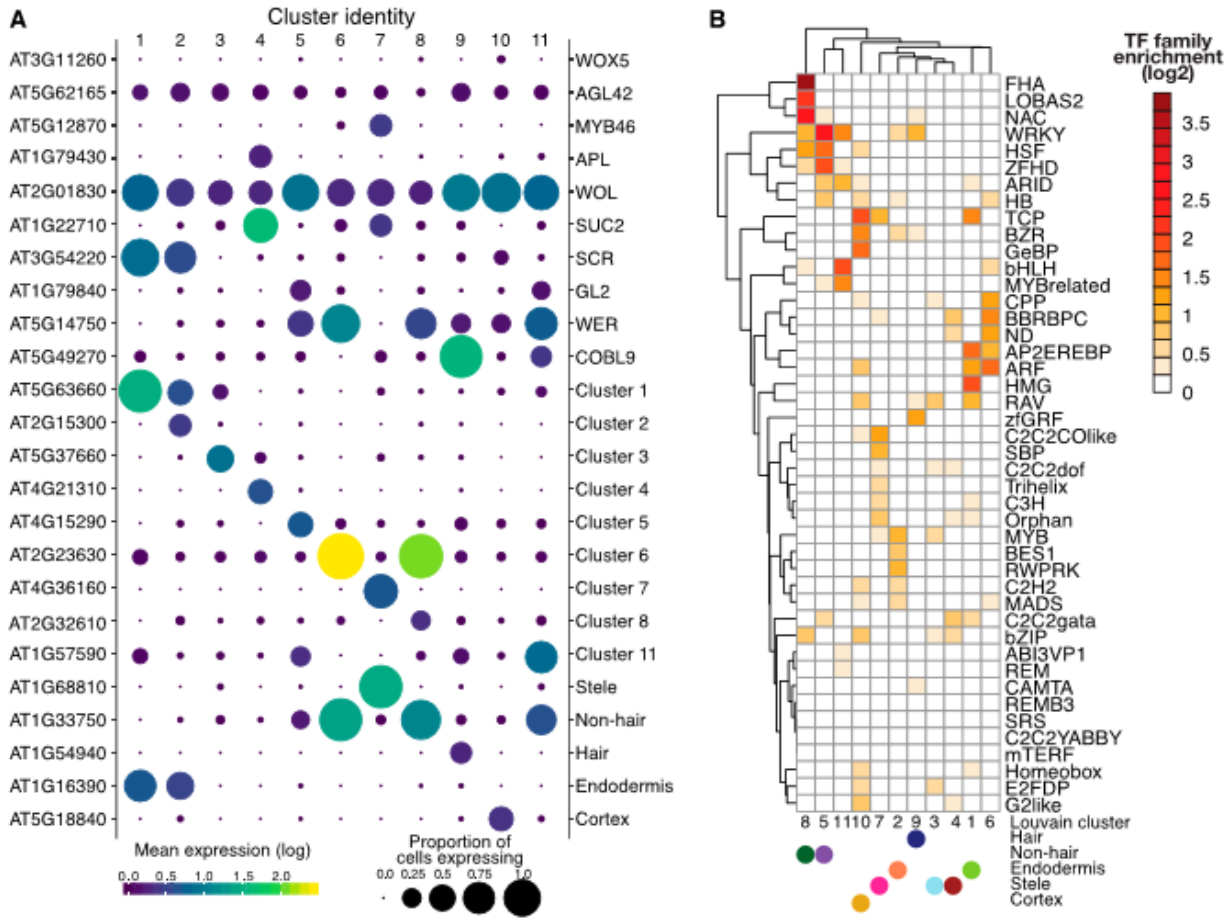


Figure 2.2. **Novel cluster-specific and tissue-specific genes and enriched transcription factor motifs.**

(A) Proportion of cells (circle size) and mean expression (circle color) of genes with cluster-specific and tissue-specific expression are shown, beginning with known marker genes labeled with their common name (right) and their systematic name (left). For novel genes, the top significant cluster-specific genes are shown, followed by the top significant tissue-specific genes; both were identified by principal graph tests (Moran's I) as implemented in Monocle 3. Note the correspondence between Louvain components and cell and tissue types. For all novel cluster-specific and tissue-specific genes, see **Supplemental Table 2.3**. **(B)** Enrichments of known transcription factor motifs (O'Malley et al., 2016) 500 bp upstream of genes with cluster-specific expression compared to genome background. Motifs are specific to transcription factor gene families rather than individual genes. The plot is clustered based on similarity in enrichments with Louvain components and cell and tissue types (filled circles) indicated.

Expression of some transcription factors shows high correlation with specific cell types

We asked whether we could identify transcription factors that may contribute to the cluster-specific expression patterns. To do so, we tested for transcription factor motif enrichments in the proximal regulatory regions of genes with cluster-specific expression, examining 500 bp upstream of the transcription start site (Alexandre et al., 2018; Sullivan et al., 2014) and a comprehensive collection of *A. thaliana* transcription factor motifs (O'Malley et al., 2016). This analysis revealed significant transcription factor motif enrichments among clusters and annotated major tissues and cell types (**Figure 2.2B**).

As transcription factors in *A. thaliana* often belong to large gene families without factor-specific motif information (Riechmann et al., 2000), it is challenging to deduce the identity of the specific transcription factor that drives cluster-specific transcription factor motif enrichment and expression. As an approximation, we examined transcription factors that were expressed in the

cluster or tissue in which a significant enrichment of their motif was found, or in neighboring cell layers (some factors move between cells (Petricka et al., 2012)) (**Supplemental Data Set 2.4**). We focused first on the small *BZR/BEH* gene family whose motif was specifically enriched in cortex cells (cluster 10). Of the six genes (BEH1/AT3G50750, BEH2/AT4G36780, BEH3/AT4G18890, BEH4/AT1G78700, BES1/AT1G19350, and BZR1/AT1G75080) the single recessive *beh4*, *bes1*, and *bzr1* mutants exhibit altered hypocotyl length (Lachowiec et al., 2018). Double mutant analysis suggests partial functional redundancy, which agrees with our observation of overlapping expression patterns for these genes across cell types (**Supplemental Figure 2.5A, B**). In contrast, neither *beh1* and *beh2* single mutants nor the respective double mutant show phenotypic defects (Lachowiec et al., 2018). However, *BEH2* was the most highly expressed *BZR/BEH* family member across clusters and annotated root tissue and cell types (**Supplemental Figure 2.5A, B**). Although *BEH4*, the most ancient family member with the strongest phenotypic impact, showed cortex-specific expression, none of *BZR/BEH* genes showed significance for cluster-specific expression, suggesting that combinations of family members, possibly as heterodimers, may result in the corresponding motif enrichment in cortex cells (**Supplemental Figure 2.5A, B**). In particular, *BES1* and *BZR* expression was highly correlated, consistent with these genes being the most recent duplicates in the family (**Supplemental Figure 2.5C**) (Lachowiec et al., 2013; Lan and Pritchard, 2016).

In contrast to the *BEH/BZR* gene family, we found stronger cluster specificity for some TCP transcription factors. The TCP motif was strongly enriched in cortex (cluster 10), endodermis (cluster 1) and stele (cluster 7). Of the 24 TCP transcription factors, we detected expression for eight. Of these, *TCP14* (AT3G47620) and *TCP15* (AT1G69690) were expressed primarily in stele (clusters 7 and 4) although this cluster-specific expression was not statistically significant (**Figure 2.2B, Supplemental Figure 2.5D, E, Supplemental Data Set 2.4**). *TCP14* and *TCP15* are class I TCP factors thought to promote development. Acting together, *TCP14*

and *TCP15* promote cell division in young internodes (Kieffer et al., 2011), seed germination (Resentini et al., 2015), cytokinin and auxin responses during gynoecium development (Lucero et al., 2015), and repression of endoreduplication (Peng et al., 2015). Both genes are expressed in stele in bulk tissue data (Brady et al., 2007; Winter et al., 2007), with *TCP14* expression also observed in the vasculature by *in situ* hybridization (Tatematsu et al., 2008). *TCP14* can affect gene expression in a non-cell-autonomous manner.

To further investigate the co-occurrence of cluster-specific transcription factor motif enrichments with transcription factor expression, we next examined the novel genes with significant cluster-specific expression. Eight of these encode transcription factors with corresponding highly enriched cluster-specific binding motifs. For one of these, *BRN2* (AT4G10350), cluster-specific expression coincided with enrichment of the NAC transcription factor family motif(cluster 8, non-hair and lateral root cap cells, **Figure 2.2B**). *BRN2* encodes a ClassIIB NAC transcription factor implicated in root cap maturation together with *BRN1* and *SMB*. Class IIB NAC transcription factors are thought to contribute to terminal cell differentiation accompanied by strong cell wall modifications (Bennett et al., 2010). In our data, *BRN2* was most highly expressed in cluster 8 (non-hair and lateral root cap cells) and less so in cluster 6 (**Supplemental Data Set 2.4**).

Clustering stele cells identifies novel genes with cell-type specific expression in the vasculature

Our initial attempts to annotate and separate cell types within stele tissue with marker gene expression or Spearman's rank correlations failed. Instead, we separately clustered stele cells to reveal 6 sub-clusters upon UMAP visualization, with 5 sub-clusters containing more than 40 cells. Their annotation via Spearman's rank correlation with sorted bulk data was not

successful; however, using well-established marker genes expression, we detected cluster-specific expression patterns (**Figure 2.3A and B**).

Cells closely related to xylem pole pericycle constituted the largest group of cells (205 cells); phloem pole pericycle cells were the second largest (84 cells). The high number of pericycle cells likely reflects our experimental procedure, as these cells reside on the exterior of the vascular bundle. Both phloem and xylem clusters showed similar numbers of cells (77 cells and 72 cells respectively); the phloem companion cells formed a distinct cluster. We observed the expected sub-cluster expression for several known genes and marker genes and identified novel genes with sub-cluster-specific expression (**Figure 2.3C, Figure 2.3D, Supplemental Data Set 2.1**). Although there was some discrepancy, especially for the *APL* gene, which is expressed in both companion and phloem cells (**Figure 2.3C)**, this is largely due to missing data.

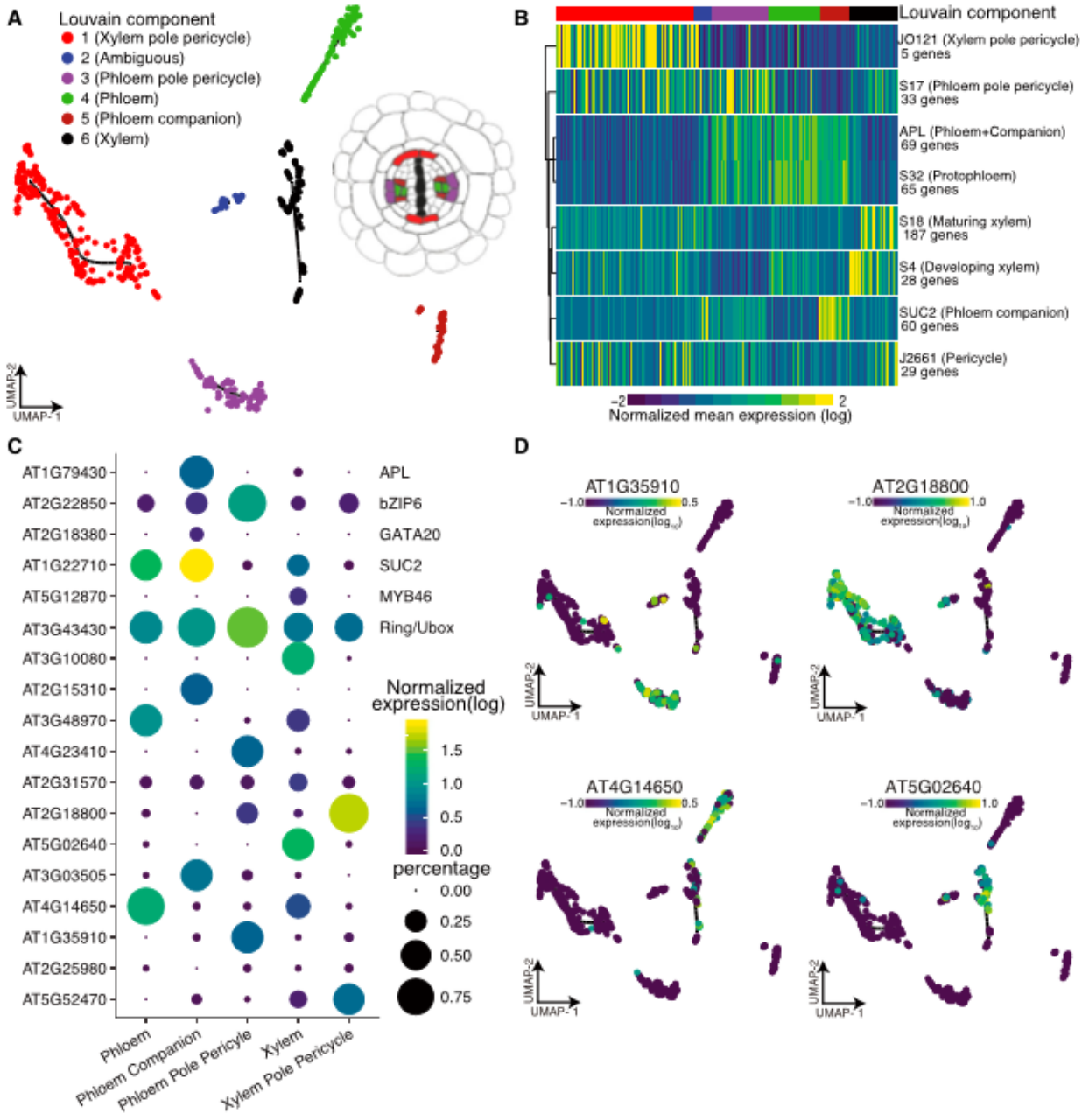


Figure 2.3. Re-clustering of stele cells yields distinct sub-clusters of vasculature cell types.

(A) Cells initially annotated as stele tissue were re-clustered, resulting in six distinct sub-clusters cells, five of which contained more than 40 cells. **(B)** Mean expression for previously identified cell-type specific genes (Cartwright et al., 2009) in each cell is shown, allowing annotation of stele sub-cluster identities as shown in **(A)**. **(C)** Proportion of cells (circle size) and mean expression (circle color) of genes with cluster-specific and tissue-specific expression are shown,

starting with known marker genes at the top, labeled with their common name (right) and their systematic name (left). Below, novel significant tissue-specific genes are shown with their systematic names, identified by principal graph tests (Moran's I) as implemented in Monocle 3. **(D)** Example expression overlays for cluster-specific genes identified by the principal graph test in (**C**).

Pseudotime trajectories coincide with the development stages of cortex, endodermis, and hair cells

We next sought to visualize the continuous program of gene expression changes that occurs as each cell type in root differentiates. Because whole roots contain a mix of cells at varying developmental stages, we reasoned that our experiment should have captured a representative snapshot of their differentiation. Monocle not only clusters cells by type but also places them in "pseudotime" order along a trajectory that describes their maturity. To make these trajectories, Monocle 3 learns an explicit principal graph from the single-cell-expression data through reversed graph embedding, an advanced machine learning method (Qiu et al., 2017a; Qiu et al., 2017b; Trapnell et al., 2014). To dissect the developmental dynamics of individual clusters, we first focused on the well-defined root-hair cells, in which combined single-cell expression values highly correlated with those from bulk protoplasts sorted for expression of the *COBL9* root-hair marker gene (**Supplemental Table 2.1**). To annotate the unsupervised trajectory that Monocle 3 created for hair cells, we used the Spearman's rank test to compare expression in all cells to bulk expression data representing 13 different developmental stages in root tissues from all the available sorted cell types (**Supplemental Figure 2.6**) (Brady et al., 2007; Cartwright et al., 2009). Each cell was assigned the developmental stage and cell type most correlated with its expression values (**Figure 2.4A)**. The hair cells with the earliest developmental stage assignment were designated as the root of the trajectory. Next,

pseudotime was calculated for all other hair cells based on their distance from the root of the trajectory (**Figure 2.4B**). We compared this calculated pseudotime with the most highly correlated developmental assignment from bulk data, finding close agreement (**Figure 2.4B**). Examples of genes that are expressed early and late in pseudotime in the UMAP hair cluster are shown in **Figure 2.4C**.

Hair cells undergo endoreduplication as they mature, resulting in up to 16N genomic copies in the developmental stages assayed (Bhosale et al., 2018). Although endoreduplication is thought to increase transcription rates (Bourdon et al., 2012), general transcription might decrease as hair-cell-specific genes become more highly expressed during hair cell differentiation. Single-cell RNA-seq affords us the opportunity to explore whether transcription rates differ across development. Single-cell RNA-seq can measure both relative expression (as in bulk RNA-seq) and the total number of RNA molecules per cell. The total amount of cellular mRNA was drastically reduced across hair cell development (**Figure 2.4D**). This result may be due to technical bias; for example, gene expression in larger, endoreduplicated cells may be more difficult to assess with this droplet-based method. If so, the observed reduction in captured transcripts should affect all genes more or less equally. Alternatively, this observation may reflect hair cell differentiation, whereby transcription of hair-cell-specific genes should remain unaffected or increase over pseudotime. Our results support the latter scenario as transcription of hair-cell-specific genes appears to increase over pseudotime, consistent with these cells undergoing differentiation towards terminally differentiated hair cells (**Figure 2.4E, Supplemental Figure 2.7A**)

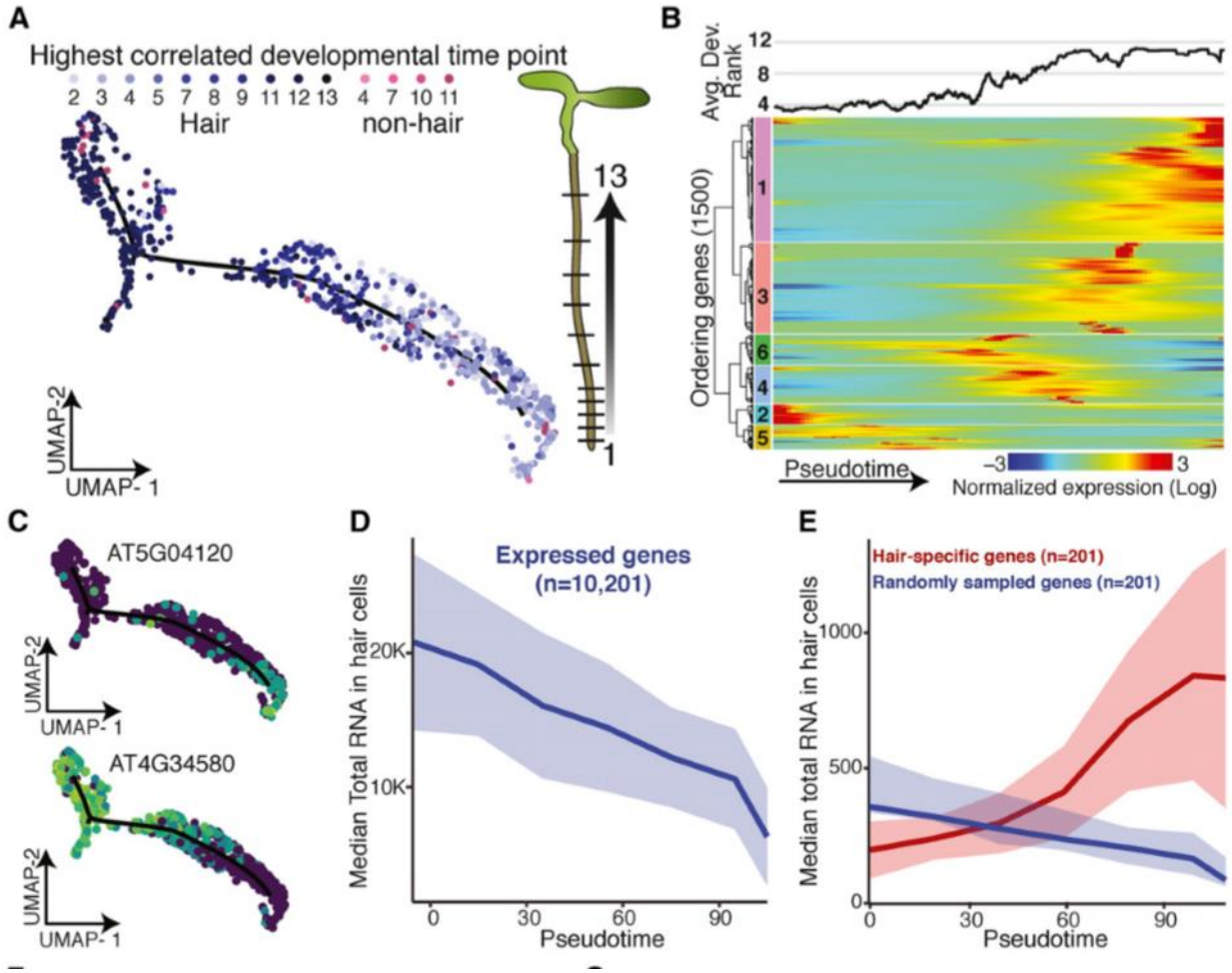


Figure 2.4. Developmental trajectory of hair cells.

(A) UMAP-clustered hair cells were assigned a developmental time point based on highest Spearman's rank correlation with bulk expression data of staged tissue (13 developmental stages) (Brady et al., 2007; Cartwright et al., 2009). Cell type and developmental time points are indicated in shades of blue (and pink). Graphic illustrates developmental stages in *A. thaliana* root (Plant Illustrations). **(B)** Cells were ordered in pseudotime; columns represent cells, rows represent expression of the 1500 ordering genes. Rows were grouped based on similarity in gene expression, resulting in 6 clusters (indicated left), with genes in clusters 2 and 5 expressed early in pseudotime and genes in cluster 1 expressed late. Hair cells with the earliest developmental signal (Brady et al., 2007; Cartwright et al., 2009) were designated as the root of the trajectory. The graph above represents the average best-correlation of developmental stage (Brady et al., 2007; Cartwright et al., 2009) in a scrolling window of 20 cells with pseudotime, showing the expected increase in developmental age with increasing pseudotime. **(C)** Examples

of an early and a late expressed hair-cell-specific gene. Gene expression in each cell is superimposed onto the UMAP cluster and trajectory, with lighter colors indicating higher gene expression. **(D)** Median total RNA captured in cells decreases across pseudotime. Number of genes included is indicated. **(E)** Comparison of median total RNA for hair-cell-specific genes (in red) to a comparable random set of genes (in blue). Number of genes is indicated (Permutation test p-value ≈ 1 x 10^{-4}).

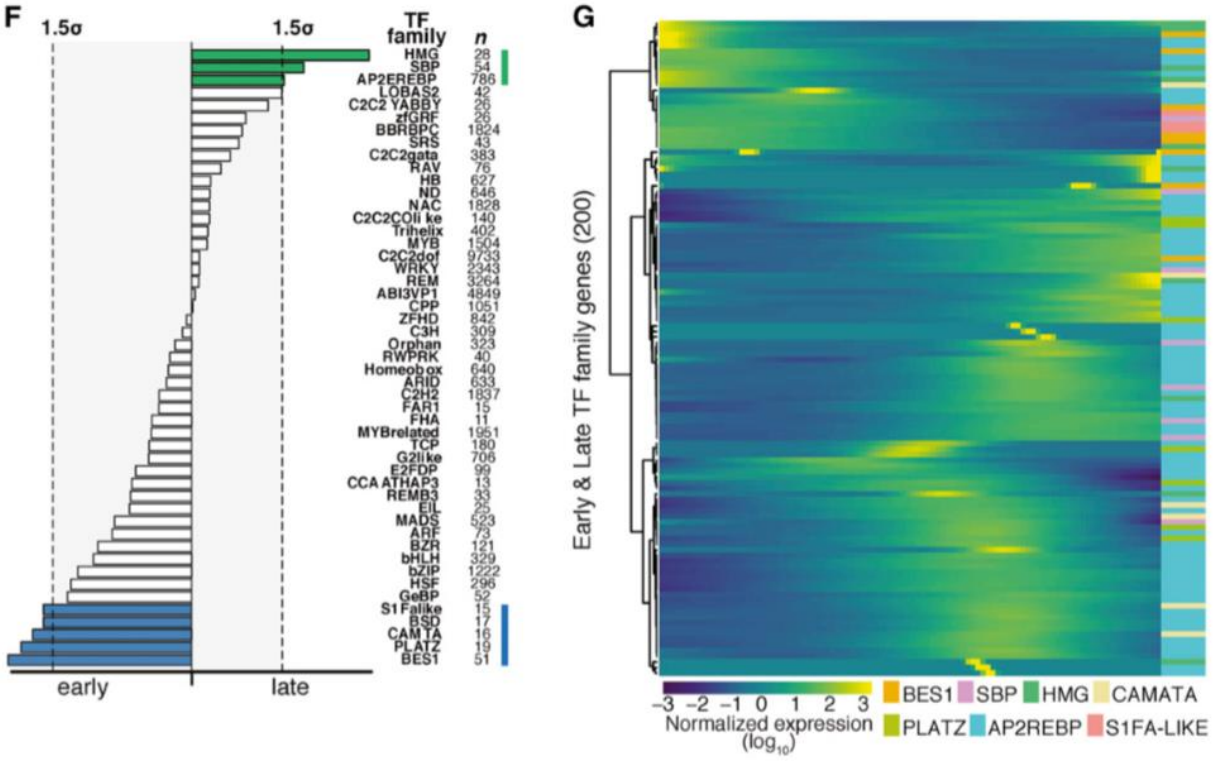


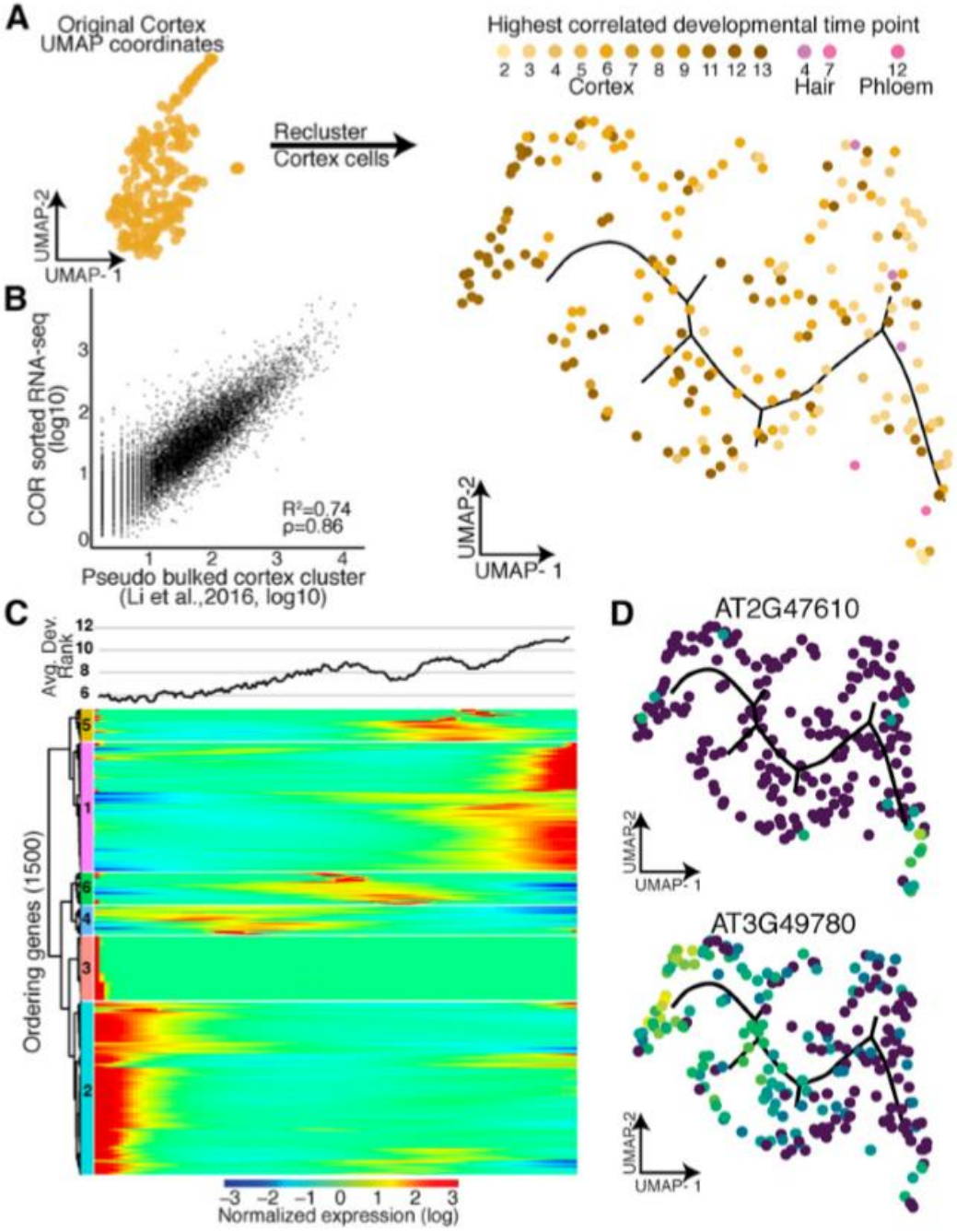
Figure 2.4. Developmental trajectory of hair cells (contd)

(**F**) Different transcription factor motifs reside in the 500 bp upstream regions of genes expressed early (clusters 2, 5) compared to genes expressed late (cluster 1). Transcription factor motifs specific to early hair cells are denoted with blue bars, those for late hair cells with green bars; bar length indicates motif frequency. Thresholds on either side (grey box, dotted lines) refer to 1.5 standard deviation above mean motif frequency. **(G)** Expression of individual members of transcription factors families highlighted in **D** across pseudotime identifies candidate factors driving early or late gene expression.

To further explore this transcriptional dynamic, we calculated RNA velocity (La Manno et al., 2018), a measure of the transcriptional rate of each gene in each cell of the hair cell cluster. RNA velocity takes advantage of errors in priming during 3' end reverse transcription to determine the splicing rate per gene and cell. It compares nascent (unspliced) mRNA to mature (spliced) mRNA; an overall relative higher ratio of unspliced to spliced transcripts indicates that transcription is increasing. In our data, only ~4% of reads were informative for annotating splicing rates, a lower percentage than what has been used in mammalian cells for velocity analyses, and thus our results may be less reliable. Based on data for 996 genes, mean RNA velocity increased across pseudotime (**Supplemental Figure 2.7B,** p = 2.2 e-16 linear model, rho = 0.73). This increase in velocity was associated with the predicted changes in endoreduplication (Bhosale et al., 2018), especially between the 4N and 8N stages (**Supplemental Figure 2.7C,** Tukey's multiple comparison p-value = 0.0477).

We also observed developmental signals in other cell types, including cortex and endodermis (**Figure 2.5A-D, Supplemental Figure 2.8**). Combined single-cell expression values for cortex cells highly correlated with those from bulk protoplasts sorted for expression of the *COR* cortex marker gene (**Figure 2.5B**, R^2= 0.74, rho=0.86). As Monocle 3 did not identify a trajectory for cortex cells in the context of all cells, we isolated the cortex cells and re-performed UMAP dimensionality reduction, clustering, and graph embedding (**Supplemental Data Set 2.1**). Each cortex cell was assigned a developmental stage based on its Spearman's rank correlation with bulk expression data (Brady et al., 2007; Cartwright et al., 2009). Cortex cells with the earliest developmental signal were designated as the root of the cortex trajectory, and pseudotime was assigned to the remaining cortex cells based on their distance from the root (**Figure 2.5A-D, Supplemental Figure 2.6**). As pseudotime increased for cortex cells, so did their age, indicating good agreement of the trajectory with developmental bulk RNA-seq data. Although we observed some decrease in total RNA expression and increased expression in

cell-type specific genes for endodermis, We did not see a clear pattern of change in total RNA across cortex pseudotime (**Supplemental Figure 2.8, Supplemental Figure 2.9**).



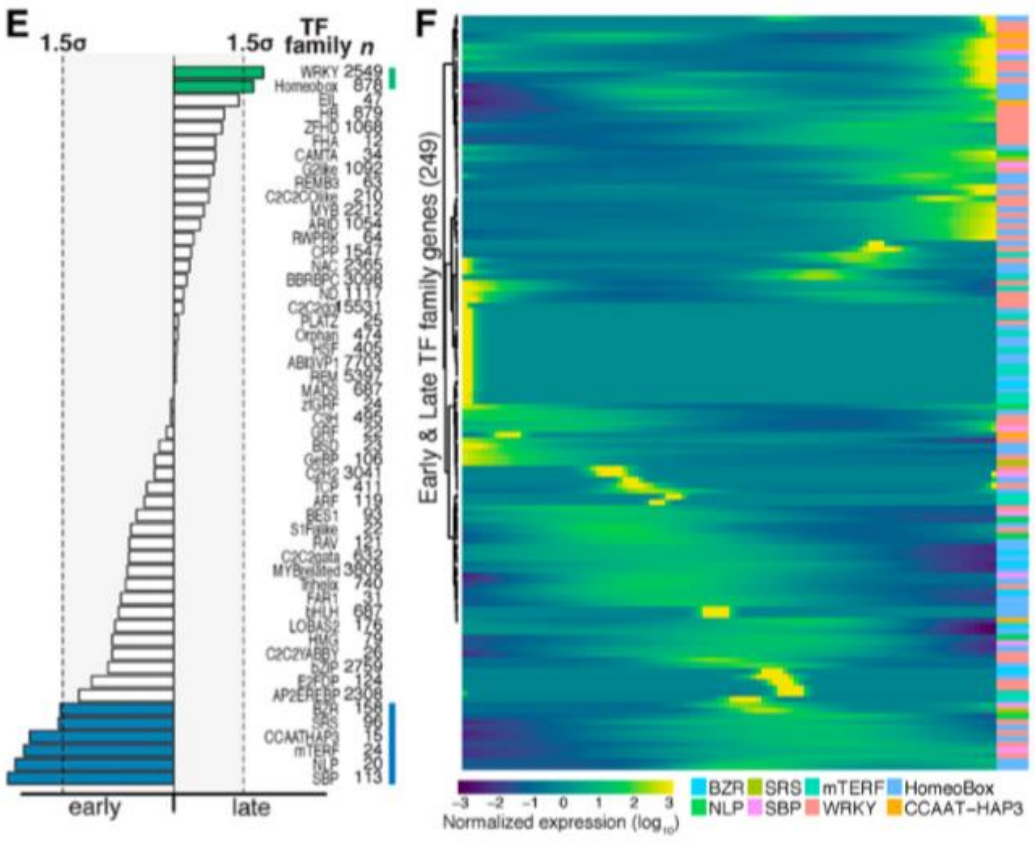


Figure 2.5. Developmental trajectory of cortex cells.

(A) Cortex cells were re-clustered to create a trajectory, in which each cell was assigned a developmental time point and identity (shades of yellow, brown, pink) based on the highest Spearman's rank correlation of a cell's gene expression with prior sorted bulk data (Brady et al., 2007; Cartwright et al., 2009). **(B)** Comparison of pseudo-bulk expression data from cells annotated as cortex cells with bulk expression data from protoplasts sorted for expression of the cortex marker gene *COR* (Li et al., 2016). **(C)** Cells were ordered in pseudotime; columns indicate cells and rows the expression of the 1500 ordering genes. Rows were grouped based on similarity in gene expression, resulting in 6 clusters (indicated left), with genes in clusters 2 and 3 expressed early in pseudotime and genes in cluster 1 expressed late. Cortex cells with the earliest developmental signal (Brady et al., 2007; Cartwright et al., 2009) were designated as the root of the trajectory. The graph above represents the average best-correlation of developmental stage (Brady et al., 2007; Cartwright et al., 2009) in a scrolling window of 20 cells with pseudotime, showing the expected increase in developmental age with increasing pseudotime. **(D)** Examples of an early and a late expressed novel cortex-cell-specific gene.

Gene expression in each cell is superimposed onto the UMAP cluster and trajectory, with lighter colors indicating higher gene expression. **(E)** Different transcription factor motifs reside in the 500 bp upstream regions of genes expressed early (clusters 2, 3) compared to genes expressed late (cluster 1). Transcription factor motifs specific to early cortex cells are denoted with blue bars, those for late cortex cells with green bars, bar length indicates motif frequency. Thresholds on either side (grey box, dotted lines) refer to 1.5 standard deviation above mean motif frequency. **(F)** Expression of individual members of transcription factors families highlighted in **D** across pseudotime identifies candidate factors driving early or late gene expression.

We asked whether we could assign the transcription factors that drive gene expression along these developmental trajectories in early and late hair, cortex, and endodermis cells. As before, we first analyzed transcription factor motif enrichments and then explored the expression of the corresponding transcription factor gene families. Indeed, for most developmentally enriched transcription factor motifs, we could pinpoint candidate transcription factors that are expressed either early or late. For example, the *AP2/EREBP* (APETALA2/ethylene responsive element binding protein) transcription factor family is one of the largest in *A. thaliana* (Riechmann et al., 2000), with nearly 80 covered in our data set; of these, only four (AT2G25820, At5G65130, AT1G36060, AT1G44830) showed strong expression in late hair cells (**Figure 2.4F, Figure 2.4G, Supplemental Figure 2.10**). One of these, AT1G36060 (Translucent Green), regulates expression of aquaporin genes (Zhu et al., 2014). Overexpression of this gene confers greater drought tolerance (Zhu et al., 2014), consistent with its expression in older hair cells. Similar examples of developmental stage-specific motif enrichments with corresponding transcription factor expression were also found for cortex and endodermis (**Figure 2.5E, Figure 2.5F**, **Supplemental Figure 2.8, Supplemental Figure 2.10**).

Branch points in developmental trajectories mark developmental decisions

Although a developmental trajectory that reflects the differentiation from early to late cells within a cell type should be branchless, we did observe some branch points, for example in Louvain component 8, affording us the opportunity to explore their biological relevance. As discussed, Louvain component 8 contains early non-hair cells and likely some lateral root cap cells. To further explore the cells within the branch, we performed a principal graph test, comparing their expression profiles to those of cells elsewhere in the cluster (**Figure 2.6A**). We found that cells within the branch were significantly enriched for expression of genes involved in cell plate formation, cytokinesis and cell cycle. We explored this enrichment for cell cycle annotations by comparing expression of previously identified core cell cycle genes (Gutierrez, 2009) in cells within the branch to cells in the rest of the cluster, finding many core cell cycle genes, in particular many G2 genes, to be specifically expressed in branch cells (**Figure 2.6B**). Among these genes were several of the cyclin-dependent kinase B family members that direct the G2 to M transition. Two cyclin-dependent kinase subunits (*CKS1* and *CKS2*), thought to interact with several *CDK* family members, were also specifically expressed in branch cells (Vandepoele et al., 2002). Other branch-cell-specific genes included *AUR1* and *AUR2*, both involved in lateral root formation and cell plate formation (**Figure 2.6C**, Van Damme et al., 2011). Louvain component 9 also showed a strong, but short branching point. We did not find any biological processes enriched in genes expressed specifically in this short branch; however, one gene whose expression is known to be affected by protoplasting was specifically expressed in these cells, perhaps reflecting that cells within this branch were more stressed by our experimental procedure (data not shown).

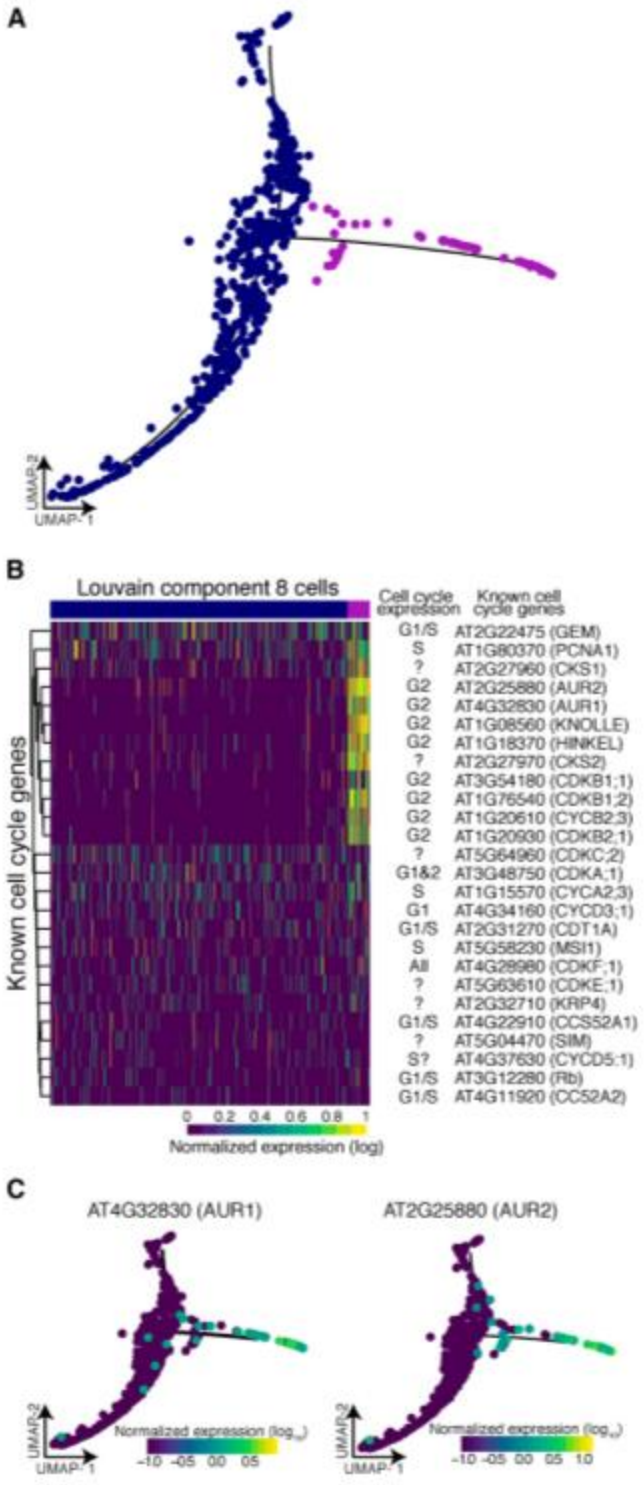


Figure 2.6. Branch analysis reveals actively dividing cells.

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

known cell cycle genes with expression in at least 5% of cells in Louvain component 8. Known cell cycle expression is denoted for each gene, if unknown '?'. (**C**) Two kinases, *AUR1* and *AUR2*, were specifically expressed in branch cells. These genes are involved in cell plate formation and lateral root formation.

Heat-shocked root cells show subtle expression differences among cell types

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

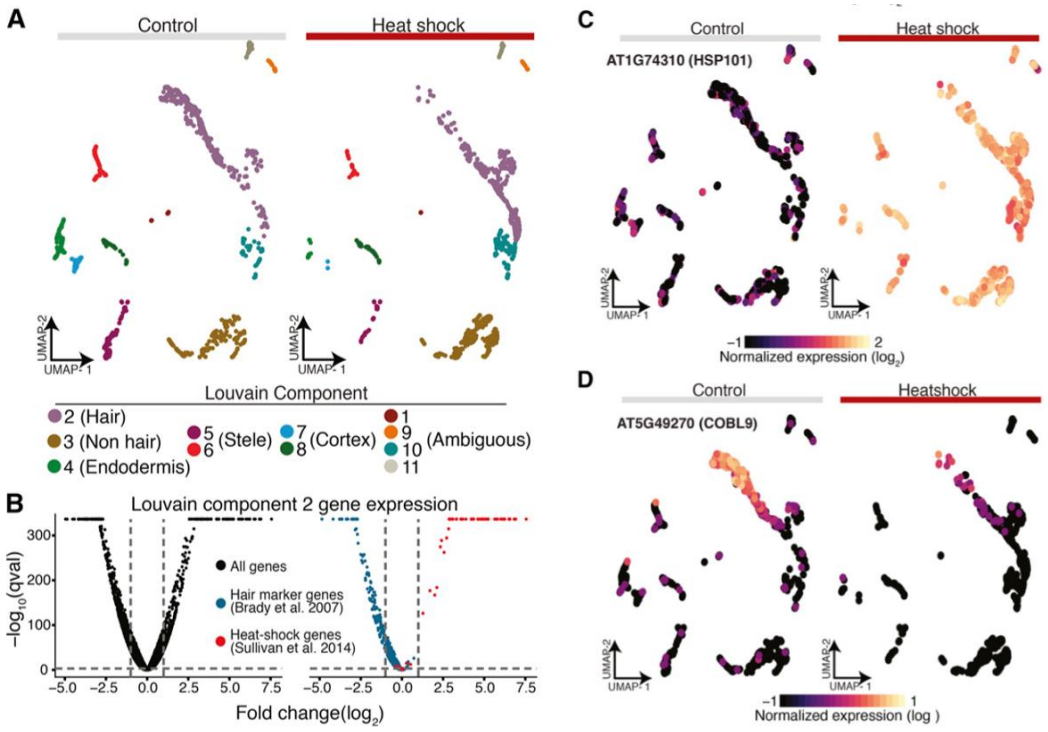
We explored whether all cells within developing roots were capable of exhibiting a typical heat shock response. To do so, we applied a standard heat stress (45 min, 38ºC) to eight-day-old seedlings, harvested their roots along with roots from age- and time-matched control seedlings, and generated protoplasts for single-cell RNA-seq of both samples. For the control sample, we captured 1076 cells, assaying expression for a median 4,079 genes per cell and a total of 22, 971 genes; 82.7% of reads mapped to the TAIR10 genome assembly. The results for these control cells were similar to those described earlier with regard to captured cell types, proportion of cell types (*e.g.* 28.8% vs. 34% annotated hair cells and 9.7% vs. 7.2% endodermis cells), and correlation of gene expression (R^2 = 0.86 for the 21,107 genes captured in both

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

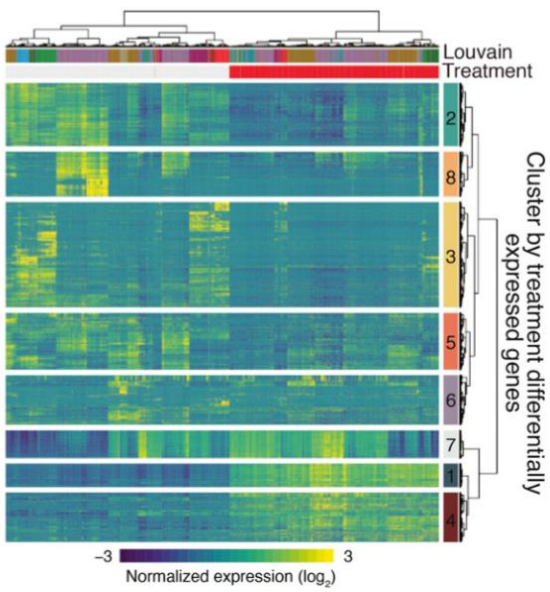
Due to global gene expression changes upon heat shock, we could not simply assign cell and tissue types as before for heat-shocked cells. The overwhelming impact of heat shock was also apparent when comparing the first and second highest cell type and developmental Spearman's rank correlations for control cells and heat-shocked cells. Upon heat shock, many cells, especially those with non-hair, phloem and columella as their highest rank, commonly showed as their second highest rank a different cell type instead of another developmental time point of the same cell type as observed in control cells (**Supplemental Figure 2.11A**). Unsurprisingly, the drastic changes in gene expression led to cells being embedded in UMAP space primarily as a function of treatment, making direct comparisons of treatment effects on any one cell type impossible (**Supplemental Figure 2.11B**). To enable such comparisons, we used a mutual nearest neighbor to embed cells conditioned on treatment in UMAP space (Haghverdi et al., 2018). The mutual nearest neighbor method was originally developed to account for batch effects by identifying the most similar cells between each batch and applying a correction to enable proper alignment of data sets. Here, we employ this technique to overcome the lack of marker expression in our heat-shock treated cells and match them to their untreated counterpart based on overall transcriptome similarity (**Figure 2.7A**). This procedure yielded corresponding clusters in control and heat-shocked cells, albeit with varying cell numbers for most (**Supplemental Figure 2.11C, Supplemental Table 2.2**).

In response to stress, organisms are thought to upregulate stress genes and to specifically downregulate genes involved in growth and development to optimize resource allocation. In response to heat stress, this presumed 'dichotomy' in gene expression is mirrored by the rapid localization of RNA polymerase II to the heat shock gene loci and its depletion

elsewhere in the genome (Abruzzi et al., 2011). Our data provide strong evidence of this regulatory trade-off at the level of individual cells. Using hair cells (Louvain component 2) as an example, we found that hair-cell-specific genes are overwhelmingly repressed and that heat shock genes are upregulated, often dramatically so (**Figure 2.7B, Figure 2.7C, Figure 2.7D)**. Indeed, *HSP101*, the most highly expressed and chromatin-accessible gene upon heat shock in previous studies (Sullivan et al., 2014), was strongly expressed across all clusters while expression of the hair marker gene *COBL9* decreased dramatically upon stress (**Figure 2.7C, Figure 2.7D**).



E


Louvain
Treatment
2
8
3
5
6
7
1
4
Cluster by treatment differentially expressed genes
−3
3
Normalized expression (log2)

F

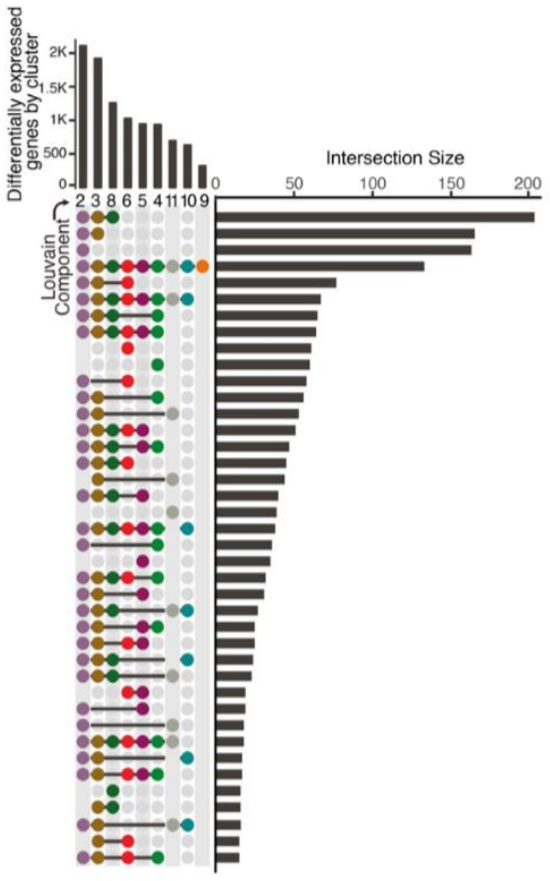

Differentially expressed genes by cluster
2K
1.5K
1K
500
0
Intersection Size
0
50
100
150
200
2 3 8 6 5 4 11 10 9
Louvain Component

Figure 2.7. Single cell RNA-seq highlights canonical and novel aspects of the heat shock response.

(A) A nearest neighbor approach aligns control and heat-shocked cells in a UMAP embedding to allow for concomitant cluster/cell type assignment. **(B)** Volcano plots of average gene expression change upon heat shock within Louvain component 2 for all genes (black), known hair marker genes (blue) and heat-shock signature genes (red). **(C)** *HSP101*, a signature heat shock gene, shows dramatic increase of expression in all cell types upon heat shock. **(D)** *COBL9*, a well-studied hair marker gene, is strongly repressed upon heat shock. **(E)** Heat map of of differentially expressed genes upon heat shock (top red bar; control, top gray bar), hierarchically clustered by both cells and genes (FDR < 0.1% and absolute value of the log2 fold change > 1). **(F)** Upset plot (Lex et al., 2014) of the number of differentially expressed genes as a function of heat shock for each Louvain cluster in our UMAP embedding (bars on top) along with the number of the intersect of differentially expressed genes between Louvain clusters (bars on the right). A surprising number of differentially expressed genes were specific to certain clusters (single dot in vertical row of dots).

Having established comparable clusters, we next identified genes that were differentially expressed as a function of treatment and cluster identity, excluding those with less than 15 cells in either control or heat shock conditions. This analysis identified 8,526 genes (FDR < 0.1%) whose expression was altered by heat shock treatment in one or more clusters; of these, 2,627 genes were up- or downregulated at least 2-fold (**Figure 2.7E, Supplementary Data Set 2.5**, FDR < 0.1% and absolute value of log2 fold change > 1). As for hair cells (**Figure 2.7B**), cell-type marker genes for all clusters were enriched among the downregulated genes upon heat shock. To identify cluster-specific differences in the response to heat shock, we compared gene expression of cells within individual clusters to the rest of the cells across treatments. We observed the largest number of cluster-specific gene expression changes in hair, non-hair and cortex cells (**Figure 2.7F**). As these cell types are the three outermost cell layers of the root, they may be exposed more directly to the heat shock and respond more quickly. Genes

differentially expressed in hair cells (Louvain component 2) upon heat shock were enriched for ribosome associated genes and RNA methylation. Stele cells (Louvain component 6) showed differential expression of genes involved in cell wall organization and biogenesis, and endodermis cells (Louvain component 4) showed differential expression of genes involved in response to external, chemical and stress stimuli as well as nitrate and anion transport (**Figure 2.7F**).

The expression of heat shock proteins protects cells from heat shock and aids their recovery (Parsell et al., 1993; Parsell and Lindquist, 1993; Queitsch et al., 2000). We were interested in whether we could detect cluster- and cell-type-specific differences in the canonical heat shock response. In principle, such differences could be exploited to alter heat shock protein expression in a cell-type-specific manner to boost plant heat and drought tolerance without pleiotropically decreasing whole-organism fitness. To address such possible differences, we focused on genes that from bulk analyses have differential expression upon heat shock (1783 genes) or reside near regulatory regions that change in accessibility upon heat shock (1730 genes) (Alexandre et al., 2018; Sullivan et al., 2014). Although these gene sets overlap (942 genes), they contain complementary information, as changes in accessibility do not necessarily translate into altered expression, and vice versa (Alexandre et al., 2018). In our single-cell expression analysis, we identified 752 of 1783 heat-responsive genes as differentially expressed upon heat shock, and 564 of 1730 genes near dynamic regulatory regions as differentially expressed. We hierarchically clustered control and heat shock-treated single-cell transcriptomes for both gene sets (**Supplemental Figure 2.12A, Supplemental Figure 2.12B**), resulting in several gene clusters with distinct expression patterns. Overall, cellular responses were dominated by the canonical heat-shock response, as visualized in cluster 4 (**Supplemental Figure 2.12A**) and cluster 2 (**Supplemental Figure 2.12B**). The 63 genes showing extreme accessibility and high expression upon heat shock (Sullivan et al., 2014) are

largely contained in these two clusters (**Figure 2.7B**, cluster 4, 49 of 63; **Supplemental Figure 2.12A**, cluster 2, 42 of 63).

Our analysis also revealed subtle but significant differences among some tissue types (**Figure 2.5B, Figure 2.5C,** *e.g.* clusters 3 and 8, **Supplemental Figure 2.12B,** *e.g.* clusters 5 and 7, **Supplemental Data Set 2.6**). Although most of these gene clusters were not enriched for specific annotations, cluster 8 genes were associated with rRNA metabolic processes (p-value=0.048) and cluster 5 genes (**Supplemental Figure 2.12B**) were enriched for transport genes (p-value=0.045). These results demonstrate both the promise and the challenges inherent in comparing single-cell data across different conditions and treatments.

2.4 Discussion

Here, we use *A. thaliana* roots to establish both experimental and analytic procedures for single-cell RNA-seq in plants. Using Monocle 3, we could assign over 3000 cells to expected cell and tissue types with high confidence. In particular, cortex, endodermis and hair cells were easily identified. However, distinguishing other cell types was challenging. For example, non-hair and columella cells had high similarity in their expression profiles, consistent with their correlation in bulk expression data (Brady et al., 2007; Cartwright et al., 2009). Similarly, it was difficult to designate cells in Louvain component 8 as early non-hair cells, as these cells showed overlapping expression signatures for early non-hair cells, lateral root caps, and epidermis cells before differentiation to hair and non-hair cells. These Louvain component 8 cells were difficult to distinguish further with the sparse expression data typical for single cell analysis, however we postulate that in fact the branch of component 8 may actually be the root of the trajectory and are cells dividing out of the epidermis/root cap precursor and cells either become root cap cells or epidermis.

We also could not initially split stele tissue into individual cell types, likely because the difficulty of digesting the cell walls of the tightly packed vascular bundle resulted in fewer cells than expected (Brady et al., 2007; Cartwright et al., 2009). However, analyzing stele cells separately yielded 6 sub-clusters, which correspond to known vasculature cell types. Our approach to annotate these sub-clusters exemplifies the ad hoc nature of current single-cell genomics studies, which require all available sources of information to be exploited to interpret the genomic data. Neither Spearman rank correlations with sorted bulk RNA-seq data nor microarray expression data yielded obvious cluster identities. However, mean expression values of genes known to be expressed in vasculature cell types allowed us to assign the stele sub-clusters.

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

We explored the relationships of endoreduplication, transcriptional rates, and differentiation to find that transcriptional rates, measured as mRNA velocity, increase with increasing ploidy. However, this transcriptional increase appears to be limited to genes specifically expressed in hair cells, as overall levels of RNA decreased over pseudotime. These

observations are consistent with hair cells becoming more specialized and moving towards a terminally differentiated state over time. However, this phenomenon of increasing specialization was not as apparent in other cell types. This difference may be due to biological causes, such as the higher rates of endoreduplication in hair cells, or to technical causes, such as the better clustering and trajectory of hair cells compared to the other cell types assayed.

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

We explored the *A. thaliana* heat shock response with single-cell RNA-seq because not all cells and tissues are equally competent to respond to stress. By identifying plant cell types that most strongly respond to abiotic stresses such as heat, drought, and nutrient starvation, ultimately we may be able to genetically manipulate relevant cell types to generate stress-tolerant crops without pleiotropically affecting plant fitness and yield. Although all heat-shocked cells showed gene expression changes typical of the canonical heat shock genes, we detected subtle but highly significant expression differences among cells and tissue types for other genes. Thus, single-cell transcriptomics across stress conditions holds potential for future crop breeding and genetic engineering. However, such analyses require much larger numbers of

cells than currently accessible by droplet-based methods. Moreover, such analyses should focus on treatments that are less overwhelmed by a strong canonical signal to increase resolution in detecting cell-type-specific differences.

In this study, we relied on the extensive and detailed expression data for bulk *A. thaliana* cell and tissue types to establish the validity of our approaches. The overwhelming correspondence of our findings with these and other data derived from traditional molecular genetics provides confidence that less well-characterized *A. thaliana* tissues and other plants, including crops, will be amenable to these approaches. Thus, continued progress on single-cell RNA-seq experiments should have a major impact on the analysis of plant development and environmental response.

2.5 Methods

Plant Material and Growth Conditions. *Arabidopsis thaliana* Col-0 seedlings were grown vertically at 22°C, on 1xMS + 1% sucrose plates covered with one layer of filter paper. Seven or eight days-old seedlings (LD, 16h light/8h dark, ~100 µmol m2 s) were collected around ZT3, and the roots/shoots excised with a sharp razor blade. For the heat-shock, seedling plates were transferred from 22°C to 38°C for 45 min (Conviron TC-26, light ~100 µmol m2 s), and the roots harvested immediately after.

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

Single-cell RNA-seq protocol. Single-cell RNA-seq was performed on fresh Arabidopsis root protoplast using the10X scRNA-seq platform, the Chromium Single Cell Gene Expression Solution (10X Genomics).

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

Data Analysis: Running Monocle 3: Dimensionality Reduction, and Cell Clustering. The output of the cellranger pipeline was parsed into R (version 3.5.0) using the cellranger R kit (version 2.0.0) and converted into a CellDataSet (cds) for further analysis using Monocle 3 alpha (version 2.99.1) (http://cole-trapnell-lab.github.io/monocle-release/monocle3/). All Monocle 3 analysis was performed on a High Performance Computing cluster using 128GB of RAM spread across 8 cores. The lower detection limit for the cds was set at 0.5, and the expression family used set to negbinomial.size().

We visualized cell clusters and trajectories using the standard Monocle workflow. Monocle internally handles all normalization needed for dimensionality reduction, visualization, and differential expression via “size factors” that control for variability in library construction efficiency across cells. After estimating the library size factors for each cell (via estimateSizeFactors), and estimating the dispersion in expression for each gene (via estimateDispersions) in the dataset, the top 1500 genes in terms of dispersion, *i.e*. 1500 genes

with the most expression variability in our dataset, were selected to order the cells into clusters. The expression values of these 1500 genes for each cell were log-transformed and projected onto the first 25 principal components via Monocle's data pre-processing function (preprocessCDS). Then, these lower-dimensional coordinates were used to initialize a nonlinear manifold learning algorithm implemented in Monocle 3 called Uniform Manifold Approximation and Projection, or UMAP (via reduceDimension) (McInnes and Healy, 2018). This allows us to visualize the data unto two or three dimensions. Specifically, we projected onto 2 components using the cosine distance metric, setting the parameters "n_neighbors" to 50, and "min_dist" to 0.1.

The Louvain method was used to detect cell clusters in our two dimensional representation of the dataset (partitionCells); this resulted in 11 cell clusters, or Louvain components. Cells were then clustered into "super" groups using a method derived from "approximate graph abstraction" (Wolf et al., 2018) and for each super group, a cell trajectory was drawn atop the projection using Monocle's reversed graph embedding algorithm, which is derived from "SimplePPT" (learnGraph) (Mao et al., 2015). This yielded 6 cell trajectories.

To further analyze the clusters we annotated as stele, Clusters 3, 4, and 7 were reclustered together and were reanalyzed using Monocle 3 as previously described except the parameter "min_dist" was changed to 0.05 when the reduceDimension function was called. This revealed 6 additional sub clusters.

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

Data Analysis: Estimating Doublets. Single-**C**ell **R**emover of Do**ublet**s (Scrublet) was used to predict doublets in our scRNA-seq data (Available at: https://github.com/AllonKleinLab/scrublet).

Using Python 3.5, Scrublet was ran using default settings as described by the example tutorial which is available as a python notebook (Available at: https://github.com/AllonKleinLab/scrublet/blob/master/examples/scrublet_basics.ipynb). The only significant change was that expected double rate was set to 0.1, in the tutorial it is 0.06.

Data Analysis: Identifying Cell Types*.* In order to categorize the cells into cell types and to apply developmental information, a deconvolved root expression map was downloaded from AREX LITE: The Arabidopsis Gene Expression Database (http://www.arexdb.org/data/decondatamatrix.zip). Using this data matrix, the Spearman's rank correlation was calculated between each cell in our dataset and each cell type and longitudinal annotation in the data matrix (3121 x 128 Spearman's rank correlations total). Specifically, we looked at the correlation of 1229 highly variable genes in our dataset. These 1229 genes represents the overlap between our 1500 highly variable genes and genes in the root expression map data matrix. Cells in our dataset were assigned a cell type and a developmental label based on the annotation with which each cell had the highest correlation. (*i.e.* if a cell correlated highest with endodermis cells in longitudinal zone 11, then it would be called as endodermis_11).

In addition to using the Spearman's rank correlation to assign cells their cell type, a set of known marker genes derived from GFP marker lines of the Arabidopsis root were used to identify cell types based on the high gene expression of these marker genes. These genes were obtained from (Brady et al., 2007; Cartwright et al., 2009). Specifically Supplemental Table 2 (Cartwright et al., 2009) was used. For the analysis comparing bulk RNA and pseudo bulk scRNA-seq data, the bulk data was obtained from Li et al. 2016 (Li et al., 2016); specifically, we used Table S5 from this study. Isoforms of each gene were averaged in order to be comparable to the pseudo bulk data. Lastly, using this same bulk RNA-seq data, the Pearson correlation was calculated between each cell in our dataset and each GFP marker line. Cells in our dataset

were assigned to a GFP marker line based on the GFP marker line with which each cell had the highest correlation.

Data Analysis: Running Monocle 3: Identifying High Specificity Genes. In order to identify differentially expressed genes between cell clusters the Moran's I test was performed on our UMAP (principalGraphTest), with the projection being broken up into 25 x 25 spatial units. Then marker genes were identified for each cluster, and each annotated grouping of clusters using a Moran's I threshold of 0.1 and a qval threshold of 0.05. In order for a gene to be considered highly specific, it must have had a specificity rating of above 0.7.

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

Data Analysis: Running Monocle 3 - Assigning Pseudotime Pseudotime analysis requires the selection of a cell as an "origin" for the pseudotime trajectory. Origin assignment was based on the Spearman's rank assignments for each cell. The following cells were used as origins for their respective cell type trajectories: cortex_2, hair_2, endodermis_2, nonHair_3. The get_correct_root_state() function was used to assign the root of a trajectory, and the orderCells() function was used to assign cells a pseudotime value.

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

Data Analysis: Calculating significance with the Permutation Test. The permutation test was used to calculate the significance of the observed trends that the total mRNA of hair marker genes and hair specific genes increased as pseudotime increased in hair cells. To do this, 10000 random samplings of 441 genes (the number of hair marker genes), and 201 genes (the number of hair specific genes) were taken respectively. Next, the median total mRNA was calculated across pseudotime for each random sampling and the slope of this data was calculated using a generalized linear model. The observed slope of the marker genes and the hair specific genes was compared to the distribution of slopes generated by 10000 random samplings. No random sampling of genes had a slope that was higher than the observed slopes generated by the hair marker genes or the hair specific genes. The significance, or the p-value, of the trend seen in the hair marker genes and the hair specific genes can then be calculated simply as the proportion of sampled permutations that have a slope that is equal to or greater than slope generated by our genes of interest. This gives us a p-value of 1/10001 or roughly 1×10^{-4}.

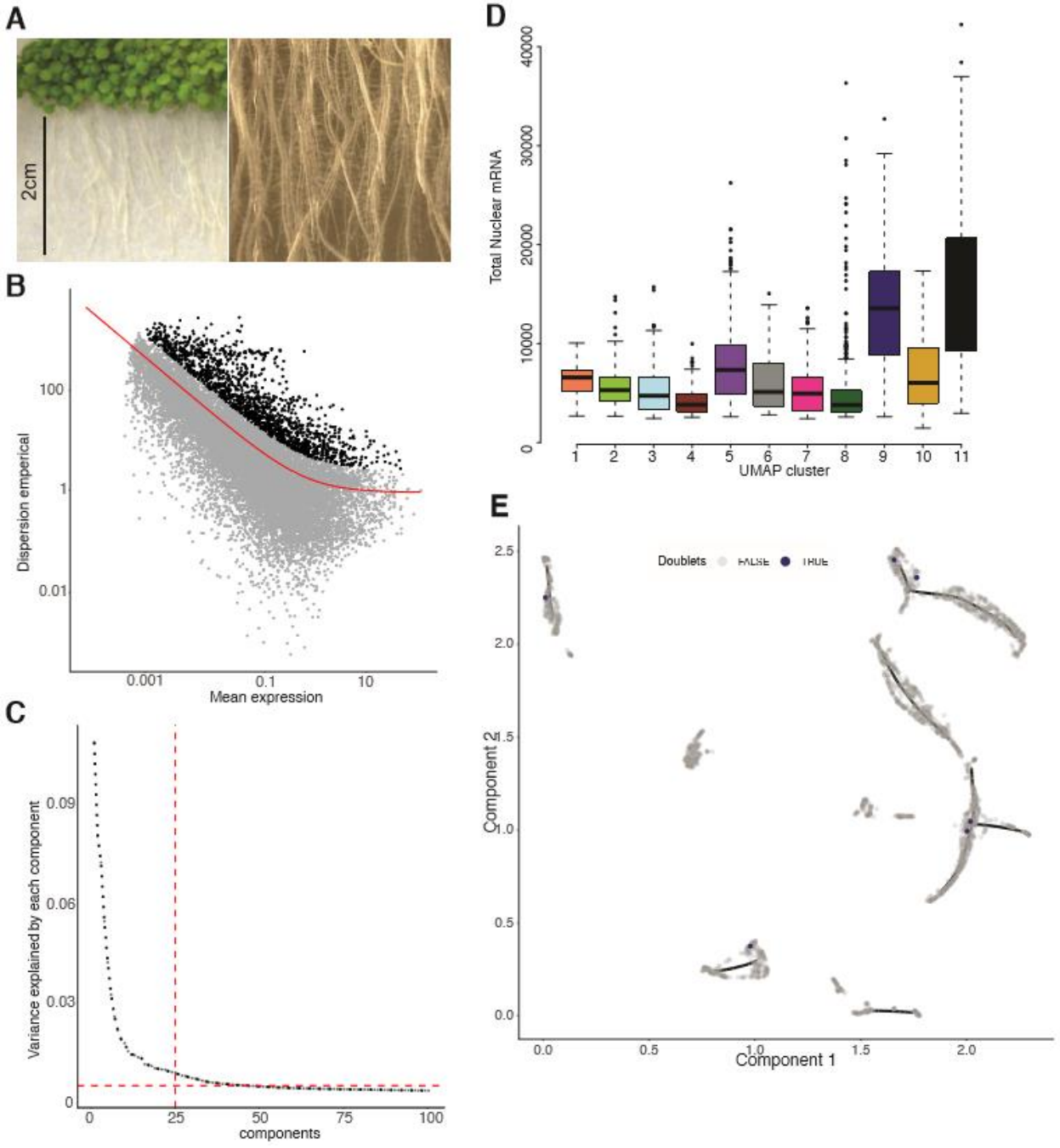
Data Analysis: Analyzing Expression Differences Between Branches of Louvain Component 8 (Early Non-Hair). To identify genes responsible for the branching in the pseudotime trajectory of Louvain component 8 (early non-hair), the principal graph test was used to identify genes with expression specific to the side branch vs. the main branch. Genes were considered specific if it had a specificity value above 0.8. Genes were removed from the analysis if they did not have expression in at least 10% of the cells considered and a mean expression greater than 0.25.

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

Data Analysis: Analysis of heat shock data. For each pair of cell types and for each gene cluster, we used a generalized linear model to determine the significance of an interaction between the effects of cell type and heat treatment on the normalized expression level of genes in that cluster. Then, to identify differentially expressed genes specific for every Louvain cluster we subsetted cells from every cluster that contained 15 or more cells in both control and treated conditions, estimated dispersions for each subset and tested for differential gene expression identified using the differentialGeneTest function in Monocle specifying a full model of Treatment cluster and a residual model of 1. FDR values per gene were then obtained across all tests using the Bejamini-Hochberg method. The overlap of differentially expressed genes as a function of heatshock treatment between clusters was visualized using an UpsetR plot. Briefly, a binary matrix of differentially expressed genes by cluster was generated were gene-cluster combinations were set to 1 (significant) or 0 (not significant). This matrix was then passed to the upset function from the UpsetR R package specifying 9 sets and ordering by frequency. To

identify whether clusters contained subtle differences in the expression of previously identified heat shock responsive genes we tested for differential gene expression across all cells and clusters and identified the intersect between differentially expressed genes obtained from single cell profiles and previously identified dynamic changes in DHS linked genes and bulk differentially expressed genes upon heat shock. Differentially expressed genes as a function of heat-shock treatment for all cells in unison where identified using the differentialGeneTest function in Monocle specifying a full model of Treatment*UMAP cluster and a residual model of UMAP cluster. Hierarchical clustering of these DHS linked and bulk differentially expressed gene sets across control and heat-shock treated cells was performed using the pheatmap function in the pheatmap R package (version 1.0.10) specifying ward.D2 as the clustering method. Genes with similar dynamics across treatment and cell types were recovered using the cutree function from the stats package in R specifying k = 8 for both DHS linked genes and bulk differentially expressed genes. To generate signatures from these 8 groups of clustered genes we log normalized expression values using a pseudocount of 1 and for each cell calculated the mean normalized expression value across genes that belong to one of the 8 gene clusters.

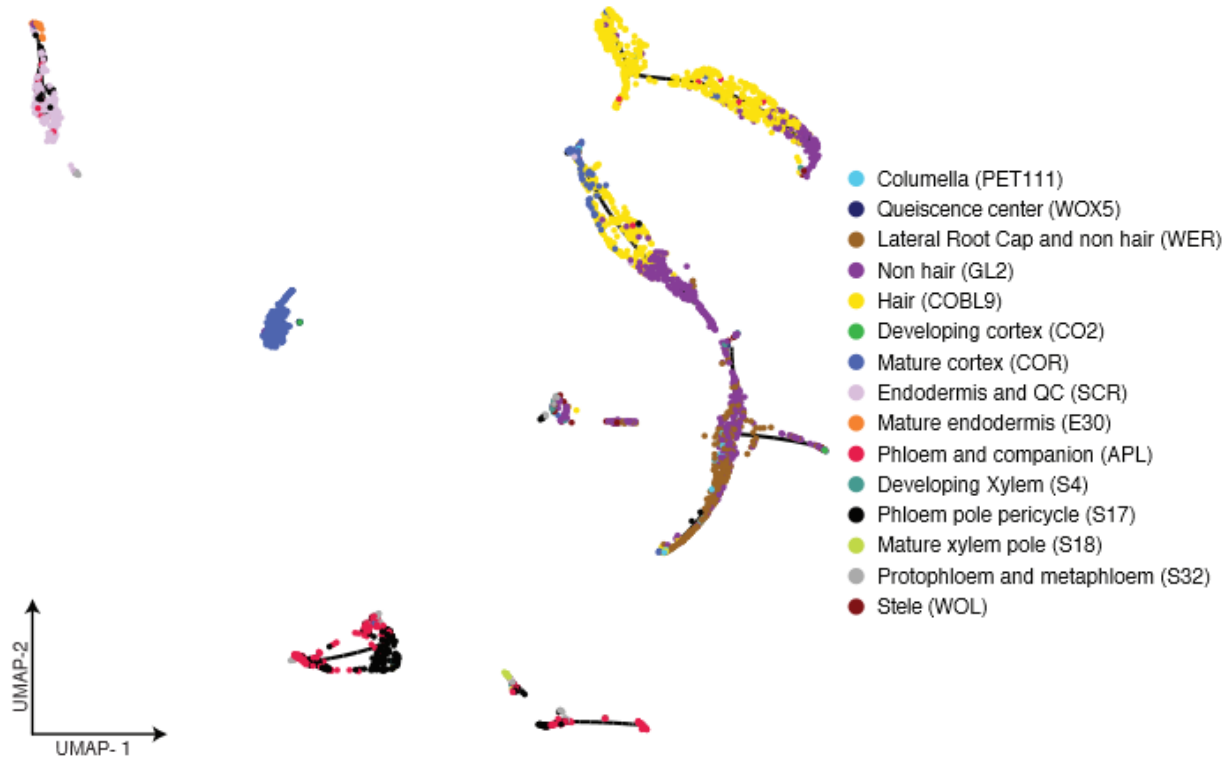
2.6 Supplementary Figures and Tables



Supplemental Figure 2.1. **General tissue and data features**.

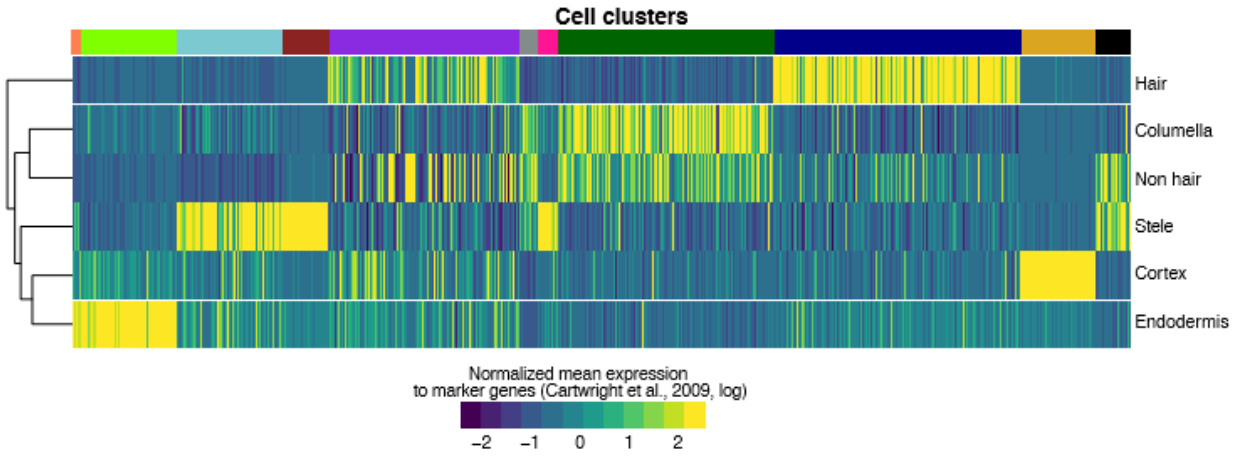
(A) Image of seven-day-old roots used for protoplast preparation. **(B)** Dispersion of gene expression across cells compared to average gene expression. The 1500 ordering genes marked in black had an empirically calculated dispersion at least 2.7 times higher than the

dispersion curve fit for the expression of all genes. **(C)** Principal components are plotted against explained variance in gene expression; 25 components were selected. **(D)** Total nuclear mRNA per Louvain component. Louvain components are colored as in **Figure 2.1A**.



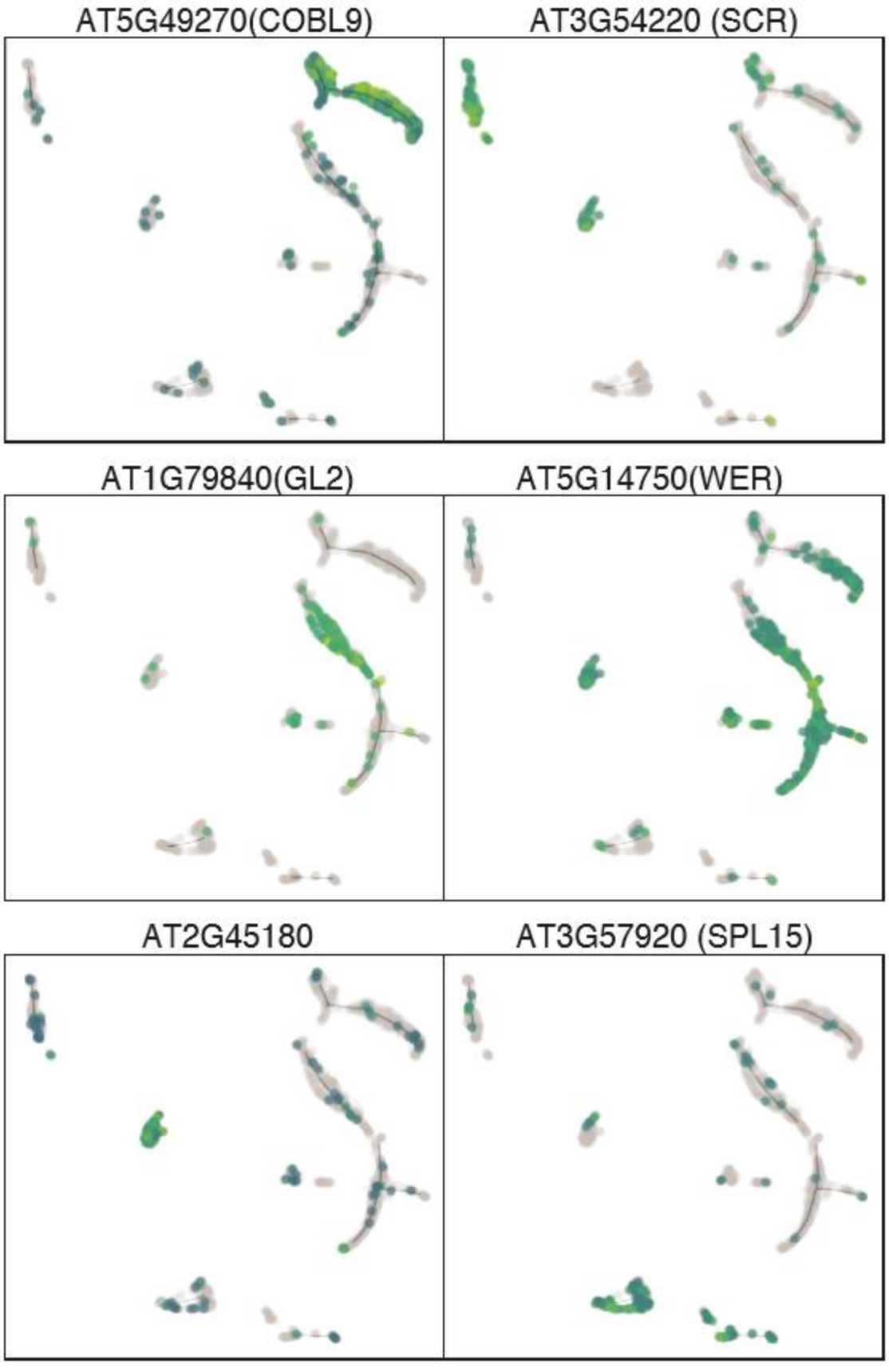
Supplemental Figure 2.2. Pearson's correlation to sorted bulk RNA-seq samples.

We correlated expression in each individual cell with previous RNA-seq expression data (Li et al., 2016), assigning cell-type annotation based on the highest Pearson's correlation.



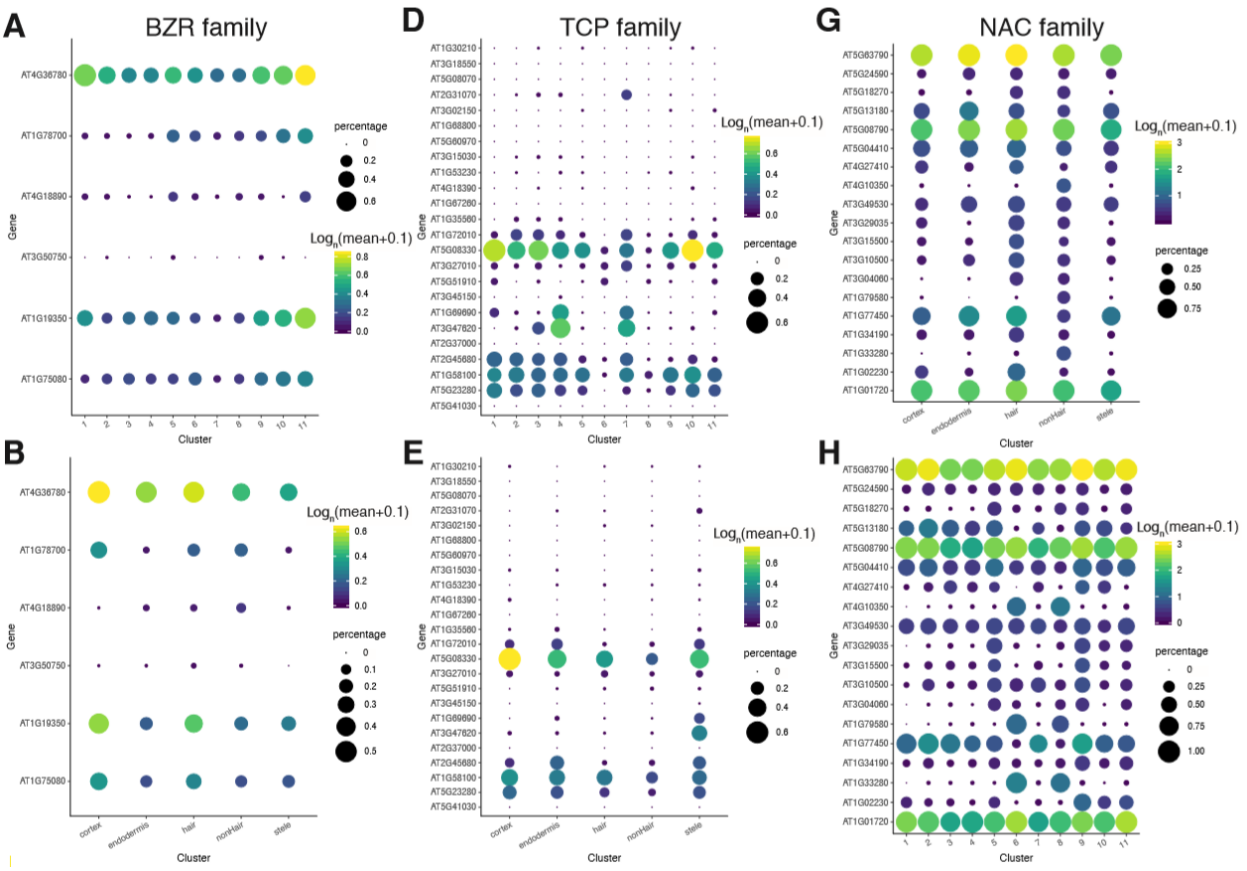
Supplemental Figure 2.3. Marker gene expression in cell type clusters.

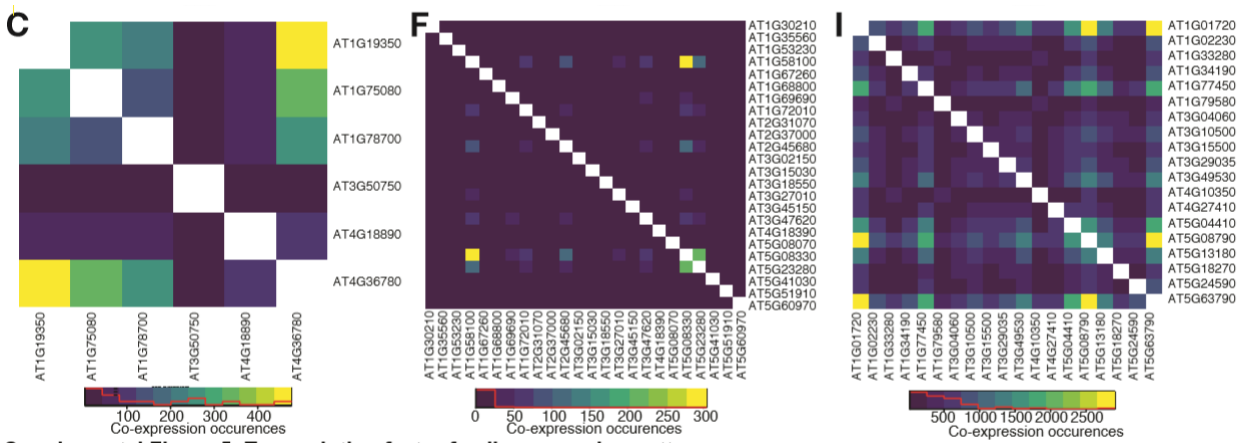
Known marker gene expression (Brady et al., 2007; Cartwright et al., 2009) was clustered based on expected respective cell types and tissues; mean gene expression is shown for each cell. Cells are clustered by Louvain components across the x-axis.


AT5G49270(COBL9)
AT3G54220 (SCR)
AT1G79840(GL2)
AT5G14750(WER)
AT2G45180
AT3G57920 (SPL15)

Supplemental Figure 2.4. Examples of tissue-specific gene expression.

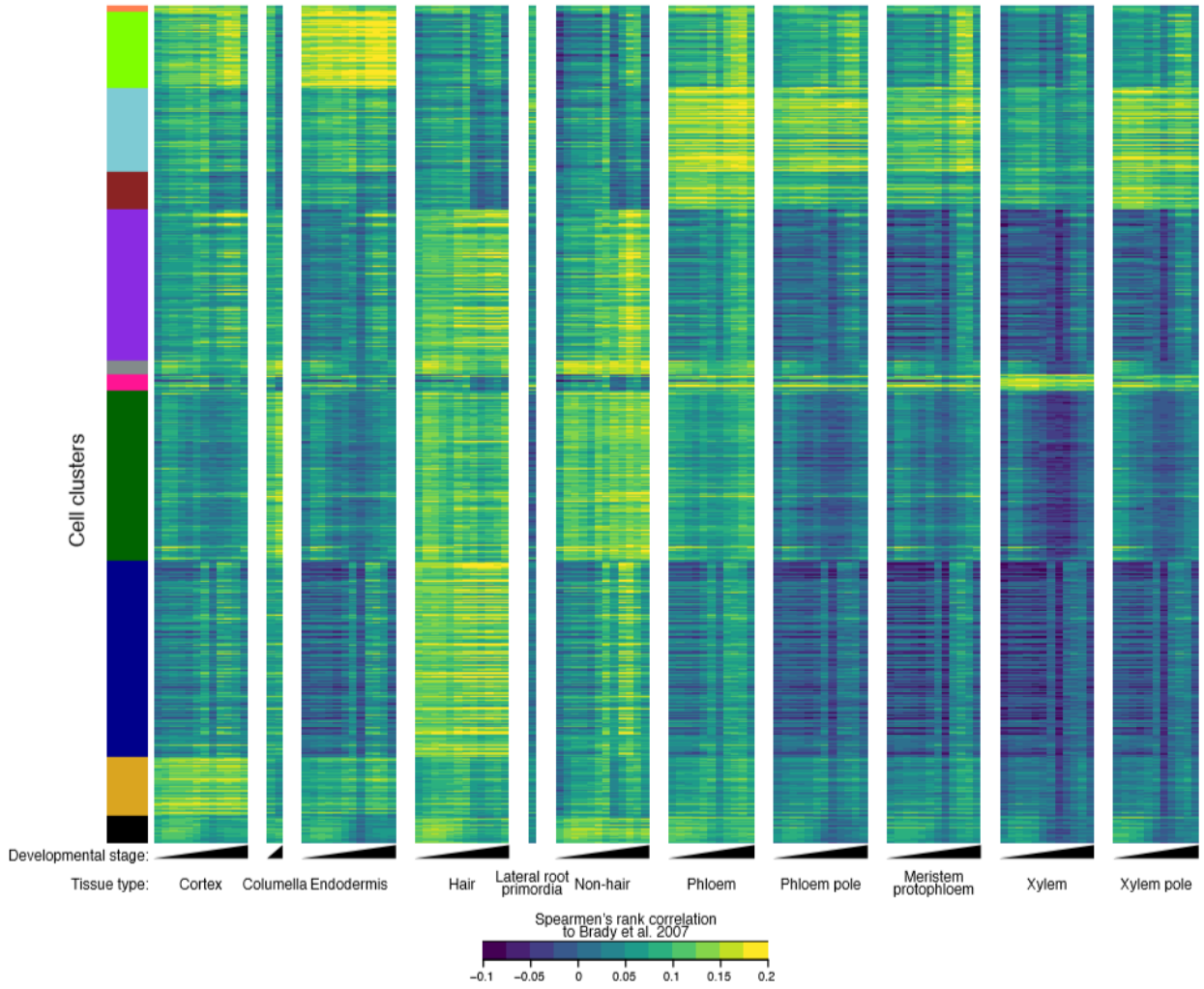
Expression of several example genes is shown superimposed onto cells across the UMAP clusters; shades of green to yellow indicate expression levels, with lighter color indicating higher expression; cells without detectable expression are in grey. Examples include known and novel genes with tissue-specific expression such as *COBL9*, *SCR*, *GL2*, *WER*, *SPL15* and *AT2G45180*.





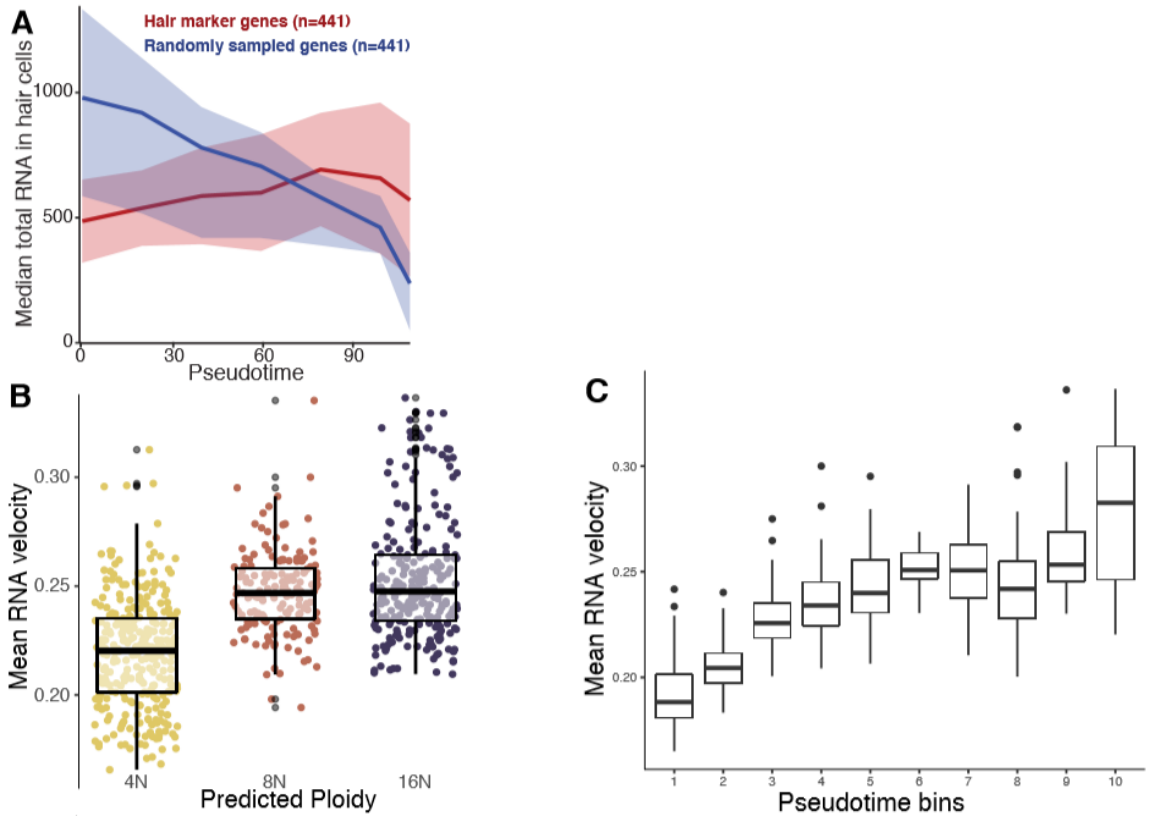
Supplemental Figure 2.5. Transcription factor family expression patterns.

(A) Louvain and **(B)** tissue-specific expression for individual members of the *BEH/BZR* transcription factor family. Circle size indicates proportions of cells expressing a given gene; circle color indicates expression level, with lighter colors corresponding to higher expression. **(C)** *BEH/BZR* gene family members co-expressed in higher numbers of cells with lighter colors. Specific patterns are identified for several genes, *e.g. BES1* (AT1G75080) and *BEH2* (AT4G36780). **(D)** Louvain and **(E)** tissue-specific expression for the TCP transcription factor family. **(F)** TCP gene family members co-expressed in higher numbers of cells with lighter colors. Specific patterns are identified for several TCP members, including for *TCP15* (AT1G69690) and *TCP14* (AT3G47620), two genes primarily expressed in stele. **(G)** Louvain and **(H)** tissue-specific expression for the NAC transcription factor family. **(I)** NAC gene family members co-expressed in higher numbers of cells with lighter colors. Specific patterns are identified for several NAC members, *e.g. ATAF1* (AT1G01720) and *ATAF2* (AT5G08790).



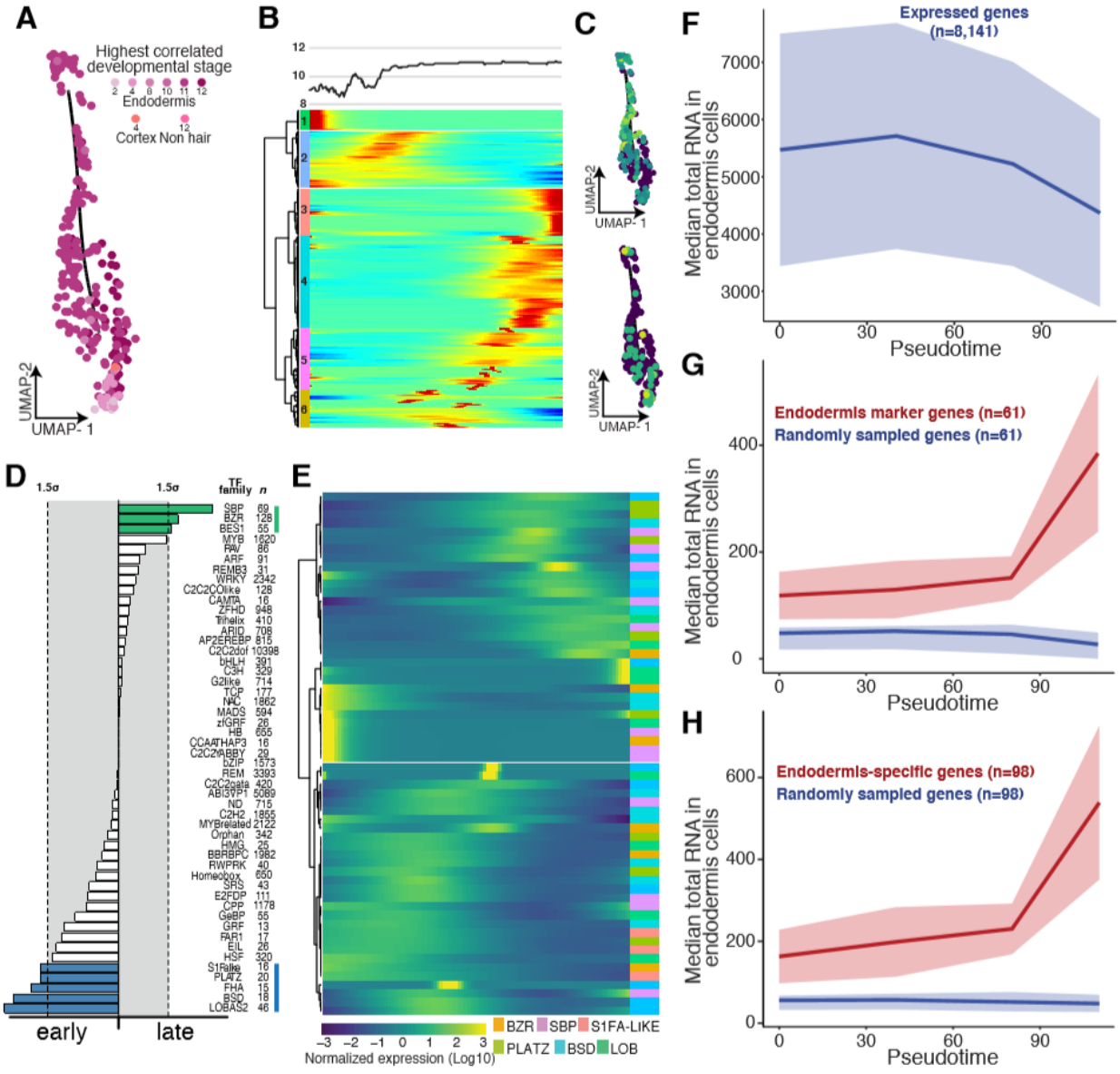
Supplemental Figure 2.6. Spearman's rank correlation for each cell's development and tissue-type.

Spearman's rank correlation coefficients determined for each tissue and developmental stage (Brady et al., 2007; Cartwright et al., 2009). Development stages are ordered left to right, youngest to oldest. Not all cell and tissue types are represented in every developmental stage; *e.g.* columella is represented only in stages 1 and 2, lateral root primordia are represented only in stage 12, and phloem is represented from stages 3 to 13. All others are represented from stages 2 to 13.



Supplemental Figure 2.7. Changes in transcription across hair cell development.

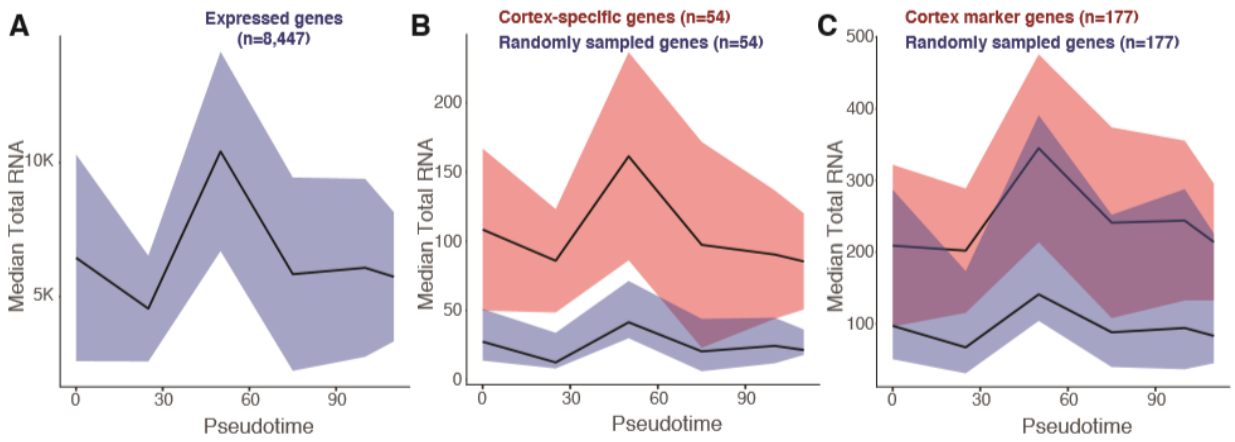
(A) Comparison of median total RNA captured for known hair marker genes (in red, Cartwright et al. 2009) versus a comparable set of random genes (in blue). Median total RNA is plotted; standard deviation is shaded in respective colors. Expression of hair marker genes differed significantly (Permutation test p-value ≈ 1 x 10^{-4}). **(B)** Using published predictions of ploidy in hair cells (Bhosale et al., 2018), and our best rank assignment of developmental time point, we assigned each cell a ploidy and compared RNA velocity across these differing predicted ploidy (Tukey's multiple comparison p-value = 0.0477). **(C)** RNA velocity, an indicator of transcriptional dynamics that relies on the ratio of mature (spliced) to nascent (unspliced) mRNA detected in standard scRNA-seq data (La Manno *et al*., Nature, 2018), increases over pseudotime in hair cells (Pearson rho = 0.73).



Supplemental Figure 2.8. Developmental trajectory of endodermal cells.

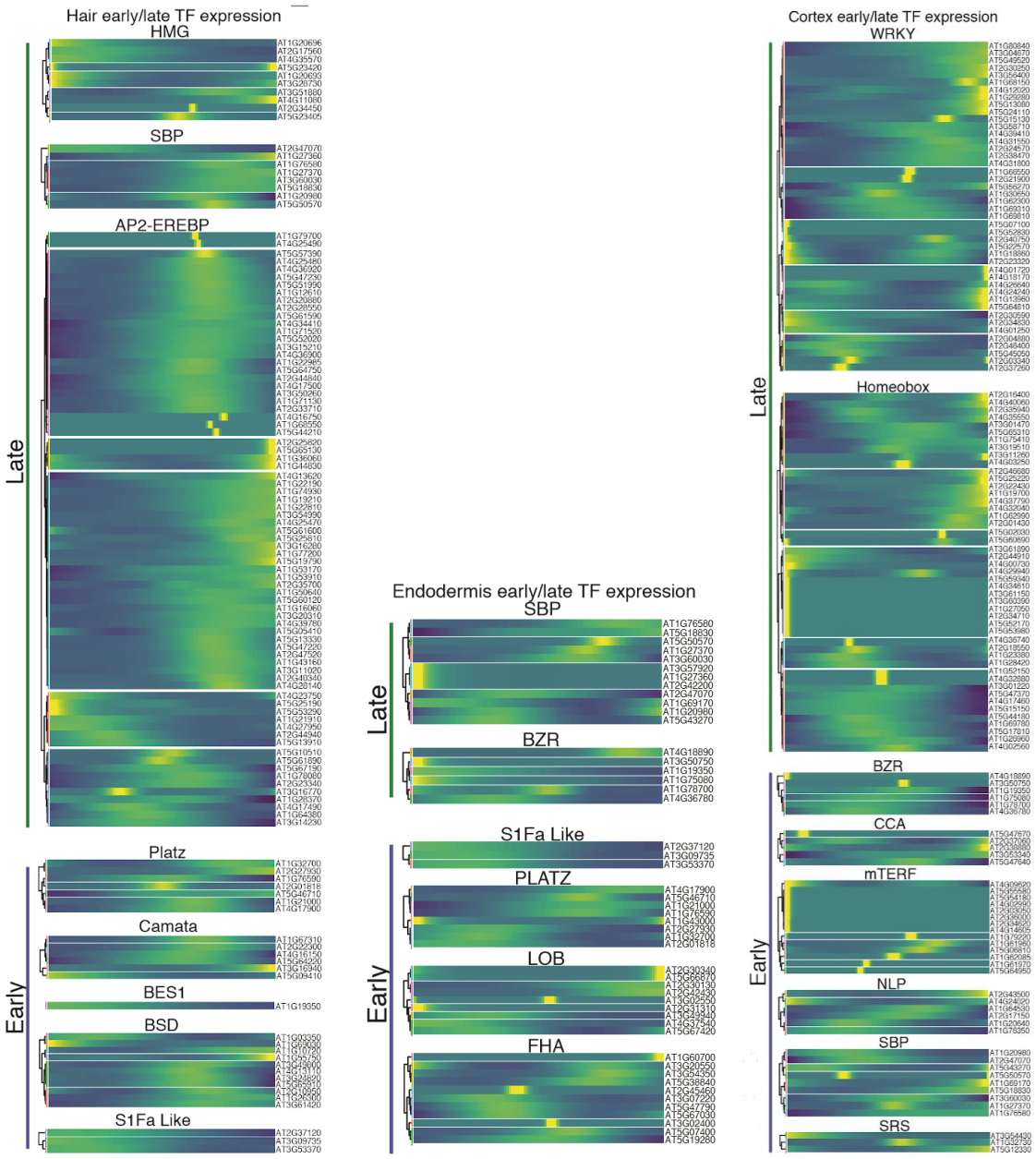
(A) UMAP-clustered endodermal cells were assigned a developmental time point based on highest Spearman's rank correlation with bulk expression data of staged tissue (13 developmental stages) (Brady et al., 2007; Cartwright et al., 2009). Cell type and developmental time point are indicated in shades of magenta. **(B)** Cells were ordered in pseudotime; columns represent cells, rows represent expression of the 1500 genes. Rows were grouped based on similarity in gene expression, resulting in 6 clusters (indicated left), with genes in clusters 1 and

2 expressed early in pseudotime and genes in clusters 3 and 4 expressed late. Endodermal cells with the earliest developmental signal (Brady et al., 2007; Cartwright et al., 2009) were designated as the root of the trajectory. The graph above represents the average best-correlation of developmental stage (Brady et al., 2007; Cartwright et al., 2009) in a scrolling window of 20 cells with pseudotime, showing the expected increase in developmental age with increasing pseudotime. **(C)** Examples of an early and a late expressed endodermis-specific gene; gene expression in each cell is superimposed onto the UMAP cluster, with lighter colors indicating higher gene expression. **(D)** Different transcription factor motifs reside in the 500 bp upstream regions of genes expressed early (clusters 1, 2) compared to genes expressed late (clusters 3, 4). Transcription factor motifs specific to early hair cells are denoted with blue bars, those for late hair cells with green bars; bar length indicates motif frequency. Thresholds on either side (grey box, dotted lines) refer to 1.5 standard deviation above mean motif frequency. **(E)** Expression of individual members of transcription factors families highlighted in **D** across pseudotime identifies candidate factors driving early or late gene expression. **(F)** Median total RNA captured in cells across endodermis pseudotime. Number of genes included is indicated. **(G)** Comparison of total RNA captured for known endodermis-cell-specific genes (in red, Cartwright et al. 2009) versus a comparable set of random genes (in blue). Median total RNA is plotted; standard deviation is shaded in respective colors. **(H)** Comparison of total RNA for endodermis-cell-specific genes (in red) to a comparable random set of genes (in blue). Number of genes is indicated.

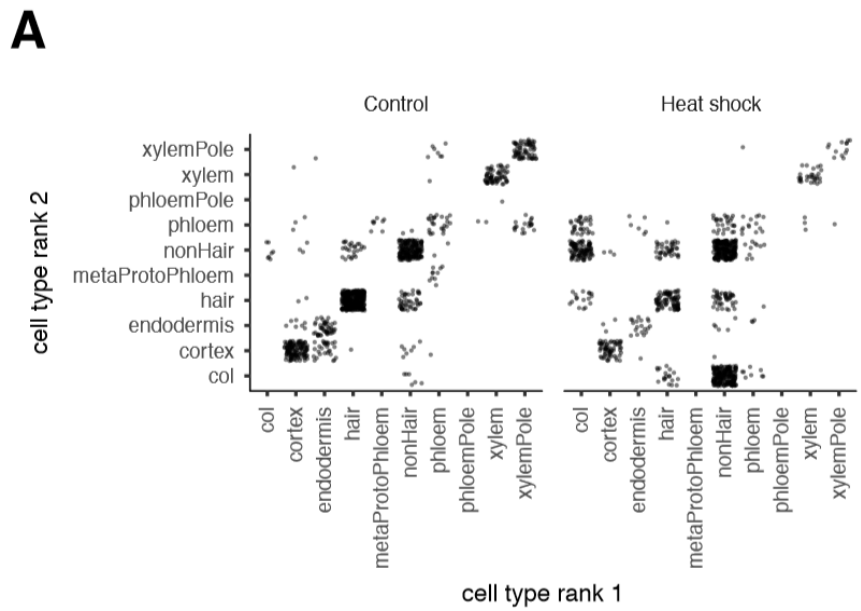


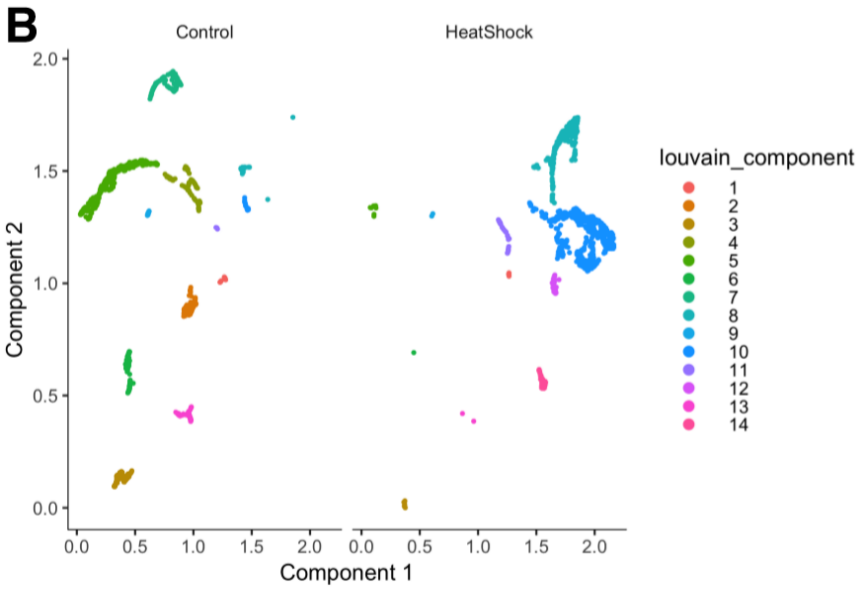
Supplemental Figure 2.9. Median total RNA in cortex cells across pseudotime.

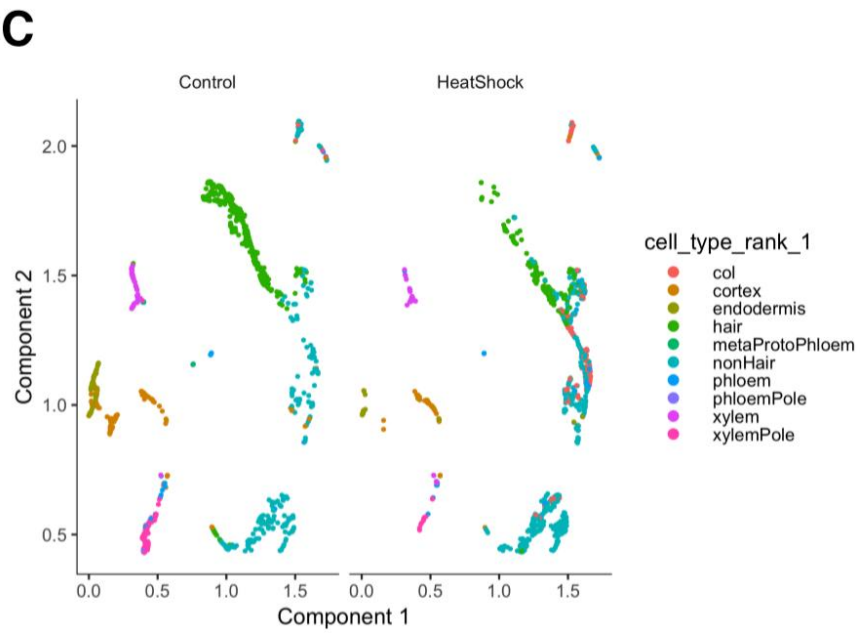
(A) Total median RNA captured in cortex cells across pseudotime. Number of genes included is indicated. **(B)** Comparison of median total RNA for cortex-cell-specific genes (in red) to a comparable random set of genes (in blue). Number of genes is indicated. **(C)** Comparison of median total RNA for known cortex genes (in red) to a comparable set of random genes (in blue). Number of genes is indicated.



Supplemental Figure 2.10. **Developmental expression of individual transcription factors.** Expression of individual members of transcription factor families with enriched early (indicated with blue bar) or late motifs (indicated in green) for hair, cortex, and endodermis cells. Lighter colors indicate higher expression; x-axis denotes pseudotime progressing from early to late as left to right.

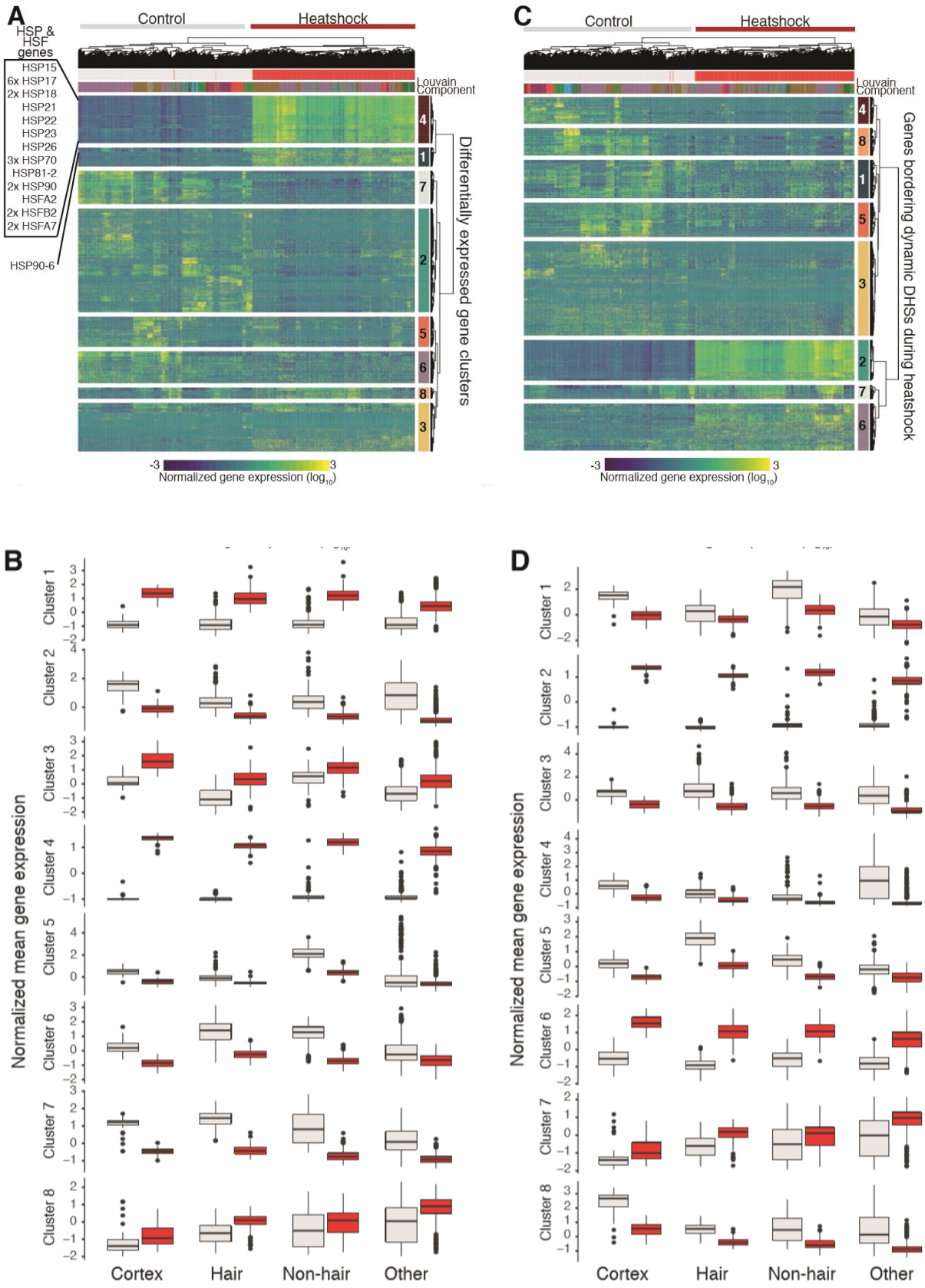

A
Control
Heat shock
cell type rank 2
xylemPole
xylem
phloemPole
phloem
nonHair
metaProtoPhloem
hair
endodermis
cortex
col
cell type rank 1


B
Control
HeatShock
Component 2
Component 1
louvain_component
1
2
3
4
5
6
7
8
9
10
11
12
13
14



Supplemental Figure 2.11. **Heat-shock clustering and expression profiling.**

(A) Upon heat shock cells are less reliably identified by cell type, as shown by their Spearman's rank correlations across cell types and development. Under control conditions, a cell's second highest correlation is most likely with the same cell type at a different developmental stage (left). Upon heat shock, this relationship between cell ranks is perturbed (right). (**B**) Conventional Monocle clustering and UMAP projection results in cells being primarily clustered by treatment (control or heat shock) rather than cell or tissue type. (**C**) A mutual nearest neighbor approach takes treatment into account to generate comparable clusters; however, some clusters change in size, become ambiguous, or are lost.


A
Control
Heatshock
HSP & HSF genes
HSP15
6x HSP17
2x HSP18
HSP21
HSP22
HSP23
HSP26
3x HSP70
HSP81-2
2x HSP90
HSFA2
2x HSFB2
2x HSFA7
HSP90-6
Louvain Component
Differentially expressed gene clusters
-3
3
Normalized gene expression (log10)
C
Control
Heatshock
Louvain Component
Genes bordering dynamic DHSs during heatshock
-3
3
Normalized gene expression (log10)
B
Normalized mean gene expression
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Cluster 7
Cluster 8
Cortex
Hair
Non-hair
Other
D
Normalized mean gene expression
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Cluster 7
Cluster 8
Cortex
Hair
Non-hair
Other

Supplemental Figure 2.12. Genes known to be associated with the heat shock response show subtle expression differences across clusters.

(A) Single-cell expression values for genes differentially regulated in response to a similar heat shock in bulk tissue (Alexandre et al., 2018) were clustered based on similarity in gene expression in control versus heat-shocked cells. Clusters are indicated at right. Cluster 4 contains the vast majority of acute heat-shock induced genes, for which chromatin accessibility extends along the gene body upon heat shock (Sullivan et al., 2014). **(B)** Pairwise comparisons of cell and tissue types reveal subtle but highly significant differences in the heat shock response for some clusters. Significance was determined with a generalized linear model. There is no cell or tissue-type specific response for cluster 4, as all cells show the canonical heat stress response typical of genes in this cluster. P-values shown in **Supplemental Table 6**. **(C)** Single-cell expression values for genes previously identified as residing near chromatin accessible sites that are dynamic in response to heat shock (Sullivan et al., 2014) clustered based on similarity in gene expression in control versus heat-shocked cells. Clusters are indicated at right. Cluster 2 contains the vast majority of acute heat-stress induced genes, for which chromatin accessibility extends along the gene body upon heat shock (Sullivan et al., 2014). **(D)** Pairwise comparisons of cell and tissue types reveal subtle but highly significant differences in the heat-shock response for clusters from genes associated with dynamic chromatin accessibility in **A**. P-values shown in **Supplemental Data Set 6.**

Bulk GFP Line	Single Cell Group	Pearson Correlation	Spearman's Rank Correlation
WOL	WOL expressing cells	0.45	0.66
WOL	Stele	0.51	0.68
WER	WER expressing cells	0.59	0.75
WER	Non-hair	0.54	0.72
SCR	SCR expressing cells	0.53	0.72
SCR	Endodermis	0.50	0.71
S4	Stele	0.37	0.57
S32	Stele	0.51	0.69
S18	Stele	0.66	0.80
Pet111	Non-hair	0.52	0.71
S17	Stele	0.66	0.80
GL2	GL2 expressing cells	0.58	0.74
GL2	Non-hair	0.62	0.78
COR	Cortex	0.74	0.86
E30	Endodermis	0.26	0.48
COBL9	COBL9 expressing cells	0.75	0.87
COBL9	Hair	0.74	0.87
APL	Stele	0.68	0.82
CO2	Cortex	0.3	0.55
Maturation	Whole root	0.49	0.70
Elongation	Whole root	0.7	0.83
Meristimatic	Whole root	0.21	0.42
Whole root	Whole root	0.52	0.71

Supplemental Table 2.1 Bulk RNA-seq comparisons to single cell RNA-seq data

Annotation / Louvain Component	Treatment ID	Number of Cells
Columella	Control	7
Columella	Heat Shock	153
Cortex	Control	149
Cortex	Heat Shock	71
Endodermis	Control	77
Endodermis	Heat Shock	26
Hair	Control	371
Hair	Heat Shock	151
MetaProtoPhloem	Control	8
MetaProtoPhloem	Heat Shock	0
NonHair	Control	256
NonHair	Heat Shock	524
Phloem	Control	46
Phloem	Heat Shock	37
PhloemPole	Control	1
PhloemPole	Heat Shock	0
Xylem	Control	70
Xylem	Heat Shock	33
XylemPole	Control	91
XylemPole	Heat Shock	14
Louvain Component 1	Control	24
Louvain Component 1	Heat Shock	5
Louvain Component 2	Control	400
Louvain Component 2	Heat Shock	415
Louvain Component 3	Control	160
Louvain Component 3	Heat Shock	283
Louvain Component 4	Control	97
Louvain Component 4	Heat Shock	18
Louvain Component 5	Control	120
Louvain Component 5	Heat Shock	30
Louvain Component 6	Control	77
Louvain Component 6	Heat Shock	26
Louvain Component 7	Control	60
Louvain Component 7	Heat Shock	2
Louvain Component 8	Control	41
Louvain Component 8	Heat Shock	62
Louvain Component 9	Control	16
Louvain Component 9	Heat Shock	20
Louvain Component 10	Control	43
Louvain Component 10	Heat Shock	132
Louvain Component 11	Control	38
Louvain Component 11	Heat Shock	16

Supplemental Table 2.2 Number of cells in the control vs. heat shock analysis

2.7 Supplementary Data Availability

The raw data have been deposited with the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo) under accession code GSE121619. Supplemental Data Sets 2.1-2.6 are available online (http://www.plantcell.org/content/31/5/993/tab-figures-data).

2.8 Project Acknowledgements

We thank members of the Trapnell Lab for input and discussion. This work was supported by NSF MCB-1516701 to CQ and NSF RESEARCH-PGR 1748843 to CQ and SF. KJB was supported by the University of Washington NIH Big Data for Genomics and Neuroscience Training Grant (T32LM012419).

Chapter 3. A single cell view of the transcriptome during lateral root initiation in *Arabidopsis thaliana*

3.1 Abstract

Root architecture is a major determinant of fitness, and is under constant modification in response to favorable and unfavorable environmental stimuli. Beyond impacts on the primary root, the environment can alter the position, spacing, density and length of secondary or lateral roots. Lateral root development is among the best-studied examples of plant organogenesis, yet there are still many unanswered questions about its earliest steps. Among the challenges faced in capturing these first molecular events is the fact that this process occurs in a small number of cells with unpredictable timing. Single-cell sequencing methods afford the opportunity to isolate the specific transcriptional changes occurring in cells undergoing this fate transition. Using this approach, we successfully captured the transcriptomes of initiating lateral root primordia, and discovered many previously unreported upregulated genes associated with this process. We developed a method to selectively repress target gene transcription in the xylem pole pericycle cells where lateral roots originate, and demonstrated that expression of several of these targets was required for normal root development. We also discovered novel subpopulations of cells in the pericycle and endodermal cell files that respond to lateral root initiation, highlighting the coordination across cell files required for this fate transition.

3.2 Introduction

Plants grow new tissues and organs throughout their lives. To enable this iterative body plan, cells must maintain mechanisms of organogenesis, proliferation, and differentiation. Iterative growth of plants is most easily observed above ground, as plants put out new branches, leaves, and flowers, but equally important is the growth of root systems below the soil. Lateral root development is essential for a plant to remain stably rooted as well as to obtain

essential water and nutrients from its surrounding environment. Plants use plasticity in their growth patterns to avoid adverse stimuli and take advantage of favorable ones. In this way, changing root architecture is one of the main mechanisms by which plants can adapt to changing environmental conditions (Khan et al., 2016).

Lateral root development proceeds through three discrete stages, specification, initiation, and emergence. Auxin response exhibits cyclic maxima in the basal meristem with a periodicity of about six hours (De Smet et al., 2007). During specification, cells in the meristematic zone of the primary root are specified as competent to form lateral roots if they transit through the basal meristem when auxin response is high (De Smet et al., 2007). Many other genes oscillate in phase with auxin response in the basal meristem—these genes may be targets of auxin signaling or independent regulators of specification (Moreno-Risueno et al., 2010). These competent cells then exhibit a sustained auxin maximum in the differentiated zone of the root, termed prebranch sites. In *Arabidopsis* and most dicot plants the pericycle cell layer within these prebranch sites is the cell layer that undergoes lateral root initiation (Beeckman et al., 2001). How this initial transient auxin response is molecularly translated to the sustained auxin response of prebranch sites leading to initiation is unknown.

The earliest morphological signal of lateral root initiation is the nuclear migration of two longitudinally adjacent pericycle cells to their shared cell wall. These cells consequently undergo the first anticlinal cell division that initiates lateral root development. *GATA TRANSCRIPTION FACTOR 23 (GATA23)* is necessary for this nuclear migration to occur (De Rybel et al., 2010). *LOB DOMAIN-CONTAINING PROTEIN 16 (LBD16)* and *LOB DOMAIN-CONTAINING PROTEIN 29 (LBD29)* are two other transcription factors shown to play a role in lateral root initiation. These genes are direct targets of *AUXIN RESPONSE FACTOR 7 (ARF7)* and *AUXIN RESPONSE FACTOR 19 (ARF19)* that promote cell division (Okushima et al., 2007). Mutants of these genes exhibit a loss of lateral root initiation, and overexpression of LBD16 rescues the

arf7arf19 mutant phenotype which also lacks lateral roots (Goh et al., 2012). Further cell divisions of specific plane orientations and structural changes of cell files exterior to the pericycle allows for lateral root emergence. The emergence process is accompanied by strong upregulation of cell wall remodelers that are also targets of auxin signaling (Lewis et al., 2013; Ramakrishna et al., 2019) and appears to be the easiest stage of lateral root development to arrest, with many mutants arresting at this stage.

Transcriptomic analyses of lateral root development have been a rich resource for determining key regulators of this developmental process (Vanneste et al., 2005). Careful temporal staging and analyses of different steps during lateral root formation have led to identification of novel regulators (Voß et al., 2015) though the complexity of the pathways regulating this process has also become more apparent. Complicating this analysis is the fact that lateral root development is not cell-autonomous, with many different cell types playing different roles and activating diverse genetic networks during this process. Cell sorting analyses on lateral root development have not been done to parse tissue-specific signals, likely because the regions of the root undergoing this fate transition are prohibitively small for such analyses. Single-cell RNA-sequencing is an alternative approach to obtain transcriptomes on the level of individual cells that requires much less tissue compared to cell sorting. In plants, single-cell RNA-sequencing has been used to characterize several plant tissue types, and single-cell analyses of root transcriptomes have identified both previously characterized and novel cell type markers (Jean-Baptiste et al., 2019; Ryu et al., 2019; Shulse et al., 2019; Shahan et al., 2020). To date, single-cell analyses of root tissue have focused on gene expression in the primary root, transcriptome changes between hair cells and non-hair cells, endodermal differentiation, and regeneration of the primary root meristem after injury.

While initiation of lateral roots is known to be regulated by auxin, only a handful of specific molecular markers of this fate switch have been identified. One reason for this scarcity

of markers may be that for any given primary root, lateral root initiation only occurs in a very small proportion of xylem pole pericycle (XPP) cells at near basal meristem, which themselves are a very small proportion of the root cells (approximately five percent) (Schmidt et al., 2014). Rarity of lateral root fate transition is further complicated by the pulsatile nature of the auxin signal, making this a highly transient event (Moreno-Risueno et al., 2010). To counteract these challenges, we microdissected sections of *Arabidopsis* roots undergoing gravity-induced lateral root initiation, and subjected the resulting protoplasts to single-cell sequencing. Using this approach, we successfully captured cells from all major cell types of the root outside the meristem. Through pseudotime analyses found that cells identified as lateral root primordia (LRP) are transcriptionally derived from those identified as xylem pole pericycle cells (XPP), consistent with previous morphological analysis (Malamy and Benfey, 1997). Differential gene analyses identified many previously unreported genes that are upregulated in LRP cells as compared to XPP cells. We validated the expression patterns of a subset of these genes using fluorescent reporters. In addition, we developed a CRISPR/dCas9 tool to specifically target the repression of these candidate genes in XPP cells and found that many of these targets shape root architecture. Finally, we were able to harness the single-cell approach to determine how cells surrounding the developing primordium, specifically endodermal cells overlaying and pericycle cells flanking LRPs, are affected by this fate transition.

3.3 Results

To examine the developmental transition of lateral root initiation, we used gravistimulation to synchronize the formation of lateral root primordia, and then dissected the region of interest at two time points and performed single-cell transcriptome analyses. Mechanical or gravitropic bending of primary roots in *Arabidopsis* causes the accumulation of auxin and the formation of a lateral root at the bend (**Figure 3.1A**) (De Smet et al., 2007; Ditengou et al., 2008) In our conditions, wild-type plants have formed a primordium at either

stage I or II by twenty hours after gravistimulation (Guseman et al., 2015) As our goal was to identify early regulators of lateral root initiation, we analyzed cells twenty hours post-bending, when initiation has just begun, and eight hours post-bending, where there are no morphological signs of lateral root development but transcriptome changes have started (Voß et al., 2015). We included a control treatment group where we did not bend the roots but cut a similar region of the primary root. We microdissected the root bend regions to maximize our yield of the rare cell types of interest.

Our experiment yielded 6658 cells with a mean number of 15987 reads per cell and a median of 1383 genes expressed. Of these 6658 cells, 1730 (~26%) cells were from the twenty-hour timepoint, 2443 (~37%) cells were from the eight-hour timepoint, and 2485 (~37%) cells were from the no bend control (**Supplemental Table 3.1**). Further analysis was performed using Monocle 3 (Trapnell et al., 2014; Qiu et al., 2017; Cao et al., 2019), which uses Principal Component Analysis (PCA) and Uniform Manifold Approximation and Projection (UMAP) to reduce the dimensionality of the dataset and to visualize the relationships among cellular transcriptomes in a two-dimensional space. Our analysis revealed five clusters. Using previously defined cell type markers (Brady et al., 2007; Cartwright et al., 2009), each cluster was assigned a label. Of the 6658 cells, 813 (~12%) cells were classified as cortex & endodermis, 1015 (~15%) cells were classified as columella/root cap, 1431 (~22%) cells were classified as epidermis, and 3399 (~51%) cells were classified as stele (**Figure 3.1B, Supplemental Figure 3.1A**). These results show that the microdissection was successful in capturing a representative sample of stele cells, as the proportion of these cells in our data was similar to that determined by imaging analyses (Shulse et al., 2019).

Xylem pole pericycle cells are precursors of mature pericycle and lateral root primordia cells.

To better understand the relationships between vasculature cell types, we re-clustered all cells identified as belonging to the stele. Of the 3399 stele cells, 216 cells were labeled phloem, 242 cells were labeled xylem, 1206 cells were labeled phloem pole pericycle (PPP), 534 cells were labeled XPP, 336 cells were labeled mature pericycle, 167 cells were LRP, and 698 cells were too ambiguous in their gene expression to assign a label (**Supplemental Figure 3.1B, Supplemental Figure 3.1C**, **Supplemental Table 3.1**). A pericycle-specific UMAP was also generated that included XPP, LRP, PPP and mature pericycle cells (**Figure 3.1C**). A cell developmental trajectory connected the pericycle sub-clusters and recapitulated the known developmental relationship between these cell types. Trajectories initiating either from the XPP or the PPP converged to mature pericycle but only the XPP (and not the PPP) population branched out toward the LRP, confirming that our results were faithfully recapitulating the known exclusive relationship of XPP to LRP cells (Malamy and Benfey, 1997). The XPP and the PPP cells are expected to differentiate at a similar time, whereas LRP and mature pericycle cells should form later. Our dataset confirmed this developmental progression, as approximately twenty percent of the XPP and PPP cell clusters were from the twenty-hour timepoint whereas approximately fifty percent of the LRP and mature pericycle clusters were from this later timepoint. (**Supplemental Figure 3.1D**).

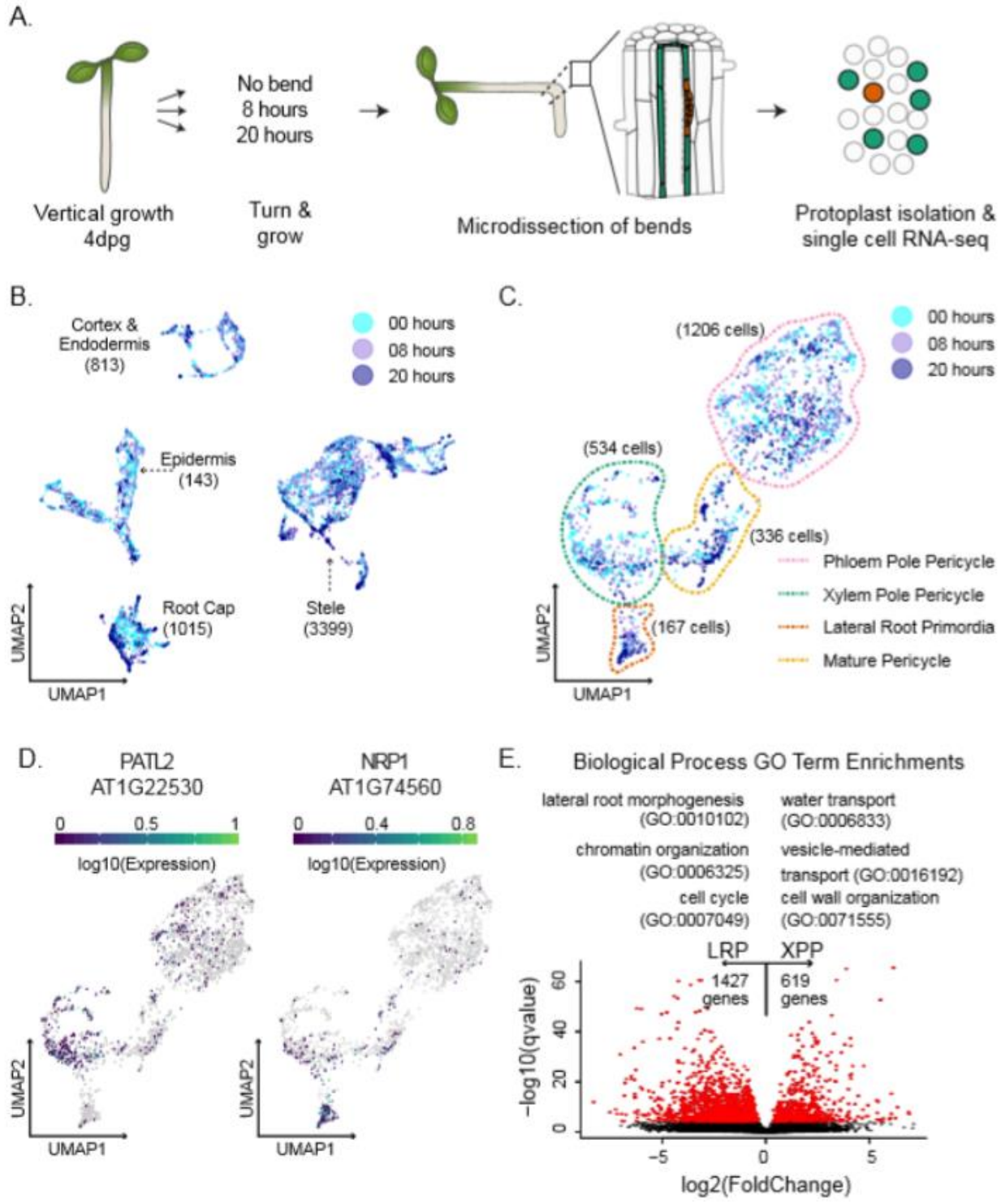


Figure 3.1. Analysis of lateral root initiation using single-cell RNA-sequencing.

(A) Experimental design: *Arabidopsis thaliana* seedlings were grown vertically for 4 days post-germination (dpg) then rotated (or marked, in case of control) and grown for an additional eight

or twenty hours. Protoplasts were prepared from microdissected root sections for single-cell RNA sequencing. **(B)** UMAP of all 6658 cells colored by experiment. Cell type identities were assigned to each partition based on a set of marker genes. The cell type identities are indicated on the UMAP with the number of cells corresponding to each cell type. **(C)** UMAP of pericycle specific cells, such as lateral root primordia (LRP), mature pericycle (MP), phloem pole pericycle (PPP), and xylem pole pericycle cells (XPP) colored by experiment. **(D)** Expression pericycle-specific UMAPs of two DEGs between LRP and XPP. E. Scatter plot of the log2 fold change of average gene expression between XPP and LRP and the Mann Whitney Wilcoxon qvalue. DEGs are colored in red. A selection of GO Terms associated with DEGs are shown above the plot.

We next attempted to identify genes expressed as XPP cells transitioned to LRP cells. To do this, we carried out a pseudotime analysis originating from the XPP population and connecting to either the LRP or the mature pericycle (**Supplemental Figure 3.2A, Supplemental Figure 3.2B**). Using a false discovery rate (FDR) cutoff of 0.1, 1892 genes were identified as changing in at least one of the trajectories. Of these genes, 878 genes were specific to the XPP to mature pericycle trajectory, 504 genes were specific to the XPP to LRP trajectory, and 510 genes were observed in both trajectories (**Supplemental Figure 3.2C, Supplemental Figure 3.2D**). As XPP cells develop, expression of genes found in both trajectories decreased and most were not expressed at high levels in the LRP cells or the mature pericycle cells. Genes uniquely upregulated across the XPP to the LRP trajectory were overrepresented with transcription factors including the well-known markers of initiation *LBD16* and *LBD29* (Okushima et al., 2007) as well as *WUSCHEL-RELATED HOMEOBOX 5* (*WOX5*), *CYTOKININ RESPONSE FACTOR 2* (*CRF2*) and *PUCHI*, all of which have previously been shown to play a role or be expressed in lateral root development (Hu and Xu, 2016; Jeon et al., 2016; Goh et al., 2019) **(Supplemental Figure 3.2E).** Other genes unique to this trajectory were regulatory kinases like *MUSTACHES* (*MUS*), *MAP KINASE KINASE 6* (*MKK6*), and *RGF INSENSITIVE 5* (*RGI5*), which have also been shown to regulate lateral root development (Zeng et al., 2011; Ou et al., 2016; Xun et al., 2020) **(Supplemental Figure 3.2E)**.

We next performed differentially expressed genes (DEG) analysis on the transcriptomes associated with these different populations. Due to the small number of LRP cells (167 cells i.e. ~2.5% of all cells), three different statistical approaches were used to perform DEG analysis: a generalized linear model (GLM), the Mann Whitney Wilcoxon (MMW) test, and a recently published packaged called Vision (DeTomaso et al., 2019). We used previously identified LRP-enriched genes as a guide to inform our use of the results from each method. *LBD16* and *LBD29* were called in all three methods, whereas *ARF19* and *GATA23* were only called in Vision (but narrowly missed with p-values of 0.001 and 0.0004 respectively in MWW method). ARF7 was not significantly different in any approach. To compile the most comprehensive list, we generated a list of DEGs that were significantly different between XPP and LRP populations in at least two approaches. We called 1427 DEGs specific to LRP and 619 DEGs specific to XPP cells. Several of these genes have been previously characterized as specific in their expression patterns, and expression maps of these genes reflect this quality (**Figure 3.1D, Supplemental Figure 3.3, Supplemental Data 3.1**) (Zhu et al., 2006; Tejos et al., 2018).

As expected, a Gene Ontology (GO) term enrichment analysis of the DEGs with higher expression in LRP cells showed a strong enrichment for terms associated with lateral root formation, lateral root morphogenesis, lateral root development, and auxin response. Terms associated with the regulation of translation initiation and RNA processing (**Supplemental Data 3.1**) were also enriched, indicating that the transition from the XPP to the LRP requires a burst of *de novo* protein production. Increased protein production is associated with a stem cell state (Himanen et al., 2004), which correlates with the transition from XPP to early LRP cells which are competent to form all the types of cells of a developing root. Other GO terms associated with the DEGs more highly expressed in LRP were lateral root morphogenesis, cell cycle, and chromatin organization (**Figure 3.1E**).

Single-cell analysis recapitulates and extends findings from previous transcriptome studies

Lateral root development has been extensively characterized with transcriptomic analysis, generated using a variety of lateral root induction models and experimental designs. We compared our LRP versus XPP DEGs with published datasets to assess differences in single-cell versus whole tissue (or population) methods. We found that roughly half of the genes (including genes regulated by auxin and genes belonging to cell cycle processes) identified by a microarray analysis of induced lateral root development (Vanneste et al., 2005) were included in our set of genes upregulated in LRP cells; only five of these genes were in our set of XPP-upregulated genes (**Supplemental Figure 3.4A**). This result suggests our approach faithfully captured genes involved in lateral root initiation and distinguished LRP and XPP cells. A detailed time course analysis of lateral root development using a similar bend assay as what we applied (Voß et al., 2015) allowed us to compare to bulk RNA sequencing data taken at similar timepoints as used in our study. Two thirds of the genes previously found upregulated at nine hours post-bending were contained within our set of genes upregulated in LRP, none of these genes were XPP-upregulated (**Supplemental Figure 3.4B**). Roughly half the genes found upregulated at twenty-one hours were within our LRP-upregulated gene list; only nine were in our XPP-upregulated gene list (**Supplemental Figure 3.4B**). At both time points, approximately one thousand genes were uniquely found in LRP DEGs from our study. The genes that were identified in the previous study but not in ours may be attributed to technical differences or likely are expressed in cell layers outside the pericycle.

We also compared our data to transcriptome assays that did not directly examine lateral root initiation through root bending, but queried related processes in the root. We compared our data to a time course analysis of primary root transcriptomes after auxin treatment (Lewis et al.,

2013), as auxin treatment strongly promotes lateral root initiation. We found that thirty-seven LRP-upregulated genes were strongly induced in this dataset in response to auxin, whereas only two XPP-upregulated genes were auxin-induced (**Supplemental Figure 3.4C**). In contrast, twenty-six XPP-upregulated genes were repressed by auxin treatment, whereas only five LRP-upregulated genes were in this repressed dataset (**Supplemental Figure 3.4C**). Another recent analysis identified genes specifically induced by ARF19-mediated auxin response (Powers et al., 2019). We found that 243 of our LRP-upregulated genes overlapped with the set of ARF19-specific auxin-induced genes, while only 19 XPP-upregulated genes did so (**Supplemental Figure 3.4D)**. We conclude that our data reflect the auxin-inducibility of lateral root initiation, and specifically that this auxin inducibility was at least in part mediated by ARF19. ARF19 is unique among the ARFs in being both auxin-responsive in its own expression pattern (Wilmoth et al., 2005) and a very strong activator of transcription itself (Lanctot et al., 2020)

Finally, we examined how our dataset compared to genes expressed in the basal meristem during lateral root specification. During specification, cells become competent to form lateral roots if they transit through the basal meristem during an auxin response maximum. Many genes exhibit similar oscillatory behavior to auxin response in the basal meristem (Moreno-Risueno et al., 2010). We found that fifty-eight genes that oscillated in phase with auxin response in the basal meristem were in our set of LRP-upregulated genes, whereas only one XPP-upregulated gene oscillated in phase with auxin (**Supplemental Figure 3.4E**). However, 213 XPP-upregulated genes oscillated antiphase to auxin in the basal meristem, while only twenty LRP-upregulated genes show antiphase oscillation. How specification and initiation are connected temporally and spatially and how competent cells “remember” their future cell fate is still unknown. Our results suggested that oscillatory behavior of some genes may predict their importance during initiation later in development, and in particular that genes with antiphase oscillation patterns may actively repress lateral root fate. We also compared our

dataset to a study that determined genes whose expression was impacted by repressing auxin response specifically in early-stage LRPs (Ramakrishna et al., 2019), and found our LRP-upregulated genes overlapped with nearly two hundred of these genes (**Supplemental Figure 3.4F**), again emphasizing the importance of auxin response for establishing lateral root fate.

Genes upregulated in LRP cells are indicative of cells undergoing fate transitions

We selected timepoints that should best capture the transition of undifferentiated XPP cells into LRPs. As one of the earliest morphological steps in this process is an asymmetric cell division, followed by many subsequent cell divisions, it is not surprising that a number of LRP-enriched DEGs are involved in cell cycle control. It is already known, for example, that expression of cyclin *CYCB1;1* marks lateral root primordia (Beeckman et al., 2001) Interestingly, one of the distinguishing features of XPP cells when compared to PPP cells is that some XPP cells are arrested in G2, whereas all PPP cells are arrested in G1 (Beeckman et al., 2001). This G2 arrest may prime these XPP cells to undergo rapid reintroduction into the cell cycle. We also found enrichment of chromatin remodeling factors in the set of LRP-upregulated genes.

Considering lateral root initiation is the first step of organogenesis, it makes sense that broad transcriptional changes, mediated by changes in the chromatin landscape, may be required, but this has not previously been reported (**Supplemental Data 3.1**). Finally, we found that genes that promote cell division, differentiation, and “stemness”, mostly transcription factors, were also enriched in the set of genes upregulated in LRP cells. Most of these genes have been characterized as regulating development in other meristems, such as the primary root meristem or the shoot apical meristem, but had not been shown to play a role in lateral root development. We chose several candidate genes from these three groups, chromatin regulators (four genes), cell cycle regulators (seven genes) and stemness regulators (six genes) to carry

out two types of validation experiments: (1) characterizing their spatial expression patterns with transcriptional reporters in wild-type Col-0 and *arf7arf19* mutant seedlings, and (2) phenotypic evaluation of lateral root development in transgenic line with cell-type specific repression of candidate genes.

A novel cell-type specific dCas9-driven repressor system can reveal drivers of lateral root fate

To explore the functional role of candidate genes in lateral root development, we devised a method to repress candidate genes only in the XPP cell lineage. We leveraged the enhancer trap line J0121 which is specifically expressed in XPP cells via a UAS-GAL4 driver system (Laplaze et al., 2005). We first introgressed J0121 into the Col-0 background (referred to hereafter as $J0121^{Col}$), and then introduced a UAS-dCas9-TPLN300 repressor (dCas9R) construct with three gene-specific sgRNAs directed to the promoter regions of candidate genes ($J0121^{Col}$>>dCas9R, **Supplemental Figure 3.5, Supplemental Data 3.2**). This cell-type specific repression system has several advantages over traditional knockdown and knockout studies. For instance, multiple guides can be used to simultaneously repress several members of the same gene family that may have redundant functions. Additionally, many of the candidate genes we identified as enriched in LRP cells also play roles in embryonic and primary root development, greatly complicating assessment of any role in lateral roots. To test the efficacy of our assay, we tested the effects of repressing expression of both *ARF7* and *ARF19*. We found that these perturbation lines have significantly reduced lateral root density compared to the empty vector control as expected although they did not fully recapitulate the full suppression of lateral roots seen in *arf7arf19* null mutants (**Supplemental Figure 3.6, Figure 3.2C**). This observation is consistent with the small reduction of lateral root number seen in a GATA23-driven CRISPR/Cas9 deletion of *ARF7* and *ARF19* (Decaestecker et al., 2019). There are at

least two likely explanations for the milder phenotype of J0121^{Col}>>dCas9R compared with the *arf7arf19* null mutant. First, repression in J0121^{Col} expression is limited to XPPs and lateral root stages I through III, so there is likely residual ARF protein that persists from expression in pre-XPP fate cells (Dubrovsky et al., 2006). Second, our synthetic repressor may not block all transcriptional activity, leading to hypermorphic rather than amorphic phenotypes. Even given these limitations, the system proved sufficiently sensitive to enable detection of cell type-specific impacts on lateral root development.

Chromatin remodeling factors influence lateral root development

Three histone deacetylases (HDACs) that are all in the same plant-specific gene family, *HISTONE DEACETYLASE 3, HISTONE DEACETYLASE 2, and HISTONE DEACETYLASE 13 (HDA3, HD2B, and HDT4*) (Li et al., 2017; Luo et al., 2017), were all enriched in LRP cells in our DEG analysis **(Figure 3.2A)**, as was the E3 ubiquitin ligase *ORTHRUS 1 (ORTH1*) which decreases DNA methylation (Kim et al., 2014). Transcriptional reporters of these genes express strongly in early stage primordia **(Figure 3.2B)**. Expression of *HDT4* and *HDA3* were specific to LRPs in the differentiated zone of the primary root, though both were also strongly expressed in the meristematic zone of the primary root **(Supplemental Figure 3.7)**. Their expression in the meristem was strongly decreased in *arf7arf19* mutants, suggesting they may be regulated by auxin **(Supplemental Figure 3.7)**. *HD2B* was also strongly expressed in LRP cells, as well as in the primary root meristem and other pericycle cells (**Supplemental Figure 3.7**). *ORTH1* was broadly expressed in the vasculature of the differentiated zone of the primary root, not only in LRP cells **(Supplemental Figure 3.7)**, which is reflected by its enrichment in mature pericycle cells in our single-cell library **(Figure 3.2A)**. Its expression was not impacted in *arf7arf19* mutant lines **(Supplemental Figure 3.7)**.

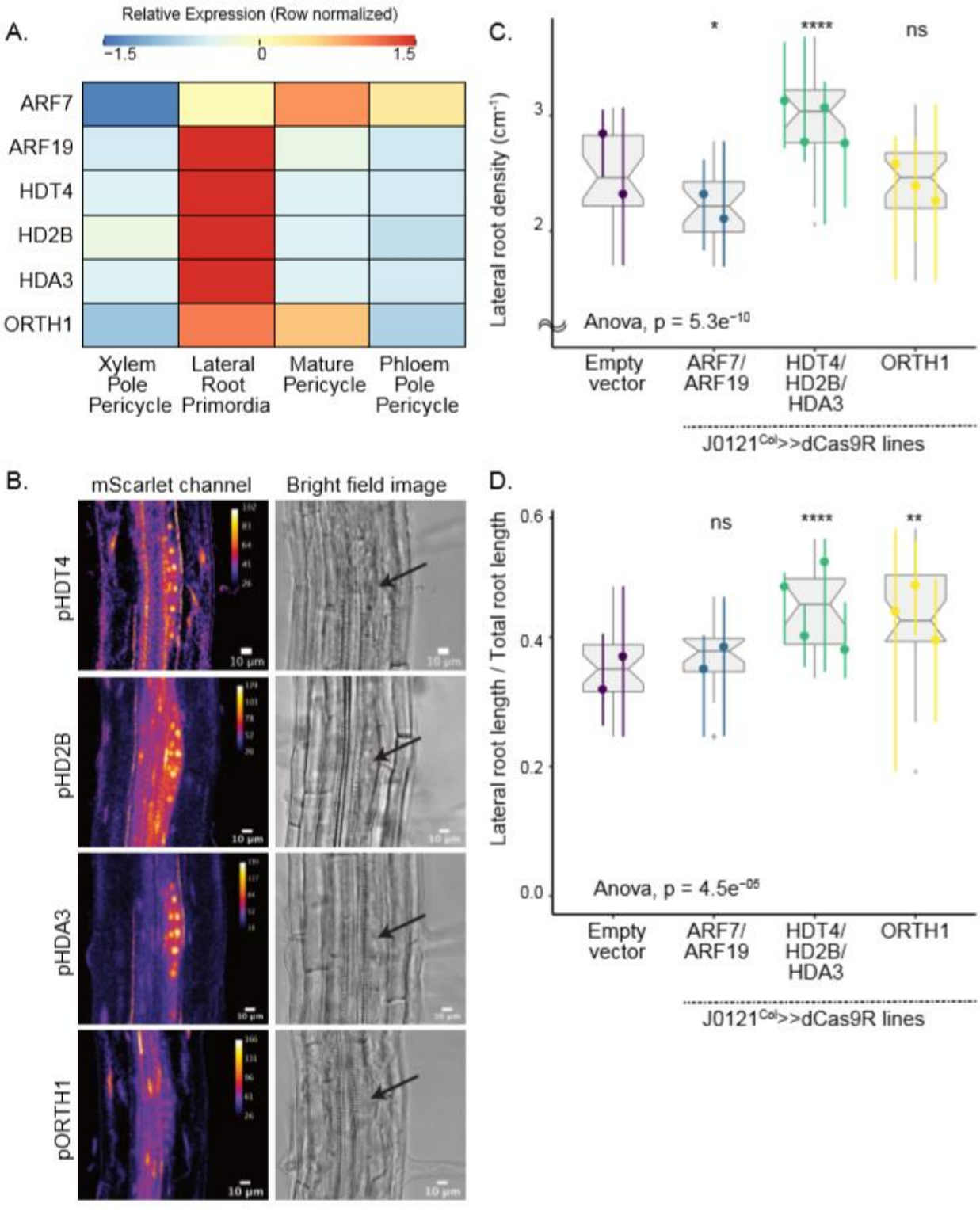


Figure 3.2. Validation experiments on chromatin modifier candidate genes.

(A) Heatmap (row-scaled) visualizing expression of candidate genes in the pericycle cell clusters from the single-cell library. Scale bar represents the z-score of the normalized expression values. **(B)** Confocal microscopy images of candidate genes' transcriptional reporters in early stage lateral root primordia (left) and bright field image of corresponding primordia (right). **(C)** Lateral root density of J0121^{Col}>>dCas9R transgenic lines of candidate genes. **(D)** Proportion of total root length contributed by lateral roots of J0121^{Col}>>dCas9R transgenic lines of candidate genes. For **C** and **D**, significance was determined by pairwise comparison with empty vector control and ANOVA. For each candidate gene, multiple independent transgenic lines have been analyzed; each colored line in the graph represents an individual line.

Using J0121^{Col}>>dCas9R, we targeted all three *HDAC* genes for repression using distinct guide RNAs for each gene. Simultaneous repression of all three genes caused a strong phenotype, where both the density of lateral roots **(Figure 3.2C)** and the proportion of total lateral root length contributed by lateral roots were significantly increased **(Figure 3.2D)**. The phenotype suggested that the HDACs may repress lateral root initiation and later stages of development. *ORTH1* repression in XPP cells did not significantly impact lateral root density **(Figure 3.2C),** but the proportion of total root length contributed by lateral roots significantly increased **(Figure 3.2D)**. Thus, *ORTH1* may repress lateral root growth only post-initiation. All of these chromatin regulators were expressed strongly in LRP cells, and repressing their function stimulates lateral root growth, suggesting they may act to coordinate cells and promote orderly development.

Cell cycle regulators are active during lateral root development

We characterized the role of five DEGs enriched in our LRP population that play a role in cell cycle regulation: *RECEPTOR FOR ACTIVATED C KINASE 1A, B* and *1C (RACK1A, RACK1B* and *RACK1C*), *NAP1-RELATED PROTEIN 1* and *2 (NRP1 and NRP2)*, and *CYCLIN-*

DEPENDENT PROTEIN KINASE INHIBITORS 6 and *11 (SMR6* and *SMR11)* **(Figure 3.3A)**. *RACK1B* and *RACK1C* interact with protein kinase C (Guo and Chen, 2008). *SMR6* and *SMR11* are cyclin-dependent kinase inhibitors (Yi et al., 2014), and *NRP1* is a histone chaperone required for the G2 to M transition (Zhu et al., 2006). We generated transcriptional fluorescent reporters of *RACK1B* and *RACK1C* and found their expression was indeed specific to early stage lateral root primordia as seen in the single-cell data **(Figure 3.3B)**. Expression of both *RACK1B* and *RACK1C* was lost in the differentiated zone of the primary root in *arf7arf19* mutants, which do not form lateral roots, indicating the specificity of their expression to LRP in this zone **(Supplementary Figure 3.8)**. Expression of these genes in the primary root meristem also was highly decreased in *arf7arf19* mutants **(Supplementary Figure 3.8)**, suggesting ARF7 and ARF19 may be the primary ARFs regulating their expression throughout the root.

We next used J0121^{Col}>>dCas9R to test for functional relevance of these cell cycle regulators. *NRP1* is highly related to *NRP2*, so we generated a repression line with guides targeting both genes. These lines did not show differences in lateral root density **(Figure 3.3C),** but did show a significantly increased proportion of summed total root length that was contributed by lateral roots reflecting longer lateral roots than in the control lines **(Figure 3.3D)**. *RACK1B* and *RACK1C* are highly related and show redundancy with *RACK1A*, so we generated two separate sets of repression lines, one with guides targeting *RACK1B* and *RACK1C* and the other with guides targeting all three *RACK1* genes. Repression of *RACK1B* and *RACK1C* caused a significant decrease in lateral root density **(Figure 3.3C)**. Interestingly, while repression of all three *RACK1* genes did not significantly impact lateral root density, this line did show a significantly increased proportion of total root length contributed by lateral roots **(Figure 3.3D)**, similar to the *NRP1/NRP2* repressed line. Singular repression of *SMR6* and concurrent repression of *SMR6* and *SMR11* did not show differences in lateral root density from

the control **(Figure 3.3C),** but again showed significantly increased proportion of total root length contributed by lateral roots **(Figure 3.3D)**.

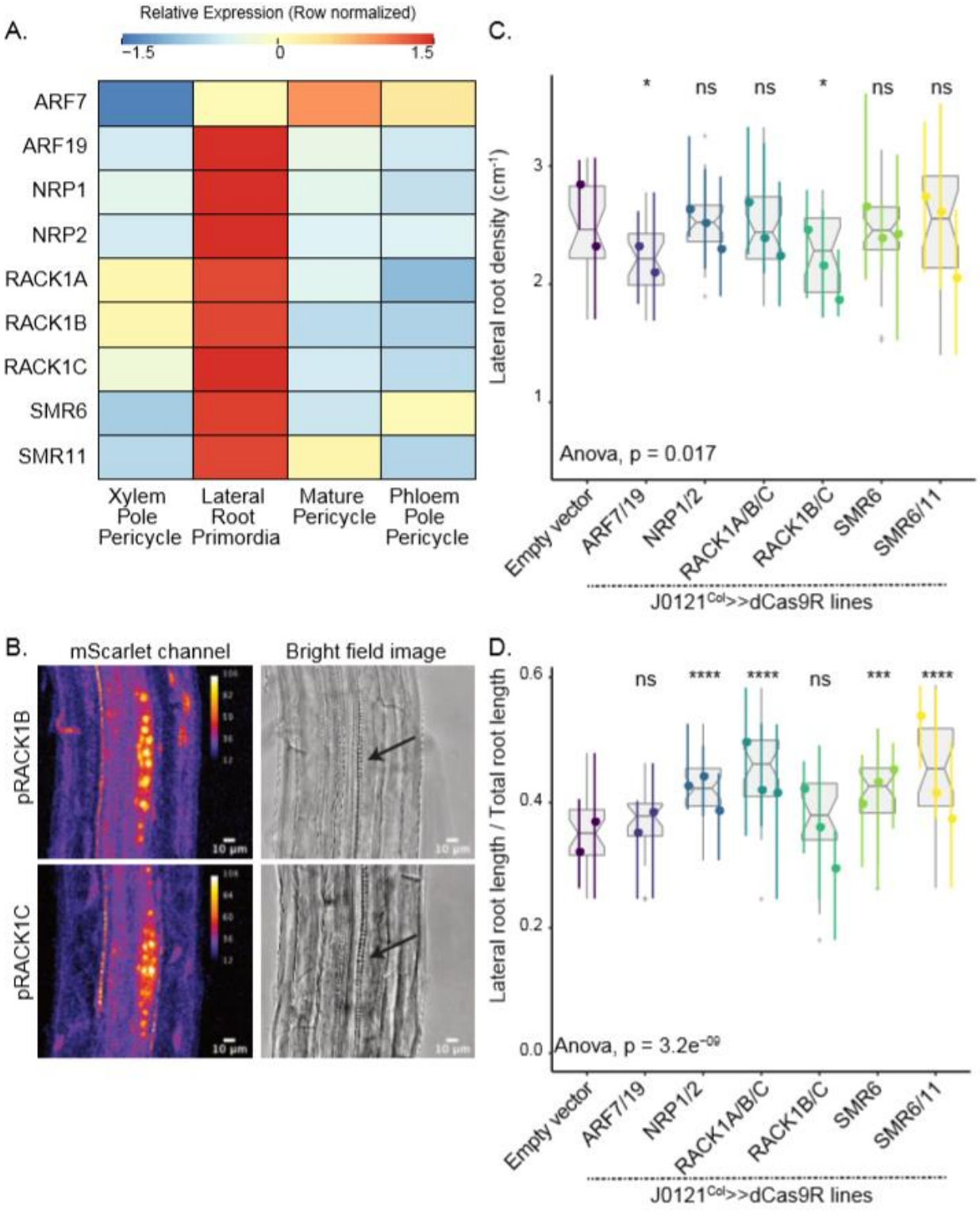


Figure 3.3. Validation experiments on cell cycle regulator candidate genes.

(A) Heatmap (row-scaled) visualizing expression of candidate genes in the pericycle cell clusters from the single-cell library. Scale bar represents the z-score of the normalized expression values. **(B)** Confocal microscopy images of candidate genes' transcriptional reporters in early stage lateral root primordia (left) and bright field image of corresponding primordia (right). **(C)** Lateral root density of J0121^{Col}>>dCas9R transgenic lines of candidate genes. **(D)**. Proportion of total root length contributed by lateral roots of J0121^{Col}>>dCas9R transgenic lines of candidate genes. For **C** and **D**, significance was determined by pairwise comparison with empty vector control and ANOVA. For each candidate gene, multiple independent transgenic lines have been analyzed; each colored line in the graph represents an individual line.

Genes that encode pluripotency and stemness are upregulated in LRP cells

We chose five DEGs known to play a role in developmental transitions for further validation studies, specifically *TARGET OF MONOPTEROS 6 (TMO6)*, a Dof-type transcription factor originally isolated as a target of ARF5 in embryos (Schlereth et al., 2010), *BREVIS RADIX-LIKE 1 (BRXL1)*, a BRX-like regulator of primary root development (Briggs et al., 2006), *LATERAL ROOT PRIMORDIUM 1 (LRP1)*, which a marker and plays a role in lateral root development (Smith and Fedoroff, 1995; Singh et al., 2020) , *OVATE FAMILY PROTEIN 8 (OFP8)*, a transcriptional repressor of KNOX family transcription factors (Wang et al., 2011), and *PLETHORA 3 (PLT3)*, a *PLETHORA* family gene that interprets auxin gradients in the primary root (Santuari et al., 2016). The presence of genes such as *LRP1*, *BRXL1*, and *PLT3*, known to regulate early stages of root development, confirmed that our dataset was isolating genes expected to be active early during lateral root initiation. BRXL proteins have recently been shown to play a role in promoting nuclear migration and asymmetric cell division in the development of stomata (Rowe et al., 2019; Muroyama et al., 2020), a process that is also essential during the very first stages of lateral root initiation. *OFP8* and *TMO6* were somewhat

unexpected discoveries, given that ovate family proteins have primarily been characterized in fruit development (Wang et al., 2016) and *TMO* genes have primarily been characterized in embryonic development (Schlereth et al., 2010).

Expression of all of these genes showed enrichment in LRP cells, though *TMO6* expression was most enriched in PPP cells (**Figure 3.4A**), consistent with its known role in phloem cell division and differentiation (Miyashima et al., 2019). Reporters of *PLT3* (Galinha et al., 2007), *LRP1* (Smith and Fedoroff, 1995) and *TMO6* (Schlereth et al., 2010) have previously been published, so we only generated transcriptional reporters of *OFP8* and *BRXL1*. Both showed strong and specific expression in lateral root primordia (**Figure 3.4B**). Expression of the *TMO6* reporter had not been previously analyzed in XPP or LRPs. We found that indeed it was strongly expressed in developing primordia, as predicted from the single-cell analysis. Nearly all of this expression was lost in *arf7arf19* mutants (**Supplemental Figure 3.9**), suggesting that, at least in the root, *TMO6* is primarily a target of these ARFs rather than ARF5. *TMO6*, *BRXL1* and *OFP8* were not expressed in the primary root meristem, (**Supplemental Figure 3.9**) making them potentially useful for targeting engineering efforts specifically to lateral roots.

Repression of these genes in XPP cells using J0121^{Col}>>dCas9R impacted root architecture but in different ways. Repression of *PLT3* caused a significant increase in both lateral root density (**Figure 3.4C**) and the proportion of total root length contributed by lateral roots (**Figure 3.4D**), suggesting *PLT3* may repress both lateral root initiation and emergence post-initiation. Repression of *TMO6* did not impact lateral root density (**Figure 3.4C**), but did cause a significant increase in the proportion of total root length contributed by lateral roots (**Figure 3.4D**), suggesting it may act on lateral root development post-initiation. Repression of *LRP1* in XPP cells significantly increased both lateral root density (**Figure 3.4C**) and the proportion of total root length that are lateral roots (**Figure 3.4D**), a phenotype that matches previously reported overexpression lines of *LRP1*, which showed decreased lateral root density

(Singh et al., 2020). Concurrent repression of *BRXL1* and its close homolog *BRXL4* did not significantly impact either lateral root density (**Figure 3.4C**) or the proportion of total root length that are lateral roots (**Figure 3.4D**), despite its strong expression in LRP cells. We did observe irregular spacing of lateral roots and shorter primary roots in *BRXL/BRXL4* repression lines, suggesting they may play a role in lateral root emergence (**Supplemental Figure 9**). Repression of *OFP8* exhibited unique behavior in our perturbation lines, as these lines showed significantly decreased lateral root density (**Figure 3.4C**). *OFP8* has not previously been characterized to play any role in root development, and this strong effect on lateral root initiation and its strongly specific expression in LRP is notable. *OFP8* repression did not impact the proportion of total root length contributed by lateral roots (**Figure 3.4D**) but these lines had shorter primary roots (**Supplemental Figure 3.6**).

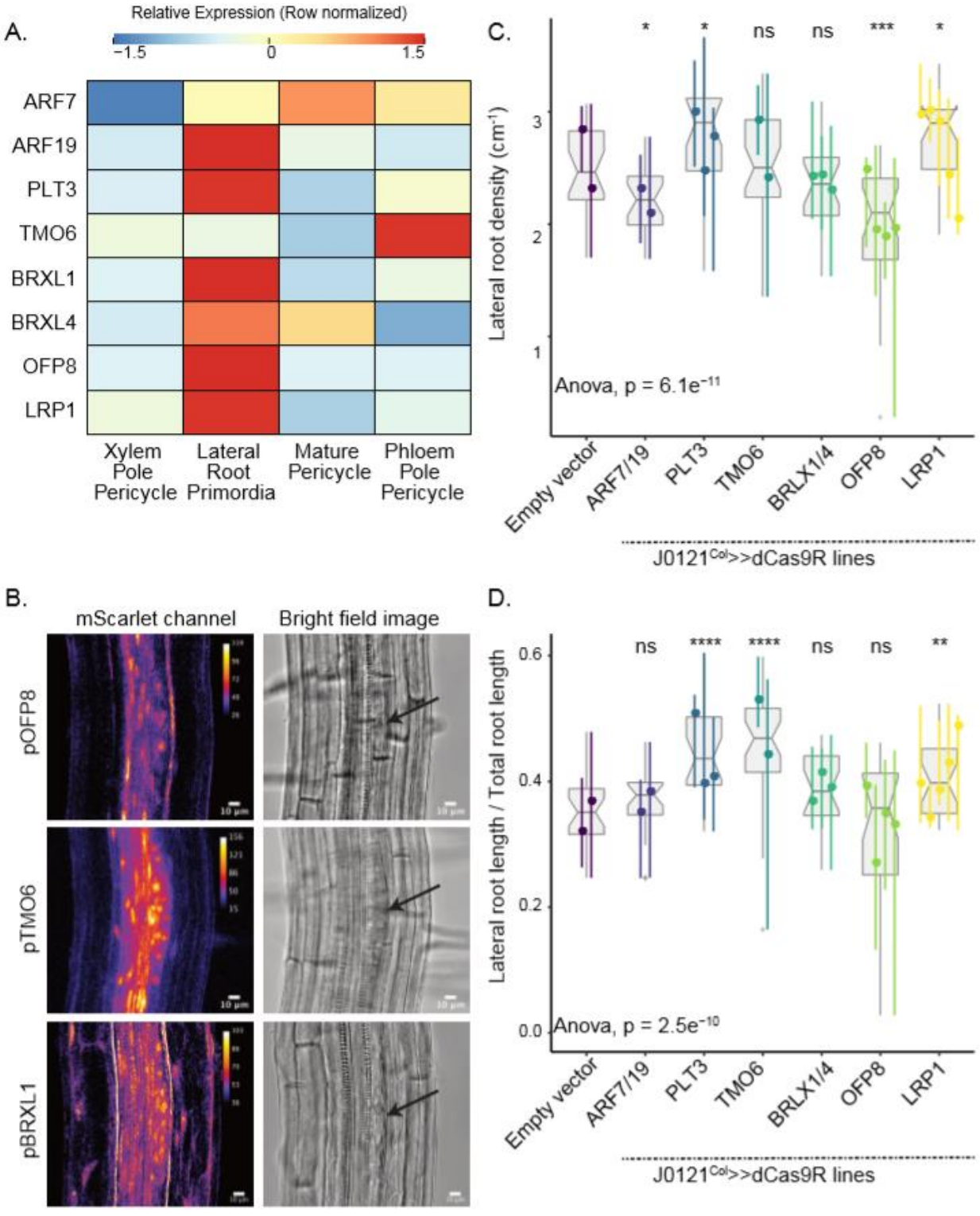


Figure 3.4. Validation experiments on stemness candidate genes.

(A) Heatmap (row-scaled) visualizing expression of candidate genes in the pericycle cell clusters from the single-cell library. Scale bar represents the z-score of the normalized expression values. **(B)** Confocal microscopy images of candidate genes' transcriptional reporters in early stage lateral root primordia (left) and bright field image of corresponding primordia (right). **(C)**. Lateral root density of J0121^{Col}>>dCas9R transgenic lines of candidate genes. **(D)**. Proportion of total root length contributed by lateral roots of J0121^{Col}>>dCas9R transgenic lines of candidate genes. For **C** and **D**, significance was determined by pairwise comparison with empty vector control and ANOVA. For each candidate gene, multiple independent transgenic lines have been analyzed; each colored line in the graph represents an individual line.

Non-LRP cells populations undergo transcriptional changes and fate transitions in response to lateral root initiation

Formation of a new lateral root is a self-organizing process during which a very limited number of competent XPP cells undergo repeated cell divisions to initiate lateral root organogenesis (Torres-Martínez et al., 2020). Continued development of the new root requires biophysical restructuring of the surrounding cell files. Signatures of lateral root development are seen outside the pericycle at pre-emergence stages of development, including during early initiation (Vermeer et al., 2014). Feedback on auxin signaling and changes in auxin transport patterns in the endodermis (Marhavý et al., 2013) and the vasculature (De Smet et al., 2007; Porco et al., 2016) are also essential for the first steps of lateral root initiation. Because our single-cell RNA sequencing dataset allowed us to examine the transcriptional state of these different cell layers independently, we examined which of the non-pericycle cell files contribute to transcriptional changes in response to this fate switch. For this we leveraged DEGs (945 genes) identified from a previous bulk RNA transcriptome study of bend-induced lateral root initiation (Voß et al., 2015) corresponding to the twenty hour post-bend timepoint in our study, and mapped the expression of these genes to our cell type resolved dataset **(Figure 3.5A)**. As

expected, most of these genes showed high expression in the LRP population and very low expression in the XPP population. In addition, we found strong enrichment of certain groups of genes in non-LRP populations, especially in those categorized as mature pericycle, endodermis cells and root cap cells.

We focused first on genes showing strong expression in the mature pericycle, as these cells could be directly in contact with the initiating primordium. To assess if the genetic circuits regulating the XPP to mature pericycle transition were linked to lateral root development at the bend, we examined the genes expressed in our pseudotime analysis of the transition of XPP cells to mature pericycle cells as compared to the genes expressed in the trajectory the transition of XPP cells to LRP cells (**Supplemental Figure 3.2A-D**). Genetic signatures in this mature pericycle trajectory suggested that these cells, acting as the outgroup for this fate transition, nonetheless were responding to lateral root development, though in a very distinct way from cells within the XPP to LRP trajectory. One gene that was expressed in XPP cells and LRP cells close to the branch point is *TRANS-ACTING siRNA3* (*TAS3)*. *TAS3* promotes lateral root development by inhibiting the class B auxin response factors *ARF2*, *ARF3* and *ARF4* (Marin et al., 2010) Consistent with this pattern, *ARF2* transcripts were found only in the mature pericycle trajectory. While *TAS3* expression is fairly consistent across XPP, LRP, and mature pericycle cells, *ARF2*, *ARF3*, and *ARF4* show greater expression in the mature pericycle cell population as compared to the XPP population, and also show high expression in PPP cells (**Figure 3.5B**). As LRPs never initiate from PPP, strong expression of these repressor ARFs in mature pericycle cells and PPP cells may indicate that these genes are repressing auxin response, and thus lateral root development, in these cells.

A growing body of research has shown that LRPs send signals to actively repress initiation of surrounding pericycle cells, allowing for the proper spacing of lateral roots along the axis of the primary root (Murphy et al., 2016; Toyokura et al., 2019; Trinh et al., 2019). The

strong expression of genes that repress auxin response in mature pericycle cells suggests one mechanism for this inhibition. Further evidence in support of this model is the expression of another repressor ARF, *ARF9*, and the auxin repressor *IAA2* only in the mature pericycle trajectory. Uniquely expressed in this trajectory are also two cytokinin response factors, *ARABIDOPSIS THALIANA RESPONSE REGULATORS 5* and *11 (ARR5* and *ARR11)*, which are known to inhibit auxin response, repress lateral root development and coordinate uniform spacing of lateral roots (Mason et al., 2005; To et al., 2004). A target of cytokinin signaling, *CFB1*, which is specifically expressed in the flanking zone of developing LRPs (Brenner et al., 2017) is also unique to this trajectory. Together, these data suggest that cells that were initially categorized as mature pericycle may be more accurately described as LRP-flanking pericycle cells, and this flanking fate is specifically induced by the initiation of a new root.

Lateral root development shows strong signatures in endodermal transcriptomes

We next examined the impact of lateral root initiation on the endodermal cell file which is immediately exterior to the pericycle. We first re-clustered the 813 cell partition labeled as cortex and endodermis cells. This analysis revealed three distinct cell populations. The first population was a set of 127 cells which expressed cortex marker genes, the second population was a set of 410 cells which expressed endodermis marker genes, and the third population of 276 cells branching out from endodermis population (**Figure 3.5C, Supplemental Figure 3.10A, Supplemental Figure 3.10B**). This third population was transcriptionally most similar to the 410 cells expressing endodermis marker genes, and twenty-six percent of the cells in this third population were from the twenty hour timepoint. As a comparison, about eight percent of the 127 cell population were from the twenty hour timepoint, while about nine percent of the 410 cell population were from the twenty hour timepoint. This third population also had the highest

expression of the NAC transcription factor, *ORESARA1* (*ORE1*) and autophagy marker gene *ATG8a* (**Figure 3.5D, Figure 3.5E**). *ORE1* is a positive regulator of programmed cell death and autophagy marker gene ATG8a are specifically expressed in cells overlying LRPs (Escamez et al., 2020). *ORE1* expressing cells eventually die in order to make space for the lateral root to emerge (Escamez et al., 2020). As such, we categorized these three populations as cortex (127 cells), endodermis (410 cells), and lateral root-responding endodermis (LRE) (276 cells) (**Supplemental Table 3.1, Supplemental Figure 3.10A**).

To aid in identifying differentially expressed genes between endodermis and LRE, both cell types were compared to cortex cells as an outgroup. As in the XPP-LRP comparison, only genes that were called significantly different in at least two methods were called as DEGs (**Supplemental Figure 3.11**). This analysis yielded over 2000 DEGs between endodermis and cortex, and over 3000 DEGs between LRE and cortex. As expected, there was large overlap between these two sets of DEGs (**Supplemental Data 3.3**). A smaller set of genes were identified that were specific to the endodermis cells (204 genes), and specific to the LRE cells (303 genes), while 240 were commonly upregulated genes in endodermis and LRE (**Figure 3.5F**). The DEGs specific to the endodermis cells were enriched for GO terms associated with cell-cell junction assembly and water transport, while the DEGs specific to the LRE cells were enriched for GO terms associated with hormone response, auxin homeostasis, cell communication, lateral root development, and multiple stress responses (**Supplemental Data 3.3**).

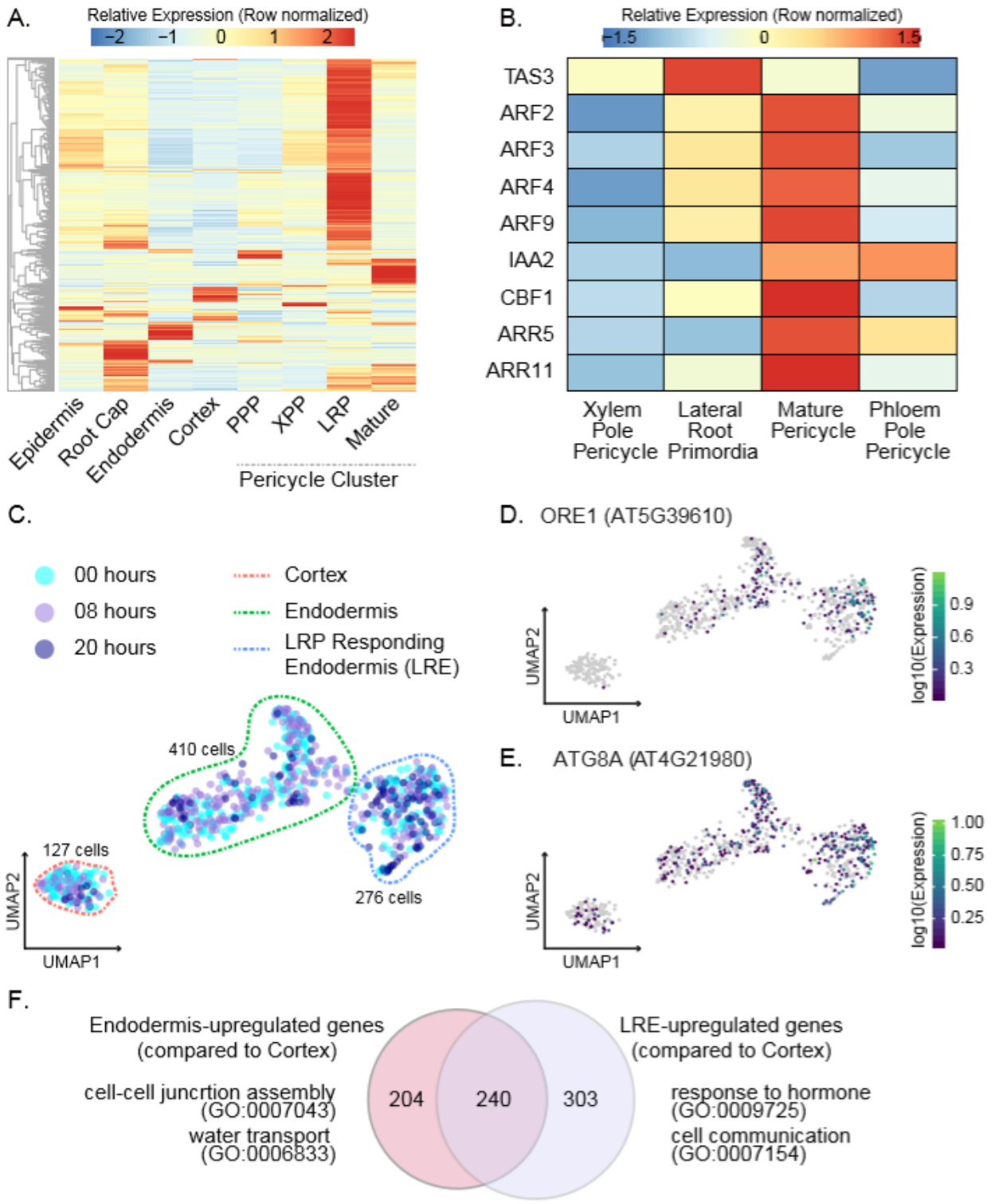


Figure 3.5. Analysis of non-LRP cells in the single-cell library.

(A) Heatmap (row-scaled) of expression patterns of previously-identified LR-specifying genes in different cell clusters. **(B)** Expression heatmap (row-scaled) of specific auxin-inducing and auxin-repressing genes in different stele cell clusters. **(C)** UMAP of Cortex, Endodermis, and Lateral Root Responding Endodermis (LRE) cells. **(D)** Expression UMAP of ORE1, a NAC transcription factor that promotes autophagy of endodermal cells overlying LRPs. **(E)** Expression UMAP of AT8GA. **(F)** Gene ontology of genes enriched in endodermis and LRE cell populations. The Venn diagram represents DEGs upregulated in endodermis in comparison to the cortex and upregulated in LRE in comparison to the cortex. A selection of GO terms for each of these sets of DEGs are shown.

Pseudotime analysis was performed to identify genes driving the transition from endodermis to the lateral root endodermis. This analysis showed a branch point between the two cell types. One branch that we termed the “main branch” was composed mostly of cells from the eight hour time point, and was transcriptionally similar to the rest of the endodermis cluster. The other branch that we called the “LRE branch” led towards the LRE cluster (**Supplemental Figure 3.12A**). Again, two separate trajectories were analyzed, one containing only endodermis cells and another containing endodermis cells below the branch point and LRE cells. This yielded a set of 2082 genes whose expression changed over the course of the cell developmental trajectory. After removing genes upregulated in cortex cells, a small set of genes were identified as specific to the main branch (154 genes) and to the LRE branch (648 genes) (**Supplemental Figure 3.12B, Supplemental Data 3.3**). GO analysis for the main branch specific genes showed enrichment for terms associated with responses to various stimuli. An example of a main branch specific gene is *DEEPER ROOT 1 (DRO1)* (**Supplemental Figure 3.12C**). Negatively regulated by auxin, *DRO1* is involved in gravity sensing in the root tip and determining lateral root branch angle (Uga et al., 2013; Guseman et al., 2017; Waite et al., 2020). *DRO1* loss of function mutants have increased horizontal lateral roots and have trouble establishing auxin gradients in response to gravistimulus. In our dataset, the majority of cells expressing *DRO1* are endodermal cells, suggesting these cells play a specific role in response

to gravity. The LRE branch genes were enriched for GO terms associated with lateral root development, auxin homeostasis, auxin transport, and biosynthetic processes. Examples of LRE branch specific genes are *WRKY75* and *PINS-LIKE 5 (PILS5)* (**Supplemental Figure 3.12D**). *WRKY75* is induced particularly in phosphate starvation (Devaiah et al., 2007), and *PILS5* is an auxin efflux carrier regulating intracellular auxin homeostasis, both independently having a role in controlling root architecture. These along with gene ontologies pertaining to several biotic and abiotic stress responses (**Supplemental Data 3.3**) indicates that LRE cell population seems to be primed to dynamically assess environment and thereby regulate lateral root emergence.

3.4 Discussion

Among the greatest mysteries in development is the process by which a stem cell begins proliferating and partitioning its progeny into increasingly determined cell fates. In plants, the initiation of lateral root development is among the best understood of these processes, yet many fundamental questions remain. Auxin is clearly a critical signal, but what other pathways interact with auxin response or regulate developmental steps downstream of auxin perception are still largely unknown. One key piece of missing information is a full accounting of transcriptional changes in the lateral root primordia during the critical window of initiation. In this study, we performed single-cell RNA sequencing at two timepoints on regions of roots where lateral root initiation was taking place. We identified cells of all cell types of the root outside the meristem in our population, including cells expressing lateral root markers. Through differential gene expression analysis, we identified genes upregulated in LRP cells as compared to XPP cells, many of which were indicative of cells undergoing cellular differentiation and organogenesis. We also identified a subset of pericycle and endodermal cells outside of the primordium itself that appear to be responding to the initiation of a new root.

We chose several genes for further study. Transcriptional reporters confirmed that most were expressed in early stage lateral roots, and that this expression was in many cases dependent on *ARF7* and *ARF19*. Several reporters showed additional expression in other parts of the root, including the primary root meristem. *OFP8* and *BRXL1* were expressed specifically in LRPs, making them potentially useful tools for studying and engineering root architecture. Repression of many of our candidate genes in XPP cells caused defects in lateral root development, a previously unknown role for some of them. For example, *HD2C* and *HD2B* have been linked to down-regulation of ribosomal biogenesis genes (Chen et al., 2018), a process we found to be strongly induced in lateral root primordia. This connection is consistent with our finding that repression of these genes led to higher density of lateral roots. It also highlights the fact that genes that are upregulated during primordia initiation may be important either for promoting developmental progression or for providing checks to keep cellular events coordinated or appropriately controlled. *OFP8* has been reported to act as a direct transcriptional repressor (Wang et al., 2011), and as a regulator of cell division patterns and organ shape via modification of the cytoskeleton (Snouffer et al., 2020).

Within the primordium itself, careful regulation of cell division and growth are essential for the formation of a primordium with normal morphology. The majority of our J0121^{Col}>>dCas9R transgenic plants exhibited increased lateral root density or increased lateral root length as a proportion of total root length. These results suggest that these target genes may normally act as repressors of lateral root initiation or emergence. When they are repressed, development is accelerated. One interpretation is that cell division and lateral root development are the default state of competent pericycle cells. This would be analogous to the situation in the root epidermis where becoming a root hair is the default state that must be actively repressed in non-hair cells (Berger et al., 1998). This hypothesis is supported by experiments where laser

ablation of surrounding tissue causes unrestricted cell division in the pericycle cell file (Marhavý et al., 2016), and where exogenous auxin treatments trigger lateral root initiation in every pericycle cell (Himanen et al., 2002).

Using cell-type specific repression of our candidates in XPP cells allowed us to avoid several problems with global mutant analyses. For example, *PLT3* and *BRXL1* play critical roles in the primary root meristem, making interpretation of any lateral root phenotypes difficult. A narrower scope of repression may also reduce the likelihood of compensation from paralogs. The repression lines decrease expression only in XPP cells and in LRP cells up to stage IV. This tissue-specific expression may explain the phenotypic difference between our *PLT3*-repressed line and *plt3 plt5 plt7* mutants, which show decreased lateral root emergence (Du and Scheres, 2017). It is possible that *PLT3* is playing opposing roles in different cell files of the root or at different points in lateral root development, depending on the interacting partners that are present in these cell files at these developmental stages. Alternatively, the phenotypes we observed could reflect feedback effects on other genes in the same gene regulatory network. The analysis performed here is likely to have missed phenotypes, especially those that rely on complex interactions with a soil environment or on a different set of metrics (Fitter, 1987; Lynch, 1995). For example, the irregular spacing and rate of growth of lateral roots in our BRXL1/BRXL4-repressed lines did not significantly impact the metrics we tested (**Supplemental Figure 3.9**). Introgression of J0121 into different accessions would allow access to a broader array of cryptotypes (Chitwood and Topp, 2015), and a more holistic view of impacts on root architecture.

We found that our single-cell experiments captured the majority of previously reported LRP-enriched transcripts in our LRP-assigned cells, and these transcripts were not enriched in our XPP-assigned cells. We additionally found many more LRP-enriched genes in our library than in bulk transcriptomes, underlining the utility of this method in examining rare

developmental events. Many of our XPP-upregulated genes oscillate antiphase to auxin response in the basal meristem. These antiphase-oscillatory genes are upregulated in cells that are *not* competent to form lateral roots. They may be actively preventing lateral root initiation in XPP cells, in alignment with the hypothesis that LRP-competency may be the default state of XPP cells. Notably, there are also 698 cells within our stele cluster that were too ambiguous in their gene expression to assign a cell label. It is possible some of these cells are precursors to stele cells, such as XPP precursors that are not yet lateral root competent. How distinct lateral root competent and non-competent XPP cells are is unknown—the only known distinguishing feature between the two is a characteristic auxin response maximum or lack thereof. Additional differential gene expression between different timepoints within the XPP cluster and this non-assigned stele cell cluster may yield novel insights.

Though lateral root development is specific to xylem pole pericycle cells, the process is not cell-autonomous. Our analysis identified a population of endodermal cells distinct from the main endodermis branch. These cells were enriched in the expression of genes falling in ontology categories for hormone and auxin response, cell-cell communication, and lateral root development, making a strong case that they are responding to developing primordia in underlying pericycle cells. They were also enriched in *ORE1* expression, a gene that has recently been shown to play a role in lateral root initiation and emergence through programmed cell death of tissue overlying LRPs (Escamez et al., 2020). Consequently, this population of endodermal cells appears to be responding to lateral root initiation in neighboring pericycle cells, and forging a path for the incipient primordium.

We were also able to identify a subset of pericycle cells that were likely directly adjacent to the primordium and responding with a distinct transcriptional program that included a combination of auxin-repressing and cytokinin-induced genes. A dynamic analysis of the gene regulatory network governing lateral root development established that early cell fate

determining genes initiate multiple genetic feedback loops that divide the developing primordium into two zones, a central proliferative core and flanking cells that have inhibited expression of meristematic genes to repress cell division (Lavenus et al., 2015). Notably, PPP cells in our dataset also show strong expression of several auxin-inhibitory genes. PPP cells never initiate lateral root development, even though these cells receive the same auxin maximum signal as XPP cells during specification. It is possible these repressors of auxin signaling act to prevent spurious root development in multiple cell files.

As human activity changes the climate and environments in which plants grow, understanding root development will help us engineer crops that are more robust to nutrient scarcities and environmental extremes. The major pathway by which eudicot plants regulate their root architecture is through modification of the position, spacing, density, and length of lateral roots. The stages of lateral root development are regulated by distinct genetic circuitry. Every stage represents an opportunity for natural and engineered modification of this developmental process. Molecular characterization of early stages of lateral root at single-cell resolution gives us a more comprehensive understanding of this fate decision and the molecular pathways that tune it.

3.5 Methods

Construction of plasmids. Each reporter plasmid is composed of the selected promoter, the red fluorescent protein mScarlet with a nuclear localization tag (Bindels et al., 2017), and the rbcS terminator (Siligato et al., 2016). The three parts were assembled using golden gate assembly in the modified pGII-Hygr vector containing compatible Golden Gate sites (Weber et al., 2011). For each of the ten constructed reporters, the promoter sequence of the reporter corresponds to the DNA sequence in 5’ of the start codon of the corresponding gene based on TAIR10 genome from http://plants.ensembl.org/. While aiming for a 2000 bp length, the lengths of promoters are usually

smaller to avoid coding sequences of other genes. The exact sequence and length of the selected promoter for each of the 10 genes can be found in supplementary material (**Supplementary Data 2**). The promoter sequence for the TMO6 reporters corresponds to the sequence used in a previous work from (Smet et al., 2019). The promoter sequences were amplified from purified *Arabidopsis thaliana* Col-0 genomic DNA using Q5 polymerase and with primer adding the specific golden gate spacer. After gel purification, each promoter part was cloned and sequence verified in a pBLUNT entry vector. Three part Golden Gate assembly was performed using the pBLUNT promoter plasmid, mScarlet, rbcS terminator to clone reporter plasmid.

For cell type specific knockdown mediated by J0121^{Col}>>dCas9R, Gibson cloning was used to replace the egg specific promoter and Cas9 from pHEE401E (Wang et al., 2015) with UAS promoter and dCas9-TPL fusion (Khakhar et al., 2018). The resulting plasmid is used as starting point to clone two or three guide RNA against the promoters of selected gene/genes (identified using CHOP CHOP (Montague et al., 2014) ranging from -200 to +100 region from the annotated TSS) using PCR and golden gate strategy described in (Wang et al., 2015)**.**

Plant growth conditions and sample preparations. For all plant experiments, *Arabidopsis* seeds were sown on 0.5× LS 0.8% agar plates, stratified at 4°C for 2 days, and grown in continuous light conditions at 22°C for respective experimental design.

Microdissection of root bend and protoplast isolation*.* For lateral root induction assays, ~150 seedling for each timepoint and treatment were rotated 90° 4 days post-germination (dpg) or in the case of the control treatment, the primary root tip was marked at this time and the plates were not turned. On the day of single-cell library preparation first the protoplasting enzyme mix was prepared adapted from (Yoo et al., 2007). Briefly, 20 mM MES (pH 5.7) containing 1.25% (wt/vol) cellulase R10 (C224 PhytoTechnology Laboratories), 0.3% (wt/vol) macerozyme R10 (M481 PhytoTechnology Laboratories), 0.4 M mannitol and 20 mM KCl was

prepared and incubated in 55°C warm water bath for 10 minutes. Upon cooling to room temperature (~25 °C),10 mM CaCl2, 1–5 and 0.1% BSA was added. Root bends (or marked region in no bend control) were microdissected using a scalpel at eight hours (control and eight hour treatment groups) and 20 hours (20 hour treatment group) post-bending, approximately 1 mm from the bend or mark in both directions. Using fine forceps dissected bend tissue was transferred into 30 mm dishes containing 1 mL of protoplasting enzyme mix and gently scored using a fresh scalpel to increase exposure of interior cell files to protoplasting enzymes. The plates were then flooded with 9 mL more protoplasting enzyme mix and incubated at room temperature for one hour with gentle shaking (75-80rpm). Protoplasting enzyme mix was filtered through 40 µm cell strainer, transferred and centrifuged in 15 ml conical tubes for five min at 500g. The supernatant was carefully removed and resuspended in 50 µl protoplasting mix without enzymes. Cell number was determined by hemocytometer and density was adjusted to ~1000cell/µL

Construction and selection of transgenic *Arabidopsis thaliana* lines. Floral dip (Clough and Bent, 1998) was used to introduce constructs into Col-0 and *arf7arf19* lines (Okushima et al., 2007). T1 seedlings were selected on 0.5X LS (Caisson Laboratories, Smithfield, UT) + 25µg/ml Hygromycin B + 0.8% bacto-agar. Plates were stratified for two days, exposed to the light for six hours, and then grown in the dark for three days (Harrison et al., 2006). Hygromycin resistant seedlings were identified by their long hypocotyl, enlarged green leaves, and long root. Transformants were transferred on soil, and T2 seeds were collected.

Lateral root bend assay and confocal microscopy. For each reporter, one Col-0 T1 line representative of other characterized T1 lines was selected to perform lateral root bend essay. For each reporter, 20 T2 seeds of the corresponding T1 line were placed on 0.5 LS+0.8% bacto-agar plate following a specific pattern to avoid seedling collision during the lateral root bend essay.

The plate was stratified during 120 hours, grown vertically for 96 hours at 22°C, rotated 90°C while keeping vertically and grown for an additional 20 hours.

Confocal microscopy of reporter lines at root bends. Seedlings were fixed at 4 dpg + 20 hours using 4% formaldehyde using vacuum infiltration followed by cleared using ClearSee solution (Kurihara et al., 2015). Fixed and cleared seedlings were mounted on microscopic slides using 50% glycerol and parafilm edges to avoid coverslips pressing on the root. Seedlings were imaged at the bend region using a SP5 confocal microscope. Images were processed using FIJI.

Comparison between Col-0 and arf7arf19 lines. To perform comparative imaging of Col-0 and *arf7arf19* reporter lines, seeds of selected T1 lines for both Col-0 and *arf7arf19* reporter lines were placed on the same 0.5 LS+0.8 phytoagar plate. The selected Col-0 and *arf7arf19* lines for each reporter is specified below the microscope images of supplemental figure 7 to 9 as being highlighted in bold. Plates were stratified for 2 days, and grown vertically at 22°C for 10 days. Then, seedlings were imaged using a Leica DMI 3000B microscope at the root tip region, at the region above the root tip corresponding to the initiation of root hair and at the lateral root initiation region. As the *arf7arf19* line does not develop lateral roots, the theoretical lateral root initiation region is identified by identifying a lateral root primordium in the Col-0 seedling and imaging at a similar region.

Lateral root phenotypes of repression lines. For cell type specific knockdown mediated by J0121^{Col}>>dCas9R we leveraged an established GAL4-UAS system (Laplaze et al., 2005) of enhancer trap line J0121. We backcrossed the J0121 line, made in the C24 background, eight times into the Col-0 background to produce a strain we refer to as J0121^{Col}. We confirmed that J0121^{Col} retained strong GFP expression in xylem pole pericycle and exhibited Col-0-like root growth dynamics. Transformants were selected as described above and T2 seeds for at least 10 lines were collected. T2 seeds were grown vertically for 10dpg in 100mm square plate on

0.8% bacto agar and were scanned on a flatbed scanner (Epson America, Long Beach, CA) for phenotyping. Since the T2 generation is a segregating population for the transformed plasmid, seedlings were genotyped for the presence of vector backbone to identify positive seeding. Roots for positive seedlings were traced using ImageJ and SmartRoot plugin (Lobet et al., 2011) and analyzed and plotted using R package archiDART (Delory et al., 2016) package plot were generated using ggplot2. Density was measured as the total number of lateral roots divided by the length of the primary root. Proportion of lateral root length was measured as the summed length of all lateral roots divided by the summed length of all lateral roots and the primary root length.

Single-cell RNA-sequencing Protocol. Single-cell RNA-Seq was performed using the 10X scRNA-Seq platform, the Chromium Single Cell Gene Expression Solution (10X Genomics). Two replicates were produced for each timepoint of the experiment for a total of six samples. We also generated two replicates from a transgenic plant line that slows the rate of degradation of IAA14 (Guseman et al., 2015), dissecting root bends in this line twenty hours after bending. This line shows delayed lateral root development, and we initially thought to compare its transcriptomes to our wild type treatment groups. Unfortunately, one of the replicates of this line failed at the 10X droplet-binding stage, so we did not obtain the same number cells from this treatment group as from our other groups. Consequently, we excluded these cells from further analysis.

Data Analysis: Estimating Gene Expression in Individual Cells. Single-cell RNA-sequencing reads were sequenced using an Illumina NextSeq 500 and then mapped to the TAIR10 *Arabidopsis* genome using the software Cellranger (v.3.0.1). Cellranger produces a matrix of UMI counts where each row is a gene and each column represents a cell. The ARAPORT gene annotation was used. For the analysis, reads from two 00 hour replicates, two 08 hour replicates, and two 20 hour replicates were aggregated using the aggr command in cellranger to

normalize to an equivalent number of mean reads per cell across samples. This resulted in a mean of 14,516 reads per cell, a median of 1,411 genes per cell, and a median of 2,873 UMIs per cell.

Data Analysis: Running Monocle 3 - Dimensionality Reduction, and Cell Clustering. The filtered output of the Cellranger pipeline (../outs/filtered_gene_bc_matrices_mex/) was parsed into R (v. 3.5.0). Particularly the matrix.mtx file was parsed using the readMM() function from the Matrix package (https://cran.r-project.org/web/packages/Matrix/Matrix.pdf), and the barcodes.tsv file and the genes.tsv file were parsed using the read.table() function. Genes that were expressed in less than 10 cells were removed from the analysis. In addition, the 346 genes induced due to protoplast generation process were also removed from the analysis (Birnbaum et al., 2003). The barcodes table was updated to label cells by Sample Number and Experiment. Finally the expression matrix, the barcode table, and the gene table were converted into a CellDataSet (CDS) using the new_cell_data_set() function in Monocle 3 (cole-trapnell-lab/monocle3, 2020) (v. 0.1.2; https://cole-trapnell-lab.github.io/monocle3/)

All Monocle 3 analysis was performed on a High Performance Computing cluster using 128 GB of RAM spread across eight cores. We visualized cell clusters and trajectories using the standard Monocle workflow. Monocle internally handles all normalization needed for dimensionality reduction, visualization, and differential expression. The CDS was normalized and pre-processed using the preprocess_cds() function with the following parameters:

```
num_dim=100,
method="PCA",
norm_method="log",
scaling=T,
residual_model_formula_str="~ Sample_Number"
```

Preprocessing involves reducing the dimensionality of the data (the number of genes) using principal component analysis (PCA). Here, we retain the first 100 PCs for further dimensionality reduction, in addition we reduce batch effect across samples. Then, the PCA matrix was used to initialize a nonlinear manifold learning algorithm implemented in Monocle 3 called Uniform Manifold Approximation and Projection **(**UMAP) (McInnes et al., 2018). This allows us to visualize the data into two or three dimensions. Specifically, we projected the data onto two dimensions using the reduce_dimension() function using the following parameters:

```
reduction_method="UMAP",
preprocess_method="PCA",
umap.metric="cosine",
umap.min_dist=0.1,
umap.n_neighbors=15L,
umap.nn_method="annoy",
umap.fast_sgd=F
```

Next, cells were clustered into partitions using PhenoGraph (Levine et al., 2015) and into clusters using the Leidan community detection algorithm (Traag et al., 2019) with the cluster_cells() function using the following parameters:

```
reduction="UMAP",
k=20,
louvain_iter=1,
partition_qval=0.05,
weight=T,
resolution=c(10^seq(-6,0)).
```

This resulted in 5 cell partitions, and 43 cell clusters. Next, for each cell partition, a cell trajectory was drawn atop the projection using Monocle's reversed graph embedding algorithm, which is derived from SimplePPT (Mao et al., 2017) using the learn_graph() function using the following parameters:

use_partition=T,

close_loop=F,

learn_graph_control=list(prune_graph=T)

To further analyze the partition we annotated as stele, those cells were reclustered together and were reanalyzed using Monocle 3 as previously described except during preprocessing, instead of performing PCA on all the genes, PCA was performed on just a set of stele cell marker genes reported earlier (Mao et al., 2017). Specifically the "use_genes" option was used in the preprocess_cds() function, and a list of gene names was provided. Lastly, cells were clustered using the cluster_cells() function, and the "resolution" parameter was set to 0.001. This produced 3 partitions, and 7 clusters.

To further analyze the clusters we annotated as Phloem Pole Pericycle, Mature Pericycle, Xylem Pole Pericycle, and Lateral Root Primordia, those cells were reclustered together and were reanalyzed using Monocle 3 as previously described except during preprocessing, instead of performing PCA on all the genes, PCA was performed on just a set of stele cell marker genes reported in (Brady et al., 2007). Again, Specifically the "use_genes" option was used in the preprocess_cds() function, and a list of gene names was provided. Next, the data was reduced onto two dimensions using the reduce dimension function but the "umap.min_dist" parameter was set to 0.01. Afterwards cell clusters were called as before using the cluster_cells() function except the "resolution" parameter was set to 0.0005. Next a cell trajectory was created using the learn_graph() function. Finally cell clusters were recalled using

the cluster_cells() function except the “resolution” parameter was set to 0.001. This produced 1 partition and 4 clusters.

To further analyze the partition we annotated as Cortex and Endodermis, those cells were reclustered together and were reanalyzed using Monocle 3 as previously described except during preprocessing, instead of performing PCA on all the genes, PCA was performed on just a set of cortex and endodermis cell marker genes reported in (Brady et al., 2007). This produced 2 partitions and 15 clusters.

Data Analysis: Estimating Doublets. Single Cell Remover of Doublets (Scrublet) was used to predict doublets in our scRNA-Seq data (Wolock et al., 2019). Using python 3.5, Scrublet was run using default settings as described by the example tutorial that is available as a Python notebook (https://github.com/AllonKleinLab/scrublet/blob/master/examples/scrublet_basics.ipynb). The only significant change was that expected double rate was set to 0.1; in the tutorial it is 0.06.

Data Analysis: Assigning Cell Types. A set of known marker genes derived from earlier studies using green fluorescent protein (GFP) marker lines of the *Arabidopsis* root were used to identify cell types (Brady et al., 2007; Cartwright et al., 2009). The average gene expression of each marker set was used to assign cell types to cells, with cells being assigned the cell type it had the highest average expression

Data Analysis: Calling Differentially Expressed Genes. *Xylem Pole Pericycle vs. Lateral Root Primordia.* Differentially expressed genes between the cluster of cells labeled Xylem Pole Pericycle and the cluster of cells labeled Lateral Root Primordia were called using three different approaches.

The first approach involved running a generalized linear model to predict the average log express of each gene as a function of the cell type label. This was done using a subsetted CDS containing only the Xylem Pole Pericycle cells and the Lateral Root Primordia cells, and the Monocle 3 function fit_models() with the following parameters:

```
model_formula_str = "~cell_type",
expression_family="negbinomial",
clean_model=T
```

where “cell_type” is a column in the dataframe returned by the colData() function that describes the cell type label associated with a cell/barcode. Using an FDR cutoff of 0.1, 1204 genes were called as differentially expressed between Xylem Pole Pericycle and Lateral Root Primordia. Of these, 424 were more highly expressed in Xylem Pole Pericycle, and 780 were more highly expressed in Lateral Root Primordia.

The second approach involved using the Mann-Whitney-Wilcoxon test to determine if the rank-sum of the normalized expression values for each gene differed between the Xylem Pole Pericycle population and the Lateral Root Primordia population. Mann-Whitney-Wicoxon test p-values were adjusted for multiple test comparisons using the Benjamini-Hochberg procedure via the R function p.adjust() from the stats package. Normalized expression values were calculated by taking the UMI matrix, obtained using Monocle 3’s counts() function and dividing by the size factors of each cell using Monocle 3’s size_factors() function. Using an adjusted p-value cutoff of 0.0001, 2088 genes were called as differentially expressed with 604 genes more highly expressed in Xylem Pole Pericycle and 1484 genes more highly expressed in Lateral Root Primordia.

The last approach involved using the analysis tool Vision. The normalised expression matrix for only XPP and LRP cell clusters was exported from Monocle. The gene signature of

pericycle cell population from (Parizot et al., 2012) were used for running Vision() and analyse() function. LRP cell population was selected in browser view mode to identify DEG against XPP population. Vision identified 4900 DEGs using an FDR of less than 0.05.

Data Analysis: Pseudotime Analysis - Xylem Pole Pericycle Cells Development. Pseudotime analysis was performed on two subsetted CDSs, one containing only Xylem Pole Pericycle cells and Mature Pericycle cells, and the other containing only Xylem Pole Pericycle cells and Lateral Root Primordia cells. Cells in both CDSs were assigned a pseudotime on the cell trajectory using Monocle 3's order_cells() function with the Xylem Pole Pericycle serving as the root of the trajectory. Genes whose expression changed as a function of pseudotime were identified using a generalized linear model. This was done on both CDSs using the fit_models() function and the following parameters:

```
model_formula_str = "~pseudotime",
expression_family="negbinomial",
clean_model=T
```

Using an FDR cutoff of 0.1, 1394 genes were identified as changing as a function of pseudotime in the CDS containing only Xylem Pole Pericycle cells and Mature Pericycle cells, and 1014 genes were identified as changing as a function of pseudotime in the CDS containing only Xylem Pole Pericycle cells and Lateral Root Primordia cells with an overlap of 510 genes.

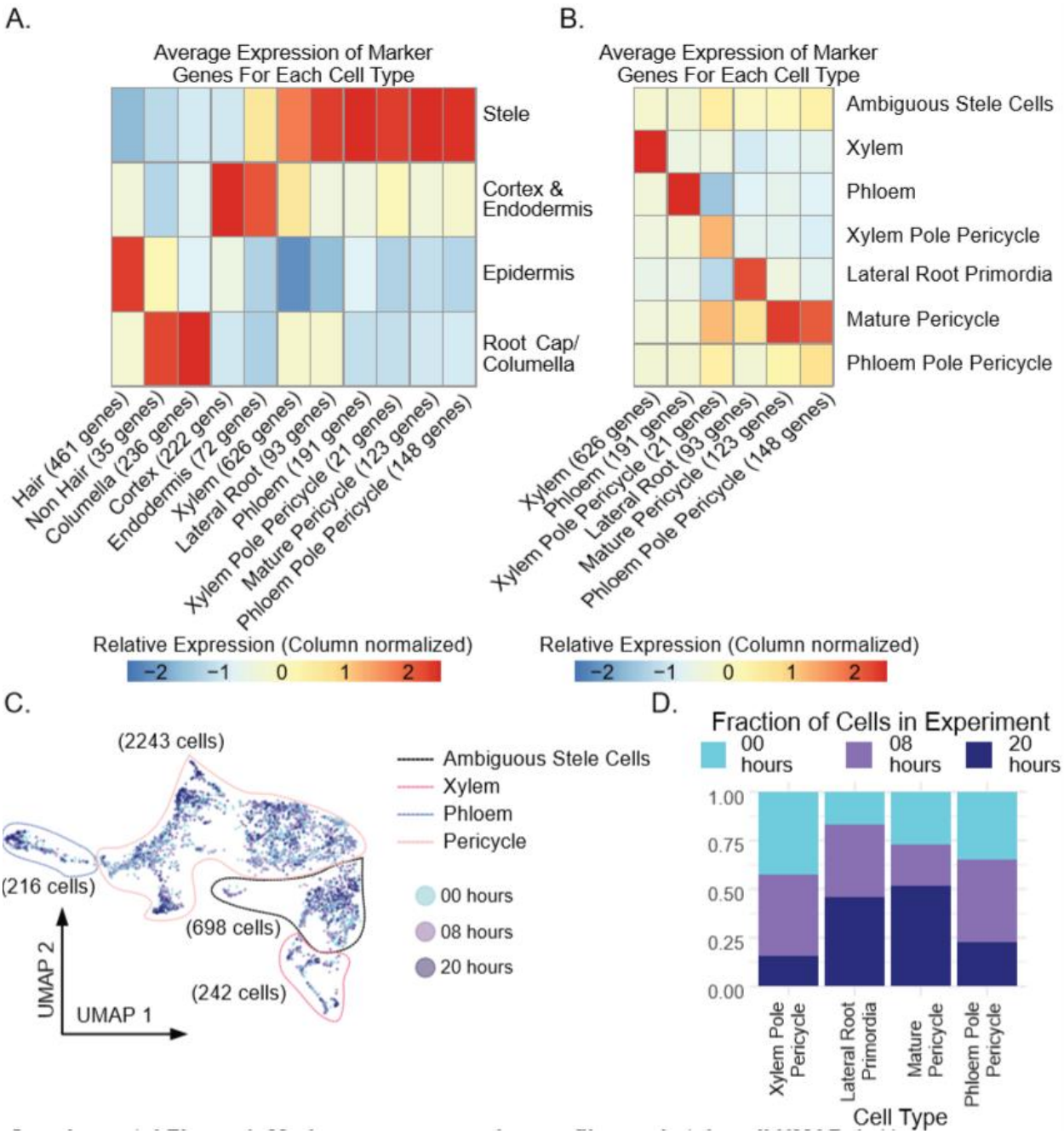
Data Analysis: Calling Differential Expressed Genes: Endodermis vs. Lateral Root Primodia Responding Endodermis. As previously described, a generalized linear model was used to identify differentially expressed genes between the cluster of cells labeled Endodermis, and the cluster of cells labeled Lateral Root Primodia Responding Endodermis. Using an FDR cutoff of 0.1, 1251 genes were identified as differentially expressed, with 748 genes being more expressed in Endodermis and 503 genes being more expressed in Lateral Root Endodermis. To

identify additional DEGs, the MMW test was performed comparing Endodermis to Cortex and Lateral Root Endodermis to Cortex.

Data Analysis: Pseudotime Analysis - Endodermis Cells. Pseudotime analysis was performed on two subsetted CDSs, one with only Endodermis cells, and the other containing only Endodermis below the branch point, and Lateral Root Primordia Responding Endodermis. As previously described, cells were assigned a pseudotime along the cell trajectory with the Endodermis cells below the branch point serving as the root. As previously described, a generalized linear model was used to identify differentially expressed genes as a function of pseudotime. Using an FDR cutoff of 0.1, 2063 genes were identified in the CDS with only Endodermis, and 2079 genes were identified in the CDS with only Endodermis and LRP Responding Endodermis with an overlap of 2060 genes.

Data Analysis: GO Term Enrichment Analysis. GO term enrichment analysis was performed using PANTHER (http://pantherdb.org/) (Mi et al., 2019). For GO term enrichments for XPP, LRP, Endodermis, and LRE, genes that were significant in at least 2 methods were used for analysis. All genes in the *Arabidopsis* genome were used as a background. Fisher's Exact test was used and the False Discovery Rate was calculated for multiple test correction. The complete annotation data set for biological process, molecular function, and cellular component GO terms were used for analysis.

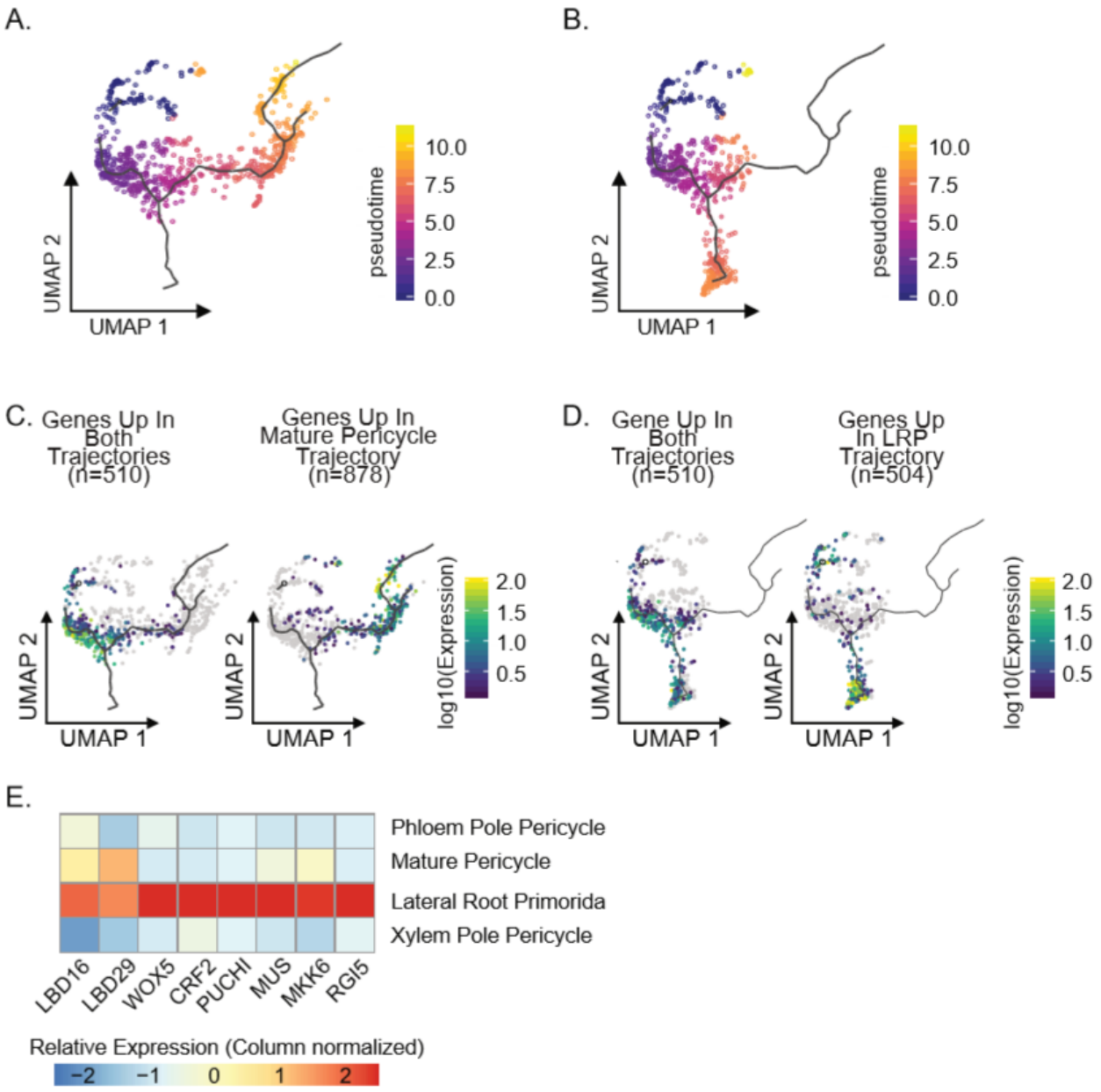
3.6 Supplementary Figures



Supplemental Figure 3.1. Marker gene expression profiles and stele cell UMAP.

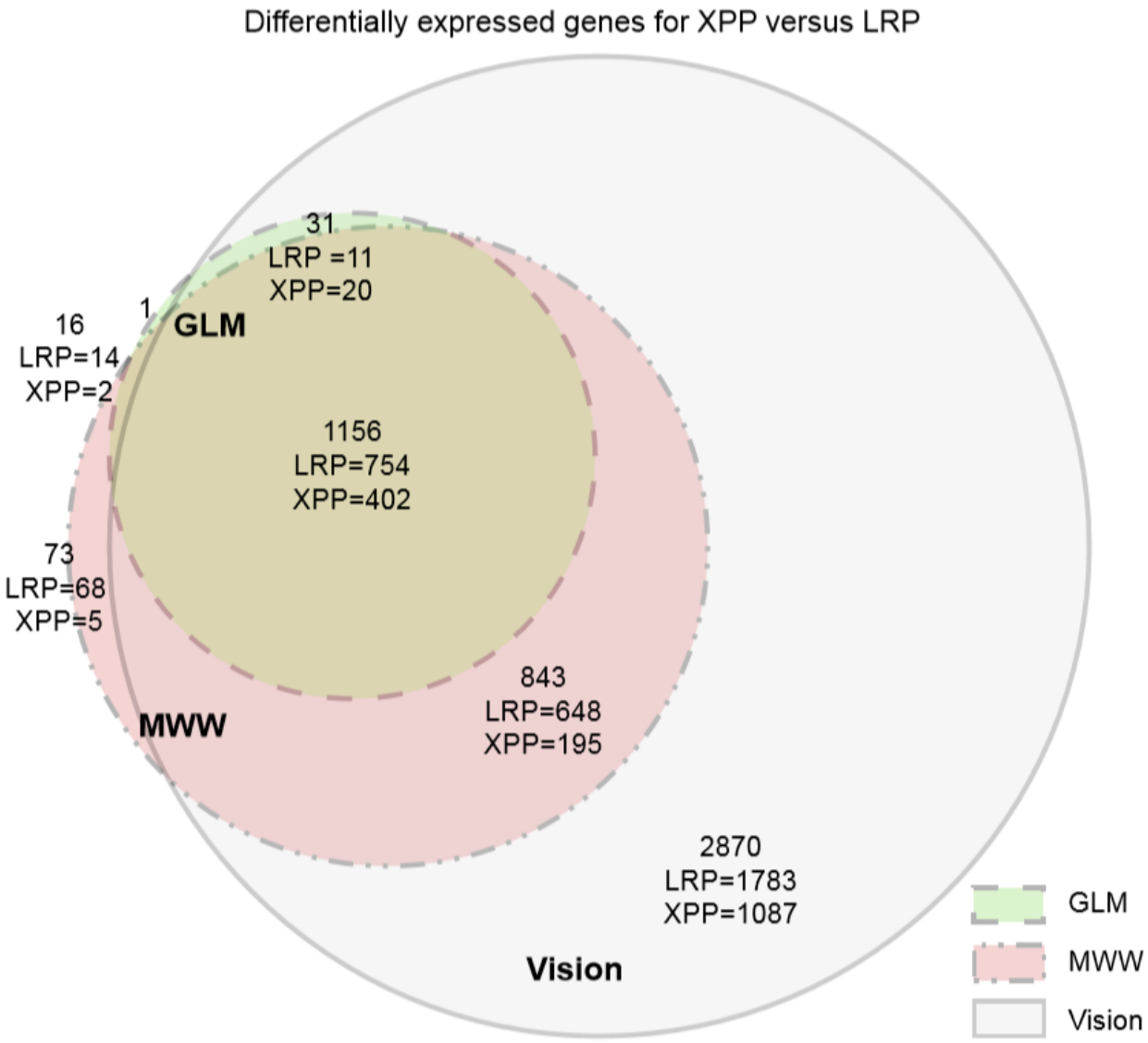
(A) Heatmap (column-scaled) visualizing average normalized expression of marker genes in the columella cells, epidermis cells, cortex & endodermis cells, and stele cells. Scale bar represents the z-score of the normalized expression values. **(B)** Heatmap (column-scaled) visualizing average normalized expression of marker genes in different stele cell types. Scale bar

represents the z-score of the normalized expression values. **(C)**. UMAP of stele cells colored by experiment. **(D)** Fraction of xylem pole pericycle (XPP), lateral root primordia (LRP), mature pericycle (MP), and phloem pole pericycle (PPP) cells from each experiment.



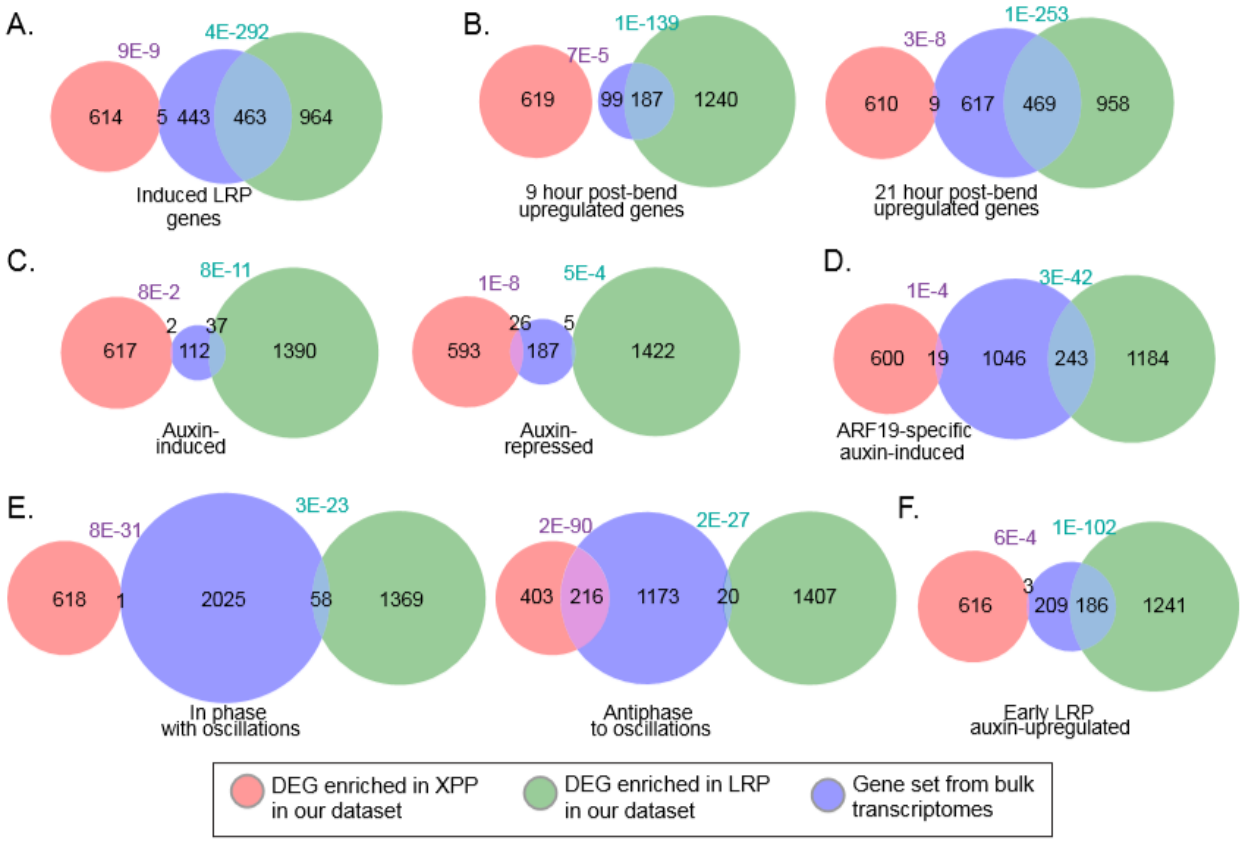
Supplemental Figure 3.2. Xylem pole pericycle developmental trajectories.

(A) UMAP of the XPP to Mature Pericycle trajectory colored by pseudotime. **(B)** UMAP of the XPP to LRP trajectory colored by pseudotime. **(C)** Expression UMAP (XPP and Mature Pericycle cells) of DEGs identified in both trajectories (XPP to Mature Pericycle and XPP to LRP) and in only the XPP to Mature Pericycle trajectory. **(D)** Expression UMAP (XPP and LRP cells) of DEGs identified in both trajectories and in only the XPP to LRP trajectory. **(E)** Heatmap (row-scaled) visualizing average normalized expression of genes identified as differentially expressed in the XPP to LRP trajectory. Scale bar represents the z-score of the normalized expression values.



Supplemental Figure 3.3. Size-adjusted Venn diagram visualizing overlap of genes between different DEG calling methods.

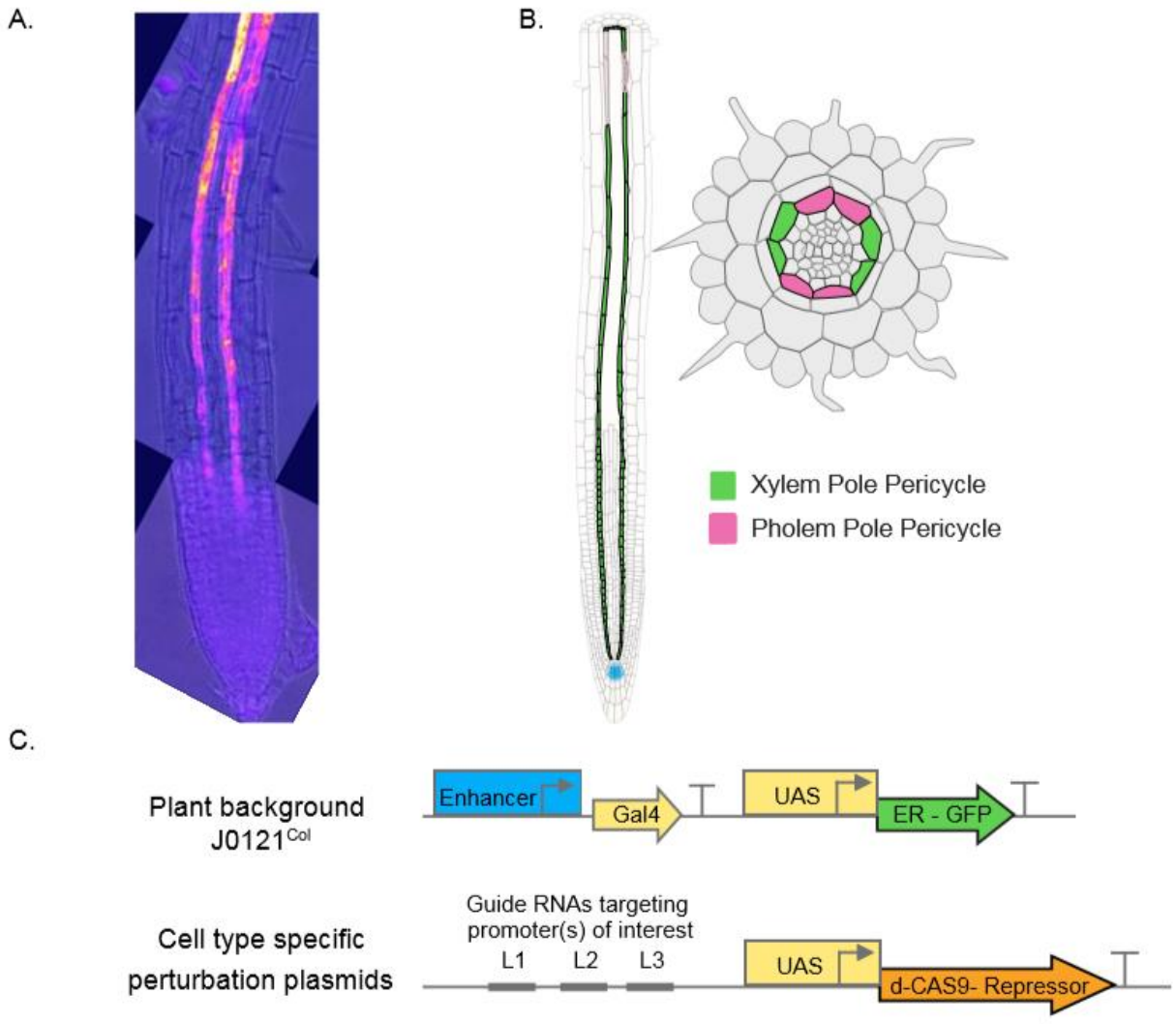
For each circle of the Venn Diagram, the total number of DEGs as well as the number of DEGs up in LRP and up in XPP are shown. All genes that were called in two or more methods were used for downstream analysis.



Supplemental Figure 3.4. Comparison of XPP and LRP DEGs from the single-cell library to bulk transcriptomes.

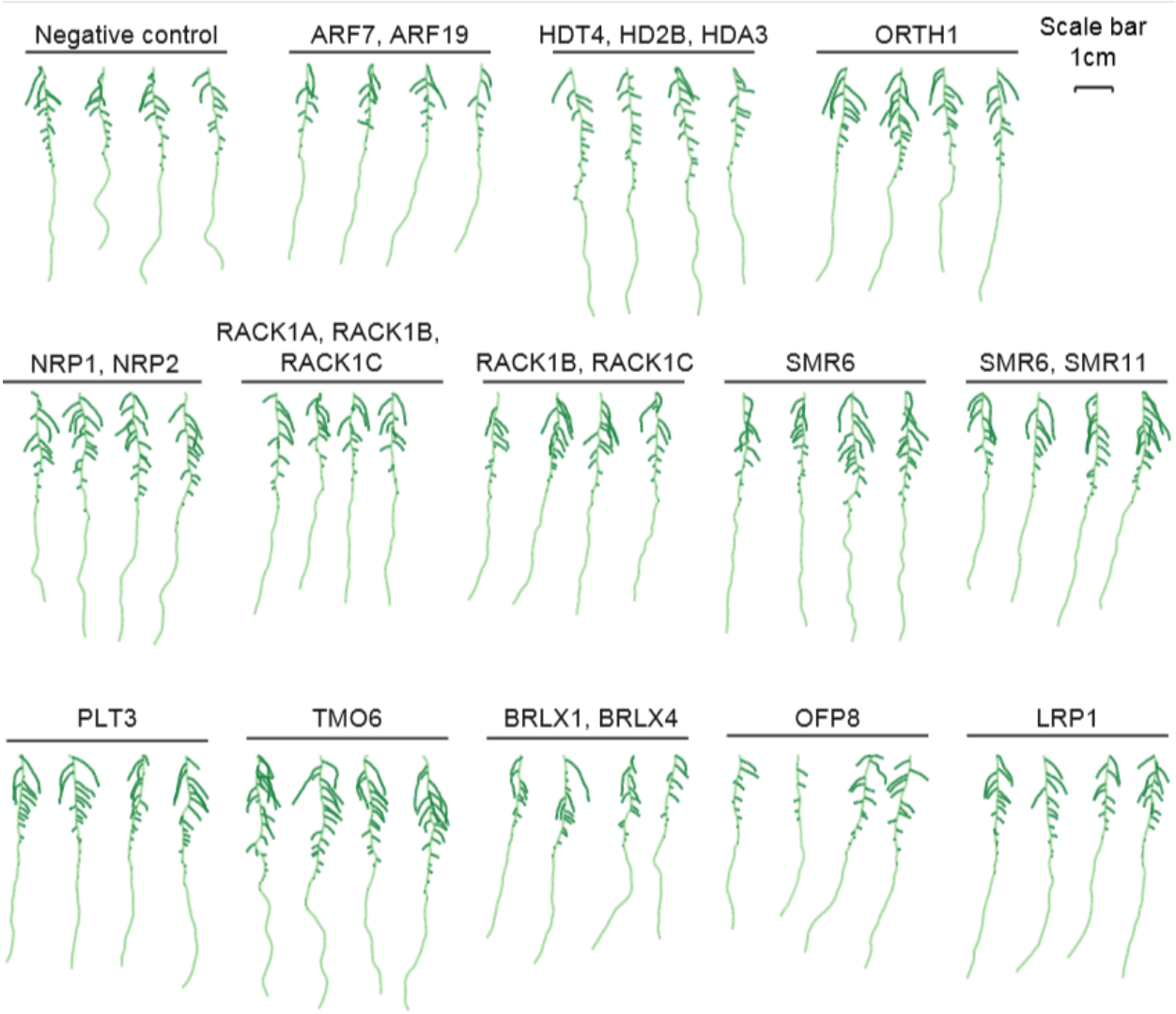
(A) Comparison to LRP-induced genes from Vanneste *et al*, 2005. **(B)** Comparison to time-course analysis of lateral root initiation at nine and twenty-one hours post-bend from Voß *et al*, 2015. **(C)** Comparison to auxin-induced and auxin-repressed genes in the root from Lewis *et al,* 2013. **(D)** Comparison to ARF19-specific auxin-induced genes from Powers *et al*, 2019. **(E)**

Comparison to genes oscillating in phase and antiphase to auxin in the basal meristem during lateral root specification from Moreno-Risueno *et al*, 2010. **(F)** Comparison to auxin-induced genes during early lateral root development from Ramakrishna *et al,* 2019. Each gene set from bulk transcriptomes is compared to the XPP and LRP DEGs with a size-adjusted Venn diagram. The number of genes in each mutually-exclusive area of each Venn diagram are specified. The values in purple and turquoise denote the hypergeometric distribution for the union XPP/bulk transcriptome and LRP/bulk transcriptome respectively.



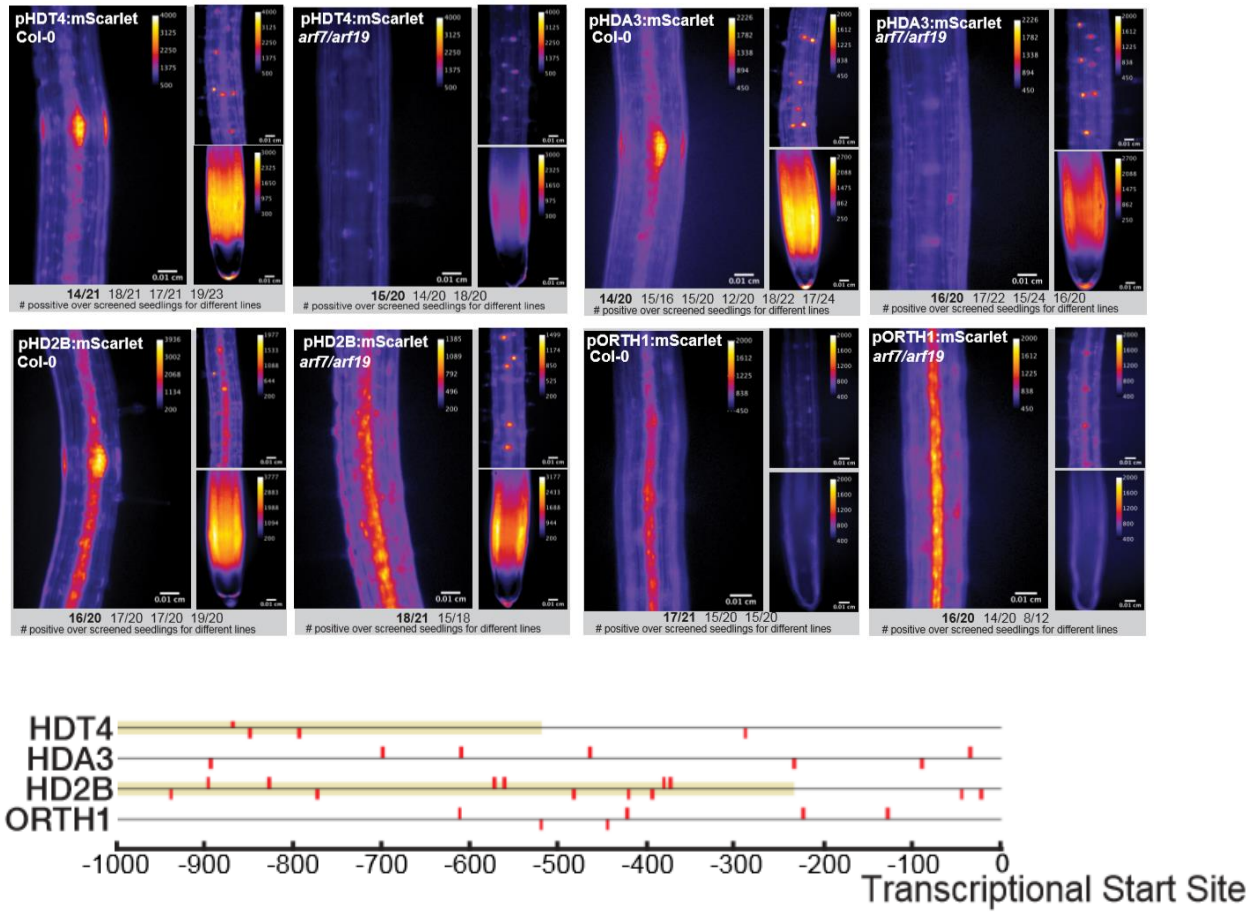
Supplemental Figure 3.5. Design of J0121^{Col}>>dCas9R system to generate cell type-specific dCas9-repressor mediated knockdown of candidate gene expression.

(A) J0121^{Col} is an enhancer trap line where the UAS-Gal4 system drives expression of GFP in the xylem pole pericycle cell file, visualized in confocal microscopy image and **(B)** labeled in cartoons (in green). **(C)** Top panel indicates the enhancer trap cassette in the J0121^{Col} and bottom panels is design of perturbation plasmids included up to three guide RNAs, tagged as location L1, L2, and L3, and a UAS promoter driving expression of dCas9-repressor cassette. Perturbation plasmids with respective cloned guide RNAs were transformed into J0121^{Col} background to drive repression of target genes specifically in xylem pole pericycle cells.



Supplemental Figure 3.6. Example seedling traces from J0121^{Col}>>dCas9R perturbation lines.

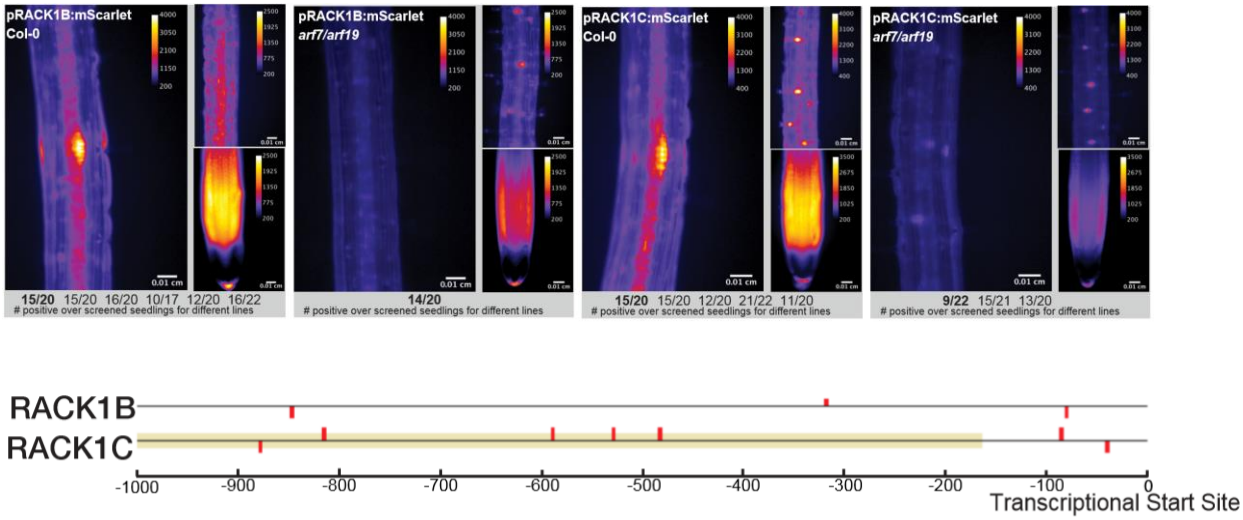
Roots from T2 perturbation line seedlings were quantified for various lateral root developmental phenotypes using SmartRoot.



Supplemental Figure 3.7. Transcriptional reporters of chromatin regulator candidate genes in wild type and *arf7arf19* roots.

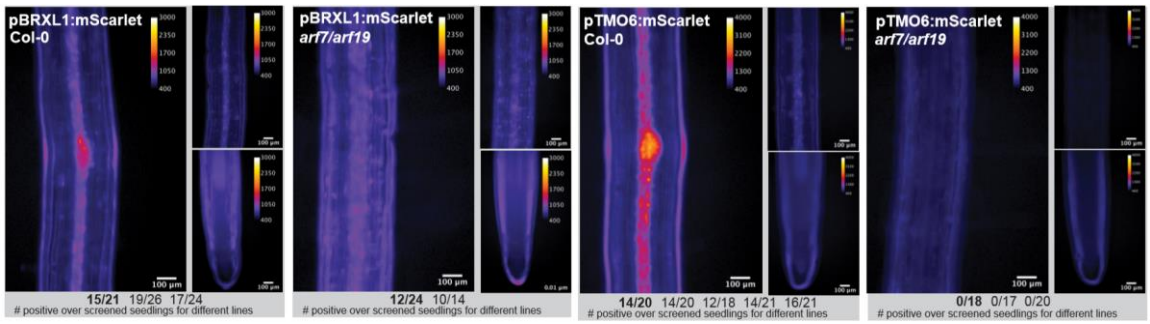
Fluorescent microscopy images of transgenic plant lines carrying transcriptional reporters of candidate genes in early stage lateral root primordia are shown in the large left-side image of each panel. Smaller images of the same transcriptional reporters in a region of the differentiated zone of the root without any developing primordia (above) and in the root apical meristem (below) are shown on the right in each panel. The same reporters in the same regions of the root are imaged in *arf7arf19* mutant background plants on the right. The number of independent transgenic lines imaged per construct and the number of plants within each line that showed expression are reported at the bottom. The lower panel represents 1000 bp upstream of the transcription start site for each gene, with auxin response elements (TGTC/GACA) highlighted

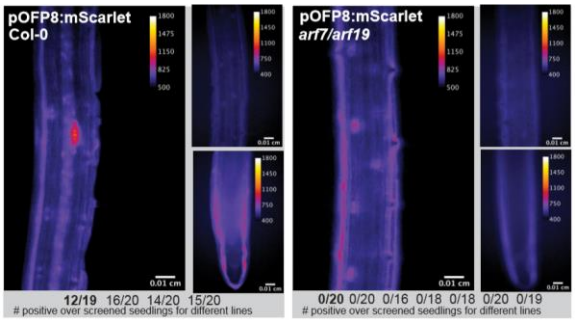
in red. Yellow bars indicate CDSs from other genes. This panel was obtained from http://bar.utoronto.ca/cistome.

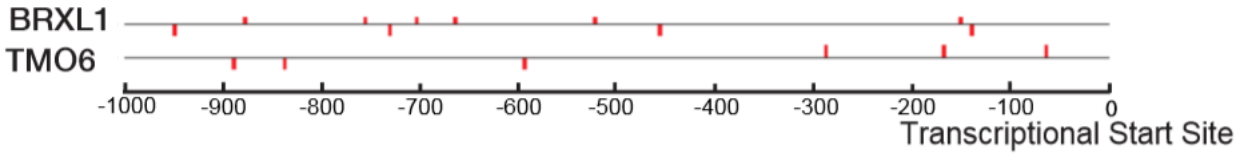


Supplemental Figure 3.8. Transcriptional reporters of cell cycle candidate genes in wild type and *arf7arf19* roots.

Fluorescent microscopy images of transgenic plant lines carrying transcriptional reporters of candidate genes in early stage lateral root primordia are shown in the large left-side image of each panel. Smaller images of the same transcriptional reporters in a region of the differentiated zone of the root without any developing primordia (above) and in the root apical meristem (below) are shown on the right in each panel. The same reporters in the same regions of the root are imaged in *arf7arf19* mutant background plants on the right. The number of independent transgenic lines imaged per construct and the number of plants within each line that showed expression are reported at the bottom. The lower panel represents 1000 bp upstream of the transcription start site for each gene, with auxin response elements (TGTC/GACA) highlighted in red. Yellow bars indicate CDSs from other genes. This panel was obtained from http://bar.utoronto.ca/cistome.



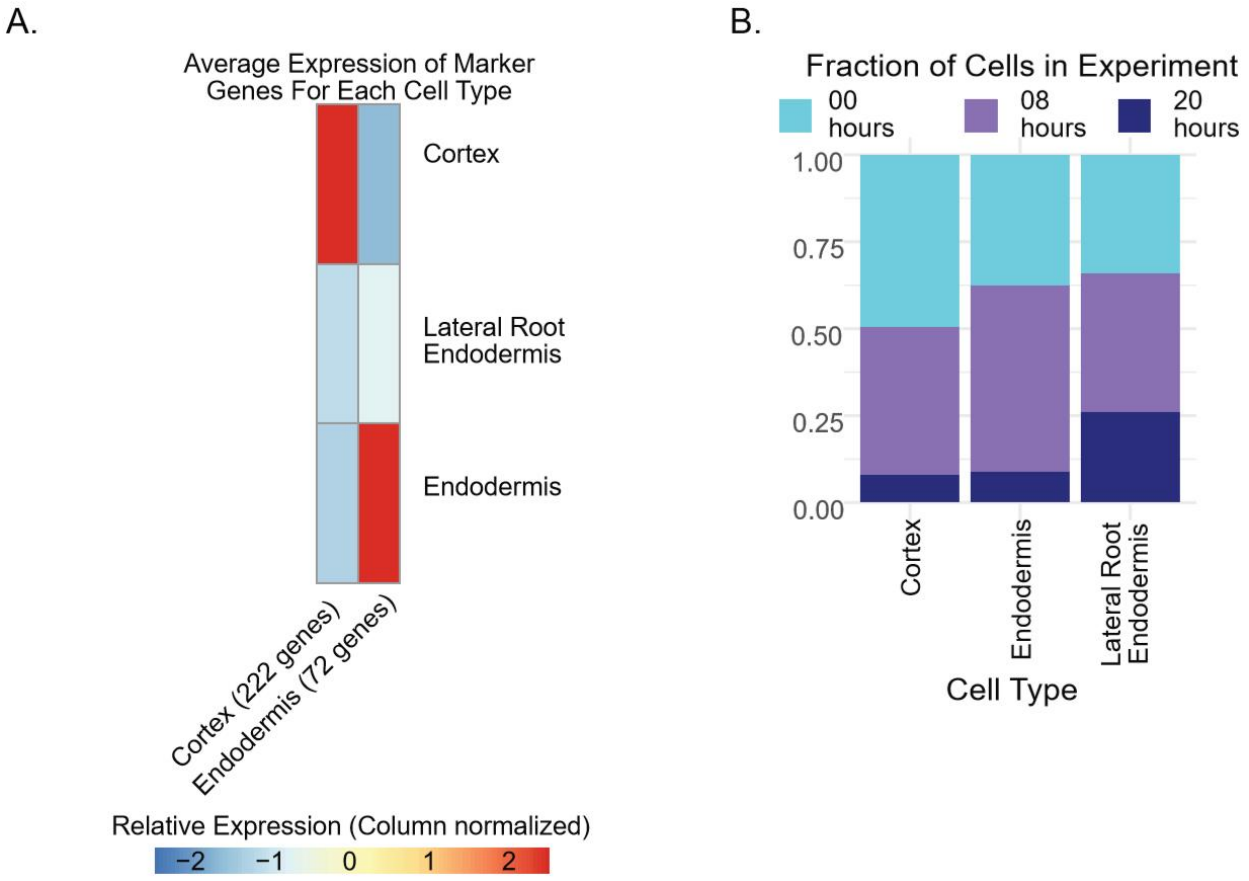




Supplemental Figure 3.9. Transcriptional reporters of stemness candidate genes in wild type and *arf7arf19* roots.

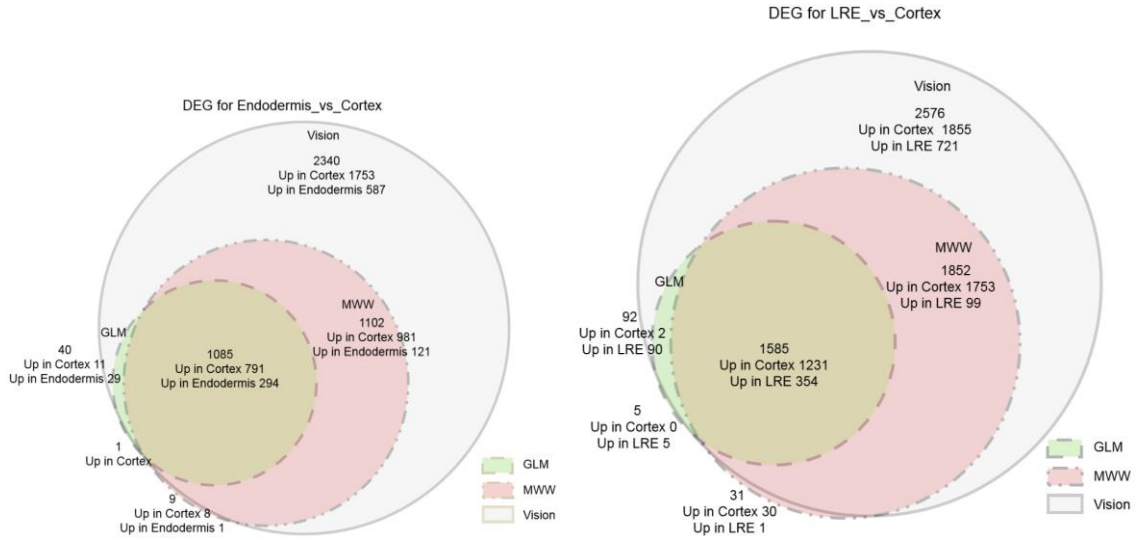
Fluorescent microscopy images of transgenic plant lines carrying transcriptional reporters of candidate genes in early stage lateral root primordia are shown in the large left-side image of each panel. Smaller images of the same transcriptional reporters in a region of the differentiated zone of the root without any developing primordia (above) and in the root apical meristem (below) are shown on the right in each panel. The same reporters in the same regions of the root are imaged in *arf7arf19* mutant background plants on the right. The number of independent transgenic lines imaged per construct and the number of plants within each line that showed

expression are reported at the bottom. The lower panel represents 1000 bp upstream of the transcription start site for each gene, with auxin response elements (TGTC/GACA) highlighted in red. Yellow bars indicate CDSs from other genes. This panel was obtained from http://bar.utoronto.ca/cistome.



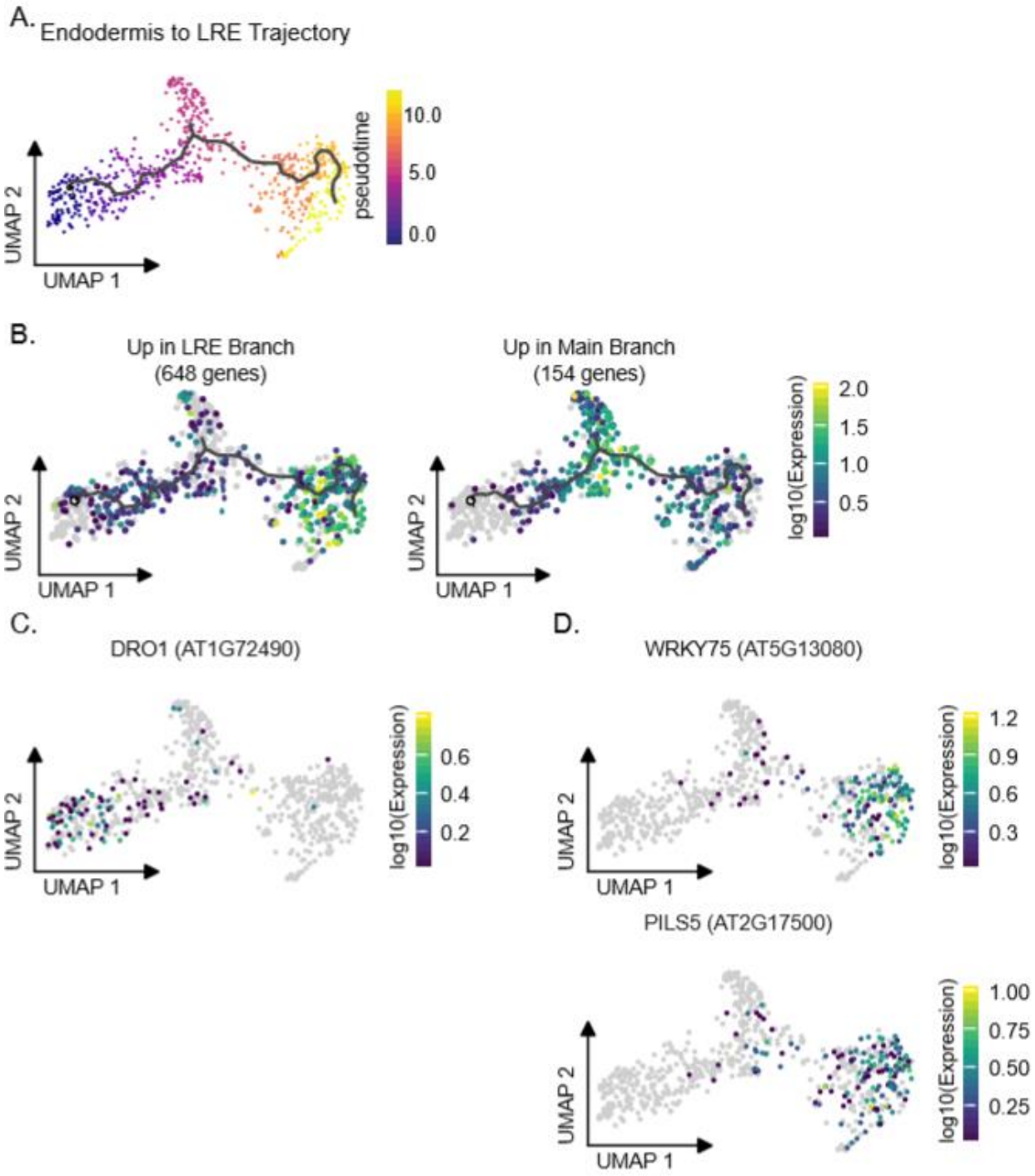
Supplemental Figure 3.10. Marker gene expression profiles and experiment breakdown of cortex, endodermis, and lateral root endodermis cells.

(A) Heatmap (column-scaled) visualizing average normalized expression of marker genes in the cortex, endodermis, lateral root endodermis (LRE) cells. **(B)** Fraction of cortex, endodermis, and LRE cells from each experiment.



Supplementary Figure 3.11. DEG overlaps with different methods for Endodermis/Lateral Root Endodermis analyses

(A) Endodermis vs cortex comparison and **(B)** LRE vs cortex comparison. For each ensemble of the Venn Diagram, the total number of DEGs, the number of DEGs up in Cortex and up in Endodermis (for **A**) and LRE (for **B**) are added to the diagram.



Supplemental Figure 3.12. Pseudotime analysis of Endodermis to Lateral Root Endodermis Cells.

(A) UMAP of the endodermis to LRE trajectory colored by pseudotime. **(B)** UMAP of the expression of gene sets that differed significantly as a function of pseudotime in the main endodermis branch and the LRE Branch. **(C)** Expression UMAP of DRO1. **(D)** Expression UMAPs of WRKY75 and PILS5.

Cell Type	00 hour cells	08 hour cells	20 hour cells
Columella/ Root Cap	379	207	429
Cortex & Endodermis	311	384	118
Epidermis	674	545	212
Stele	1121	1307	971
Total	2485	2443	1730

Cell Type	00 hour cells	08 hour cells	20 hour cells
Ambiguous Stele Cells	212	267	219
Lateral Root Primordia	28	62	77
Mature Pericycle	92	71	173
Phloem	58	68	90
Phloem Pole Pericycle	418	515	273
Xylem	85	102	55
Xylem Pole Pericycle	228	222	84
Total	1121	1307	971

Cell Type	00 hour cells	08 hour cells	20 hour cells
Cortex	63	54	10
Endodermis	154	220	36
Lateral Root Endodermis	94	110	72
Total	311	385	108

Supplementary Table 3.1. Breakdown of cell types by experiment.

3.7 Supplementary Data Availability

The raw data have been deposited with the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo) under accession code GSE158761. Supplemental Data Sets 3.1-3.3 are available online (https://www.biorxiv.org/content/10.1101/2020.10.02.324327v1).

3.8 Project Acknowledgements

We would like to thank members of the Nemhauser, Queitsch, Trapnell, Steinbrenner and Imaizumi groups for helpful discussions and technical guidance. We are grateful for the generosity of Dr. Dolf Weijers and his group in hosting A.L. to work on TMO6, as well as for Dr. Bert De Rybel's sharing of TMO6 reporter lines. This work was supported by the National Institutes of Health (R01-GM107084; J.L.N.; R01-GM079712, C.Q. and J.C.), the Howard Hughes Medical Institute Faculty Scholars Program (J.L.N.), and the National Science Foundation (IOS-1748843; C.Q. and J.C.). A.L. was supported by an NSF Graduate Research Fellowship (DGE-1256082), and S.G. was supported by an EMBO Postdoctoral Award (ALTF 409-2019).

Chapter 4. The Role of Chromatin Accessibility, Chromosome Conformation, and Efficient H2A.Z Incorporation in Acquired Thermotolerance in *Arabidopsis thaliana*

4.1 Abstract

Previous work, exploring chromatin accessibility in Arabidopsis after a lethal heat stress, revealed roughly 60 heat accessible genes whose entire gene bodies become highly accessible. To determine if this phenomenon occurred during more biologically relevant heat stresses, we performed ATAC-seq during multiple stages of acquired thermotolerance, a conserved mechanism by which organisms temporarily obtain the ability to survive severe, normally lethal heat stress shortly after experiencing a mild, sublethal heat stress. Here, we show that heat accessible gene bodies do emerge during non-lethal heat stresses, but this heat induced accessibility reverts back to normal levels within a few hours. Aside for a few exceptions, most of the previously identified heat accessible genes were accessible in any heat stress. We also identified additional genes that had heat induced accessibility, with a small number of genes displaying heat induced accessibility only in plants with acquired thermotolerance. Next, we explored the role H2A.Z incorporation plays in heat induced gene body accessibility using the mutant *arp6*, which we found to have defective thermomemory. When assayed for long term acquired thermotolerance, most genes had slightly lower gene body accessibility in *arp6* compared to wildtype plants. Lastly, to determine if heat stress altered chromosome conformation, we performed Hi-C. We found that most interactions between loci in the nuclear genome did not change significantly during heat stress. Instead, we found that lethal heat stress increased the number of interactions between the chloroplast genome and the nuclear genome. Acquired thermotolerance, however, prevented this increase in chloroplast interactions.

4.2 Introduction

Previous work using DNase-seq to examine chromatin accessibility in *Arabidopsis thaliana* at 22 °C and at 45 °C revealed several notable features about how a lethal heat stress changes chromatin accessibility (Sullivan et al., 2014a). The first notable feature was that chromatin accessibility remains largely unchanged in response to a lethal heat stress. The second notable feature was that the few loci that became less accessible after a lethal heat stress were enriched near growth, transport, and metabolic genes. The third notable feature was the loci that became more accessible after a lethal heat stress were predominantly located either in distal intergenic regions or in protein coding regions, often making entire gene bodies accessible. Transcription factor motif analysis showed that the intergenic heat released loci were enriched for the Heat Shock Element (HSE) motif. The heat accessible genes were predominantly heat shock response (HSR) genes such as *HEAT STRESS TRANSCRIPTION FACTOR A2* (*HSFA2* [AT2G26150]), *HEAT SHOCK PROTEIN 90* (*HSP90*) family genes, and *HSP70* family genes. The gene with the most accessibility was *HSP101* [AT1G74310], which was previously found to be necessary but not sufficient for acquired thermotolerance in *Arabidopsis thaliana (Queitsch et al., 2000)*. This observation prompted further exploration into the mechanism of acquired thermotolerance in plants.

Acquired thermotolerance is a conserved mechanism that allows an organism to survive a lethal heat stress after the organism has been pretreated by a milder heat stress at a lower, non lethal temperature (Carper et al., 1987; Lindquist and Craig, 1988; Morimoto et al., 1990; Nover, 1991). In plants, acquired thermotolerance depends upon the expression of *HSP101* to obtain this increased level of thermotolerance (Queitsch et al., 2000). The mere presence of HSP101 seems to increase a plant's basal thermotolerance, i.e. an organism's ability to survive higher temperatures without prior pretreatment. The presence of HSP101 in germinating seeds confers high basal thermotolerance for upwards of 48 hours after imbibition, and, in older

seedlings, constitutively overexpressing *HSP101* allows a plant to survive a lethal heat stress for short periods of time (~ 60 minutes) without prior pretreatment (Queitsch et al., 2000; Hong and Vierling, 2001; Mc Lellan et al., 2007). Despite the importance of HSP101, high levels of HSP101 alone is not sufficient to mimic the acquired thermotolerance phenomenon and does not provide the same degree of protection. Plants that have experienced a proper pretreatment can obtain higher levels of *HSP101* expression than *HSP101* constitutive expression lines, implying that there are additional components that regulate *HSP101* expression and acquired thermotolerance (Queitsch et al., 2000; Mc Lellan et al., 2007).

Additional studies examining long-term acquired thermotolerance have revealed epigenetic and transcriptional regulatory components serve as a form of memory for previous heat stresses, or thermomemory. One necessary, but not sufficient, transcription factor for thermomemory is HSFA2. After pretreatment, HSFA2 promotes the expression of other HSR genes (Charng et al., 2007). HSFA2 target genes have been shown to accumulate H3K4me2/H3K4me3 and exhibit prolonged expression after a pretreatment via a sublethal heat stress; both the accumulation H3K4 methylation and the prolonged expression are dependent on HSFA2. Genes that accumulated H3K4 methylation after an initial heat stress, showed hyper-induction upon subsequent heat stresses (Lämke et al., 2016).

Despite how much is known about acquired thermotolerance in Arabidopsis at the protein, transcriptional, and histone marks levels, the complete scope of regulatory loci involved in the phenomenon remains unclear. To identify potential HSR loci involved in acquired thermotolerance and to determine if heat induced gene body accessibility occurs outside lethal heat stresses, we performed ATAC-seq in *Arabidopsis thaliana.* Here we describe the dynamics of chromatin accessibility during various stages of acquired thermotolerance in genes and throughout the genome. We also describe these dynamics in the H2A.Z incorporation mutant, *ACTIN RELATED PROTEIN 6* (*arp6*). H2A.Z has been shown to be anticorrelated with DNA

methylation, to be incorporated into the gene bodies of stress responsive genes prior to stress, and to be evicted from gene bodies in response to stress (Zilberman et al., 2008; Sadeghi et al., 2011; Coleman-Derr and Zilberman, 2012; Sura et al., 2017; Nguyen and Cheong, 2018). Finally, we describe the effect heat stress has on chromosome conformation.

4.3 Results

Gene bodies become accessible during non-lethal heat stresses

To identify regions that experience accessibility changes in the process of acquired thermotolerance, ATAC-seq was performed on seven day old Col-0 seedlings for six treatments: Control (CN), Pretreatment (PT), Recovery (RC), Short-Term Acquired Thermotolerance (SAT), Long-Term Acquired Thermotolerance (LAT), and Lethal (LT) (**Figure 4.1A**).

The set of genes (n = 61 genes) previously identified by Sullivan *et al.* 2014 as becoming accessible under lethal heat stress also become accessible during non-lethal heat stresses (**Figure 4.1B**). This gene body accessibility, however, is only temporary. During RC, within two hours after the plant returns to normal temperatures, these heat accessible genes lose their PT accessibility and revert back to CN levels of accessibility. Furthermore, for many of these heat accessible genes (42/61), including as *HSP101* and *HSFA2*, their LAT accessibility is greater than their LT accessibility (**Figure 4.1C, Figure 4.1D**). Thirty-three of these 42 genes had over a 50% increase in accessibility during LAT compared to LT. This suggests that a hallmark of thermomemory in pretreated seedlings is increasing accessibility at specific heat accessible genes in response to severe heat stress of 45 ℃ relative to non-pretreated seedlings.

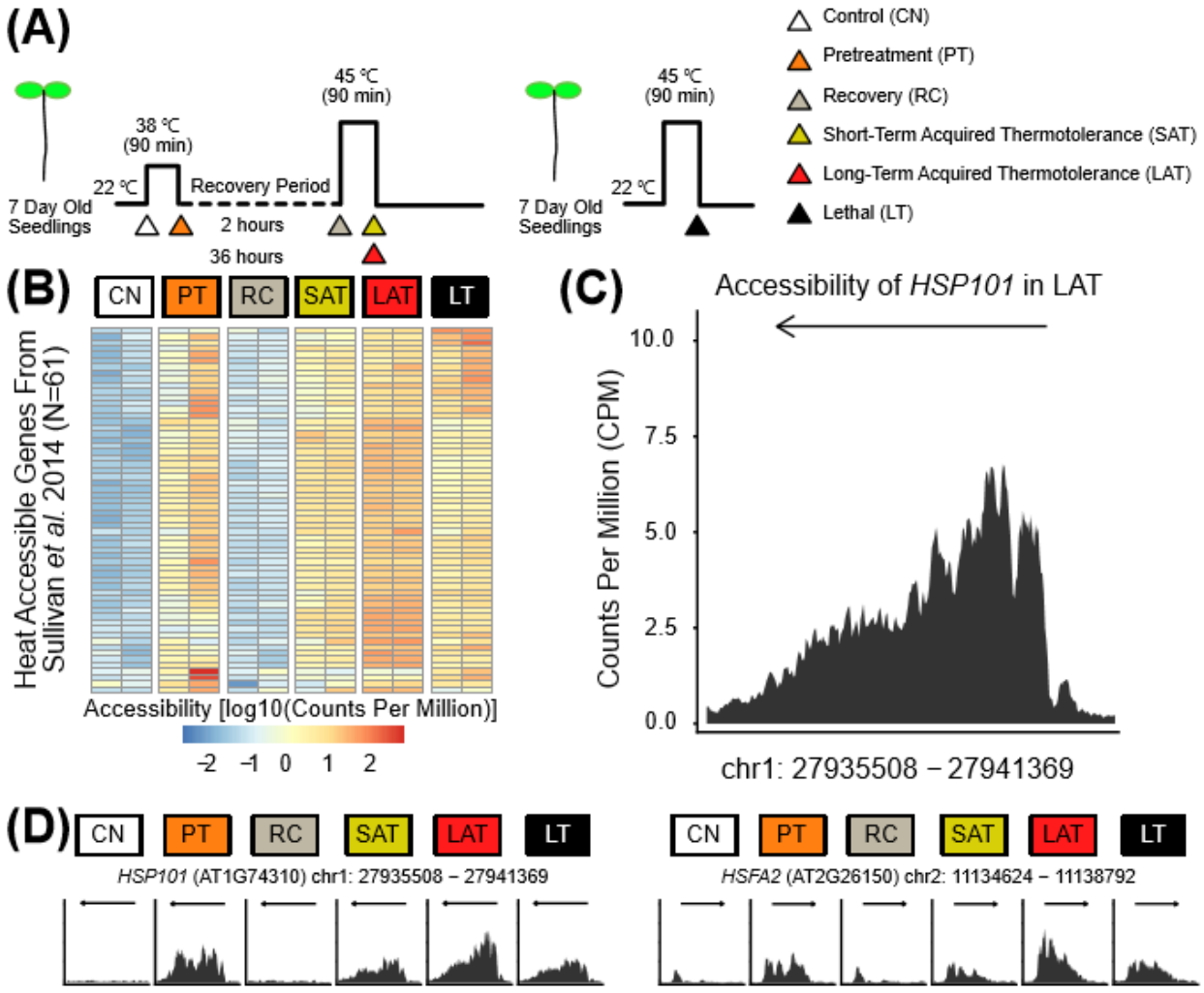


Figure 4.1. Gene body accessibility is present in non-lethal heat stresses.

(A) Seven day old Arabidopsis seedlings, grown under long day conditions (16 hour light, 22 ℃; 8 h dark, 22 ℃), were subjected to several treatment conditions, after which nuclei were extracted using the INTACT method, and ATAC-seq was performed. As a control (CN) ATAC-seq was performed on non heat stressed Arabidopsis seedlings at 22 ℃. For pretreatment (PT), plants were shifted from 22 ℃ to 38 ℃ for 90 minutes. For recovery (RC), plants were shifted from 22 ℃ to 38 ℃ for 90 minutes and allowed to recover at 22 ℃ for 2 hours. For short-term acquired thermotolerance (SAT), plants were shifted from 22 ℃ to 38 ℃ for 90 minutes, allowed to recover at 22 ℃ for 2 hours, and then were shifted from 22 ℃ to 45 ℃ for 90 minutes. For long-term acquired thermotolerance (LAT), plants were shifted from 22 ℃ to 38 ℃ for 90 minutes, allowed to recover at 22 ℃ for 36 hours, and then were shifted from 22 ℃ to 45 ℃ for

90 minutes. Lastly, as a lethal heat shock (LT), plants were shifted from 22 ℃ to 45 ℃ for 90 minutes. **(B)** Accessibility heatmap (row-scaled) of heat accessible genes identified by Sullivan et al. 2014 across treatments. Normalized read counts, Counts Per Million (CPM), within a gene were summed for visualization (See Methods: ATAC-seq analysis). **(C)** Density curve plot visualizing accessibility of HSP101 (as well as 1kb upstream and 1kb downstream of the gene) in LAT. For the y-axis, the replicate average is visualized. For the x-axis, each "point" on the density curve is a 20bp window. The arrow indicates the gene body as well as the direction of transcription. **(D)** Accessibility of HSP101 and HSFA2 across treatments. As with (C), the y-axes of these plots range from 0 to 10, however for visualization, axis labeling has been removed for this and subsequent visualizations of accessibility via density curve plots. Instead, the y-axes will be fixed at a range of 0 to 10, and genomic regions will be provided as a header.

The 61 heat accessible genes had also previously been shown to have dramatically increased expression. This trend held true during mild heat treatment (38C) for key thermotolerance genes, HSP101 and HSFA2, and that LAT conditions evoked higher expression than LT conditions for 58/61 of these genes. For *HSP101* and *HSFA2*, the correlation between accessibility and expression during a mild heat stress of 38 °C was high (*HSP101* Pearson correlation ~0.8; *HSFA2* Pearson correlation > 0.9). For both genes, expression and accessibility peak 40-60 minutes into a mild heat stress (**Supplemental Figure 4.1A, Supplemental Figure 4.1B, Supplemental Figure 4.1C**). Expression of *HSP101* also correlates with temperature (Pearson correlation > 0.8) (**Supplemental Figure 4.1D, Supplemental Figure 4.1E**). To further explore the relationship between accessibility and expression in the other previously identified heat accessible genes, Global Run-On sequencing (GRO-seq) data from a study performed by Liu *et al.* 2020 examining nascent RNA expression in pretreated and non-pretreated plants was obtained for reanalysis (**Supplemental Figure 4.2A**) (Liu et al., 2020). Of the 61 heat accessible genes, 58 genes had higher nascent RNA expression in pretreated plants compared to non-pretreated plants in response to brief exposure

to 45 °C (**Supplemental Figure 4.2B**). Both *HSP101* and *HSFA2* were among the genes with increased nascent RNA expression in pretreated plants (**Supplemental Figure 4.2C, Supplemental Figure 4.2D**). Similar to the findings observed at 38 °C for *HSP101* and *HSFA2*, 48 of the 61 heat accessible genes had a high correlation between accessibility and nascent RNA expression (Pearson correlations > 0.8). Together, these results suggest that for many of the previously identified 61 heat accessible genes there is a strong correlation between their accessibility and the rate of RNA expression.

A small number of heat accessible HSPs and HSFs become inaccessible in certain heat stress conditions

Moving beyond the 61 previously identified heat accessible genes, the gene body accessibilities of each of the *HSP90* genes, the *HSF* genes, the *HSP70* genes, and various small *HSP* genes (sometimes referred to as *sHSPs* or *HSP20-LIKE CHAPERONES*) were examined across all six treatments. This revealed two things. The first was that some *HSPs*, regardless of the treatment, do not display the gene body accessibility phenomenon observed in LT. The second was that if a gene is accessible in LT, it is typically accessible in the other heat stresses as well (**Figure 4.2, Supplemental Figure 4.3**). There are some notable exceptions to this second observation.

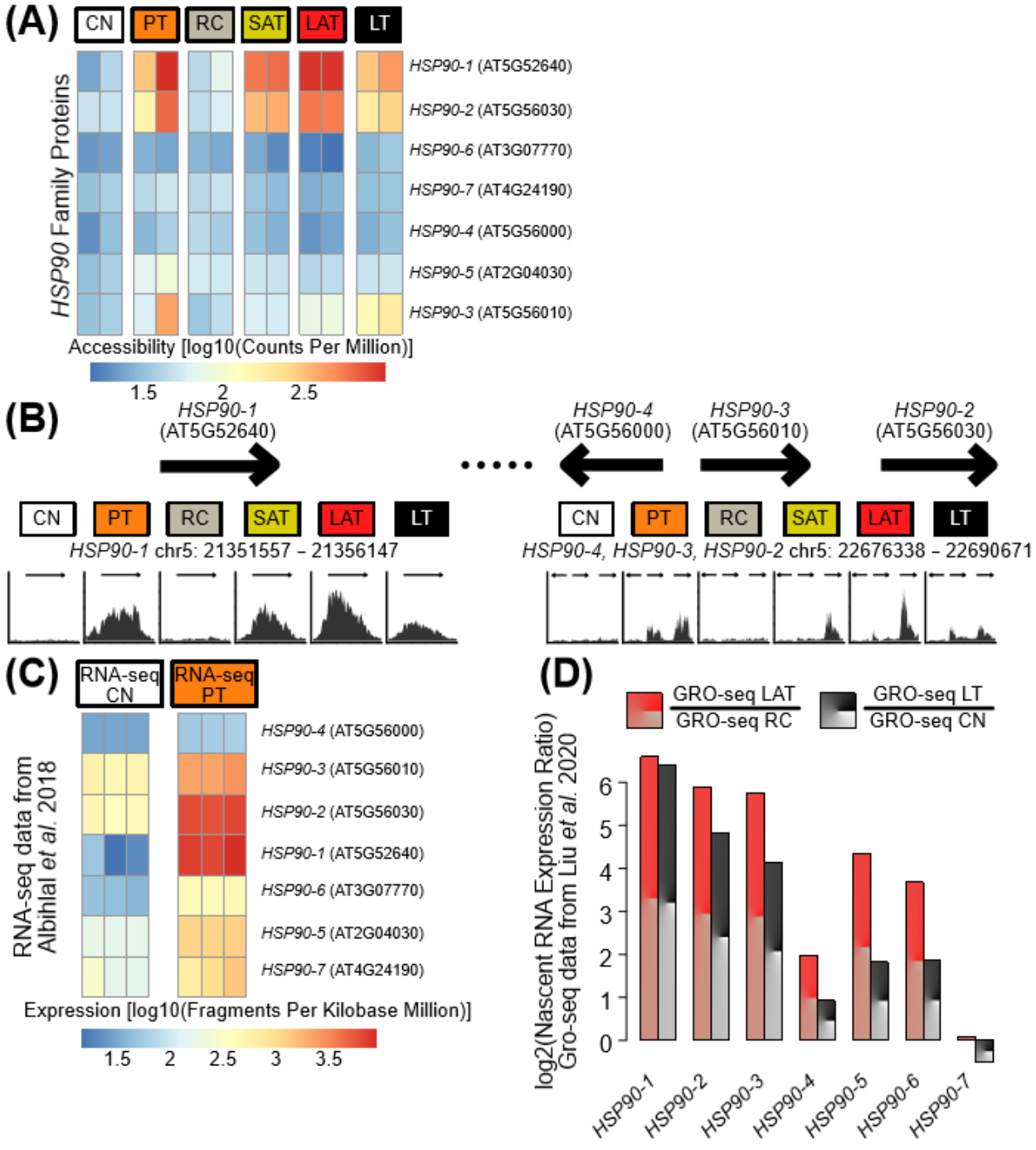


Figure 4.2. *HSP90-3* does not have heat induced gene body accessibility during SAT and LAT.

(A) Accessibility heatmap of *HSP90* family genes across treatment replicates. **(B)** Accessibility of *HSP90-1*, *HSP90-2*, *HSP90-3*, and *HSP90-4* across treatments. **(C)** Expression heatmap of *HSP90* family genes from RNA-seq data generated Albihlal et al. 2018. Five week old

Arabidopsis plants were grown under short day conditions (8 hour light; 16 hour dark) at 23 °C. For RNA-seq, tissue was collected from non-heat stressed plants (RNA-seq CN) or from pretreated plants that just experienced a 37 °C heat stress for 30 minutes (RNA-seq PT). Three replicates were generated for both conditions. **(D)** Nascent RNA expression ratios (GRO-seq LAT/ GRO-seq RC; GRO-seq LT/ GRO-seq CN) for *HSP90* genes. For details on GRO-seq experiments performed by Liu *et al.* 2020, see **Supplemental Figure 4.2A**.

One notable exception to the trend of being accessible in all heat stresses was *HSP90-3* [AT5G56010]. Of the seven *HSP90* genes in the Arabidopsis genome, only three of the *HSP90* genes displayed gene body accessibility in LT: *HSP90-1* [AT5G52640], *HSP90-2* [AT5G56030], and *HSP90-3* (**Figure 4.2A**). These genes were among the 61 previously identified heat accessible genes. Like many other previously identified heat accessible genes, *HSP90-1* and *HSP90-2* had gene body accessibility in PT, SAT, and LAT with both genes being most accessible in LAT. *HSP90-3* is distinct in that, of the four heat stress conditions, it is most accessible in PT, and is least accessible in SAT and LAT (**Figure 4.2B**).

Expression was also correlated with degree of gene-body accessibility for these genes. For example, expression of the three heat accessible *HSP90s* increased after a mild heat stress (**Figure 4.2C**), with *HSP90-1*, the most accessible of the trio, was also the most expressed *HSP90* gene, while *HSP90-3*, the least accessible of the trio, was also the least expressed of the trio (Albihlal et al., 2018). As seen with *HSP101* and *HSFA2*, there seems to be a high correlation between PT accessibility and total RNA levels, in response to mild heat stress, for these heat accessible *HSP90* genes (Pearson correlations ~0.9). Examining the GRO-seq data published by Liu *et al.* 2020, one would expect that the reduced accessibility of *HSP90-3* observed in pretreated plants in response to severe heat stress would also be reflected in the nascent RNA expression of *HSP90-3.* In terms of function, downregulating the expression of *HSP90-3* would help improve thermotolerance. *HSP90-3* overexpression has been shown to

inhibit heat tolerance in Arabidopsis and to delay the expression of *HSP101*, *HSP17s*, and *HSP21 (Xu et al., 2010)*. Nascent expression of *HSP90-3*, however, was highest in the pretreated plants briefly exposed to 45 °C, experiencing over a 60-fold increase in nascent RNA expression (**Figure 4.2D**). Although heat stress reduced nascent RNA expression genome wide, the number of genes that were significantly downregulated was much smaller than the number of significantly upregulated genes. From this observation, Liu *et al.* 2020 conclude that downregulation must be slower than upregulation and downregulation would probably be captured after a severe heat stress longer than five minutes. Given this observation and that the ATAC-seq data described here is captured after a 90 minute heat stress, this may explain the mismatch between *HSP90-3* accessibility and nascent RNA expression in the available GRO-seq data.

From examining the relationship between accessibility and expression in the *HSP90* genes, it is important to note that the low gene body accessibility does not mean the absence of gene expression, as seen with *HSP90-3* in control conditions. In addition, low gene body accessibility after a heat stress does not mean the absence of heat induced expression as seen with the other non heat accessible *HSP90s* (**Figure 4.2C, Figure 4.2D**).

Another gene that does not have gene body accessibility in all heat stresses is *HSFA7A.* Like the *HSP90* genes, not all *HSFs* displayed the gene body phenomenon (**Supplemental Figure 4.3A**). Aside from *HSFA2*, *HSFA7A* [AT3G51910] was the only other gene to have gene body accessibility during non-lethal heat stresses. Both genes were among the 61 previously identified heat accessible genes, but *HSFA7A* is distinct in that it has little PT accessibility and only becomes accessible in response to severe heat stress (**Supplemental Figure 4.3B**). Like *HSP101* and *HSFA2, HSFA7A* has been shown to be important for thermotolerance with *hsfa7a* mutants having reduced survivability to high temperatures even after being pretreated (Larkindale and Vierling, 2008). As for the other *HSF* genes, their expression differs depending

on the abiotic stress, but *HSFA2* and *HSFA7A* are not the only heat inducible *HSF*s (Swindell et al., 2007; Sugio et al., 2009; Lin et al., 2018).

Seven of the *HSP70* genes were previously identified as being heat accessible. Including the other treatment conditions, two additional *HSP70s* showed a slight increase in gene body accessibility (**Supplemental Figure 4.3C**). *HSP70-11* [AT5G28540] (also annotated as *BINDING IMMUNOGLOBULIN PROTEIN 1* or *BIP1*) only showed a slight increase in accessibility after a mild heat stress, while *HSP70-6* [AT4G24280] only showed a slight increase in gene body accessibility while assaying for acquired thermotolerance (**Supplemental Figure 4.3D**). HSP70-6 proteins localize to the chloroplast and have been shown to be important for thermotolerance in germinating seedlings (Su and Li, 2008). In older seedlings, *HSP70-6* was also shown not to be essential for acquired thermotolerance. Knockouts did not have reduced survivability when assayed for acquired thermotolerance, but they did have reduced chlorophyll levels (Su and Li, 2008). Like the *HSF*s, heat stress induces varying responses in the *HSP70* genes, with many genes being heat inducible, but not necessarily accessible (Lin et al., 2001; Sung et al., 2001).

All but four of the small *HSPs* were previously identified as heat accessible during a severe heat stress. These same *HSPs* were accessible in all the other heat stress conditions as well (**Supplemental Figure 4.3E**). A newly identified heat accessible small *HSP* was *HSP21* [AT4G27670], which becomes most accessible when assaying for acquired thermotolerance (**Supplemental Figure 4.3F**). *HSP21* is a known target for *HSFA2* and has been shown to play a key part in thermomemory maintenance in pretreated plants (Sedaghatmehr et al., 2016). The remaining three small *HSPs* that showed no gene body accessibility in any treatment were *HSP14.7* [AT5G47600], *HSP15.4* [AT4G21870], and *HSP21.7* [AT5G54660]. Unlike the other small *HSPs* which have heat inducible expression, *HSP15.4* has been shown to be downregulated in response to heat stress, while *HSP14.7* and *HSP21.7* are constitutively

expressed but their transcript levels are relatively unchanged during heat stress (Swindell et al., 2007; Siddique et al., 2008).

Collectively, these findings show that the most heat accessible genes are accessible in any heat stress with the few exceptions potentially reflecting biological function. Although heat induced accessibility often does correlate with the magnitude of heat induced expression, the lack of heat induced accessibility does not necessitate the lack of gene expression.

Identifying Differentially Accessible Chromatin Regions (dACRs) between treatments

In order to identify more heat accessible genes and other Accessible Chromatin Regions (ACRs), peak calling was performed on the ATAC-seq data (**Supplemental Data 4.1**). As expected, most ACRs identified resided in intergenic regions and promoters (**Supplemental Figure 4.4A**). Each treatment also had comparable replicate variances, and comparable fractions of reads in ACRs (**Supplemental Figure 4.4B, Supplemental Table 4.1**). To identify ACRs with differential accessibility between treatments (dACRs), a number of pairwise comparisons were performed: (1) CN vs. RC; (2) SAT/LAT vs. LT; (3) SAT vs. LAT; (4) PT vs. SAT/LAT and (5) PT vs. LT (**Figure 4.3A, Supplemental Figure 4.5A, Supplemental Data 4.2**). CN was compared to RC to identify regions whose accessibility could potentially “mark” a plant as having received a mild heat stress in the recent past. SAT and LAT were compared to LT to identify accessibility differences in the response to a severe heat stress in plants with and without acquired thermotolerance. SAT was compared to LAT to identify accessibility differences between short-term and long-term acquired thermotolerance response to severe heat stress. PT was compared to SAT and LAT to potentially identify loci that only become accessible with acquired thermotolerance as well as loci that are specifically accessible in PT. Finally, PT was compared to LT to identify accessibility differences between mild and severe

heat stress. Depending on the comparison, different regions of the genome were more accessible with the majority of differences residing in intergenic and promoter regions (**Figure 4.3B, Supplemental Figure 4.5B**). These comparisons also found additional genes, outside of the *HSPs*, *HSFs*, and the previously identified heat accessible genes, that display gene body accessibility. Many of the heat accessible genes and genes nearby dACR were previously associated with abiotic stress response and tolerance, but some had not been associated with heat stress before.

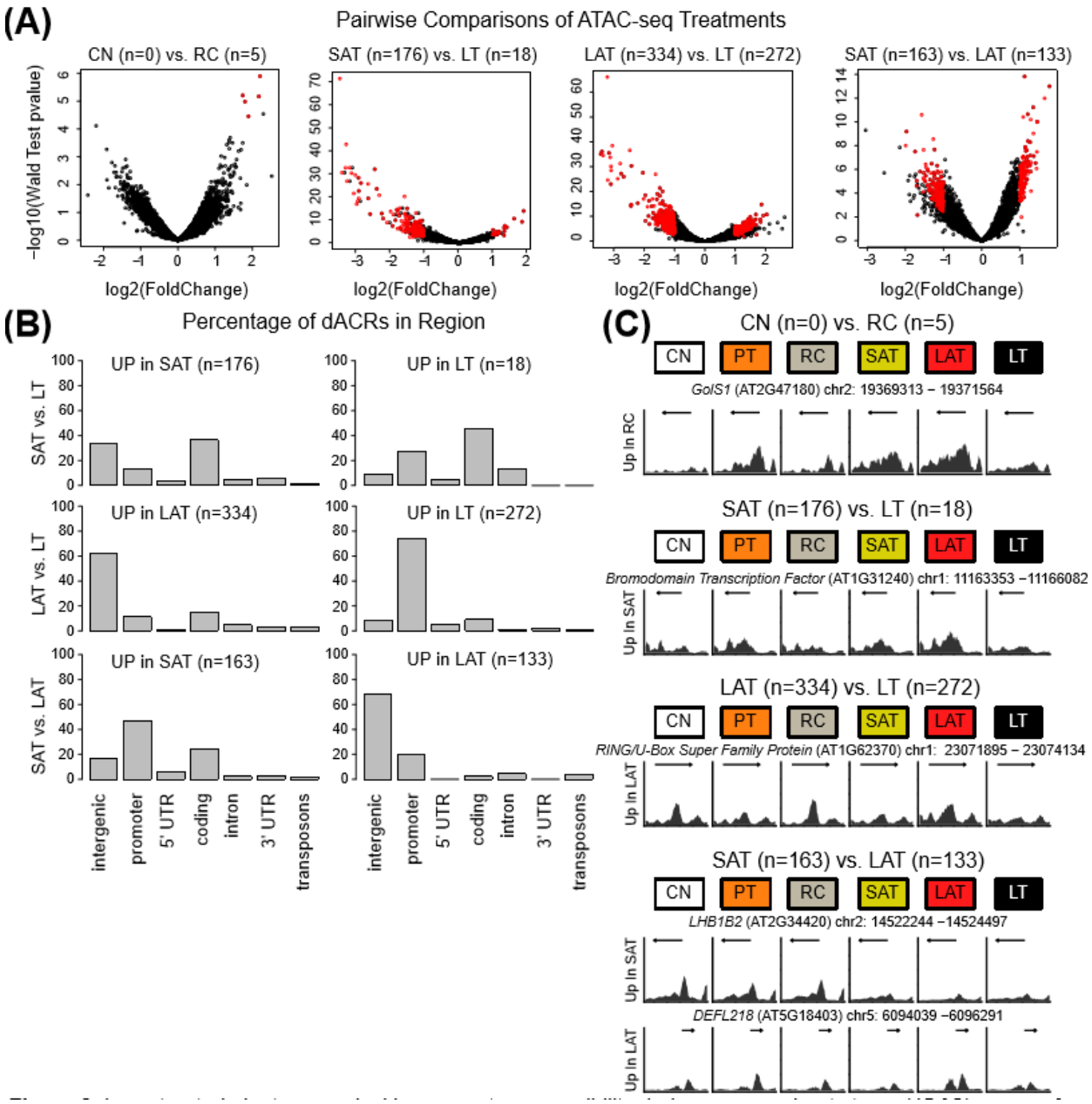


Figure 4.3. In pretreated plants, more loci have greater accessibility during a severe heat stress (45 °C); many of these loci reside in intergenic regions.

(A) Volcano plots comparing the following pairs of treatments: CN vs. RC, SAT vs LT, LAT vs LT, and SAT vs LAT. ACRs with an adjusted p-value less than 0.1 and an absolute fold change greater than 2 are colored in red. **(B)** Percentage of dACRs from the SAT vs. LT, LAT vs. LT, and SAT vs LAT comparisons that fall into the following regions: intergenic, promoter, 5' UTR, coding, intron, 3' UTR, and transposons. **(C)** Visualization of dACRs from the aforementioned

pairwise comparisons across treatments. For context, 1kb upstream and downstream of the dACR is visualized, and the name of any gene that falls within that interval is provided with the coordinates.

As with the heat accessible genes, most heat induced accessible loci outside of gene-bodies quickly revert back to becoming inaccessible within hours after a mild heat stress. The accessibility profile of a RC sample was nearly identical to the accessibility profile of a CN sample (Pearson correlation = 0.91). There were only five dACRs between CN and RC, all of which were more accessible in RC (**Figure 4.3A**). Despite the small number, these ACRs had an enrichment for *HSF* motifs (**Supplemental Data 4.2**). An example dACR that retains some of it's PT accessibility is the promoter region of *GALACTINOL SYNTHASE 1* (*GolS1* [AT2G47180]) (**Figure 4.3C**). Overexpression of *GolS1* leads to increased tolerance to heat, salt, and cold stress (Panikulangara et al., 2004; Nishizawa et al., 2008). It is also important to note that the gene body of *GolS1* becomes accessible in PT, SAT, and LAT. Another dACR lies in an intergenic region approximately 700 bps upstream of the TSS of the *ETHYLENE RESPONSE FACTOR 19 (ERF19* [AT1G22810]*)*. Although the gene body of *ERF19* does not display gene body accessibility in any treatment, it is known that overexpression of *ERF19* leads to delayed senescence and improves drought tolerance (Scarpeci et al., 2017). The last three dACRs lie in the promoters of *HSP17.4* [AT3G46230], *HSP23.6* [AT4G25200], and *HSP17.7* [AT5G12030] (also annotated as *HSP17.6A*). *HSP17.4* responds to heat stress and plays an important role in desiccation tolerance in seeds, and the expression of *HSP17.4*, *HSP23.6-MITO*, and *HSP17.7* have been shown to peak one hour into a heat shock, and nine hours into a recovery period (Wehmeyer and Vierling, 2000; Swindell et al., 2007).

A greater number of loci are more accessible in pretreated plants than non-pretreated plants after a severe heat stress. There were nine times as many ACRs with increased

accessibility in the SAT treatment (176 dACRs) as compared to the LT (18 dACRs), while the number of dACRs between LAT (334 dACRs) and LT (272 dACRs) was much more even (**Figure 4.3A**). Motif enrichment analysis of the SAT dACRs and the LAT ACRs revealed an enrichment for the *HSF* family motif (**Supplemental Data 4.2**). Neither of the LT dACR sets had motif enrichments. As expected, Gene Ontology (GO) term enrichment analysis of genes near each set of dACRs showed an enrichment for genes associated with abiotic stress response (**Supplemental Data 4.2**). Looking at the regional distribution of the dACRs, a higher percentage of the SAT dACRs and the LAT dACRs resided in intergenic regions while a higher percentage of the LT dACRs were in promoter regions (**Figure 4.3B**). Two genes whose gene bodies were more accessible in SAT and LAT compared to LT, were *Bromodomain Transcription Factor* [AT1G31240] and *RING/U-Box Super Family Protein* [AT1G62370] (**Figure 4.3C**). Neither of these genes have known roles in HSR, but genes from their respective gene families play roles in gene regulation and abiotic stress response. Bromodomain proteins are known to play roles in chromatin remodeling and recruiting other transcription factors for transcription (Josling et al., 2012). RING/U-box proteins have been shown to be upregulated during heat stress, with one interacting with *HSPs* and promoting protein degradation (Yan et al., 2003; Hatakeyama and Nakayama, 2003; Kilian et al., 2007; Yee and Goring, 2009).

Most large differences in accessibility between SAT and LAT occur outside of genes with no gene being accessible in one treatment but not the other. Comparing SAT and LAT, more of the dACRs identified had higher accessibility in SAT (163 vs 133) (**Figure 4.3A**). These accessibility differences largely reside in intergenic and promoter regions with a higher fraction of SAT dACRs being in promoters and a higher fraction of LAT dACRs being in intergenic regions (**Figure 4.3B**). Although the SAT dACRs had no motif enrichments, LAT dACRs had an enrichment for *DNA BINDING WITH ONE FINGER (Dof)* family motifs, and a motif bound by B-BOX ZINC FINGER FAMILY PROTEIN (BBX31 [AT3G21890]) (**Supplemental Data 4.2**). *Dof*

family motifs are known to regulate phytochromes and BBX31 is a known repressor of photomorphogenesis (Park et al., 2003; Ward et al., 2005; Heng et al., 2019). GO term enrichment analysis of the SAT dACRs showed an enrichment for genes associated with cell-cell junctions while the LAT dACRs showed an enrichment for genes associated with transcription regulation and DNA binding (**Supplemental Data 4.2**). Some of the SAT dACRs were regions that were accessible in CN, PT, and RC, lose some accessibility in SAT, and become completely inaccessible in LAT, and LT. An example of this is seen with the gene body and promoter of *PHOTOSYSTEM II LIGHT HARVESTING COMPLEX GENE B1B2* (*LHB1B2* [AT2G34420]) (**Figure 4.3C**). Although LHB proteins have been shown to play a role in thermotolerance via signaling and thermosensing, loci like this seem to become inaccessible specifically during severe heat stress (Xu et al., 2012a; Dickinson et al., 2018). Most heat accessible genes are more accessible in LAT than SAT, but only a small number of genes had more than a twofold increase in accessibility. One gene with a twofold increase in accessibility was *DEFENSIN-LIKE PROTEIN 218* (*DEFL218* [AT5G18403]) (**Figure 4.3C**). Although there are over 300 *DEFLs* in Arabidopsis, little is actually known about their function in relation to abiotic stress response (Silverstein et al., 2005; Tesfaye et al., 2013).

Most differences between PT and SAT/LAT seem to reside in the promoter regions, with a greater number of promoters being more accessible in PT. There were nearly twice as many ACRs with more accessibility in PT (107 dACRs) as compared to the SAT (57 dACRs) while the number of dACRs between PT (311 dACRs) and LAT (218 dACRs) were somewhat closer (**Supplemental Figure 4.5A**). In both pairwise comparisons, a higher percentage of PT dACRs were located in promoter regions, while more SAT dACRs were within coding regions and more LAT dACRs were within intergenic regions (**Supplemental Figure 4.5B**). Although the PT dACRs identified in the PT vs. SAT comparison had no motif enrichments, the PT dACRs from the PT vs. LAT comparison had an enrichment for *AP2/ERF* motifs. The SAT dACRs had no

motif enrichments. but the LAT dACRs showed an enrichment for *Dof* transcription factor motifs and *AT-HOOK MOTIF NUCLEAR LOCALIZING* (*AHL)* transcription factor motifs (**Supplemental Table 4.2**). *AHL* transcription factors are known to repress hypocotyl elongation in plants grown in light via phytochrome suppression (Weigel et al., 2000; Street et al., 2007; Zhao et al., 2013; Favero et al., 2016). Genes near PT dACRs from both comparisons had an enrichment for GO terms associated with photosynthesis and light response. Genes near SAT dACRs and LAT dACRs had an enrichment for GO terms associated with heat stress response and protein folding.

The PT vs. SAT/LAT comparisons revealed a small number of heat accessible genes that were only accessible in SAT and LAT. Two of these genes, *MYO-INOSITOL-1-PHOSPHATE SYNTHASE 2* (*MIPS2* [AT2G22240]) and *ASCORBATE PEROXIDASE 2* (*APX2* [AT3G09640]) are targets for *HSFA2* and play roles in acquired thermotolerance (Panchuk et al., 2002; Nishizawa et al., 2006; Charng et al., 2007; Shahnejat-Bushehri et al., 2012; Lämke et al., 2016; Liu et al., 2018). Other examples of the genes with gene body accessibility only in SAT and LAT are *ZINC FINGER 3* (*ZF3* [AT5G43170]) and *NFU1 IRON-SULFUR CLUSTER PROTEIN* [AT5G07330] (**Supplemental Figure 4.5C**). Neither of these genes are none to play roles in thermotolerance. *ZF3* encodes a transcriptional repressor which downregulates the transactivation of other transcription factors. *ZF3* is lowly expressed in response to abiotic stress and is specifically expressed in Arabidopsis roots (Sakamoto et al., 2004). Iron-sulfur cluster proteins are known to localize to chloroplast, and in some species, they interact with *HSP70* proteins (Dorn et al., 2010; Przybyla-Toscano et al., 2018).

The PT vs. SAT/LAT comparison also revealed a set of genes that are only heat accessible during a severe heat stress. Examples of such genes were *RAB GTPASE HOMELOG A2B* (*RABA2B* [AT1G07410]) and *GLUTATHIONE S-TRANSFERASE TAU 1* (*GSTU1* [AT2G29490]) (**Supplemental Figure 4.5C**). Both genes were among the previously

identified 61 heat accessible genes, but their gene bodies do not become accessible after a mild heat stress. This gene body inaccessibility seems to be limited to these genes and not the entire region as adjacent genes, *HSP17.8* [AT1G07400] and *HSP17.6B* [AT2G29500], do display gene body accessibility in all heat stress conditions. The precise cause of this accessibility pattern in *RABA2B* and *GSTU1* is unclear. For *GSTU1*, its gene body inaccessibility during a mild stress may be a consequence of HSFA2 regulation and the role HSFA2 plays in *GSTU1* expression. In *hsfa2* knockout mutant cells, *GSTU1* has been shown to be overexpressed under control conditions, and has increased expression in response to stress (Zupanska et al., 2019). In wildtype plants, *GSTU1* expression is downregulated during a mild heat stress, while *GSTU1* expression is upregulated during a severe heat stress (Sullivan et al., 2014b; Albihlal et al., 2018). For *RABA2B*, there must be another explanation for its PT gene body inaccessibility since *RABA2B* has increased total RNA levels after any heat stress (Sullivan et al., 2014b; Albihlal et al., 2018).

The PT vs. LT comparison showed that a greater number of loci are more accessible in PT compared to LT. There were nearly five times as many ACRs with higher accessibility in PT (262 dACRS) than in LT (51 dACRs) (**Supplemental Figure 4.5A**). PT dARCs had motif enrichments for *HSF* motifs, and *APETALA 2/ ETHYLENE RESPONSE FACTOR* (*AP2/ERF*) motifs (**Supplemental Table 4.2**). Like the *HSFs, AP2/ERF* transcription factors are responsive to multiple environmental signals including heat stress (Feng et al., 2005). Genes near PT dACRs showed an enrichment for GO terms associated with various abiotic stresses (**Supplemental Table 4.2**). LT dACRs had no motif enrichments and genes near LT dACRs did not show any GO term enrichments. Both sets of dACRs primarily landed in intergenic and promoter regions, but PT had more dACRs within genes (**Supplemental Figure 4.5B**). An example of a gene with a PT dACR is *EPIDERMAL PATTERNING FACTOR 2* (*EPF2* [AT1G34245]) (**Supplemental Figure 4.5C**). *EPF2* encodes a secretory peptide that controls

asymmetric cell division in stomatal development and epidermal cell density, but is not known to play a role in HSR or acquired thermotolerance (Hara et al., 2009; Hunt and Gray, 2009). It is important to note that this ACR is present in CN and RC. This shows that PT dACRs are composed of not only loci that are heat accessible, but also loci that are able to remain accessible despite the heat stress.

Taken together, these pairwise comparisons revealed five things. First, most heat induced accessibility changes are transient, with only a small number of loci serving as temporary accessibility "markers" of prior heat stress. These loci are upstream of the genes *GolS1*, *ERF19*, *HSP17.4*, *HSP23.6-MITO*, and *HSP17.7.* Second, a greater number of loci either remain or become accessible during non-lethal heat stresses compared to lethal ones. Third, recently pretreated plants have more promoter accessibility in response to severe stress compared to plants that had longer recovery periods. Fourth, a greater number of loci are accessible in PT than in either SAT or LAT, with promoter regions showing this trend most strongly. Fifth, heat accessible intergenic regions seem to be a hallmark of long-term thermomemory.

H2A.Z is evicted from HSR genes after a temperature increase

One possible explanation for the gene body accessibility changes observed after heat stress is the modification of histones. A number of studies have linked H2A.Z to abiotic stress stress response in Arabidopsis (Smith et al., 2010; Kumar and Wigge, 2010; Coleman-Derr and Zilberman, 2012; Cortijo et al., 2017; Sura et al., 2017; Nguyen and Cheong, 2018). H2A.Z has

also been described to play a role in transcriptional memory in yeast (Guillemette et al., 2005; Brickner, 2009; Light et al., 2010). Reexamining ChIP-seq data generated by Cortijo *et al.* 2007 showed that H2A.Z is quickly evicted from the gene bodies of 22 heat accessible genes in response to increasing temperatures (**Supplemental Figure 4.6A, Supplemental Figure 4.6B**). These 22 genes were the only genes in the ChIP data that overlapped with the 61 heat accessible genes found by Sullivan *et al.* 2014. This eviction of H2A.Z within 15 minutes of a temperature increase seems to be specific to these 22 HSR genes and H2A.Z is incorporated back into these genes within an hour after the initial temperature increase. Four hours after a temperature increase, all gene sets have increased levels of H2A.Z downstream of the TSS, and decreased levels of H2A.Z upstream of the TSS. To study the role this H2A.Z incorporation may have with acquired thermotolerance, we compared the phenotypes of wild type Col-0 seedlings to that of the H2A.Z incorporation mutant *ACTIN RELATED PROTEIN 6* (*arp6)* under acquired thermotolerance conditions.

The arp6 mutant has diminished acquired thermotolerance 36 hours after pretreatment

To measure the effectiveness of acquired thermotolerance in the *arp6* mutant, seven-day-old *arp6* mutants were exposed to a pretreatment at 38 °C for 90 minutes, allowed to recover at 22 °C for either 2 hours, 8 hours, 24 hours, 36 hours, and then subjected to a 45 °C heat stress for 90 minutes. The number of leaves the *arp6* mutant had were measured five days after the 45 °C heat stress. For controls, the number of leaves *arp6* possessed was measured at 22 °C, 5 days after a 90 minute pretreatment at 38 °C, and 5 days after a 90 minute lethal treatment at 45 °C (**Figure 4.4A**). For comparison, the number of leaves were measured for Col-0, and the *HSP101* loss of function mutant, *hot1-3,* under the same conditions (Hong and Vierling, 2001). In working with *hot1-3*, it was found that the mutant was tetraploid. DAPI

staining of chromosomes in *hot1-3* revealed approximately 20 chromosomes (**Supplemental Figure 4.7A**). Col-0 and *arp-6* each have 10 chromosomes, as expected. Pollen area and seed size were also typically larger for *hot1-3*, compared to Col-0 (**Supplemental Figure 4.7B, Supplemental Figure 4.7C**). Despite its ploidy, *hot1-3* was still useful as a phenotypic control for the absence of acquired thermotolerance.

Although *arp6* generally grows fewer leaves than Col-0, the largest disparity in growth can be seen when both genotypes are given a 36 hour recovery period between a mild heat stress and a severe heat stress (**Figure 4.4B, Figure 4.4C**). At 22 °C, *arp6*, on average, has one less leaf than Col-0. So in the absence of heat stress, *arp6*, on average, has approximately a 13% decrease in leaf number in comparison to Col-0. This 13% decrease in leaf number is seen after a mild heat stress, and during a test of acquired thermotolerance with either 2 hours or 8 hours of recovery between heat stresses. The acquired thermotolerance of *arp6* begins to be less effective after 24 hours. The *arp6* mutant has an 18% decrease in leaf number with a 24 hour recovery period between heat stresses, and has a 39% decrease in leaf number with a 36 hour recovery. For comparison, *hot1-3,* has a 6%-10% decrease in leaf number under control conditions and after a mild heat stress. With a shorter recovery period of 2 hours and 8 hours, *hot1-3* has approximately a 25% decrease in leaf growth. With a longer recovery period of 24 hours and 36 hours, *hot1-3* has a 40%-45% decrease in leaf growth (**Figure 4.4D**). These results show that *arp6* begins to lose acquired thermotolerance after 24 hours and after 36 hours its thermotolerance is comparable to *hot1-3*.

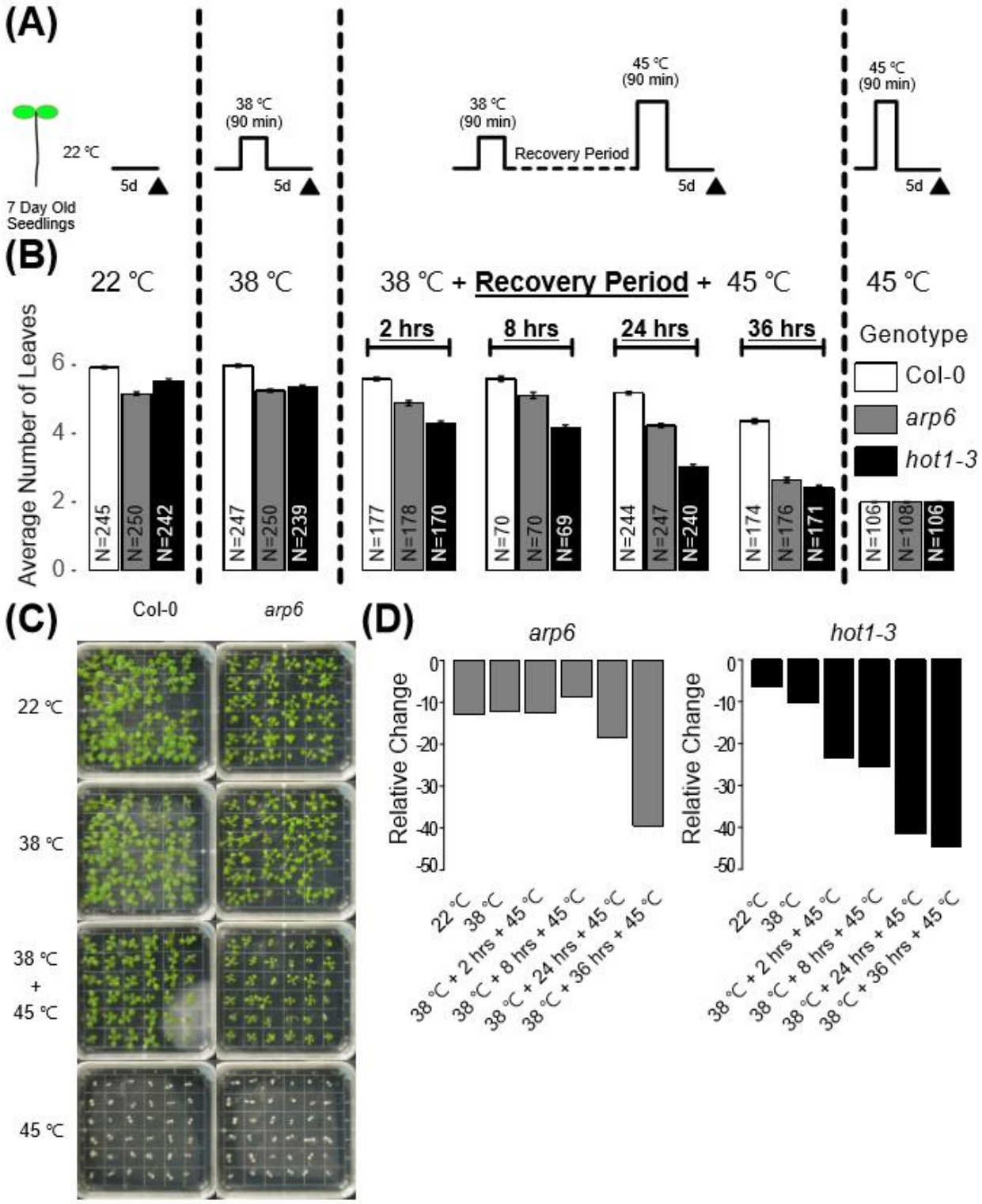


Figure 4.4. *arp6* has defective long-term acquired thermotolerance.

(A) Experimental design of leaf counting experiments to assay plant growth after various heat shock treatments. Leaves were counted five days after their respective treatments. **(B)** Barplot of the number of leaves in Col-0, *arp6*, and *hot1-3* after various heat shock treatments. **(C)** Photos of Col-0 and arp6 seedlings after various heat shock treatments. **(D)** Relative change of the number of leaves in *arp6* and *hot1-3* compared to Col-0 [(Genotype - Col-0)/Col-0].

To study the role chromatin accessibility may play in the shortened acquired thermotolerance of *arp6*, the same ATAC-seq experiments that were performed in Col-0 were performed for *arp6*. Like the Col-0 samples, the *arp6* samples had a similar range of fraction of reads in ACRs and treatment replicates were highly correlated (**Supplemental Table 4.3, Supplemental Table 4.4**).

The arp6 mutant has slightly lower gene body accessibility than Col-0 after LAT in key HSR genes

As a control, the accessibility difference between *arp6* and Col-0 was examined under CN conditions. In the process of identifying these accessibility differences, it was discovered that *arp6* has a number of large deletions in its genome relative to Col-0, resulting in the loss of five percent of the genome (**Supplemental Data 4.3**). This difference is comparable to the genomic differences between Arabidopsis ecotypes. Outside these large scale deletions, there were only 17 dACRs between Col-0 and *arp6* under control conditions (**Supplemental Figure 4.8A**). Most of these differences resided in intergenic regions (**Supplemental Figure 4.8B**). Looking at other treatments, *arp6* did exhibit gene body accessibility in PT, SAT, LAT, and LT. There were no significant differences, however, between *arp6* and Col-0 in PT, RC, and SAT that were not seen in CN. Therefore, accessibility differences at these stages of acquired thermotolerance do not explain the defect *arp6* has at maintaining acquired thermotolerance.

Consistent with this, the correlation between *arp6* accessibility and Col-0 accessibility for each treatment was quite high for all treatments except LAT (Pearson correlation ~0.95). The LAT correlation between the Col-0 and *arp6* was slightly lower (Pearson correlation = 0.87).

In the LAT, however, there were accessibility differences between *arp6* and Col-0. Generally, *arp6* has lower gene body accessibility than Col-0 at key acquired thermotolerance genes such as *HSP101* and *HSFA2* after LAT (**Figure 4.5A**). Most of these genic accessibility differences are not significant, but there are a small number of significant accessibility differences between *arp6* and Col-0 after LAT (**Figure 4.5B**). Most of these dACRs reside in intergenic and promoter regions, with seven dACRs being more accessible in *arp6* and 13 dACRs being more accessible in Col-0 (**Figure 4.5C**). Both sets of dACRs did not have any significant motif enrichments. Two Col-0 dACRs were in the genes *HSP70-3* [AT3G09440] and *HSP17.6A* [AT1G59860] (**Figure 4.5D**). As *HSP*s, these two genes play roles in stress response but their role in thermomemory is unclear. Furthermore, it appears that their accessibility during LAT is dependent on efficient HSA.Z incorporation into the nucleosomes.

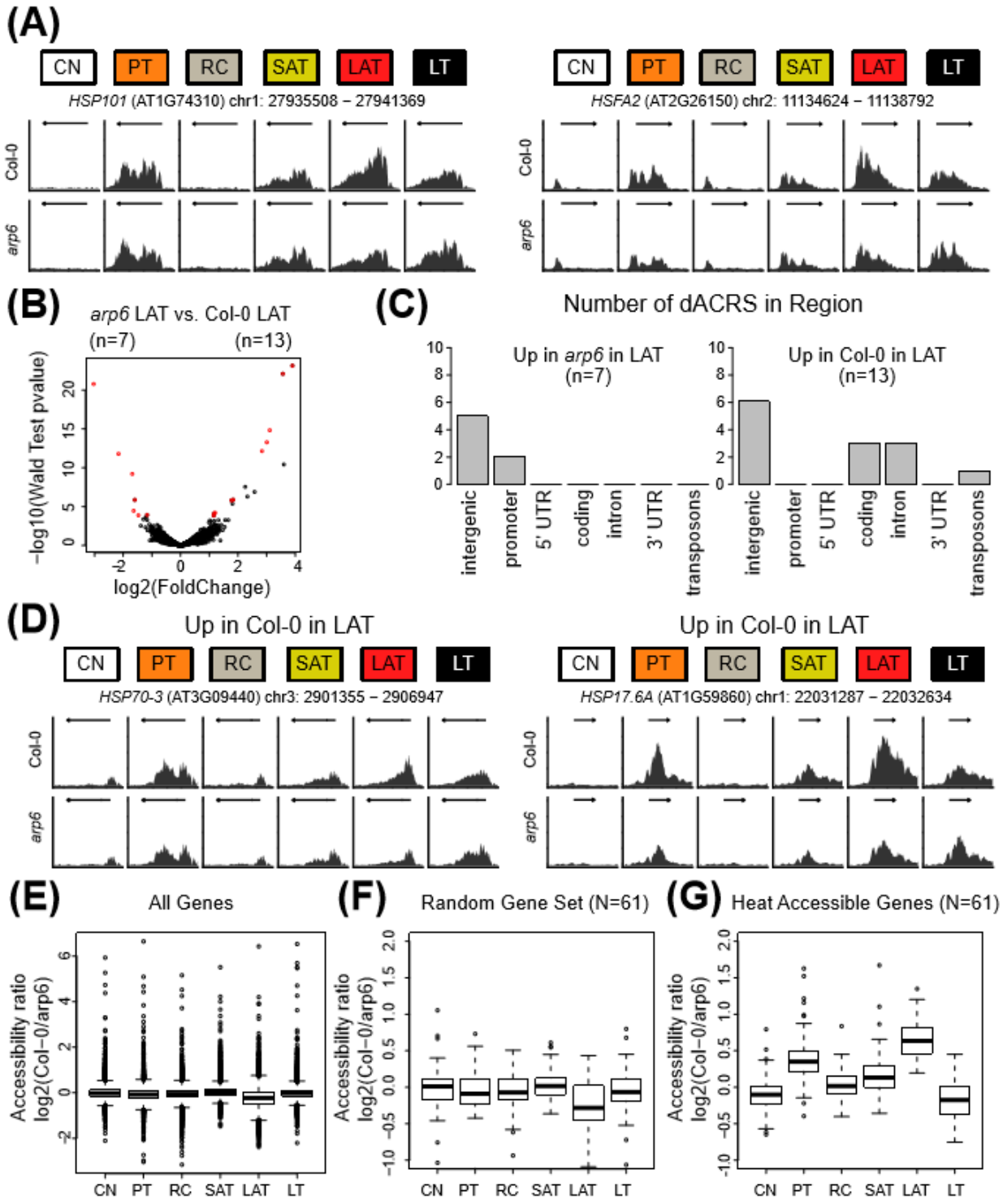


Figure 4.5. There are few significant accessibility differences between Col-0 and *arp6* in response to heat stress.

(A) Accessibility of HSP101 and HSFA2 in Col-0 and *arp6*. **(B)** Volcano plot and **(C)** regional distribution of dACRS between genotypes in LAT. **(D)** Visualization of dACRs located in *HSP70-3* and *HSP17.6A*. Log2 fold change of gene body accessibility between Col-0 and *arp6* for **(E)** all genes, **(F)** a random set of genes, and **(G)** heat accessible genes across treatments.

We then sought to confirm that the differences in apparent accessibility at LAT between arp6 and Col-0 were not due to random technical differences in the preparation of the LAT samples. First, we plotted the ratio of gene body accessibility in Col-0 and gene body accessibility in *arp6* was calculated for all genes for each treatment (**Figure 4.5E**). On average, when considering the remaining genes, the gene body accessibility seen in Col-0 and *arp6* are about the same in all treatments except LAT, in which the ratio drops slightly. This trend can be seen in a random set of 61 genes (**Figure 4.5F**). This ratio drop seems to be driven by differences in genes with relatively low accessibility (<30 CPM in *arp6*). Next, the gene body accessibility ratio was examined for the heat accessible genes. Notably, the ratio increases substantially in PT and LAT, indicating that at key HSR genes, Col-0 has slightly higher gene body accessibility than *arp6* (**Figure 4.5G**). This suggests that our observations of higher accessibility in Col-0 than *arp6* at LAT are biologically accurate and not experimental artifacts.

Since one of the Col-0 dACRs was located in the small *HSP* gene, *HSP17.6A*, this result prompted an examination of the accessibility of a number of small HSPs and genes targeted by HSFA2 for increased H3K4 methylation (Lämke et al., 2016). These genes were *HSP18.2* [AT5G59720], *HSP21* [AT4G27670], *HSP22.0* [AT4G10250], *HSA32* [AT4G21320], *MIPS2* and *APX2* (**Supplemental Figure 4.8D**). After LAT, as with *HSP101* and *HSFA2*, *arp6* exhibited slightly lower accessibility in comparison to Col-0 at these HSFA2 target genes with the exception of *HSP21* and *HSA32*. *HSP21* has been shown to play a role in the maintenance of thermomemory, but a difference in accessibility at this gene does not seem to explain the

acquired thermotolerance defect of *arp6* (Sedaghatmehr et al., 2016). Interestingly, *HSA32* does not exhibit the accessible gene body phenomenon despite the fact that *HSA32* has been shown to be essential for long term acquired thermotolerance and the fact that it is expressed in response to heat stress (Charng et al., 2006).

In summary, although there were few large differences in accessibility between Col-0 and *arp6*, the thermomemory defect observed in *arp6* seemed to be reflected in the lower gene body accessibility of key HSR genes in *arp6* compared to Col-0 at the LAT condition. This would suggest higher gene body accessibility in key HSR genes is indicative of more effective thermomemory.

To test if the long-term acquired thermotolerance defect in *arp6* was caused by lack of protein or an accelerated HSP degradation rate, western blots for large HSPs such as HSP101, HSP90, and HSP70 were performed on *arp6* plants up to 72 hours after a pretreatment. In comparison to Col-0, HSP degradation occurred at a similar rate in *arp6*, with both genotypes slowly losing HSPs over time after a mild heat stress (**data not shown**). In addition pretreated *arp6*, like Col-0, was able to generate HSPs after experiencing a severe heat stress.

Acquired thermotolerance reduces the number of interactions with the chloroplast genome in response to severe heat stress

To explore the role chromosome conformation plays in the HSR and in acquired thermotolerance, Hi-C data was collected from Col-0 and *arp6* in the following treatments: CN, PT, SAT, and LT. Differential Hi-C interaction analysis between CN and LT using a subtraction matrix approach revealed that heat shock does not greatly change chromosome conformation (**Figure 4.6A**). Additional differential Hi-C interaction analysis using ACCOST revealed a small number of differential interactions between CN and LT however these differential interactions

were typically between genomic bins that possessed low mappability (<30%) and mapped to centromeric or telomeric regions (Cook et al., 2020). Similar results were observed when comparing chromosome conformation in CN to PT and to SAT. Like Col-0, *arp6* does not display any major changes in chromosome conformation in response to heat stress.

Although most the frequency of most interactions do not change significantly upon heat stress, the number of interactions observed between the nuclear genome and the chloroplast genome increases greatly in LT compared to CN for both Col-0 and *arp6* (**Figure 4.6B**). Because Hi-C is intended to be done on intact nuclei, this result indicates rupture of the nuclear (and chloroplast) membranes during the LT condition. This drastic increase in chloroplast interactions in the Hi-C data is not observed in PT or in SAT. In the Hi-C data, the chloroplast genome interacts with a large number of loci in the nuclear genome in nonlethal conditions, but over 95% of bins do have a chloroplast interaction in LT (**Figure 4.6C**). These results show that severe heat stress can cause a sharp increase in chloroplast interactions, but acquired thermotolerance helps to prevent this large increase. These chloroplast interactions are spread out across the nuclear genome with more bins coming in contact with the chloroplast genome in LT.

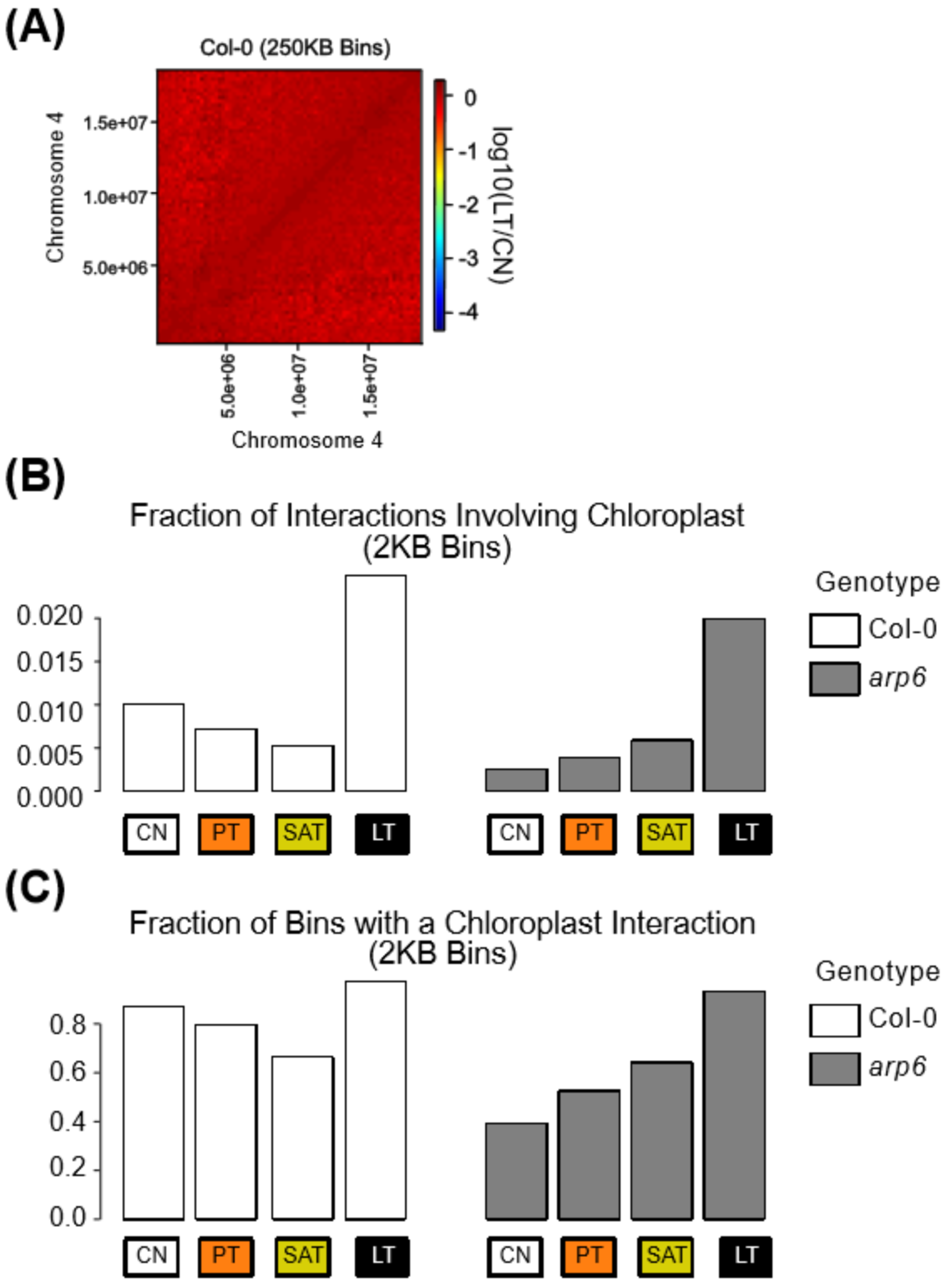


Figure 4.6. The number of interactions involving chloroplast increases sharply in a lethal heat stress.

(A) Subtraction matrix comparing LT interactions to CN interactions in Chromosome 4. Barplots of **(B)** the fraction of interactions involving chloroplast and **(C)** the fraction of bins with a chloroplast interaction.

4.4 Discussion

Chromatin accessibility undergoes distinct changes at various stages of acquired thermotolerance

As a follow up to Sullivan *et al.* 2014, which found that lethal heat stress triggered a drastic increase of gene body accessibility at known HSR genes, we explored chromatin accessibility during several stages of acquired thermotolerance. Here we showed that the gene body accessibility phenomenon occurs during any heat stress and that the increased accessibility is temporary. This increased accessibility peaks within 40-60 minutes of a heat stress and is lost within 2 hours during a recovery period. A mild heat stress at 38 ℃ triggers more gene body accessibility than a severe heat stress at 45 ℃, and genes are typically most accessible when assaying for long term acquired thermotolerance. Using *HSP101* and *HSFA2* as examples, we show that accessibility and expression are highly correlated. Previous work examining stress memory in *Arabidopsis* through the lens of drought stress and gene expression, showed that plants that were repeatedly stressed showed memory of these prior stresses via an increased rate of transcription and elevated transcript levels of a subset of stress response genes (Ding et al., 2012). Indeed, the increased gene body accessibility seen during long term acquired thermotolerance also coincides with increased levels of nascent RNA transcription for many of the genes with heat induced gene body accessibility, suggesting that that the gene body accessibility serves as part of acquired thermotolerance (Liu et al., 2020).

Although most of the previously identified heat accessible genes were accessible in any heat stress, we presented examples of genes that broke this general trend (i.e. *HSP90-3*, and

HSFA7A) as well as additional *HSPs* that become accessible during non-lethal heat stresses (i.e. *HSP70-6, HSP70-11,* and *HSP21*). Notably, many *HSPs* and *HSFs* did not display gene body accessibility in any condition. This may be a consequence of either redundancy or specificity in function. For plant *HSF* genes, although there are distinct classes of *HSFs,* many of them function redundantly while others are specifically upregulated in response to distinct abiotic stresses (Nover et al., 2001; Scharf et al., 2012). Similarly within the *HSP90* and the *HSP70* gene families, there are distinctions in intracellular localization as well as tissue specificity, but protein similarity within a family is often quite high (> 90% similarity in some cases) with many *HSP*s having redundant function (Yabe et al., 1994; Milioni and Hatzopoulos, 1997; Lin et al., 2001; Sung et al., 2001; Xu et al., 2012b; Leng et al., 2017).

Comparing different stages of acquired thermotolerance revealed a number of things. (1) Most heat induced accessibility is temporary but there are a small number of promoter regions that retain their accessibility during a recovery period and may potentially “mark” a plant as pretreated. Although there are only five of these regions, they do have an enrichment for *HSF* motifs. (2) Generally, a greater number of regions are more accessible during non-lethal heat stresses than during a lethal one. As expected many of these regions are promoter regions with motifs for transcription factors that play roles in stress response or reside in stress response genes. (3) In response to heat stress, a greater number of promoter regions are more accessible in recently pretreated plants compared to plants with longer recoveries. (4) Although acquired thermotolerance protects from lethality, a severe heat stress still causes many loci to become inaccessible, particularly promoter regions. (5) Heat accessible intergenic regions seem to be a hallmark of thermomemory. The exact function of these intergenic regions is unclear, but they may serve as good candidates in STARR-seq experiments to assay for potential enhancer activity. As to what drives the accessibility of these regions, ChIP-seq experiments assaying for

the binding of *Dof* transcription factors may prove insightful since *Dof* family motifs were enriched in regions more accessible during the LAT treatment.

These treatment comparisons revealed additional genes that displayed this unusual phenomenon of heat induced gene-body accessibility outside of the previously identified heat accessible genes. Perhaps the most important of these genes in terms of acquired thermotolerance were a small set of genes that had heat induced gene body accessibility in pretreated plants only (i.e. *MIPS2, APX2, ZF3,* and *NFU1 IRON-SULFUR CLUSTER PROTEIN*). Further study is needed to characterize the roles *ZF3* and *NFU1 IRON-SULFUR CLUSTER PROTEIN* play in thermomemory and acquired thermotolerance.

Efficient H2A.Z incorporation is needed for effective thermomemory

We showed that the H2A.Z incorporation mutant *arp6* has a defect in thermomemory, suggesting that H2A.Z plays a role in acquired thermotolerance. H2A.Z is evicted from the promoters and gene bodies of heat accessible genes in response to increasing temperatures. Several hours after the initial temperature increase, H2A.Z is incorporated back into these promoters and gene bodies, resulting in a net gain in H2A.Z compared to H2A.Z levels at 22 ℃. This is consistent with work in yeast showing that H2A.Z is incorporated into recently transcribed genes and is necessary for the rapid reinduction of a given gene's expression (Brickner et al., 2007). As discussed previously, heat accessible genes have higher levels of nascent RNA transcription in pretreated plants. In terms of heat induced gene body accessibility, *arp6* had a decrease in the accessibility of heat accessible genes compared to Col-0 in LAT. This result would suggest that the extreme levels of gene body accessibility observed in LAT in Col-0 is caused by the eviction of increased levels of H2A.Z. The *arp6* mutant cannot obtain these extreme levels of gene body accessibility in LAT because there are

not as many H2A.Z containing nucleosomes to evict; *arp6* could not efficiently incorporate H2A.Z during its recovery period. Additional H2A.Z ChIP-seq experiments would be needed to confirm that levels of H2A.Z pre-heat stress directly correlate with the degree of heat induced gene body accessibility.

Heat stress did not change the chromosome conformation of the nuclear genome

We found that heat stress did not alter chromosome conformation of the nuclear genome in Arabidopsis. This result is in contrast with prior work in *Drosophila* which showed that heat stress altered Topologically Associated Domains (TADs) in the *Drosophila* genome (Li et al., 2015). Although some plant genomes have TADs, multiple Hi-C datasets including our own have shown that TADs are not a characteristic of the Arabidopsis genome (Feng et al., 2014; Grob et al., 2014; Wang et al., 2015; Liu et al., 2016, 2017; Concia et al., 2020). In addition, a subsequent study of chromosome conformation in *Drosophila* saw no change in chromosome conformation in response to heat stress (Ray et al., 2019). Prior work in mammalian cell lines have shown that the nuclear localization of the Hsp70 gene did change in response to heat stress, so it is reasonable to expect to observe heat related changes in chromosome conformation (Hu et al., 2010). Of the known examples where transcription has altered chromosome conformation, these transcriptional changes seem to affect TADs (van Steensel and Furlong, 2019). Since Arabidopsis does not contain TADs, Hi-C may not be the best approach to capture heat induced changes in nuclear organization.

Our data did capture interactions between the chloroplast genome and the nuclear genome. Contact between the chloroplast genome and the nuclear genome have been previously shown in plants (Stegemann et al., 2003; Huang et al., 2003; Lister et al., 2003; Martin, 2003). These are the interactions observed in the Hi-C data at 22 ℃. Heat stress can

cause damage to chloroplast causing the organelle to burst and for the plant leaves to lose color (Cullis et al., 2009; Sun and Guo, 2016; Hu et al., 2020). This chloroplast bursting would free additional fragments of the chloroplast genome to interact with the nuclear genome which is what we observe in our Hi-C data under lethal conditions. Non-lethal heat stress does not seem to cause this same degree of chloroplast bursting as the number of chloroplasts interactions is more comparable to conditions. Acquired thermotolerance most likely functions in part by preventing the chloroplast from bursting via recently translated HSR proteins serving as molecular chaperones for chloroplast proteins.

4.5 Methods

Plant Materials. A transgenic INTACT line harboring the UBQ10: NTF construct and ACT2: BirA construct was created for the Col-0 genetic background (Deal and Henikoff 2010; Deal and Henikoff 2011; Sullivan et al. 2014). An *arp6-1* T-DNA insertion mutant (SAIL_599_G03) was obtained from the ABRC stock center. This *arp6-1* mutant was then crossed with the Col-0 INTACT line to generate an *arp6-1* INTACT line. The *hot1-3* T-DNA insertion mutant (CS16284) was also obtained from the ABRC stock center**.**

Sample Preparation/ Growth Conditions. Seeds were surface sterilized and plated onto 1x MS + 1% sucrose. Plates were sealed with micropore tape, double wrapped with aluminum foil and stratified for 3 days at 4 ℃. Stratified plates were unwrapped and moved to Long Day conditions (16 hour light, 22 ℃; 8 h dark, 22 ℃) in a growth chamber (Conviron CMP5090, Controlled environment ltd. Winnipeg, Manitoba, Canada) and grown for 7 days.

Heat Stress Regime, Nuclei Extraction, and ATAC-seq. For each genotype, a plate of seven day old seedlings was subjected to one of the following treatments: Control [CN]: 22 ℃; Pretreatment [PT]: 90 minutes at 38 ℃; Recovery [RC]: 90 minutes at 38 ℃, followed by 2 hours at 22 ℃; Acquired Thermotolerance: 90 minutes at 38 ℃ followed by either 2 hours [SAT], 8

hours, 24 hours, or 36 hours [LAT], or 48 hours at 22 ℃ followed by 90 minutes at 45 ℃; Lethal [LT] 90 minutes at 45 ℃.

For ATAC-seq, two replicates for each genotype and treatment were prepared. Nuclei were extracted from 2 grams of whole seedling plant tissue using the INTACT method as described in the protocol available at: http://www.plantregulome.org/media/protocols/data_assay/dnasei_arabidopsis/FINAL_whole_seedling_DNaseI_protocol.pdf (Deal and Henikoff 2010).

After extracting nuclei, ATAC-seq was performed, as described in Buenrostro et al. 2013, except transposase treatment of INTACT-purified bead-bound nuclei was performed at 37 ℃ for 30 minutes. Paired end reads were sequenced on an Illumina Nextseq.

ATAC-seq analysis. Paired End Reads ATAC-seq reads were trimmed down to 36 bp sequences using FASTQ trimmer (FASTX-Toolkit: http://hannonlab.cshl.edu/fastx_toolkit/). These trimmed reads were then aligned to the Col-0 reference genome (TAIR10) using Bowtie 2 version 2.2.3 with default settings (Langmead and Salzberg 2012; Berardini et al. 2015). Read alignments with MAPQ score less than two were filtered out. ChrC reads, ChrM reads, centromeric regions from Clark et al. 2007 (Chr1: 13,698,788–15,897,560; Chr2: 2,450,003–5,500,000; Chr3: 11,298,763–14,289,014; Chr4: 1,800,002–5,150,000; Chr5: 10,999,996–13,332,770) and duplicate reads were also filtered out using SAMtools (Li et al. 2009).

For each genotype and treatment, one of the replicates was downsampled to seven million reads and accessibility summits were called using the callpeak command in MACS2 (Zhang et al. 2008) using the following parameters:

-f BAM --nomodel --shift -75 --extsize 150 --call-summits -g 1.3e08 -q 0.0001

Each accessibility summit was padded with 75 bp on both sides to form a set of 150 bp Accessible Chromatin Regions (ACRs). The union of all ACRs from each genotype and treatment was generated using the merge operation in BEDOPS (Neph et al. 2012), which also combines overlapping ACRs into one larger ACR (**Supplemental Table 4.1, Supplemental Table 4.2**). For each replicate, the number of reads overlapping each ACR was counted using the BEDOPS command: bedmap --echo --sum.

Due to large deletions in the *arp6* line, ACRs that were within these deleted regions were removed from analysis. In addition, for both Col-0 and *arp6*, the bottom quartile of ACRs, in terms of total raw read counts per kilobase across all samples, were removed from the analysis in order to remove small ACRs from the analysis.

Differential ACRs (dACRs) between *arp6* and Col-0, and between treatments in Col-0 were identified using DESeq2 (Love et al. 2014). For the DESeq2 analysis, the raw read counts were used. ACRs were considered differentially expressed if the baseMean counts were above 100, the absolute value of the log2 Fold Change was greater than 1, and the false discovery rate/adjusted p-value (FDR) was less than 0.1.

For visualization of ATAC-seq data via heatmaps, the raw read counts in each ACR (or gene) for each sample were first normalized for the ACR (or gene) length by dividing the raw read counts by the ACR (or gene) length in kilobases to get reads per kilobase (RPK). Next, these RPK values were then summed for each sample and the RPK value of each ACR (or gene) in each sample was divided by the total RPK value from their respective sample. Lastly, these quotients were multiplied by 1 million to get counts per million (CPM). Most accessibility heatmaps visualize these CPM values, with the exception of Figure 1B, which is row-scaled. For visualization of ATAC-seq data via density plots (i.e. Figure 1C, 1D, etc), the CPM values between two replicates were averaged. The y-axis ranges from 0 to 10 CPM. For the x-axis,

each “point” on the density curve represents a 20bp window. When visualizing ACRs (or genes), typically 1kb upstream and downstream of the loci of interest was visualized.

For relating the accessibility data to expression data, RNA-seq data was obtained from GEO accession: GSE85655. GRO-seq data was obtained from GEO accession: GSE128698. For relating the accessibility data to H2A.Z occupancy, ChIP-seq data was obtained from GEO accession: GSE79355.

For motif enrichment analysis, AME was used (McLeay and Bailey 2010). As a background, the filtered union ACRs were filtered for base pairs that overlapped with genes using the difference command in BEDOPS (Neph et al. 2012). If an ACR was entirely within a gene it was removed as well. Input ACRs also were filtered in the same manner (**Supplemental Table 4.4**). Sequences were retrieved using the BEDTools command get-fasta and the TAIR10 genome (Quinlan and Hall 2010; Lamesch et al. 2012; Berardini et al. 2015).

AME was run using the following parameters:

--scoring avg --method fisher --hit-lo-fraction 0.25 --evalue-report-threshold 10.0

Two motif databases were used, the Arabidopsis PBM motifs (ArabidopsisPBM_20140210.meme) and the Arabidopsis DAP motifs (ArabidopsisDAPv1.meme) (Franco-Zorrilla et al. 2014; O'Malley et al. 2016).

GO Term Enrichment analysis was performed with Panther using all the Arabidopsis genes as a background (http://pantherdb.org/) (Mi et al. 2019; Mi et al. 2019). For each ARC, the closest gene was identified using the BEDOPS command closest-features and the Araport annotation (Cheng et al. 2016). The complete annotation data sets for biological process, molecular function, and cellular components were used. Fisher’s exact test with false discovery rate correction was run.

Hi-C. Whole seedlings were crosslinked immediately after heatshock (1% Formaldehyde for 15min at room temperature, then quenched with 2.5M Glycine for 15min on ice and washed with ddH2O), flash-frozen in LN2 and stored at -80C until processing. Hi-C prep and library construction was performed by Phase Genomics, Inc (Seattle, WA) on 0.1- 0.2g of crosslinked material. The restriction enzyme MboI/DpnII (GATC) was used for chromatin digestion. Sequencing was done on the Illumina NextSeq platform.

Hi-C analysis. Hi-C Reads were analyzed using Hi-C Pro version 2.7.6. (Servant et al. 2015). Using Hi-C Pro, reads were aligned to the repeat masked Col-0 reference genome (TAIR10) with Bowtie 2 version 2.2.3 (Berardini et al. 2015; Langmead and Salzberg 2012) and the following alignment options:

FORMAT = phred 33

MIN_MAPQ = 10

BOWTIE2_GLOBAL_OPTIONS = --very-sensitive -L 30 --score-min L,-0.6,-0.2 --end-to-end --reorder

BOWTIE2_LOCAL_OPTIONS = --very-sensitive -L 20 --score-min L,-0.6,-0.2 --end-to-end --reorder

To account for differences in sequencing depth, Hi-C samples were downsampled to have similar numbers of total Hi-C reads (**Supplemental Table 4.5**). The Arabidopsis genome was split into 2kb, 10kb, 50kb, and 250kb bins for contact maps. Interaction matrices were ice normalized (Imakaev et al. 2012). Differential contact analysis was performed using a subtraction method approach and using ACCOST (Cook et al. 2020).

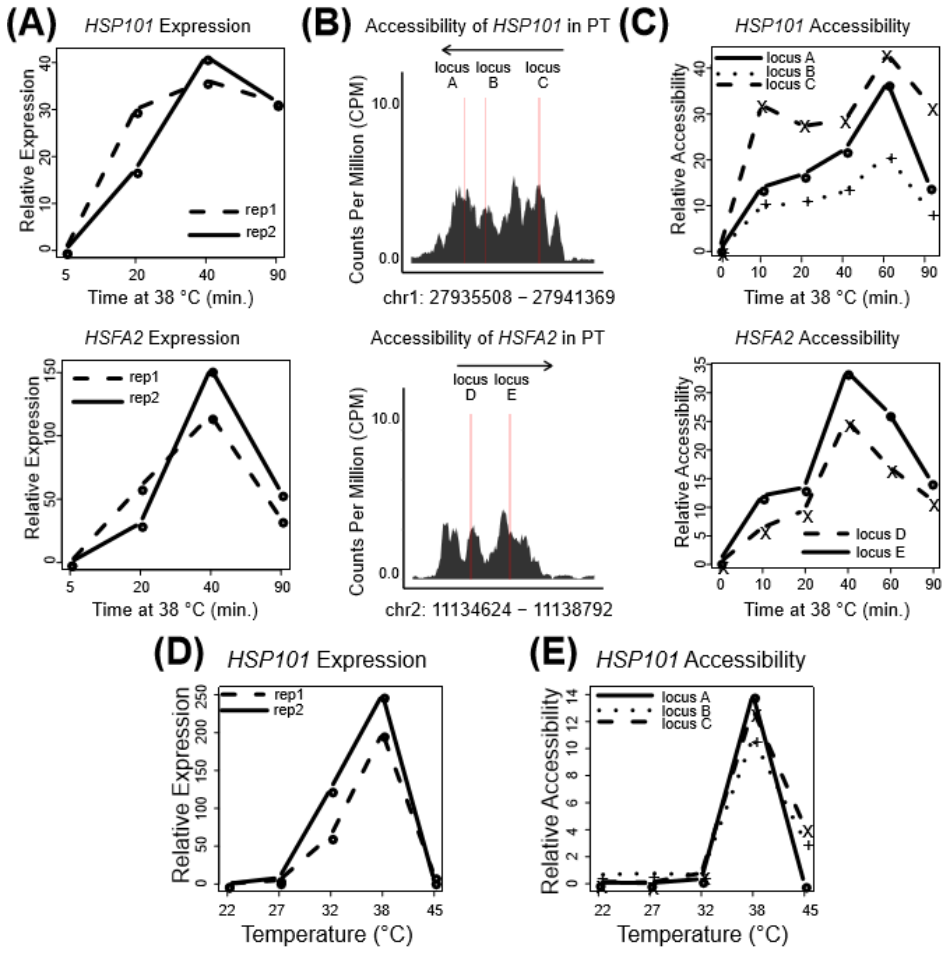
RT-qPCR, and RT-qPCR analysis. RNA was extracted from whole seedlings with Trizol (Invitrogen) and DNAse I-treated (Ambion DNase I). oligo-d(T) cDNA was generated from 0.5ug of total RNA using the Revert Aid First Strand cDNA synthesis kit (Thermo Scientific). Relative expression was determined by qPCR using UBC/AT1G14400 as the reference gene. All data

was normalized to *UBIQUITIN-CONJUGATING ENZYME E2 1* (*UBC* [AT1G14400]) from their respective treatment/ time point using the **ΔΔ**CT method. All primers used are listed in (**Supplemental Table 4.6**).

ATAC-qPCR, and ATAC-qPCR analysis. ATAC libraries were built as described for ATAC-seq, only instead of sequencing, fragment enrichment over regions of interest was surveyed by PCR.

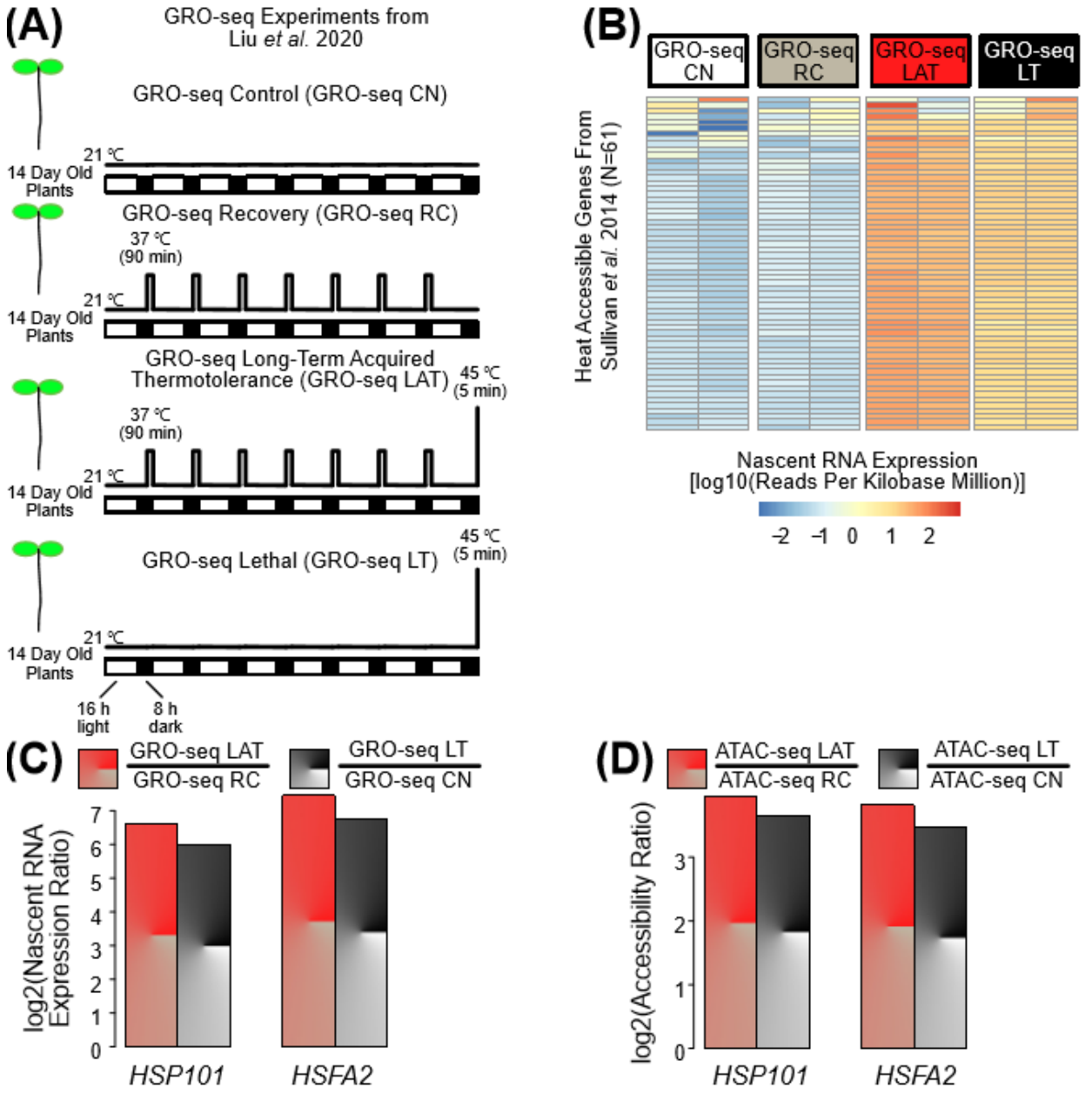
As with the ATAC-seq, plant nuclei were isolated using the INTACT method and treated with transposase. All data was normalized to the control region from their respective treatment/time point using the **ΔΔ**CT method. All primers used are listed in (**Supplemental Table 4.6**).

4.6 Supplementary Figures and Tables



Supplemental Figure 4.1. Expression and accessibility of HSP101 and HSFA2 correlate over time and across temperatures.

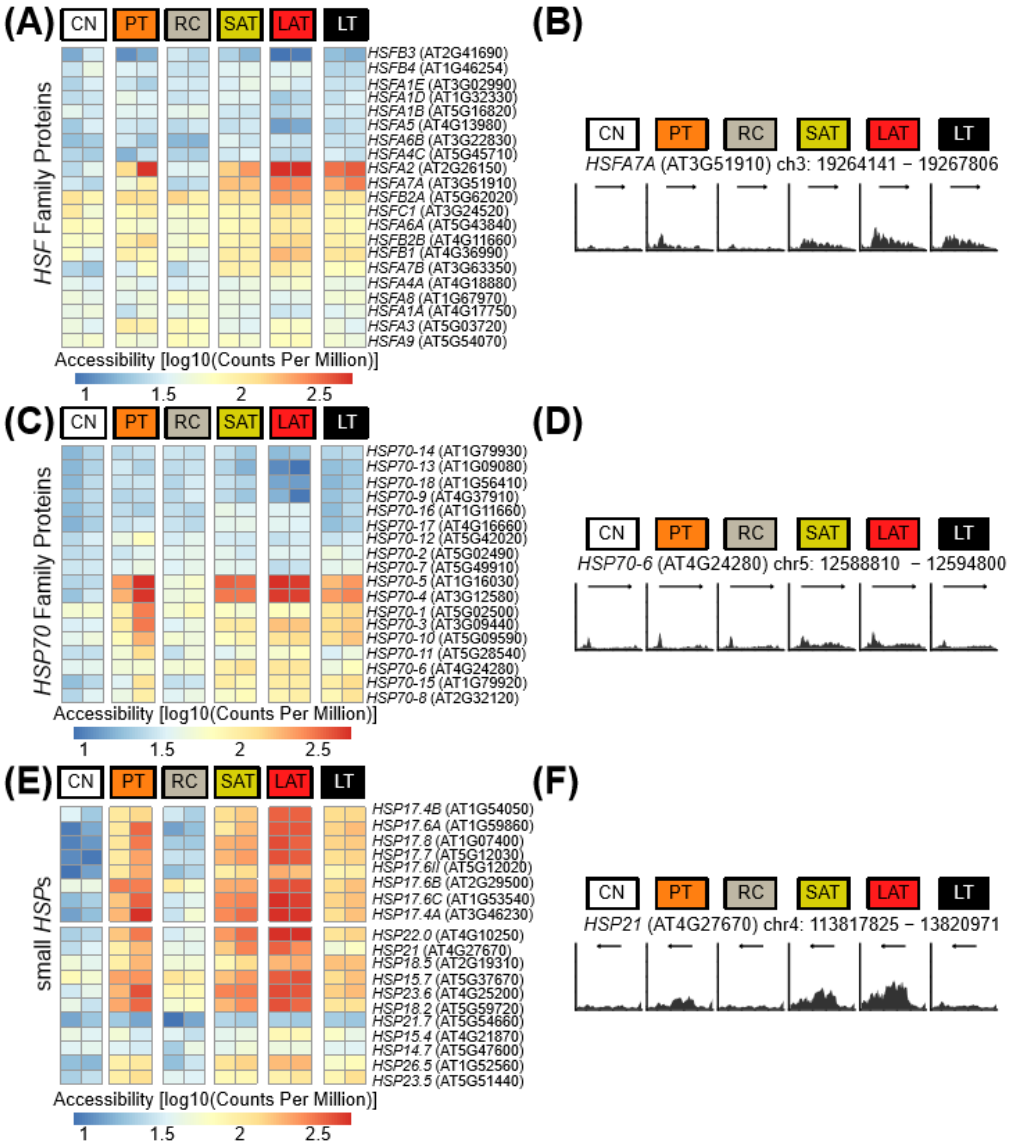
(A) Relative expression of HSP101 and HSFA2 in 38 ℃ over time. **(B)** Loci amplified for ATAC-qPCR**.** PT accessibility of HSP101 and HSFA2 from the ATAC-seq data is shown to display how accessible each loci becomes after 90 minutes: (HSP101: locus C > locus A > locus B; HSFA2: locus E > locus D). **(C)** Relative accessibility of HSP101 and HSFA2 in 38 ℃ over time (*See Methods: ATAC-qPCR, and ATAC-qPCR analysis*). **(D)** Relative expression of HSP101 vs. treatment temperature. **(E)** Relative accessibility of HSP101 vs. treatment temperature.



Supplemental Figure 4.2. Nascent RNA expression of heat accessible genes, in response to heat stress, is higher in pretreated plants.

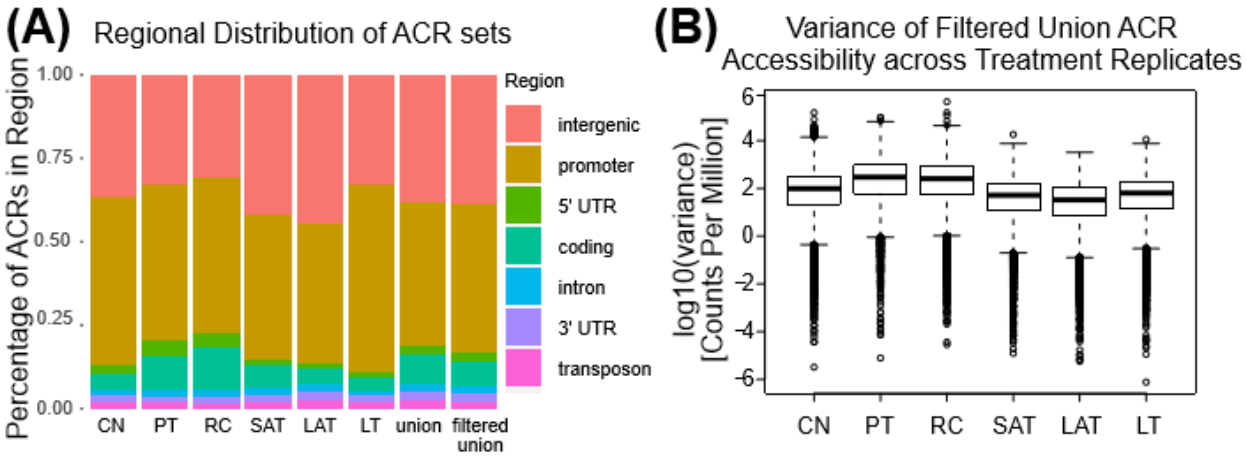
(A) Experimental design of GRO-seq experiments performed by Liu et al. 2020. Two week old Arabidopsis plants, grown at 21 °C under long day conditions, were pretreated at 37 °C for 90 minutes before dawn for seven days. A day after the last pretreatment, leaf tissue was collected in the dark for GRO-seq either at 21 °C (GRO-seq Recovery [GRO-seq RC]) or after a 5 minute

heat stress at 45 °C (GRO-seq Long-Term Acquired Thermotolerance [GRO-seq LAT]). As controls, leaf tissue was collected in the dark for GRO-seq in three week old plants either at 21 °C (GRO-seq Control [GRO-seq CN]) or after a 5 minute heat stress at 45 °C (GRO-seq Lethal [GRO-seq LT]). These three week old plants were not pretreated. **(B)** Heatmap (row scaled) of nascent RNA expression of heat accessible genes across GRO-seq treatment replicates. **(C)** Nascent RNA expression ratios (GRO-seq LAT/ GRO-seq RC; GRO-seq LT/ GRO-seq CN) for HSP101 and HSFA2. **(D)** Accessibility ratios (LAT/RC; LT/CN) for HSP101 and HSFA2.



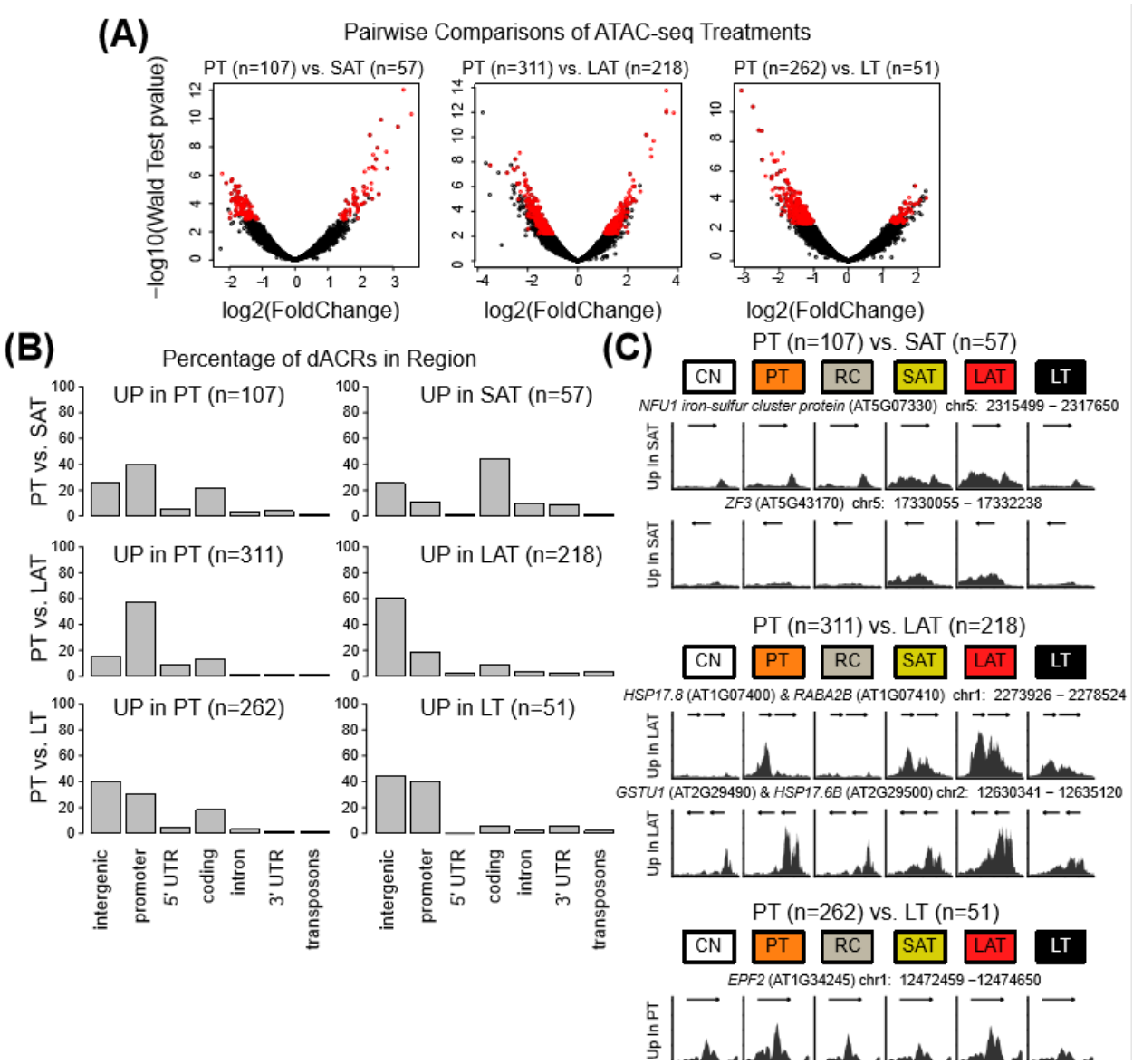
Supplemental Figure 4.3. With a few exceptions, HSPs and HSFs with heat induced accessibility are accessible in all heat stress conditions.

Accessibility heatmaps of **(A)** HSF family proteins, **(C)** HSP70 family proteins, and **(E)** small HSPs across treatment replicates. Accessibility of **(B)** HSFA7A, **(D)** HSP70-11, and **(F)** HSP21.



Supplemental Figure 4.4. Most ACRs reside in intergenic and promoter regions.

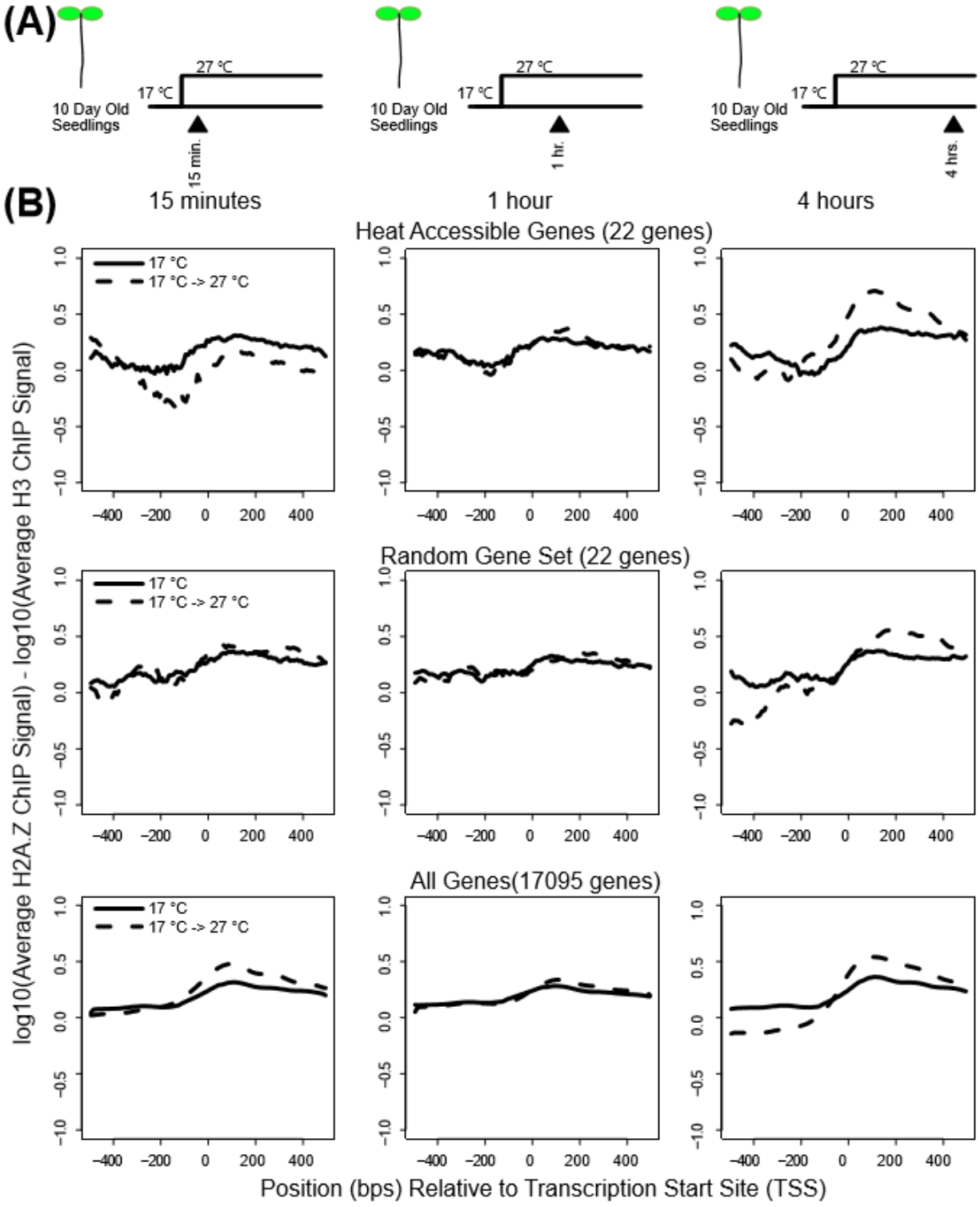
(A) Percentage of ACRs called from each treatment that fall into the following regions: intergenic, promoter, 5' UTR, coding, intron, 3' UTR, and transposons. **(B)** Variance of filtered union ACRs across treatment replicates.



Supplemental Figure 4.5. More loci have greater accessibility during a mild heat stress (38 °C) compared to a severe heat stress (45 °C).

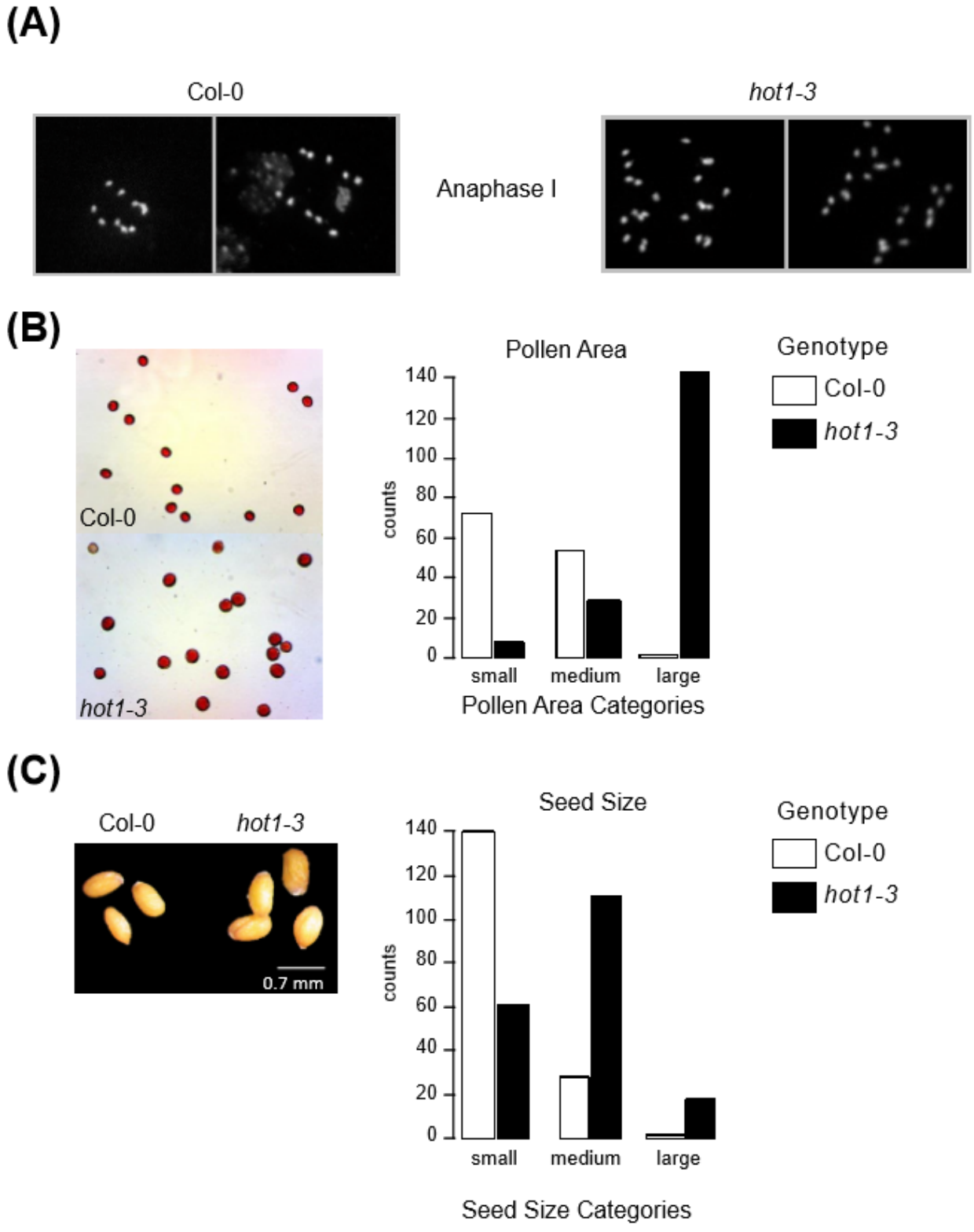
(A) Volcano plots comparing the following pairs of treatments: PTvs. SAT, PTvs LAT, and PTvs LT.ACRs with an adjusted p-value less than 0.1 and an absolute fold change greater than 2 are colored in red. **(B)** Percentage of dACRs from the aforementioned pairwise comparisons that fall into the following regions: intergenic, promoter, 5' UTR, coding, intron, 3' UTR, and transposons. **(C)** Visualization of dACRs from the aforementioned pairwise comparisons across treatments.

For context, 1kb upstream and downstream of the dACR is visualized, and the name of any gene that falls within that interval is provided with the coordinates.The exception to this is the PTvs. LAT comparison.To further contextualize the dACRs in RABA2B and GSTU1, the accessibility of nearby genes, HSP17.8 and HSP17.6 respectively, are included as well.



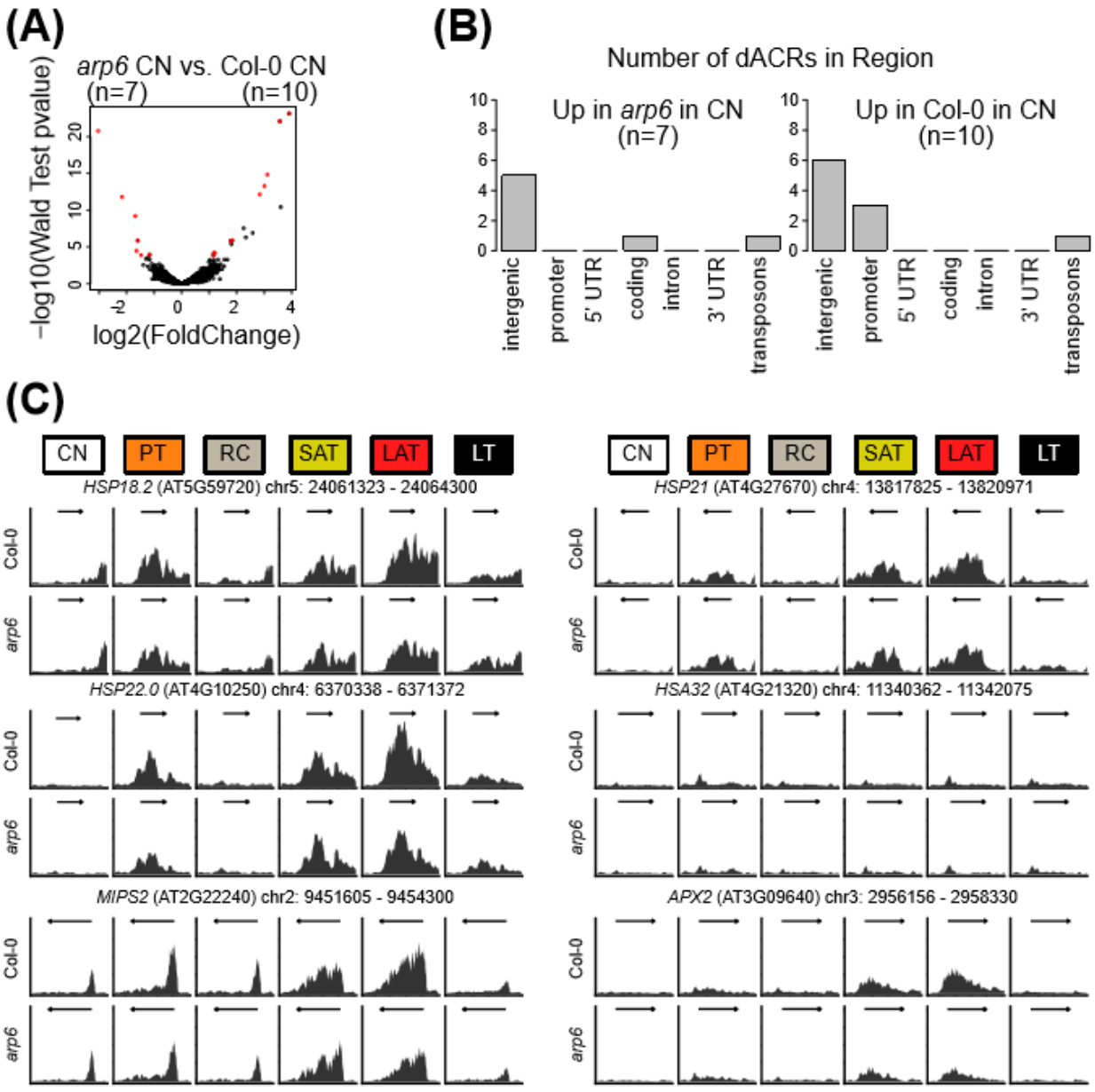
Supplemental Figure 4.6. Heat accessible genes quickly lose H2A.Z occupancy with increased temperature but reincorporate H2A.Z over time.

(A) Experimental design of ChIP-seq experiments performed by Cortijo *et al.* 2017. Ten day old Col-0 seedlings were grown at 17 °C and plants either remained at 17 °C, or were shifted to 27 °C. Samples were collected 15 min, 1 hr, or 4hrs into the shift for ChIP-seq. **(B)** H2A.Z ChIP signal normalized to H3 around TSS for heat accessible genes, a random set, and all genes.



Supplemental Figure 4.7. The *hot1-3* mutant is a tetraploid.

(A) Microscope images of Col-0 chromosomes and *hot1-3* chromosomes during anaphase I. **(B)** Pollen sizes of Col-0 and *hot1-3*. **(C)** Seed sizes of Col-0 and *hot1-3*.



Supplemental Figure 4.8. Col-0 and *arp6* have non-heat induced accessibility differences and do not differ significantly at key genes associated with acquired thermotolerance memory.

(A) Volcano plot, **(B)** Regional Distribution of dACRs between Col-0 and *arp6* in CN. **(C)** Accessibility of *HSP18.2*, *HSP21*, *HSP22.0*, *HSA32*, *MIPS2*, *APX2* in Col-0 and *arp6*.

Sample	Number of ACRs Called for Respective Treatment	Fraction of Reads in ACRs Called for Respective Treatment	Fraction of Reads in Union ACRs (16508 ACRs)	Fraction of Reads in Filtered Union ACRs (12352 ACRs)
Col-0 CN rep1	8565	0.12	0.22	0.20
Col-0 CN rep2	8565	0.08	0.15	0.13
Col-0 PT rep1	8277	0.09	0.19	0.17
Col-0 PT rep2	8277	0.13	0.24	0.22
Col-0 RC rep1	5725	0.07	0.17	0.16
Col-0 RC rep2	5725	0.08	0.18	0.16
Col-0 SAT rep1	5868	0.05	0.14	0.13
Col-0 SAT rep2	5868	0.07	0.18	0.17
Col-0 LAT rep1	10454	0.16	0.26	0.24
Col-0 LAT rep2	10454	0.15	0.26	0.23
Col-0 LT rep1	6239	0.07	0.17	0.16
Col-0 LT rep2	6239	0.06	0.17	0.15

Supplemental Table 4.1. Fraction of Reads in ACRs in Col-0 ATAC-seq samples. Number of ACRs called, and Fraction of Reads in ACRs for Col-0 samples.

Sample	Number of ACRs Called for Respective Treatment	Fraction of Reads in ACRs Called for Respective Treatment	Fraction of Reads in Union ACRs (16508 ACRs)	Fraction of Reads in Filtered Union ACRs (12352 ACRs)
arp6 CN rep1	8338	0.14	0.24	0.23
arp6 CN rep2	8338	0.07	0.15	0.13
arp6 PT rep1	5287	0.08	0.19	0.17
arp6 PT rep2	5287	0.07	0.18	0.16
arp6 RC rep1	1936	0.04	0.19	0.17
arp6 RC rep2	1936	0.03	0.13	0.11
arp6 SAT rep1	3941	0.07	0.21	0.19
arp6 SAT rep2	3941	0.05	0.17	0.15
arp6 LAT rep1	7563	0.10	0.22	0.19
arp6 LAT rep2	7563	0.10	0.20	0.18
arp6 LT rep1	5522	0.07	0.17	0.16
arp6 LT rep2	5522	0.06	0.18	0.16

Supplemental Table 4.2. Number of ACRs called, and Fraction of Reads in ACRs for *arp6* samples.

Condition	Col-0	*arp6*
Control (CN)	0.86	0.90
Pretreatment (PT)	0.76	0.88
Recovery (RC)	0.75	0.91
Short-Term Acquired Thermotolerance (SAT)	0.90	0.91
Long-Term acquired Thermotolerance (LAT)	0.97	0.83
Lethal (LT)	0.93	0.82

Supplemental Table 4.3. Pearson Correlation of ATAC-seq reps (Counts Per Million)

Analysis	Up In Set	Number of ACRs used in Motif Enrichment Analysis
CN vs RC	Up In RC	5
SAT vs LT	Up In SAT	91
SAT vs LT	Up In LT	3
LAT vs LT	Up In LAT	245
LAT vs LT	Up In LT	209
SAT vs LAT	Up In SAT	102
SAT vs LAT	Up In LAT	117
PT vs SAT	Up In PT	62
PT vs SAT	Up In SAT	23
PT vs LAT	Up In PT	213
PT vs LAT	Up In LAT	182
PT vs. LT	Up In PT	178
PT vs. LT	Up In LT	49
CN: Col-0 vs. *arp6*	Up In Col-0	10
CN: Col-0 vs. *arp6*	Up In *arp6*	4
LAT: Col-0 vs. *arp6*	Up In Col-0	10
LAT: Col-0 vs. *arp6*	Up In *arp6*	6

Supplemental Table 4.4. Number of ACRs used in various motif enrichment analyses after removing base pairs in genes.

Treatment	Genotype	2KB Bin Interactions	10KB Bin Interactions	50KB Bin Interactions	250KB Bin Interactions	Total Hi-C Reads
CN	Col-0	13415044	8858987	2359604	115912	17188189
PT	Col-0	14690431	10316380	2624291	115921	18925469
SAT	Col-0	13205761	8791452	2395319	115880	17214045
LT	Col-0	12662987	7723923	2169894	115778	17889968
CN	*arp6*	12657726	9779053	2500011	115860	15347288
PT	*arp6*	12395021	9493426	2469232	115847	15281047
SAT	*arp6*	11450100	8300498	2378744	115840	14253916
LT	*arp6*	10365875	6938267	2167697	115810	15134088

Supplemental Table 4.5. Number of Interactions and Total number of reads used in Hi-C down sampling.

Gene/Name	Forward	Reverse	Experiment
HSP101	AGGGACATCATGGTGTGCGAATC	CTTTATCCGGTAAATGCCGACCAG	RT-qPCR
HSFA2	GCAGCGTTGGATGTGAAAGTGG	TTGGCTGTCCCAATCCAAAGGC	RT-qPCR
UBC	GACCAAGATATTCCATCCTA	GTTAAGAGGACTGTCCG	RT-qPCR
Locus A (*HSP101*)	AGGATCAAGCTTCCCTGCTT	GTGCTTCAGGGGACACAAAT	ATAC-qPCR
Locus B (*HSP101*)	CTTCAAGCGTTGTAGCACCA	GGTCGATGGATGCAGCTAAT	ATAC-qPCR
Locus C (*HSP101*)	TGCCAAAGCAACTCCTCTTT	CCCCTTTCACATGACCAGTT	ATAC-qPCR
***HSP101* control region**	ACATAACCATTGGGCTTTGG	GGCTAATCGATGAGGGATGT	ATAC-qPCR
Locus D (*HSFA2*)	CAGCAGCAGATGATGACGTT	CATAACCGCAAACTGCTGAA	ATAC-qPCR
Locus E (*HSFA2*)	TGGTTCTGTAGCGGCTTCTT	TTCGTTAAGCCCTTCCATTG	ATAC-qPCR
***HSFA2* control region**	CACGTGAGAGACCTGTTTGG	TATCTCGTGCGTGTCTCCAC	ATAC-qPCR

Supplemental Table 4.6. Primers for RT-qPCR and ATAC-qPCR experiments.

4.7 Supplementary Data Availability

This data has not been made publicly available yet, but a pre-print on bioRxiv will be made available in the near future. Supplemental Data tables will be available as additional files with this thesis submission.

4.8 Project Acknowledgements

This work was supported by the National Science Foundation (RESEARCH-PGR grant 568 17488843) to C.Q. and the University of Washington National Institutes of Health (Big Data for Genomics and Neuroscience Training Grant T32LM012419) to K.J.-B.

Chapter 5. Conclusions

Here I will share some thoughts and reflections on my dissertation, some potential future directions, and lingering questions.

Classification of cells in single cell sequencing experiments

Because of the novelty of using scRNA-seq to study root development in Arabidopsis, one of the challenges I faced when analyzing the data was classifying cells. In chapter 2, I discussed how I ultimately used the spearman rank correlation between a given cell's transcriptome and the bulk transcriptomics data for a specific cell type to label cells in the scRNA-seq data. Two other root scRNA-seq papers published around the same time dealt with the classification problem in a similar manner by calculating an Index of Cell Identity, a score which described who much a given cell type "contributed" to a cell's transcriptome (Denyer et al. 2019; Shulse et al. 2019). Due to the nature of single cell experiments, there will typically be an issue of classifying each cell in the data. New experiments, however, can now rely on previously generated scRNA-seq data to aid in classifying their datasets. This can take the form of using cell type specific markers, or comparing their cells to labeled scRNA-seq data. This has motivated work to create single cell expression atlases by combining the cells from different scRNA-seq experiments (Shahan et al. 2020). This leads to a more comprehensive transcriptome, allows one to better resolve ambiguous cell types, and allows one to discover developmental trends that are harder to describe with a smaller number of cells. Each additional scRNA-seq experiment in roots provides another random sampling of developing root cells. So perhaps the solution to the classification problem is just more scRNA-seq right? Not quite.

Part of the work described in Chapter 2 was the use of scRNA-seq in the hopes of resolving cell type specific responses to heat stress. Since heat stress upregulates stress response genes and downregulates other genes, cell type specific markers were not expressed.

This made classifying the cells more difficult. Ultimately, to classify the cells, a mutual nearest neighbor approach was used to map heat stressed cells to their nearest neighbor in control conditions. This successfully separated cells from different cell layers, but may not accurately resolve subcell types. The lateral root formation work described in chapter 3 was only possible because microdissections enriched for stele cells, a cell type that was underrepresented in the work described in chapter 2. Depending on the design of the experiment, the classification problem can be circumvented by performing scRNA-seq on an isolated cell type or a sample enriched for a known cell type. This approach would also be appropriate for isolating rare cells, such as the quiescent center. Another emerging single cell technology that will help ease the classification process is the addition of spatial information to scRNA-seq (Ståhl et al. 2016). If the relative position of root cells can be preserved, then classifying cells based on spatial information alone may be viable in roots. This would also enable more transcriptomics studies on cell type specific stress responses.

Acquired thermotolerance follow up work

Some potential future directions for the acquired thermotolerance work discussed in Chapter 4 is to evaluate the function, if any, of the accessible loci identified at different stages of acquired thermotolerance. Many of the heat accessible loci were in the upstream promoter regions of genes, but a substantial number of the heat accessible loci were intergenic regions. When comparing stages of acquired thermotolerance, intergenic regions were most accessible when assaying for long term acquired thermotolerance and thermomemory. Since these regions do not overlap with transposable elements there is a possibility that they may be enhancers. This can be assayed using STARR-seq. There has been evidence that showed the stress response genes have faster induction upon repeated exposure to stress and this faster induction may be partially driven by enhancers (Lämke et al. 2016). If these intergenic regions are indeed enhancer elements then the question becomes what is binding them. One candidate

could be *Dof* transcription factors, whose transcription factors were enriched in the regions with more accessibility when assaying long term acquired thermotolerance. There are however 36 *Dof* transcription factors in Arabidopsis, and that would be a large number of ChIP-seq experiments to perform (Yanagisawa and Schmidt 1999). A smaller candidate list of *Dof* transcription factors can be generated by examining which *Dof* transcription factor genes have heat induced expression or have increased levels of nascent RNA after a pretreatment. Another transcription factor candidate for ChIP-seq is HSFA2. Knowing exactly where HSFA2 binds during long term acquired thermotolerance can prove generally useful to the study of acquired thermotolerance in plants, but if these heat accessible intergenic regions are HSFA2 targets then it would help solidify their role in thermomemory.

In relation to the heat accessible genes, H2A.Z ChIP-seq data would be helpful in correlating the level of heat induced accessibility with H2A.Z occupancy prior to heat stress. If this was true, then the heat accessible gene body phenomenon may simply be the eviction of H2A.Z containing nucleosomes. Another set of experiments should be performed to characterize the genes that only had heat induced gene body accessibility in plants with acquired thermotolerance. This would most likely take the form of knockout lines of these genes, but a knockout might not have obvious effects in control or under heat stress conditions. Despite this challenge, characterizing these genes may unveil more of how thermomemory works. If the precise function of these genes cannot be easily ascertained, other relevant measurements would be expression data confirming that these genes have faster induction during repeated heat stress, or ChIP-seq data showing that these genes are HSFA2 targets.

The last followup experiment has to do with how heat stress may affect nuclear organization. Microscopy based approaches may be able to better capture heat induced changes like localizing to the nuclear periphery than a high throughput method like Hi-C. If pretreatment did change nuclear organization than thermomemory may be a function of proximity to the nuclear periphery for a subset of genes.

References

Chapter 1

Becker, J. and Craig, E.A. (1994). Heat-shock proteins as molecular chaperones. Eur. J. Biochem. 219: 11–23.

Birnbaum, K., Shasha, D.E., Wang, J.Y., Jung, J.W., Lambert, G.M., Galbraith, D.W., and Benfey, P.N. (2003). A gene expression map of the Arabidopsis root. Science 302: 1956–1960.

Brady, S.M., Orlando, D.A., Lee, J.-Y., Wang, J.Y., Koch, J., Dinneny, J.R., Mace, D., Ohler, U., and Benfey, P.N. (2007a). A high-resolution root spatiotemporal map reveals dominant expression patterns. Science 318: 801–806.

Brady, S.M., Song, S., Dhugga, K.S., Rafalski, J.A., and Benfey, P.N. (2007b). Combining expression and comparative evolutionary analysis. The COBRA gene family. Plant Physiol. 143: 172–187.

Brickner, D.G., Cajigas, I., Fondufe-Mittendorf, Y., Ahmed, S., Lee, P.-C., Widom, J., and Brickner, J.H. (2007). H2A.Z-Mediated Localization of Genes at the Nuclear Periphery Confers Epigenetic Memory of Previous Transcriptional State. PLoS Biol. 5: e81.

Carlsbecker, A. et al. (2010). Cell signalling by microRNA165/6 directs gene dose-dependent root cell fate. Nature 465: 316–321.

Casimiro, I., Marchant, A., Bhalerao, R.P., Beeckman, T., Dhooge, S., Swarup, R., Graham, N., Inzé, D., Sandberg, G., Casero, P.J., and Bennett, M. (2001). Auxin transport promotes Arabidopsis lateral root initiation. Plant Cell 13: 843–852.

Charng, Y.-Y., Liu, H.-C., Liu, N.-Y., Chi, W.-T., Wang, C.-N., Chang, S.-H., and Wang, T.-T. (2007). A heat-inducible transcription factor, HsfA2, is required for extension of acquired thermotolerance in Arabidopsis. Plant Physiol. 143: 251–262.

Chen, G., Ning, B., and Shi, T. (2019). Single-Cell RNA-Seq Technologies and Related Computational Data Analysis. Front. Genet. 10: 317.

Clowes, F.A.L. (1958). DEVELOPMENT OF QUIESCENT CENTRES IN ROOT MERISTEMS. New Phytol. 57: 85–88.

Dannel, F., Pfeffer, H., and Römheld, V. (1998). Compartmentation of boron in roots and leaves of sunflower as affected by boron supply. J. Plant Physiol. 153: 615–622.

De Smet, I. et al. (2007). Auxin-dependent regulation of lateral root positioning in the basal meristem of Arabidopsis. Development 134: 681–690.

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

Hawes, M.C., Gunawardena, U., Miyasaka, S., and Zhao, X. (2000). The role of root border cells in plant defense. Trends Plant Sci. 5: 128–133.

Iijima, M., Morita, S., and Barlow, P.W. (2008). Structure and Function of the Root Cap. Plant Prod. Sci. 11: 17–27.

Ishikawa, H. and Evans, M.L. (1995). Specialized zones of development in roots. Plant Physiol. 109: 725–727.

Laplaze, L., Parizot, B., Baker, A., Ricaud, L., Martinière, A., Auguy, F., Franche, C., Nussaume, L., Bogusz, D., and Haseloff, J. (2005). GAL4-GFP enhancer trap lines for genetic manipulation of lateral root development in Arabidopsis thaliana. J. Exp. Bot. 56: 2433–2442.

Laskowski, M.J., Williams, M.E., Nusbaum, H.C., and Sussex, I.M. (1995). Formation of lateral root meristems is a two-stage process. Development 121: 3303–3310.

Lee, J.-Y., Colinas, J., Wang, J.Y., Mace, D., Ohler, U., and Benfey, P.N. (2006). Transcriptional and posttranscriptional regulation of transcription factor expression in Arabidopsis roots. Proc. Natl. Acad. Sci. U. S. A. 103: 6055–6060.

Lee, M.M. and Schiefelbein, J. (2002). Cell pattern in the Arabidopsis root epidermis determined by lateral inhibition with feedback. Plant Cell 14: 611–618.

Lin, C.Y., Roberts, J.K., and Key, J.L. (1984). Acquisition of Thermotolerance in Soybean Seedlings : Synthesis and Accumulation of Heat Shock Proteins and their Cellular Localization. Plant Physiol. 74: 152–160.

Lindquist, S. and Craig, E.A. (1988). The heat-shock proteins. Annu. Rev. Genet. 22: 631–677.

Li, S., Yamada, M., Han, X., Ohler, U., and Benfey, P.N. (2016). High-Resolution Expression Map of the Arabidopsis Root Reveals Alternative Splicing and lincRNA Regulation. Dev. Cell 39: 508–522.

McCully, M.E. (1999). ROOTS IN SOIL: Unearthing the Complexities of Roots and Their Rhizospheres. Annu. Rev. Plant Physiol. Plant Mol. Biol. 50: 695–718.

Miyashima, S., Koi, S., Hashimoto, T., and Nakajima, K. (2011). Non-cell-autonomous microRNA165 acts in a dose-dependent manner to regulate multiple differentiation status in the Arabidopsis root. Development 138: 2303–2313.

Miyashima, S. and Nakajima, K. (2011). The root endodermis: a hub of developmental signals and nutrient flow. Plant Signal. Behav. 6: 1954–1958.

Nawy, T., Lee, J.-Y., Colinas, J., Wang, J.Y., Thongrod, S.C., Malamy, J.E., Birnbaum, K., and Benfey, P.N. (2005). Transcriptional profile of the Arabidopsis root quiescent center. Plant Cell 17: 1908–1925.

Oparka, K.J. (1990). What Is Phloem Unloading? Plant Physiol. 94: 393–396.

Packer, J.S., Zhu, Q., Huynh, C., Sivaramakrishnan, P., Preston, E., Dueck, H., Stefanik, D., Tan, K., Trapnell, C., Kim, J., Waterston, R.H., and Murray, J.I. (2019). A lineage-resolved molecular atlas of C. elegans embryogenesis at single-cell resolution. Science 365.

Parizot, B. et al. (2008). Diarch symmetry of the vascular bundle in Arabidopsis root

encompasses the pericycle and is reflected in distich lateral root initiation. Plant Physiol. 146: 140–148.

Parizot, B., Roberts, I., Raes, J., Beeckman, T., and De Smet, I. (2012). In silico analyses of pericycle cell populations reinforce their relation with associated vasculature in Arabidopsis. Philos. Trans. R. Soc. Lond. B Biol. Sci. 367: 1479–1488.

Parsell, D.A., Kowal, A.S., Singer, M.A., and Lindquist, S. (1994). Protein disaggregation mediated by heat-shock protein Hsp104. Nature 372: 475–478.

Queitsch, C., Hong, S.W., Vierling, E., and Lindquist, S. (2000). Heat shock protein 101 plays a crucial role in thermotolerance in Arabidopsis. Plant Cell 12: 479–492.

Ross-Elliott, T.J. et al. (2017). Phloem unloading in Arabidopsis roots is convective and regulated by the phloem-pole pericycle. Elife 6.

Sanchez, Y. and Lindquist, S.L. (1990). HSP104 required for induced thermotolerance. Science 248: 1112–1115.

Schiefelbein, J.W., Shipley, A., and Rowse, P. (1992). Calcium influx at the tip of growing root-hair cells of Arabidopsis thaliana. Planta 187: 455–459.

Shendure, J., Balasubramanian, S., Church, G.M., Gilbert, W., Rogers, J., Schloss, J.A., and Waterston, R.H. (2017). DNA sequencing at 40: past, present and future. Nature 550: 345–353.

Sullivan, A.M. et al. (2014). Mapping and Dynamics of Regulatory DNA and Transcription Factor Networks in A. thaliana. Cell Rep. 8: 2015–2030.

Takano, J., Noguchi, K., Yasumori, M., Kobayashi, M., Gajdos, Z., Miwa, K., Hayashi, H., Yoneyama, T., and Fujiwara, T. (2002). Arabidopsis boron transporter for xylem loading. Nature 420: 337–340.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A., Lao, K., and Surani, M.A. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. Nat. Methods 6: 377–382.

Trent, J.D., Gabrielsen, M., Jensen, B., Neuhard, J., and Olsen, J. (1994). Acquired thermotolerance and heat shock proteins in thermophiles from the three phylogenetic domains. J. Bacteriol. 176: 6148–6152.

Verbelen, J.-P., De Cnodder, T., Le, J., Vissenberg, K., and Baluska, F. (2006). The Root Apex of Arabidopsis thaliana Consists of Four Distinct Zones of Growth Activities: Meristematic Zone, Transition Zone, Fast Elongation Zone and Growth Terminating Zone. Plant Signal. Behav. 1: 296–304.

Zhang, H., Roberts, D.N., and Cairns, B.R. (2005). Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss. Cell 123: 219–231.

Zheng, G.X.Y. et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nat. Commun. 8: 14049.

Chapter 2

Alexandre, C.M., Urton, J.R., Jean-Baptiste, K., Huddleston, J., Dorrity, M.W., Cuperus, J.T., Sullivan, A.M., Bemm, F., Jolic, D., Arsovski, A.A., et al. (2018). Complex Relationships between Chromatin Accessibility, Sequence Divergence, and Gene Expression in Arabidopsis thaliana. Mol Biol Evol 35, 837-854.

Bennett, T., van den Toorn, A., Sanchez-Perez, G.F., Campilho, A., Willemsen, V., Snel, B., and Scheres, B. (2010). SOMBRERO, BEARSKIN1, and BEARSKIN2 regulate root cap maturation in Arabidopsis. Plant Cell 22, 640-654.

Bhosale, R., Boudolf, V., Cuevas, F., Lu, R., Eekhout, T., Hu, Z., van Isterdael, G., Lambert, G., Xu, F., Nowack, M.K., et al. (2018). A spatiotemporal DNA endoploidy map of the Arabidopsis root reveals roles for the endocycle in root development and stress adaptation. Plant Cell.

Birnbaum, K., Shasha, D.E., Wang, J.Y., Jung, J.W., Lambert, G.M., Galbraith, D.W., and Benfey, P.N. (2003). A gene expression map of the Arabidopsis root. Science 302, 1956-1960.

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment 2008.

Bourdon, M., Pirrello, J., Cheniclet, C., Coriton, O., Bourge, M., Brown, S., Moise, A., Peypelut, M., Rouyere, V., Renaudin, J.P., et al. (2012). Evidence for karyoplasmic homeostasis during endoreduplication and a ploidy-dependent increase in gene transcription during tomato fruit growth. Development 139, 3817-3826.

Brady, S.M., Orlando, D.A., Lee, J.Y., Wang, J.Y., Koch, J., Dinneny, J.R., Mace, D., Ohler, U., and Benfey, P.N. (2007). A high-resolution root spatiotemporal map reveals dominant expression patterns. Science 318, 801-806.

Cao, J., Packer, J.S., Ramani, V., Cusanovich, D.A., Huynh, C., Daza, R., Qiu, X., Lee, C., Furlan, S.N., Steemers, F.J., et al. (2017). Comprehensive single-cell transcriptional profiling of a multicellular organism. Science 357, 661-667.

Cartwright, D.A., Brady, S.M., Orlando, D.A., Sturmfels, B., and Benfey, P.N. (2009). Reconstructing spatiotemporal gene expression data from partial observations. Bioinformatics 25, 2581-2587.

Dura, J.M. (1981). Stage dependent synthesis of heat shock induced proteins in early embryos of Drosophila melanogaster. Mol Gen Genet 184, 381-385.

Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics 27, 1017-1018.

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

Irish, V.F. (1991). Cell lineage in plant development. Curr Opin Cell Biol 3, 983-987.

Katayama, H., Iwamoto, K., Kariya, Y., Asakawa, T., Kan, T., Fukuda, H., and Ohashi-Ito, K. (2015). A Negative Feedback Loop Controlling bHLH Complexes Is Involved in Vascular Cell Division and Differentiation in the Root Apical Meristem. Curr Biol 25, 3144-3150.

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

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

Lachowiec, J., Mason, G.A., Schultz, K., and Queitsch, C. (2016). Redundancy, feedback, and robustness in the Arabidopsis thaliana BZR/BEH gene family. bioRxiv.

Lachowiec, J., Mason, G.A., Schultz, K., and Queitsch, C. (2018). Redundancy, feedback, and robustness in the Arabidopsis thaliana BZR/BEH gene family. bioRxiv 053447; doi: https://doiorg/101101/053447.

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

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

Li, S., Yamada, M., Han, X., Ohler, U., and Benfey, P.N. (2016). High-Resolution Expression Map of the Arabidopsis Root Reveals Alternative Splicing and lincRNA Regulation. Dev Cell 39, 508-522.

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

Mao, Q., Yang, L., Wang, L., Goodison, S., and Sun, Y. (2015). SimplePPT: A Simple Principal Tree Algorithm Proceedings of the 2015 SIAM International Conference on Data Mining, 792–800.

McInnes, L., and Healy, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. ArXiv e-prints 180203426.

Morange, M., Diu, A., Bensaude, O., and Babinet, C. (1984). Altered expression of heat shock proteins in embryonal carcinoma and mouse early embryonic cells. Mol Cell Biol 4, 730-735.

Mortimer, J.C., Miles, G.P., Brown, D.M., Zhang, Z., Segura, M.P., Weimar, T., Yu, X., Seffen, K.A., Stephens, E., Turner, S.R., et al. (2010). Absence of branches from xylan in Arabidopsis gux mutants reveals potential for simplification of lignocellulosic biomass. Proc Natl Acad Sci U S A 107, 17409-17414.

O'Malley, R.C., Huang, S.C., Song, L., Lewsey, M.G., Bartlett, A., Nery, J.R., Galli, M., Gallavotti, A., and Ecker, J.R. (2016). Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. Cell 165, 1280-1292.

Ohashi-Ito, K., Saegusa, M., Iwamoto, K., Oda, Y., Katayama, H., Kojima, M., Sakakibara, H., and Fukuda, H. (2014). A bHLH complex activates vascular cell division via cytokinin action in root apical meristem. Curr Biol 24, 2053-2058.

Packer, J., and Trapnell, C. (2018). Single-Cell Multi-omics: An Engine for New Quantitative Models of Gene Regulation. Trends Genet 34, 653-665.

Peng, Y., Chen, L., Lu, Y., Wu, Y., Dumenil, J., Zhu, Z., Bevan, M.W., and Li, Y. (2015). The ubiquitin receptors DA1, DAR1, and DAR2 redundantly regulate endoreduplication by modulating the stability of TCP14/15 in Arabidopsis. Plant Cell 27, 649-662.

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

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

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

Resentini, F., Felipo-Benavent, A., Colombo, L., Blazquez, M.A., Alabadi, D., and Masiero, S. (2015). TCP14 and TCP15 mediate the promotion of seed germination by gibberellins in Arabidopsis thaliana. Mol Plant 8, 482-485.

Riechmann, J.L., Heard, J., Martin, G., Reuber, L., Jiang, C., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O.J., Samaha, R.R., et al. (2000). Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. Science 290, 2105-2110.

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

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

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

Sullivan, A.M., Arsovski, A.A., Lempe, J., Bubb, K.L., Weirauch, M.T., Sabo, P.J., Sandstrom, R., Thurman, R.E., Neph, S., Reynolds, A.P., et al. (2014). Mapping and dynamics of regulatory DNA and transcription factor networks in A. thaliana. Cell Rep 8, 2015-2030.

Tan, T.T., Endo, H., Sano, R., Kurata, T., Yamaguchi, M., Ohtani, M., and Demura, T. (2018). Transcription Factors VND1-VND3 Contribute to Cotyledon Xylem Vessel Formation. Plant Physiol 176, 773-789.

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

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

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

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

Wolf, F.A., Hamey, F., Plass, M., Solana, J., Dahlin, J.S., Gottgens, B., Rajewsky, N., Simon, L., and Theis, F.J. (2018). Graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. bioRxiv, doi: https://doiorg/101101/208819

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

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

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

Chapter 3

Beeckman, T., Burssens, S., and Inzé, D. (2001). The peri-cell-cycle in Arabidopsis. J. Exp. Bot. 52: 403–411.

Berger, F., Hung, C.-Y., Dolan, L., and Schiefelbein, J. (1998). Control of Cell Division in the Root Epidermis ofArabidopsis thaliana. Dev. Biol. 194: 235–245.

Bindels, D.S., Haarbosch, L., van Weeren, L., Postma, M., Wiese, K.E., Mastop, M., Aumonier, S., Gotthard, G., Royant, A., Hink, M.A., and Gadella, T.W.J. (2017). mScarlet: a bright monomeric red fluorescent protein for cellular imaging. Nat. Methods 14: 53–56.

Birnbaum, K., Shasha, D.E., Wang, J.Y., Jung, J.W., Lambert, G.M., Galbraith, D.W., and Benfey, P.N. (2003). A gene expression map of the Arabidopsis root. Science 302: 1956–1960.

Brady, S.M., Orlando, D.A., Lee, J.-Y., Wang, J.Y., Koch, J., Dinneny, J.R., Mace, D., Ohler, U., and Benfey, P.N. (2007). A high-resolution root spatiotemporal map reveals dominant expression patterns. Science 318: 801–806.

Brenner, W.G., Leuendorf, J.E., Cortleven, A., Martin, L.B.B., Schaller, H., and Schmülling, T. (2017). Analysis of CFB, a cytokinin-responsive gene of Arabidopsis thaliana encoding a novel F-box protein regulating sterol biosynthesis. J. Exp. Bot. 68: 2769–2785.

Briggs, G.C., Mouchel, C.F., and Hardtke, C.S. (2006). Characterization of the Plant-Specific BREVIS RADIX Gene Family Reveals Limited Genetic Redundancy Despite High Sequence Conservation. Plant Physiol. 140: 1306–1316.

Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F.J., Trapnell, C., and Shendure, J. (2019). The single-cell transcriptional landscape of mammalian organogenesis. Nature 566: 496–502.

Cartwright, D.A., Brady, S.M., Orlando, D.A., Sturmfels, B., and Benfey, P.N. (2009). Reconstructing spatiotemporal gene expression data from partial observations. Bioinforma. Oxf. Engl. 25: 2581–2587.

Chen, X., Lu, L., Qian, S., Scalf, M., Smith, L.M., and Zhong, X. (2018). Canonical and Noncanonical Actions of Arabidopsis Histone Deacetylases in Ribosomal RNA Processing. Plant Cell 30: 134–152.

Chitwood, D.H. and Topp, C.N. (2015). Revealing plant cryptotypes: defining meaningful phenotypes among infinite traits. Curr. Opin. Plant Biol. 24: 54–60.

Clough, S.J. and Bent, A.F. (1998). Floral dip: a simplified method for Agrobacterium -mediated transformation of Arabidopsis thaliana. Plant J. 16: 735–743.

cole-trapnell-lab/monocle3 (2020). (cole-trapnell-lab).

De Rybel, B. et al. (2010). A Novel Aux/IAA28 Signaling Cascade Activates GATA23-Dependent Specification of Lateral Root Founder Cell Identity. Curr. Biol. 20: 1697–1706.

De Smet, I. et al. (2007). Auxin-dependent regulation of lateral root positioning in the basal meristem of Arabidopsis. Dev. Camb. Engl. 134: 681–690.

Decaestecker, W., Buono, R.A., Pfeiffer, M.L., Vangheluwe, N., Jourquin, J., Karimi, M., Van Isterdael, G., Beeckman, T., Nowack, M.K., and Jacobs, T.B. (2019). CRISPR-TSKO: A Technique for Efficient Mutagenesis in Specific Cell Types, Tissues, or Organs in Arabidopsis. Plant Cell 31: 2868–2887.

Delory, B.M., Baudson, C., Brostaux, Y., Lobet, G., du Jardin, P., Pagès, L., and Delaplace, P. (2016). archiDART: an R package for the automated computation of plant root architectural traits. Plant Soil 398: 351–365.

DeTomaso, D., Jones, M.G., Subramaniam, M., Ashuach, T., Ye, C.J., and Yosef, N. (2019). Functional interpretation of single cell similarity maps. Nat. Commun. 10: 4376.

Devaiah, B.N., Karthikeyan, A.S., and Raghothama, K.G. (2007). WRKY75 transcription factor is a modulator of phosphate acquisition and root development in Arabidopsis. Plant Physiol. 143: 1789–1801.

Ditengou, F.A., Teale, W.D., Kochersperger, P., Flittner, K.A., Kneuper, I., Graaff, E. van der, Nziengui, H., Pinosa, F., Li, X., Nitschke, R., Laux, T., and Palme, K. (2008). Mechanical induction of lateral root initiation in Arabidopsis thaliana. Proc. Natl. Acad. Sci. 105: 18818–18823.

Du, Y. and Scheres, B. (2017). PLETHORA transcription factors orchestrate de novo organ patterning during Arabidopsis lateral root outgrowth. Proc. Natl. Acad. Sci. 114: 11709–11714.

Dubrovsky, J.G., Gambetta, G.A., Hernández-Barrera, A., Shishkova, S., and González, I. (2006). Lateral root initiation in Arabidopsis: developmental window, spatial patterning, density and predictability. Ann. Bot. 97: 903–915.

Escamez, S., André, D., Sztojka, B., Bollhöner, B., Hall, H., Berthet, B., Voß, U., Lers, A., Maizel, A., Andersson, M., Bennett, M., and Tuominen, H. (2020). Cell Death in Cells Overlying Lateral Root Primordia Facilitates Organ Growth in Arabidopsis. Curr. Biol. CB 30: 455–464.e7.

Fitter, A.H. (1987). An Architectural Approach to the Comparative Ecology of Plant Root Systems. New Phytol. 106: 61–77.

Galinha, C., Hofhuis, H., Luijten, M., Willemsen, V., Blilou, I., Heidstra, R., and Scheres, B. (2007). PLETHORA proteins as dose-dependent master regulators of Arabidopsis root development. Nature 449: 1053–1057.

Goh, T. et al. (2019). Lateral root initiation requires the sequential induction of transcription factors LBD16 and PUCHI in Arabidopsis thaliana. New Phytol. 224: 749–760.

Goh, T., Joi, S., Mimura, T., and Fukaki, H. (2012). The establishment of asymmetry in Arabidopsis lateral root founder cells is regulated by LBD16/ASL18 and related LBD/ASL proteins. Development 139: 883–893.

Guo, J. and Chen, J.-G. (2008). RACK1 genes regulate plant development with unequal genetic redundancy in Arabidopsis. BMC Plant Biol. 8: 108.

Guseman, J.M., Hellmuth, A., Lanctot, A., Feldman, T.P., Moss, B.L., Klavins, E., Calderón Villalobos, L.I.A., and Nemhauser, J.L. (2015). Auxin-induced degradation dynamics set the pace for lateral root development. Dev. Camb. Engl. 142: 905–909.

Guseman, J.M., Webb, K., Srinivasan, C., and Dardick, C. (2017). DRO1 influences root system architecture in Arabidopsis and Prunus species. Plant J. Cell Mol. Biol. 89: 1093–1105.

Harrison, S.J., Mott, E.K., Parsley, K., Aspinall, S., Gray, J.C., and Cottage, A. (2006). A rapid and robust method of identifying transformed Arabidopsis thaliana seedlings following floral dip transformation. Plant Methods 2: 19.

Himanen, K., Boucheron, E., Vanneste, S., de Almeida Engler, J., Inzé, D., and Beeckman, T. (2002). Auxin-Mediated Cell Cycle Activation during Early Lateral Root Initiation. Plant Cell 14: 2339–2351.

Himanen, K., Vuylsteke, M., Vanneste, S., Vercruysse, S., Boucheron, E., Alard, P., Chriqui, D., Montagu, M.V., Inzé, D., and Beeckman, T. (2004). Transcript profiling of early lateral root initiation. Proc. Natl. Acad. Sci. 101: 5146–5151.

Hu, X. and Xu, L. (2016). Transcription Factors WOX11/12 Directly Activate WOX5/7 to Promote Root Primordia Initiation and Organogenesis. Plant Physiol. 172: 2363–2373.

Jean-Baptiste, K., McFaline-Figueroa, J.L., Alexandre, C.M., Dorrity, M.W., Saunders, L., Bubb, K.L., Trapnell, C., Fields, S., Queitsch, C., and Cuperus, J.T. (2019). Dynamics of Gene Expression in Single Root Cells of Arabidopsis thaliana. Plant Cell 31: 993–1011.

Jeon, J., Cho, C., Lee, M.R., Van Binh, N., and Kim, J. (2016). CYTOKININ RESPONSE FACTOR2 (CRF2) and CRF3 Regulate Lateral Root Development in Response to Cold Stress in Arabidopsis. Plant Cell 28: 1828–1843.

Khakhar, A., Leydon, A.R., Lemmex, A.C., Klavins, E., and Nemhauser, J.L. (2018). Synthetic hormone-responsive transcription factors can monitor and re-program plant development. eLife 7.

Khan, M.A., Gemenet, D.C., and Villordon, A. (2016). Root System Architecture and Abiotic Stress Tolerance: Current Knowledge in Root and Tuber Crops. Front. Plant Sci. 7.

Kim, J., Kim, J.H., Richards, E.J., Chung, K.M., and Woo, H.R. (2014). Arabidopsis VIM Proteins Regulate Epigenetic Silencing by Modulating DNA Methylation and Histone Modification in Cooperation with MET1. Mol. Plant 7: 1470–1485.

Kurihara, D., Mizuta, Y., Sato, Y., and Higashiyama, T. (2015). ClearSee: a rapid optical clearing reagent for whole-plant fluorescence imaging. Dev. Camb. Engl. 142: 4168–4179.

Lanctot, A., Taylor-Teeples, M., Oki, E.A., and Nemhauser, J.L. (2020). Specificity in Auxin Responses Is Not Explained by the Promoter Preferences of Activator ARFs. Plant Physiol. 182: 1533–1536.

Laplaze, L., Parizot, B., Baker, A., Ricaud, L., Martinière, A., Auguy, F., Franche, C., Nussaume, L., Bogusz, D., and Haseloff, J. (2005). GAL4-GFP enhancer trap lines for genetic manipulation of lateral root development in Arabidopsis thaliana. J. Exp. Bot. 56: 2433–2442.

Lavenus, J. et al. (2015). Inference of the Arabidopsis Lateral Root Gene Regulatory Network Suggests a Bifurcation Mechanism That Defines Primordia Flanking and Central Zones[OPEN]. Plant Cell 27: 1368–1388.

Levine, J.H. et al. (2015). Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. Cell 162: 184–197.

Lewis, D.R., Olex, A.L., Lundy, S.R., Turkett, W.H., Fetrow, J.S., and Muday, G.K. (2013). A Kinetic Analysis of the Auxin Transcriptome Reveals Cell Wall Remodeling Proteins That Modulate Lateral Root Development in Arabidopsis. Plant Cell 25: 3329–3346.

Li, H., Torres-Garcia, J., Latrasse, D., Benhamed, M., Schilderink, S., Zhou, W., Kulikova, O., Hirt, H., and Bisseling, T. (2017). Plant-Specific Histone Deacetylases HDT1/2 Regulate GIBBERELLIN 2-OXIDASE2 Expression to Control Arabidopsis Root Meristem Cell Number[OPEN]. Plant Cell 29: 2183–2196.

Lobet, G., Pagès, L., and Draye, X. (2011). A Novel Image-Analysis Toolbox Enabling Quantitative Analysis of Root System Architecture1[W][OA]. Plant Physiol. 157: 29–39.

Luo, M., Cheng, K., Xu, Y., Yang, S., and Wu, K. (2017). Plant Responses to Abiotic Stress Regulated by Histone Deacetylases. Front. Plant Sci. 8.

Lynch, J. (1995). Root Architecture and Plant Productivity. Plant Physiol. 109: 7–13.

Malamy, J.E. and Benfey, P.N. (1997). Organization and cell differentiation in lateral roots of Arabidopsis thaliana. Development 124: 33–44.

Mao, Q., Wang, L., Tsang, I.W., and Sun, Y. (2017). Principal Graph and Structure Learning Based on Reversed Graph Embedding. IEEE Trans. Pattern Anal. Mach. Intell. 39: 2227–2241.

Marhavý, P., Montesinos, J.C., Abuzeineh, A., Van Damme, D., Vermeer, J.E.M., Duclercq, J., Rakusová, H., Nováková, P., Friml, J., Geldner, N., and Benková, E. (2016). Targeted cell elimination reveals an auxin-guided biphasic mode of lateral root initiation. Genes Dev. 30: 471–483.

Marhavý, P., Vanstraelen, M., De Rybel, B., Zhaojun, D., Bennett, M.J., Beeckman, T., and Benková, E. (2013). Auxin reflux between the endodermis and pericycle promotes lateral root initiation. EMBO J. 32: 149–158.

Marin, E., Jouannet, V., Herz, A., Lokerse, A.S., Weijers, D., Vaucheret, H., Nussaume, L., Crespi, M.D., and Maizel, A. (2010). miR390, Arabidopsis TAS3 tasiRNAs, and Their AUXIN RESPONSE FACTOR Targets Define an Autoregulatory Network Quantitatively Regulating Lateral Root Growth. Plant Cell 22: 1104–1117.

Mason, M.G., Mathews, D.E., Argyros, D.A., Maxwell, B.B., Kieber, J.J., Alonso, J.M., Ecker, J.R., and Schaller, G.E. (2005). Multiple type-B response regulators mediate cytokinin signal transduction in Arabidopsis. Plant Cell 17: 3007–3018.

McInnes, L., Healy, J., Saul, N., and Großberger, L. (2018). UMAP: Uniform Manifold Approximation and Projection. J. Open Source Softw. 3: 861.

Mi, H., Muruganujan, A., Ebert, D., Huang, X., and Thomas, P.D. (2019). PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Res. 47: D419–D426.

Miyashima, S. et al. (2019). Mobile PEAR transcription factors integrate positional cues to prime cambial growth. Nature 565: 490–494.

Montague, T.G., Cruz, J.M., Gagnon, J.A., Church, G.M., and Valen, E. (2014). CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. Nucleic Acids Res. 42: W401–W407.

Moreno-Risueno, M.A., Van Norman, J.M., Moreno, A., Zhang, J., Ahnert, S.E., and Benfey, P.N. (2010). Oscillating gene expression determines competence for periodic Arabidopsis root branching. Science 329: 1306–1311.

Muroyama, A., Gong, Y., and Bergmann, D.C. (2020). Opposing, Polarity-Driven Nuclear Migrations Underpin Asymmetric Divisions to Pattern Arabidopsis Stomata. Curr. Biol. CB.

Murphy, E. et al. (2016). RALFL34 regulates formative cell divisions in Arabidopsis pericycle during lateral root initiation. J. Exp. Bot. 67: 4863–4875.

Okushima, Y., Fukaki, H., Onoda, M., Theologis, A., and Tasaka, M. (2007). ARF7 and ARF19 Regulate Lateral Root Formation via Direct Activation of LBD/ASL Genes in Arabidopsis. Plant Cell 19: 118–130.

Ou, Y. et al. (2016). RGF1 INSENSITIVE 1 to 5, a group of LRR receptor-like kinases, are essential for the perception of root meristem growth factor 1 in Arabidopsis thaliana. Cell Res. 26: 686–698.

Parizot, B., Roberts, I., Raes, J., Beeckman, T., and De Smet, I. (2012). In silico analyses of pericycle cell populations reinforce their relation with associated vasculature in Arabidopsis. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 367: 1479–1488.

Porco, S. et al. (2016). Lateral root emergence in Arabidopsis is dependent on transcription factor LBD29 regulation of auxin influx carrier LAX3. Development 143: 3340–3349.

Powers, S.K. et al. (2019). Nucleo-cytoplasmic Partitioning of ARF Proteins Controls Auxin Responses in Arabidopsis thaliana. Mol. Cell 76: 177–190.e5.

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

Ramakrishna, P. et al. (2019). EXPANSIN A1-mediated radial swelling of pericycle cells positions anticlinal cell divisions during lateral root initiation. Proc. Natl. Acad. Sci. 116: 8597–8602.

Rowe, M.H., Dong, J., Weimer, A.K., and Bergmann, D.C. (2019). A Plant-Specific Polarity Module Establishes Cell Fate Asymmetry in the Arabidopsis Stomatal Lineage. bioRxiv: 614636.

Ryu, K.H., Huang, L., Kang, H.M., and Schiefelbein, J. (2019). Single-Cell RNA Sequencing Resolves Molecular Relationships Among Individual Plant Cells. Plant Physiol. 179: 1444–1456.

Santuari, L. et al. (2016). The PLETHORA Gene Regulatory Network Guides Growth and Cell Differentiation in Arabidopsis Roots. Plant Cell 28: 2937–2951.

Schlereth, A., Möller, B., Liu, W., Kientz, M., Flipse, J., Rademacher, E.H., Schmid, M., Jürgens, G., and Weijers, D. (2010). MONOPTEROS controls embryonic root initiation by regulating a mobile transcription factor. Nature 464: 913–916.

Schmidt, T., Pasternak, T., Liu, K., Blein, T., Aubry-Hivet, D., Dovzhenko, A., Duerr, J., Teale, W., Ditengou, F.A., Burkhardt, H., Ronneberger, O., and Palme, K. (2014). The iRoCS Toolbox--3D analysis of the plant root apical meristem at cellular resolution. Plant J. Cell Mol. Biol. 77: 806–814.

Shahan, R., Hsu, C.-W., Nolan, T.M., Cole, B.J., Taylor, I.W., Vlot, A.H.C., Benfey, P.N., and Ohler, U. (2020). A single cell Arabidopsis root atlas reveals developmental trajectories in wild type and cell identity mutants. bioRxiv: 2020.06.29.178863.

Shulse, C.N., Cole, B.J., Ciobanu, D., Lin, J., Yoshinaga, Y., Gouran, M., Turco, G.M., Zhu, Y., O'Malley, R.C., Brady, S.M., and Dickel, D.E. (2019). High-Throughput Single-Cell Transcriptome Profiling of Plant Cell Types. Cell Rep. 27: 2241–2247.e4.

Siligato, R. et al. (2016). Multisite gateway-compatible cell type-specific gene-inducible system for plants. Plant Physiol. 170: 627–641.

Singh, S., Yadav, S., Singh, A., Mahima, M., Singh, A., Gautam, V., and Sarkar, A.K. (2020). Auxin signaling modulates LATERAL ROOT PRIMORDIUM1 (LRP1) expression during lateral root development in Arabidopsis. Plant J. 101: 87–100.

Smet, W. et al. (2019). DOF2.1 Controls Cytokinin-Dependent Vascular Cell Proliferation Downstream of TMO5/LHW. Curr. Biol. 29: 520–529.e6.

Smith, D.L. and Fedoroff, N.V. (1995). LRP1, a gene expressed in lateral and adventitious root primordia of arabidopsis. Plant Cell 7: 735–745.

Snouffer, A., Kraus, C., and van der Knaap, E. (2020). The shape of things to come: ovate family proteins regulate plant organ shape. Curr. Opin. Plant Biol. 53: 98–105.

Tejos, R., Rodriguez-Furlán, C., Adamowski, M., Sauer, M., Norambuena, L., and Friml, J. (2018). PATELLINS are regulators of auxin-mediated PIN1 relocation and plant development in Arabidopsis thaliana. J. Cell Sci. 131.

To, J.P.C., Haberer, G., Ferreira, F.J., Deruère, J., Mason, M.G., Schaller, G.E., Alonso, J.M., Ecker, J.R., and Kieber, J.J. (2004). Type-A Arabidopsis response regulators are partially redundant negative regulators of cytokinin signaling. Plant Cell 16: 658–671.

Torres-Martínez, H.H., Hernández-Herrera, P., Corkidi, G., and Dubrovsky, J.G. (2020). From one cell to many: Morphogenetic field of lateral root founder cells in Arabidopsis thaliana is built by gradual recruitment. Proc. Natl. Acad. Sci. 117: 20943–20949.

Toyokura, K. et al. (2019). Lateral Inhibition by a Peptide Hormone-Receptor Cascade during Arabidopsis Lateral Root Founder Cell Formation. Dev. Cell 48: 64–75.e5.

Traag, V.A., Waltman, L., and van Eck, N.J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. Sci. Rep. 9: 5233.

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

Trinh, D.-C. et al. (2019). PUCHI regulates very long chain fatty acid biosynthesis during lateral root and callus formation. Proc. Natl. Acad. Sci. U. S. A. 116: 14325–14330.

Uga, Y. et al. (2013). Control of root system architecture by DEEPER ROOTING 1 increases rice yield under drought conditions. Nat. Genet. 45: 1097–1102.

Vanneste, S. et al. (2005). Cell Cycle Progression in the Pericycle Is Not Sufficient for SOLITARY ROOT/IAA14-Mediated Lateral Root Initiation in Arabidopsis thaliana. Plant Cell 17: 3035–3050.

Vermeer, J.E.M., Wangenheim, D. von, Barberon, M., Lee, Y., Stelzer, E.H.K., Maizel, A., and Geldner, N. (2014). A Spatial Accommodation by Neighboring Cells Is Required for Organ Initiation in Arabidopsis. Science 343: 178–183.

Voß, U. et al. (2015). The circadian clock rephases during lateral root organ initiation in Arabidopsis thaliana. Nat. Commun. 6: 7641.

Waite, J.M., Collum, T.D., and Dardick, C. (2020). AtDRO1 is nuclear localized in root tips under native conditions and impacts auxin localization. Plant Mol. Biol. 103: 197–210.

Wang, S., Chang, Y., and Ellis, B. (2016). Overview of OVATE FAMILY PROTEINS, A Novel Class of Plant-Specific Growth Regulators. Front. Plant Sci. 7.

Wang, S., Chang, Y., Guo, J., Zeng, Q., Ellis, B.E., and Chen, J.-G. (2011). Arabidopsis ovate family proteins, a novel transcriptional repressor family, control multiple aspects of plant growth and development. PloS One 6: e23896.

Wang, Z.-P., Xing, H.-L., Dong, L., Zhang, H.-Y., Han, C.-Y., Wang, X.-C., and Chen, Q.-J. (2015). Egg cell-specific promoter-controlled CRISPR/Cas9 efficiently generates homozygous mutants for multiple target genes in Arabidopsis in a single generation. Genome Biol. 16: 144.

Weber, E., Engler, C., Gruetzner, R., Werner, S., and Marillonnet, S. (2011). A Modular Cloning System for Standardized Assembly of Multigene Constructs. PLOS ONE 6: e16765.

Wilmoth, J.C., Wang, S., Tiwari, S.B., Joshi, A.D., Hagen, G., Guilfoyle, T.J., Alonso, J.M., Ecker, J.R., and Reed, J.W. (2005). NPH4/ARF7 and ARF19 promote leaf expansion and auxin-induced lateral root formation. Plant J. Cell Mol. Biol. 43: 118–130.

Wolock, S.L., Lopez, R., and Klein, A.M. (2019). Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. Cell Syst. 8: 281-291.e9.

Xun, Q., Wu, Y., Li, H., Chang, J., Ou, Y., He, K., Gou, X., Tax, F.E., and Li, J. (2020). Two receptor-like protein kinases, MUSTACHES and MUSTACHES-LIKE, regulate lateral root development in Arabidopsis thaliana. New Phytol. 227: 1157–1173.

Yi, D. et al. (2014). The Arabidopsis SIAMESE-RELATED Cyclin-Dependent Kinase Inhibitors SMR5 and SMR7 Regulate the DNA Damage Checkpoint in Response to Reactive Oxygen Species[W]. Plant Cell 26: 296–309.

Yoo, S.-D., Cho, Y.-H., and Sheen, J. (2007). Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. Nat. Protoc. 2: 1565–1572.

Zeng, Q., Sritubtim, S., and Ellis, B.E. (2011). AtMKK6 and AtMPK13 are required for lateral root formation in Arabidopsis. Plant Signal. Behav. 6: 1436–1439.

Zhu, Y., Dong, A., Meyer, D., Pichon, O., Renou, J.-P., Cao, K., and Shen, W.-H. (2006). Arabidopsis NRP1 and NRP2 encode histone chaperones and are required for maintaining postembryonic root growth. Plant Cell 18: 2879–2892.

Chapter 4

Albihlal, W.S., Obomighie, I., Blein, T., Persad, R., Chernukhin, I., Crespi, M., Bechtold, U., and Mullineaux, P.M. (2018). Arabidopsis HEAT SHOCK TRANSCRIPTION FACTORA1b regulates multiple developmental genes under benign and stress conditions. J. Exp. Bot. 69: 2847–2862.

Berardini, T.Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., and Huala, E. (2015). The Arabidopsis information resource: Making and mining the “gold standard” annotated reference plant genome. Genesis 53: 474–485.

Brickner, D.G., Cajigas, I., Fondufe-Mittendorf, Y., Ahmed, S., Lee, P.-C., Widom, J., and Brickner, J.H. (2007). H2A.Z-Mediated Localization of Genes at the Nuclear Periphery Confers Epigenetic Memory of Previous Transcriptional State. PLoS Biol. 5: e81.

Brickner, J.H. (2009). Transcriptional memory at the nuclear periphery. Curr. Opin. Cell Biol. 21: 127–133.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods 10: 1213–1218.

Carper, S.W., Duffy, J.J., and Gerner, E.W. (1987). Heat shock proteins in thermotolerance and other cellular processes. Cancer Res. 47: 5249–5255.

Charng, Y.-Y., Liu, H.-C., Liu, N.-Y., Chi, W.-T., Wang, C.-N., Chang, S.-H., and Wang, T.-T. (2007). A heat-inducible transcription factor, HsfA2, is required for extension of acquired thermotolerance in Arabidopsis. Plant Physiol. 143: 251–262.

Charng, Y.-Y., Liu, H.-C., Liu, N.-Y., Hsu, F.-C., and Ko, S.-S. (2006). Arabidopsis Hsa32, a novel heat shock protein, is essential for acquired thermotolerance during long recovery after acclimation. Plant Physiol. 140: 1297–1305.

Cheng, C.-Y., Krishnakumar, V., Chan, A., Schobel, S., and Town, C.D. (2016). Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. bioRxiv: 047308.

Clark, R.M. et al. (2007). Common sequence polymorphisms shaping genetic diversity in Arabidopsis thaliana. Science 317: 338–342.

Coleman-Derr, D. and Zilberman, D. (2012). Deposition of histone variant H2A.Z within gene bodies regulates responsive genes. PLoS Genet. 8: e1002988.

Concia, L. et al. (2020). Wheat chromatin architecture is organized in genome territories and transcription factories. Genome Biol. 21: 104.

Cook, K.B., Hristov, B.H., Le Roch, K.G., Vert, J.P., and Noble, W.S. (2020). Measuring significant changes in chromatin conformation with ACCOST. Nucleic Acids Res. 48: 2303–2311.

Cortijo, S., Charoensawan, V., Brestovitsky, A., Buning, R., Ravarani, C., Rhodes, D., van Noort, J., Jaeger, K.E., and Wigge, P.A. (2017). Transcriptional Regulation of the Ambient Temperature Response by H2A.Z Nucleosomes and HSF1 Transcription Factors in Arabidopsis. Mol. Plant 10: 1258–1273.

Cullis, C.A., Vorster, B.J., Van Der Vyver, C., and Kunert, K.J. (2009). Transfer of genetic material between the chloroplast and nucleus: how is it related to stress in plants? Ann. Bot. 103: 625–633.

Deal, R.B. and Henikoff, S. (2010). A simple method for gene expression and chromatin profiling of individual cell types within a tissue. Dev. Cell 18: 1030–1040.

Deal, R.B. and Henikoff, S. (2011). The INTACT method for cell type-specific gene expression

and chromatin profiling in Arabidopsis thaliana. Nat. Protoc. 6: 56–68.

Dickinson, P.J., Kumar, M., Martinho, C., Yoo, S.J., Lan, H., Artavanis, G., Charoensawan, V., Schöttler, M.A., Bock, R., Jaeger, K.E., and Wigge, P.A. (2018). Chloroplast Signaling Gates Thermotolerance in Arabidopsis. Cell Rep. 22: 1657–1665.

Ding, Y., Fromm, M., and Avramova, Z. (2012). Multiple exposures to drought “train” transcriptional responses in Arabidopsis. Nat. Commun. 3: 740.

Dorn, K.V., Willmund, F., Schwarz, C., Henselmann, C., Pohl, T., Hess, B., Veyel, D., Usadel, B., Friedrich, T., Nickelsen, J., and Schroda, M. (2010). Chloroplast DnaJ-like proteins 3 and 4 (CDJ3/4) from Chlamydomonas reinhardtii contain redox-active Fe-S clusters and interact with stromal HSP70B. Biochem. J 427: 205–215.

Favero, D.S., Jacques, C.N., Iwase, A., Le, K.N., Zhao, J., Sugimoto, K., and Neff, M.M. (2016). SUPPRESSOR OF PHYTOCHROME B4-#3 Represses Genes Associated with Auxin Signaling to Modulate Hypocotyl Growth. Plant Physiol. 171: 2701–2716.

Feng, J.-X., Liu, D., Pan, Y., Gong, W., Ma, L.-G., Luo, J.-C., Deng, X.W., and Zhu, Y.-X. (2005). An annotation update via cDNA sequence analysis and comprehensive profiling of developmental, hormonal or environmental responsiveness of the Arabidopsis AP2/EREBP transcription factor gene family. Plant Mol. Biol. 59: 853–868.

Feng, S., Cokus, S.J., Schubert, V., Zhai, J., Pellegrini, M., and Jacobsen, S.E. (2014). Genome-wide Hi-C analyses in wild-type and mutants reveal high-resolution chromatin interactions in Arabidopsis. Mol. Cell 55: 694–707.

Franco-Zorrilla, J.M., López-Vidriero, I., Carrasco, J.L., Godoy, M., Vera, P., and Solano, R. (2014). DNA-binding specificities of plant transcription factors and their potential to define target genes. Proc. Natl. Acad. Sci. U. S. A. 111: 2367–2372.

Grob, S., Schmid, M.W., and Grossniklaus, U. (2014). Hi-C Analysis in Arabidopsis Identifies the KNOT, a Structure with Similarities to the flamenco Locus of Drosophila. Mol. Cell 55: 678–693.

Guillemette, B., Bataille, A.R., Gévry, N., Adam, M., Blanchette, M., Robert, F., and Gaudreau, L. (2005). Variant histone H2A.Z is globally localized to the promoters of inactive yeast genes and regulates nucleosome positioning. PLoS Biol. 3: e384.

Hara, K., Yokoo, T., Kajita, R., Onishi, T., Yahata, S., Peterson, K.M., Torii, K.U., and Kakimoto, T. (2009). Epidermal cell density is autoregulated via a secretory peptide, EPIDERMAL PATTERNING FACTOR 2 in Arabidopsis leaves. Plant Cell Physiol. 50: 1019–1031.

Hatakeyama, S. and Nakayama, K.-I.I. (2003). U-box proteins as a new family of ubiquitin ligases. Biochem. Biophys. Res. Commun. 302: 635–645.

Heng, Y., Lin, F., Jiang, Y., Ding, M., Yan, T., Lan, H., Zhou, H., Zhao, X., Xu, D., and Deng, X.W. (2019). B-Box Containing Proteins BBX30 and BBX31, Acting Downstream of HY5, Negatively Regulate Photomorphogenesis in Arabidopsis. Plant Physiol. 180: 497–508.

Hong, S.W. and Vierling, E. (2001). Hsp101 is necessary for heat tolerance but dispensable for

development and germination in the absence of stress. Plant J. 27: 25–35.

Huang, C.Y., Ayliffe, M.A., and Timmis, J.N. (2003). Direct measurement of the transfer rate of chloroplast DNA into the nucleus. Nature 422: 72–76.

Hunt, L. and Gray, J.E. (2009). The signaling peptide EPF2 controls asymmetric cell divisions during stomatal development. Curr. Biol. 19: 864–869.

Hu, S., Ding, Y., and Zhu, C. (2020). Sensitivity and Responses of Chloroplasts to Heat Stress in Plants. Front. Plant Sci. 11: 375.

Hu, Y., Plutz, M., and Belmont, A.S. (2010). Hsp70 gene association with nuclear speckles is Hsp70 promoter specific. J. Cell Biol. 191: 711–719.

Imakaev, M., Fudenberg, G., McCord, R.P., Naumova, N., Goloborodko, A., Lajoie, B.R., Dekker, J., and Mirny, L.A. (2012). Iterative correction of Hi-C data reveals hallmarks of chromosome organization. Nat. Methods 9: 999–1003.

Josling, G.A., Selvarajah, S.A., Petter, M., and Duffy, M.F. (2012). The role of bromodomain proteins in regulating gene expression. Genes 3: 320–343.

Kilian, J., Whitehead, D., Horak, J., Wanke, D., Weinl, S., Batistic, O., D'Angelo, C., Bornberg-Bauer, E., Kudla, J., and Harter, K. (2007). The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. Plant J. 50: 347–363.

Kumar, S.V. and Wigge, P.A. (2010). H2A.Z-containing nucleosomes mediate the thermosensory response in Arabidopsis. Cell 140: 136–147.

Lamesch, P. et al. (2012). The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res. 40: D1202–D1210.

Lämke, J., Brzezinka, K., Altmann, S., and Bäurle, I. (2016). A hit-and-run heat shock factor governs sustained histone methylation and transcriptional stress memory. EMBO J. 35: 162–175.

Langmead, B. and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods 9: 357–359.

Larkindale, J. and Vierling, E. (2008). Core genome responses involved in acclimation to high temperature. Plant Physiol. 146: 748–761.

Leng, L., Liang, Q., Jiang, J., Zhang, C., Hao, Y., Wang, X., and Su, W. (2017). A subclass of HSP70s regulate development and abiotic stress responses in Arabidopsis thaliana. J. Plant Res. 130: 349–363.

Light, W.H., Brickner, D.G., Brand, V.R., and Brickner, J.H. (2010). Interaction of a DNA zip code with the nuclear pore complex promotes H2A.Z incorporation and INO1 transcriptional memory. Mol. Cell 40: 112–125.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G.,

Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079.

Li, L., Lyu, X., Hou, C., Takenaka, N., Nguyen, H.Q., Ong, C.-T., Cubeñas-Potts, C., Hu, M., Lei, E.P., Bosco, G., Qin, Z.S., and Corces, V.G. (2015). Widespread rearrangement of 3D chromatin organization underlies polycomb-mediated stress-induced silencing. Mol. Cell 58: 216–231.

Lin, B.L., Wang, J.S., Liu, H.C., Chen, R.W., Meyer, Y., Barakat, A., and Delseny, M. (2001). Genomic analysis of the Hsp70 superfamily in Arabidopsis thaliana. Cell Stress Chaperones 6: 201–208.

Lindquist, S. and Craig, E.A. (1988). The heat-shock proteins. Annu. Rev. Genet. 22: 631–677.

Lin, K.-F., Tsai, M.-Y., Lu, C.-A., Wu, S.-J., and Yeh, C.-H. (2018). The roles of Arabidopsis HSFA2, HSFA4a, and HSFA7a in the heat shock response and cytosolic protein response. Bot. Stud. 59: 15.

Lister, D.L., Bateman, J.M., Purton, S., and Howe, C.J. (2003). DNA transfer from chloroplast to nucleus is much rarer in Chlamydomonas than in tobacco. Gene 316: 33–38.

Liu, C., Cheng, Y.-J., Wang, J.-W., and Weigel, D. (2017). Prominent topologically associated domains differentiate global chromatin packing in rice from Arabidopsis. Nat Plants 3: 742–748.

Liu, C., Wang, C., Wang, G., Becker, C., Zaidem, M., and Weigel, D. (2016). Genome-wide analysis of chromatin packing in Arabidopsis thaliana at single-gene resolution. Genome Res. 26: 1057–1068.

Liu, H.-C., Lämke, J., Lin, S.-Y., Hung, M.-J., Liu, K.-M., Charng, Y.-Y., and Bäurle, I. (2018). Distinct heat shock factors and chromatin modifications mediate the organ-autonomous transcriptional memory of heat stress. Plant J. 95: 401–413.

Liu, M., Zhu, J., and Dong, Z. (2020). Immediate transcriptional responses of Arabidopsis leaves to heat shock. J. Integr. Plant Biol.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15: 550.

Martin, W. (2003). Gene transfer from organelles to the nucleus: frequent and in big chunks. Proc. Natl. Acad. Sci. U. S. A. 100: 8612–8614.

McLeay, R.C. and Bailey, T.L. (2010). Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. BMC Bioinformatics 11: 165.

Mc Lellan, C.A., Turbyville, T.J., Kithsiri Wijeratne, E.M., Kerschen, A., Vierling, E., Queitsch, C., Whitesell, L., and Leslie Gunatilaka, A.A. (2007). A Rhizosphere Fungus Enhances Arabidopsis Thermotolerance through Production of an HSP90 Inhibitor1. Plant Physiology 145: 174–182.

Mi, H., Muruganujan, A., Ebert, D., Huang, X., and Thomas, P.D. (2019a). PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Res. 47: D419–D426.

Mi, H., Muruganujan, A., Huang, X., Ebert, D., Mills, C., Guo, X., and Thomas, P.D. (2019b). Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). Nat. Protoc. 14: 703–721.

Milioni, D. and Hatzopoulos, P. (1997). Genomic organization of hsp90 gene family in Arabidopsis. Plant Mol. Biol. 35: 955–961.

Morimoto, R.I., Tissières, A., and Georgopoulos, C. (1990). 1 The Stress Response, Function of the Proteins, and Perspectives. Cold Spring Harbor Monograph Archive 19: 1–36.

Neph, S. et al. (2012). BEDOPS: high-performance genomic feature operations. Bioinformatics 28: 1919–1920.

Nguyen, N.H. and Cheong, J.-J. (2018). H2A.Z-containing nucleosomes are evicted to activate AtMYB44 transcription in response to salt stress. Biochem. Biophys. Res. Commun. 499: 1039–1043.

Nishizawa, A., Yabuta, Y., and Shigeoka, S. (2008). Galactinol and raffinose constitute a novel function to protect plants from oxidative damage. Plant Physiol. 147: 1251–1263.

Nishizawa, A., Yabuta, Y., Yoshida, E., Maruta, T., Yoshimura, K., and Shigeoka, S. (2006). Arabidopsis heat shock transcription factor A2 as a key regulator in response to several types of environmental stress. Plant J. 48: 535–547.

Nover, L. (1991). Heat Shock Response (CRC Press).

Nover, L., Bharti, K., Döring, P., Mishra, S.K., Ganguli, A., and Scharf, K.D. (2001). Arabidopsis and the heat stress transcription factor world: how many heat stress transcription factors do we need? Cell Stress Chaperones 6: 177–189.

O'Malley, R.C., Huang, S.-S.C., Song, L., Lewsey, M.G., Bartlett, A., Nery, J.R., Galli, M., Gallavotti, A., and Ecker, J.R. (2016). Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. Cell 165: 1280–1292.

Panchuk, I.I., Volkov, R.A., and Schöffl, F. (2002). Heat stress- and heat shock transcription factor-dependent expression and activity of ascorbate peroxidase in Arabidopsis. Plant Physiol. 129: 838–853.

Panikulangara, T.J., Eggers-Schumacher, G., Wunderlich, M., Stransky, H., and Schöffl, F. (2004). Galactinol synthase1. A novel heat shock factor target gene responsible for heat-induced synthesis of raffinose family oligosaccharides in Arabidopsis. Plant Physiol. 136: 3148–3158.

Park, D.H., Lim, P.O., Kim, J.S., Cho, D.S., Hong, S.H., and Nam, H.G. (2003). The Arabidopsis COG1 gene encodes a Dof domain transcription factor and negatively regulates phytochrome signaling. Plant J. 34: 161–171.

Przybyla-Toscano, J., Roland, M., Gaymard, F., Couturier, J., and Rouhier, N. (2018). Roles and maturation of iron-sulfur proteins in plastids. J. Biol. Inorg. Chem. 23: 545–566.

Queitsch, C., Hong, S.W., Vierling, E., and Lindquist, S. (2000). Heat shock protein 101 plays a

crucial role in thermotolerance in Arabidopsis. Plant Cell 12: 479–492.

Quinlan, A.R. and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841–842.

Ray, J., Munn, P.R., Vihervaara, A., Lewis, J.J., Ozer, A., Danko, C.G., and Lis, J.T. (2019). Chromatin conformation remains stable upon extensive transcriptional changes driven by heat shock. Proc. Natl. Acad. Sci. U. S. A. 116: 19431–19439.

Sadeghi, L., Bonilla, C., Strålfors, A., Ekwall, K., and Svensson, J.P. (2011). Podbat: a novel genomic tool reveals Swr1-independent H2A.Z incorporation at gene coding sequences through epigenetic meta-analysis. PLoS Comput. Biol. 7: e1002163.

Sakamoto, H., Maruyama, K., Sakuma, Y., Meshi, T., Iwabuchi, M., Shinozaki, K., and Yamaguchi-Shinozaki, K. (2004). Arabidopsis Cys2/His2-type zinc-finger proteins function as transcription repressors under drought, cold, and high-salinity stress conditions. Plant Physiol. 136: 2734–2746.

Scarpeci, T.E., Frea, V.S., Zanor, M.I., and Valle, E.M. (2017). Overexpression of AtERF019 delays plant growth and senescence, and improves drought tolerance in Arabidopsis. J. Exp. Bot. 68: 673–685.

Scharf, K.-D., Berberich, T., Ebersberger, I., and Nover, L. (2012). The plant heat stress transcription factor (Hsf) family: structure, function and evolution. Biochim. Biophys. Acta 1819: 104–119.

Sedaghatmehr, M., Mueller-Roeber, B., and Balazadeh, S. (2016). The plastid metalloprotease FtsH6 and small heat shock protein HSP21 jointly regulate thermomemory in Arabidopsis. Nat. Commun. 7: 12439.

Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.-J., Vert, J.-P., Heard, E., Dekker, J., and Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol. 16: 259.

Shahnejat-Bushehri, S., Mueller-Roeber, B., and Balazadeh, S. (2012). Arabidopsis NAC transcription factor JUNGBRUNNEN1 affects thermomemory-associated genes and enhances heat stress tolerance in primed and unprimed conditions. Plant Signal. Behav. 7: 1518–1521.

Siddique, M., Gernhard, S., von Koskull-Döring, P., Vierling, E., and Scharf, K.-D. (2008). The plant sHSP superfamily: five new members in Arabidopsis thaliana with unexpected properties. Cell Stress Chaperones 13: 183–197.

Silverstein, K.A.T., Graham, M.A., Paape, T.D., and VandenBosch, K.A. (2005). Genome organization of more than 300 defensin-like genes in Arabidopsis. Plant Physiol. 138: 600–610.

Smith, A.P., Jain, A., Deal, R.B., Nagarajan, V.K., Poling, M.D., Raghothama, K.G., and Meagher, R.B. (2010). Histone H2A.Z regulates the expression of several classes of phosphate starvation response genes but not as a transcriptional activator. Plant Physiol. 152: 217–225.

van Steensel, B. and Furlong, E.E.M. (2019). The role of transcription in shaping the spatial organization of the genome. Nat. Rev. Mol. Cell Biol. 20: 327–337.

Stegemann, S., Hartmann, S., Ruf, S., and Bock, R. (2003). High-frequency gene transfer from the chloroplast genome to the nucleus. Proc. Natl. Acad. Sci. U. S. A. 100: 8828–8833.

Street, I.H., Shah, P.K., Smith, A.M., Avery, N., and Neff, M.M. (2007). The AT-hook-containing proteins SOB3/AHL29 and ESC/AHL27 are negative modulators of hypocotyl growth in Arabidopsis. The Plant Journal 54: 1–14.

Sugio, A., Dreos, R., Aparicio, F., and Maule, A.J. (2009). The cytosolic protein response as a subcomponent of the wider heat shock response in Arabidopsis. Plant Cell 21: 642–654.

Sullivan, A.M. et al. (2014a). Mapping and Dynamics of Regulatory DNA and Transcription Factor Networks in A. thaliana. Cell Rep. 8: 2015–2030.

Sullivan, A.M. et al. (2014b). Mapping and Dynamics of Regulatory DNA and Transcription Factor Networks in A. thaliana. Cell Rep. 8: 2015–2030.

Sun, A.-Z. and Guo, F.-Q. (2016). Chloroplast Retrograde Regulation of Heat Stress Responses in Plants. Front. Plant Sci. 7: 398.

Sung, D.Y., Vierling, E., and Guy, C.L. (2001). Comprehensive expression profile analysis of the Arabidopsis Hsp70 gene family. Plant Physiol. 126: 789–800.

Su, P.-H. and Li, H.-M. (2008). Arabidopsis stromal 70-kD heat shock proteins are essential for plant development and important for thermotolerance of germinating seeds. Plant Physiol. 146: 1231–1241.

Sura, W., Kabza, M., Karlowski, W.M., Bieluszewski, T., Kus-Slowinska, M., Pawełoszek, Ł., Sadowski, J., and Ziolkowski, P.A. (2017). Dual Role of the Histone Variant H2A.Z in Transcriptional Regulation of Stress-Response Genes. Plant Cell 29: 791–807.

Swindell, W.R., Huebner, M., and Weber, A.P. (2007). Transcriptional profiling of Arabidopsis heat shock proteins and transcription factors reveals extensive overlap between heat and non-heat stress response pathways. BMC Genomics 8: 125.

Tesfaye, M. et al. (2013). Spatio-temporal expression patterns of Arabidopsis thaliana and Medicago truncatula defensin-like genes. PLoS One 8: e58992.

Wang, C., Liu, C., Roqueiro, D., Grimm, D., Schwab, R., Becker, C., Lanz, C., and Weigel, D. (2015). Genome-wide analysis of local chromatin packing in Arabidopsis thaliana. Genome Res. 25: 246–256.

Ward, J.M., Cufr, C.A., Denzel, M.A., and Neff, M.M. (2005). The Dof transcription factor OBP3 modulates phytochrome and cryptochrome signaling in Arabidopsis. Plant Cell 17: 475–485.

Wehmeyer, N. and Vierling, E. (2000). The expression of small heat shock proteins in seeds responds to discrete developmental signals and suggests a general protective role in desiccation tolerance. Plant Physiol. 122: 1099–1108.

Weigel, D. et al. (2000). Activation Tagging in Arabidopsis. Plant Physiology 122: 1003–1014.

Xu, X., Song, H., Zhou, Z., Shi, N., Ying, Q., and Wang, H. (2010). Functional characterization

of AtHsp90.3 in Saccharomyces cerevisiae and Arabidopsis thaliana under heat stress. Biotechnol. Lett. 32: 979–987.

Xu, Y.-H., Liu, R., Yan, L., Liu, Z.-Q., Jiang, S.-C., Shen, Y.-Y., Wang, X.-F., and Zhang, D.-P. (2012a). Light-harvesting chlorophyll a/b-binding proteins are required for stomatal response to abscisic acid in Arabidopsis. J. Exp. Bot. 63: 1095–1106.

Xu, Z.-S., Li, Z.-Y., Chen, Y., Chen, M., Li, L.-C., and Ma, Y.-Z. (2012b). Heat shock protein 90 in plants: molecular mechanisms and roles in stress responses. Int. J. Mol. Sci. 13: 15706–15723.

Yabe, N., Takahashi, T., and Komeda, Y. (1994). Analysis of tissue-specific expression of Arabidopsis thaliana HSP90-family gene HSP81. Plant Cell Physiol. 35: 1207–1219.

Yan, J., Wang, J., Li, Q., Hwang, J.R., Patterson, C., and Zhang, H. (2003). AtCHIP, a U-box-containing E3 ubiquitin ligase, plays a critical role in temperature stress tolerance in Arabidopsis. Plant Physiol. 132: 861–869.

Yee, D. and Goring, D.R. (2009). The diversity of plant U-box E3 ubiquitin ligases: from upstream activators to downstream target substrates. J. Exp. Bot. 60: 1109–1121.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based Analysis of ChIP-Seq (MACS). Genome Biol. 9: R137.

Zhao, J., Favero, D.S., Peng, H., and Neff, M.M. (2013). Arabidopsis thaliana AHL family modulates hypocotyl growth redundantly by interacting with each other via the PPC/DUF296 domain. Proc Natl Acad Sci U S A 110: E4688–E4697.

Zilberman, D., Coleman-Derr, D., Ballinger, T., and Henikoff, S. (2008). Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. Nature 456: 125–129.

Zupanska, A.K., LeFrois, C., Ferl, R.J., and Paul, A.-L. (2019). HSFA2 Functions in the Physiological Adaptation of Undifferentiated Plant Cells to Spaceflight. Int. J. Mol. Sci. 20.

Chapter 5

Denyer, T., Ma, X., Klesen, S., Scacchi, E., Nieselt, K., and Timmermans, M.C.P. (2019). Spatiotemporal Developmental Trajectories in the Arabidopsis Root Revealed Using High-Throughput Single-Cell RNA Sequencing. Dev. Cell 48: 840–852.e5.

Lämke, J., Brzezinka, K., Altmann, S., and Bäurle, I. (2016). A hit-and-run heat shock factor governs sustained histone methylation and transcriptional stress memory. EMBO J. 35: 162–175.

Shahan, R., Hsu, C.-W., Nolan, T.M., Cole, B.J., Taylor, I.W., Vlot, A.H.C., Benfey, P.N., and Ohler, U. (2020). A single cell Arabidopsis root atlas reveals developmental trajectories in wild type and cell identity mutants. Cold Spring Harbor Laboratory: 2020.06.29.178863.

Shulse, C.N., Cole, B.J., Ciobanu, D., Lin, J., Yoshinaga, Y., Gouran, M., Turco, G.M., Zhu, Y., O'Malley, R.C., Brady, S.M., and Dickel, D.E. (2019). High-Throughput Single-Cell Transcriptome Profiling of Plant Cell Types. Cell Rep. 27: 2241–2247.e4.

Ståhl, P.L. et al. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science 353: 78–82.

Yanagisawa, S. and Schmidt, R.J. (1999). Diversity and similarity among recognition sequences of Dof transcription factors. Plant J. 17: 209–214.