Spatiotemporal Regulation of Gene Expression during Plant Meristem Development

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer.nat.)

vorgelegt von

Xiaoli Ma

aus

Hebei, China

Tübingen

2019

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation: 06.09.2019

Dekan: Prof. Dr. Wolfgang Rosenstiel

1. Berichterstatter: Prof. Dr. Marja Timmermans

2. Berichterstatter: Prof. Dr. Eric Kemen

ACKNOWLEDGEMENT

Time fly so fast. It was amazing to come to Tuebingen four years ago. Thanks to my supervisor Prof. Dr. Timmermans for giving me this opportunity to study in Germany. I also would like to thank her for taking me into a meristem world and learning this cool science. Your passion to science and open-mind in life are going to affect me for long time. Thanks for all your guidance, support, and help during my PhD study and dissertation writing.

I would like to thank my TAC meeting members, Dr. Liu and Dr. Markmann for your encouragement and advices. I also would like to thank Chang for giving me an opportunity to participate in his project. Additionally, thanks for all collaborators during my study. I would like to thank my reviewers for spending time reading my dissertation and my and examiners for coming my defense.

I also would like to thank Tuebingen BinAC team for computing server support. Most of my studies were finished on this high-performance computing system.

I would like to thank my colleagues in Timmermans Lab, especially Tom Denyer and Antje Feller. Tom not only worked together with me in scRNA-Seq project, he also offered many inputs for my dissertation. Antje and I not only worked on the LCM, she also offered lots of help on German translation stuff including my summary. I would also like to thank Kristine Hill for reading my introduction and offering some advices. All the other members (I will not list here one by one), there are lot of fun with you, exploring eating in different restaurant, canoeing, and playing card games. Thanks to all people who offered help to me during my PhD study, if I forgot to mention when I wrote here.

I would like to thank my friends who traveled together with me to explore this amazing world, which made my PhD life colorful and wonderful. I would like to thank some people for caring about me in some way in somewhere.

Finally, I would like to give my sincere appreciations to my parents, thanks for your understanding and supporting for my PhD study. Of course, I also would like to thank myself who worked hard and still would like to grow every day.

Table of Contents

ABBREVIATIONS

NGS	Next generation sequencing
TF	Transcript Factor
miRNA	microRNA
PPI	Protein-protein-interaction
ChIP	Chromatin immunoprecipitation
GRN	Gene regulatory network
nt	Nucleotide
pri-miRNA	Primary miRNA
pre-miRNA	Precursor hairpin miRNA
RT-PCR	Reverse transcription polymerase chain reaction
SNP	Single nucleotide polymorphism
Indel	Insertion-deletion
LCM	Laser capture microdissection
FACS	Fluorescence-activated cell sorting
TRAP	Translating ribosome affinity purification
INTACT	Isolation of nuclei tagged in specific cell type
ST	Spatial transcriptomic
GFP	Green fluorescent protein
scRNA-Seq	Single cell RNA sequencing
RNA-Seq	RNA sequencing
UMI	unique molecular identifier
STAR	Spliced transcript alignment algorithm
RPKM	Reads per kilobase per million
FPKM	Fragments per kilobase per million
RPM	Reads per million
HVG	Highly variable gene
PCA	Principle component analysis
PC	Principle component
KNN	K-nearest neighbor
TI	Trajectory inference
t-SNE	t-Distributed stochastic neighbor embedding
LRT	Likely ratio test

GO   Gene ontology

IGV   Integrative genomics viewer

ChIP-Seq   Chromatin immunoprecipitation with DNA sequencing

BS-Seq   Bisulfite sequencing

DAP-Seq   DNA affinity purification sequencing

SAM   Shoot apical meristem

RAM   Root apical meristem

QC   Quiescent center

CZ   Center zone

OC   Organizing center

SCN   stem cell niche

LRC   Later root cap

Col   Columella

MZ   Meristem zone

EZ   Elongation zone

DZ   Differential zone

RZ   Rib zone

V/Vas   Vasculature

L1   Layer1

L2   Layer2

LIST OF FIGURES

Appendix I

- **Figure 1. Sequenced single cells cluster by identity.**
- **Figure 2. Cell identity and developmental stage are reflected in the (sub)clustering.**
- **Figure 3. An unbiased, marker- gene-selection method is validated with transcriptional fusions.**
- **Figure 4. scRNA-Seq data of QC cells identifies novel QC marker genes.**
- **Figure 5. Trichoblast development is guided by progressive waves of gene expression.**
- **Figure 6. A Gene Regulatory Network (GRN) predicts key regulators during trichoblast cell differentiation.**
- **Figure S1. ScRNA-Seq of the *Arabidopsis* root reveals distinct cell-type specific clusters; Related to Figure 1.**
- **Figure S2. Key *Arabidopsis* genes reveal distinct expression profiles across clusters. Related to Figure 1.**
- **Figure S3. A localised GFP marker and mutant line validate cluster calling; Related to Figure 2.**
- **Figure S4. Discreet cell identities can be found within select subclusters; Related to Figure 1.**
- **Figure S5. Developmental stage-specific clusters can be identified. Related to Figure 1.**
- **Figure S6. Cortex development is guided by distinct waves of gene expression; Related to Figure 5.**
- **Movie S1. ScRNA-Seq of the *Arabidopsis* root reveals distinct clusters; Related to Figure 1.**

Appendix II

- **Figure 1. Tissue-specific genes contributing to cell identity. (*A*) Longitudinal section of a 14-day-old B73 seedling apex.**
- **Figure 2. Dynamic expression of individual gene family members determines cell identity.**
- **Figure 3. Divergent gene sets define functional SAM domains in maize versus *Arabidopsis.***

- **Figure 4. Combinatorial effects of multiple TFs distinguish cell identities.**
- **Figure 5. Dynamically-expressed meristem genes modulate important architectural traits.**
- **Figure S1. Precision of LCM.**
- **Figure S2. Predominant differential expression of individual gene family members distinguishes cell identities.**
- **Figure S3. CAST clustering and transcriptional conservation.**
- **Figure S4. Combinatorial effects of diverse TFs distinguishes cell identities.**
- **Figure S5. Dynamically expressed genes shape morphological variation.**

LIST OF TABLES

(In separate sheets)

4.2_Dataset

- **Dataset S1. RNA-Seq data revealing pri-miRNA and mature miRNA expression in the maize shoot apex.**
  Dataset S2. RNA-Seq data revealing the expression of enzymes involved in miRNA biology in the maize shoot apex.
- **Dataset S3. RNA-Seq data revealing gene expression in CZ of dcl1-2 and wildtype**
  Dataset S4. RNA-Seq data revealing miRNA-target gene expression in the maize shoot apex. Dataset S5. The regulatory mechanism of miRNAs.

Appendix I_Table

- **Table S1. RNA sequencing data generated in this study and Differentially Expressed Genes (DEGs) in bulk RNA-seq of protoplasted and un-protoplasted root tissues; Related to Figure 1 and Figure S1.**
- **Table S2. DEGs between clusters or between subclusters and marker genes used for cluster cell-identity calling; Related to Figure 1.**
- **Table S3. Cluster specific marker genes and DEGs in the QC over meristematic cells; Related to Figures 3 and 4.**
- **Table S4. Dispersed genes across trichoblast and cortex pseudotime and trichoblast Gene Regulatory Network (GRN) interactions; Related to Figures 5 and 6.**
- **Table S5. Full names of genes discussed in the text.**
- **Movie S1. ScRNA-Seq of the *Arabidopsis* root reveals distinct clusters; Related to Figure 1.**

Appendix II_Table:

- **Table S1. RNA-Seq metrics**
- **Table S2. Gene annotations and mean expression values (RPM)**
- **Table S3. Shannon entropy scores of all expressed genes**

Chapter 1. SUMMARY

Plants are products of meristems. The root apical meristem (RAM) gives rise to the below-ground root system, while the shoot apical meristem (SAM) generates all the above-ground organs. The RAM and SAM contain highly organized stem cell niches, characterized by the presence of distinct cell types at various developmental stages. Meristems are thus ideal model tissues to study molecular regulatory mechanisms during development, via the generation of high-resolution expression atlases. In this dissertation, I aim to study the spatiotemporal regulation of gene expression during plant meristem development at a genome-wide level, including transcriptional regulation and post-transcriptional regulation.

In my first study, high-throughput single cell RNA sequencing (scRNA-Seq) was used to build a cellular resolution gene expression atlas of the *Arabidopsis* root that includes all major cell types. In total, 4,727 single cell profiles were generated and analyzed. Developmental trajectories along root development were built. These depict a cascade of developmental progressions from stem cell to final differentiation. New regulators and downstream genes that define cell types or control cell state transition during the development were identified. This study demonstrates the power of applying scRNA-Seq to plants, and provides a unique spatiotemporal perspective of root cell differentiation.

In my second study, a high-resolution maize shoot apex expression atlas in appendix II (Knauer et al., 2019) was used to investigate the spatiotemporal action of microRNAs (miRNAs) during development. Expression patterns of miRNA precursors and mature miRNA accumulation were examined, revealing that miRNA accumulation is regulated at both transcriptional and post-transcriptional level. Examples of the latter included effects on miRNA processing and/or stability in the vasculature and the stem cell population at the SAM tip, as well as the movement of miRNA within developing leaf primordia. By integrating data from RNA-Seq and degradome-Seq, a system was devised to predict the regulatory mechanism employed by miRNAs on their targets. This study provides a first comprehensive investigation of how the activity of the miRNAs that are critical to developmental pattern are regulated across space and time, revealing inputs from processes of regulating transcription, processing, stability, mobilities and miRNA efficacy.

Chapter 2. ZUSAMMENFASSUNG

Pflanzen sind Produkte von Meristemen. Das apikale Wurzelmeristem (RAM) führt zum unterirdischen Wurzelsystem, während das apikale Sprossmeristem (SAM) alle oberirdischen Organe erzeugt. RAM und SAM enthalten hochorganisierte Stammzellnischen, die sich durch das Vorhandensein verschiedener Zelltypen in verschiedenen Entwicklungsstadien auszeichnen. Meristeme sind somit ideale Modellgewebe, um molekulare Regulierungsmechanismen während der Entwicklung zu untersuchen; unter anderem durch die Erzeugung hochauflösender Expressionsatlanten. In dieser Dissertation möchte ich die raumzeitliche Regulierung der Genexpression während der Entwicklung von Pflanzenmeristemen auf genomweiter Ebene untersuchen, einschließlich der transkriptionellen - und der posttranskriptionellen Regulierung.

In meiner ersten Studie wurde die Hochdurchsatz-Einzelzellen-RNA-Sequenzierung (scRNA-Seq) verwendet, um einen Genexpressionsatlas der Arabidopsiswurzel zu erstellen, der alle wichtigen Zelltypen umfasst. Insgesamt wurden 4.727 Einzelzellenprofile generiert und analysiert. Entwicklungspfade entlang der Wurzelentwicklung wurden aufgebaut. Diese stellen eine Kaskade von Entwicklungsprozessen von der Stammzelle bis zur endgültigen Differenzierung dar. Neue Regulatoren und Zielgene, die Zelltypen definieren oder den Übergang von Zellzuständen während der Entwicklung steuern, wurden identifiziert. Diese Studie zeigt die Leistungsfähigkeit der Anwendung von scRNA-Seq in Pflanzen und bietet eine einzigartige raumzeitliche Perspektive der Wurzelzelldifferenzierung.

In meiner zweiten Studie wurde ein hochauflösender Maisspross-Apex-Expressionsatlas in Anhang I (Knauer et al., 2019) verwendet, um die raumzeitliche Wirkung von microRNAs (miRNAs) während der Entwicklung zu untersuchen. Expressionsmuster von miRNA-Vorläufern und reifer miRNA-Akkumulation wurden untersucht, was zeigt, dass die miRNA-Akkumulation sowohl auf transkriptioneller als auch auf post-transkriptioneller Ebene reguliert wird. Beispiele für Letztere beinhaltet Effekte auf die miRNA-Verarbeitung und/oder Stabilität im Gefäßsystem und in der Stammzellpopulation an der SAM-Spitze sowie die Bewegung der miRNA innerhalb der sich entwickelnden Blattprimordien. Durch die Integration von Daten aus RNA-Seq und Degradome-Seq wurde ein System entwickelt, um den

Regulierungsmechanismus der miRNAs auf ihren Targets vorherzusagen. Diese Studie bietet eine erste umfassende Untersuchung, wie die Aktivität der für das Entwicklungsmuster kritischen miRNAs über Raum und Zeit reguliert wird, und zeigt Inputs aus Prozessen der Regulierung von Transkription, Verarbeitung, Stabilität, Mobilitäten und miRNA-Wirksamkeit.

Chapter 3. Introduction

This dissertation describes three studies of plant meristems, where next generation sequencing (NGS) datasets have been analyzed, offering spatiotemporal insights into transcriptional and post-transcriptional regulation of gene expression. This chapter introduces some general concepts of regulation of transcript accumulation during plant development, presents structures of plant meristems, and describes established technologies for measuring gene expression level globally and spatiotemporally. In addition, this chapter also describes NGS platforms and analysis methods used in this work. Subsequent chapters include background information relevant to the studies presented in each chapter. Finally, objectives of this dissertation are presented.

3.1 Gene regulation during plant development

Plant development reflects an elaborate program of regulatory interactions that control gene expression in precise spatial and temporal patterns. Over the years, our understanding of the actions of Transcription Factors (TFs), microRNAs (miRNAs), and other signaling molecules, in the control of gene expression has increased, exponentially. In particular, the development of large-scale genomics approaches has been instrumental. These have revealed a surprising complexity in how the dynamic patterns of gene expression, which characterize the developmental progressions from stem cell to final differentiation, come about.

3.1.1 Transcription Factors

A substantial number of genes show a dynamic pattern of expression during plant development that contributes to the specification and differentiation of distinct cell and tissue types. Gene expression is typically regulated at multiple levels. On a transcriptional level, TFs ensure that genes are expressed properly either quantitatively, spatially, or temporally, during plant development.

TFs form complexes together with other co-regulators to provide their activation and repression activity. TF members within a given complex can either directly or indirectly cooperate and confer differential occupancy, resulting in the complex behaving dynamically at any given promoter, allowing functional specificity (Biggin, 2011). Direct

cooperativity such as direct protein–protein-interactions (PPIs) between the functional combination TFs can enhance the level of binding and determine cell types. For example, leaf polarity is typically determined by AS1-AS2 cooperation at the level of DNA binding (Phelps-Durr et al., 2005; Xu, 2003). In signaling pathways, TFs employ different partner TFs to regulate different target genes in different cellular contexts (Doerks et al., 2002). On the other side, indirect combination such as collaborative binding - multiple TFs recognize closely spaced binding sites within enhancers independently - can compete TF binding sequences at sites of accessible DNA with nucleosome, forming the cell-type-specific regulation (Reiter et al., 2017). Consequently, a TF complex regulates the expression of the same target genes differently in different cell types.

TFs can be either direct or indirect targets in tissue-specific regulation (Biggin, 2011). In signaling pathways, TFs (like activator insufficiency or cooperative activation) are regulated by developmental cell signaling to control target gene transcription in the proper context, preventing their expression in all other cells (Doerks et al., 2002). However, TFs typically recognize short, degenerate DNA sequences that as a result occur quite frequently in most genes. The action of TFs in gene regulation is therefore far-more complex than often depicted in basic textbooks. Data from genome-wide chromatin immunoprecipitation (ChIP) analysis indicates that TFs occupy just a fraction of all possible sites. Binding site selection is in part determined by chromatin accessibility, although this does not fully explain the observed patterns of TF occupancy. On the other hand, TFs are found to occupy many hundreds or even thousands of sites where they have no apparent effect on transcriptional output.

In addition, feedback, feedforward, and cross-regulatory loops among TFs leads to the formation of complex gene regulatory networks (GRNs). These guides gene cascades to control tissue development (Li and Davidson, 2009). In the process of plant evolution, the emergence of species-specific tissues and novel traits are a result of the reorganization of GRNs. As such, GRNs play an important role in plant developmental biology. There are two types of regulatory networks: continuous networks and discrete networks. Continuous networks connect each TF to all the genes, and show a weight on the connection to indicate the regulation degree, including many weak transcriptional regulatory interactions. Meanwhile, discrete

networks connect a TF to a subset of directed target genes, without weight on the connection (Biggin, 2011).

In Chapter 4.1, a continuous GRN based on dynamically expressed TFs along the *Arabidopsis* trichoblast developmental trajectory is described. This reveals key players and suggests their regulatory reaction. Complex wiring between components, including many weaker connections, mediates the transition of cells from the stem cell niche to the root maturation zone. In addition, the network also implies a high degree of feedback regulation toward to the meristem, not only from the elongation zone but also from the genes in the maturation zone (Denyer, Ma et al., 2019). We constructed a discrete GRN based on the occurrences of TF binding motifs within the promoters of tissues-specific genes and revealed the combinatorial effects of diverse TFs promoting stem cell fates (Appendix II). TFs' cell-type-specific cascades contribute hugely to spatiotemporal gene expression, even though some TF complex are expressed more broadly than the target genes they regulated. In addition, GRNs show hierarchical transcriptional regulation. In highly connected networks, mis-regulation of the top can lead to a collapse of the entire GRN, such as KN1 (Steffen et al., 2019).

3.1.2 microRNAs

miRNAs are small regulatory RNAs of about 21-24 nucleotides (nt) in length. MiRNAs play a role in almost every aspect of plant development including, embryogenesis, meristem organization, leaf development, flowering, and flower organ identity (D'Ario et al., 2017). Their importance in development is obvious from the severe defects observed from mutants within the genes required for miRNA biogenesis. MIR genes are transcribed into primary miRNAs (pri-miRNAs) containing a stem-loop structure which is processed into the precursor hairpin miRNA (pre-miRNA) by the endonuclease DICER-LIKE1 (DCL1) protein complex (Chen, 2009). Pre-miRNAs are further processed by DCL1 to generate the miRNA/miRNA* duplex, which is methylated on the ribose of the last nucleotide by the methyltransferase HEN1 to stabilize the immediate (Huang et al., 2009). After removal of the complimentary miRNA* sequence, the mature single-stranded miRNA is incorporated into an ARGONAUTE (AGO) effector complex to regulate mRNA targets (Chen, 2009; Rogers and Chen, 2013). Different from TFs that function at transcriptional level, miRNAs regulate gene expression post-transcriptionally. miRNAs bind in a homology-

dependent manner to target mRNAs and trigger the degradation through transcript cleavage, and/or their translational inhibition.

microRNAs are ideally suited to diversify gene expression programs and contribute to specification of distinct cell types during development by regulating genes post-transcriptionally. Base pairing between small RNAs and target transcripts provides an exceptionally high degree of specificity, particularly in plants, and the direct repression of targets allows for rapid cell fate transitions. Small RNA regulation also confers sensitivity and robustness onto gene networks (Plavskin et al., 2016; Schmiedel et al., 2015), and their movement through plasmodesmata is regulated independently from mobile proteins (Skopelitis et al., 2018). Small RNAs are mobile and represent a class of signalling molecules with properties distinct from other developmental signals, such as hormones, transcription factors, and peptides ligands. Moreover, recent findings indicate that the concentration gradients generated by the movement of small RNAs from a fixed source can generate sharp domains of target gene expression through a morphogen-like readout. This reflects a non-linear, threshold-based readout, which is a highly sensitive to the relative ratio of small RNA-to-target levels (Skopelitis et al., 2017). When this ratio is above a given threshold, small RNAs can clear target transcripts or protein accumulation. Contrastingly, below this threshold, small RNAs can reduce target gene expression noise and show a rheostat behaviour. Thus, there are multiple types of small RNA-mediated regulatory modes in development (Skopelitis et al., 2017).

Consistent with miRNAs having important roles in development, many display dynamic spatiotemporal patterns of expression/activity. How these patterns come about is addressed in Chapter 4.2, where I show how the spatiotemporal action of plant miRNAs during development is regulated at multiple levels through the regulation of miRNA transcription, processing/stability, and activity.

3.2 Opportunities provided by high-throughput sequencing to understand gene regulation

In 1869, Friedrich Meischer discovered DNA in Tübingen (Dahm, 2005). Much later, in 1943, DNA was proposed as the genetic material storing general heritable information (Avery et al., 1943). Hypotheses on how this might work followed from the

discovery of the double helix structure of DNA by Watson and Crick in 1953 (Watson and Crick, 1953). However, it was 1958, when Francis Crick stated the central dogma of molecular biology, which explained the flow of genetic information across all living organisms (Cobb, 2017; Crick, 1970): *DNA is transcribed into RNA that is translated into protein*.

Accurate quantification of gene expression across specific tissues or cells is required for a comprehensive understanding of the complexity of gene regulation during plant development. In the past few decades, the technologies available to do this have developed rapidly. While methods such as the classical northern blot and reverse transcription polymerase chain reaction (RT-PCR) allowed the comparative quantitative analysis of gene expression for select genes (Alwine et al., 1977; Nick and John, 1993), development of DNA microarrays took gene expression quantification onto a high-throughput level (Shchena et al., 1995). However, the subsequent development of high-throughput parallel sequencing technologies revolutionized the way of quantifying gene expression. These sequencing-based technologies make it much easier and faster to measure gene expression, and also reveal information inaccessible by microarrays, such as expression of rare or novel RNA variates (Levy and Myers, 2016). A brief timeline of development of gene expression quantification is shown in Fig 1.
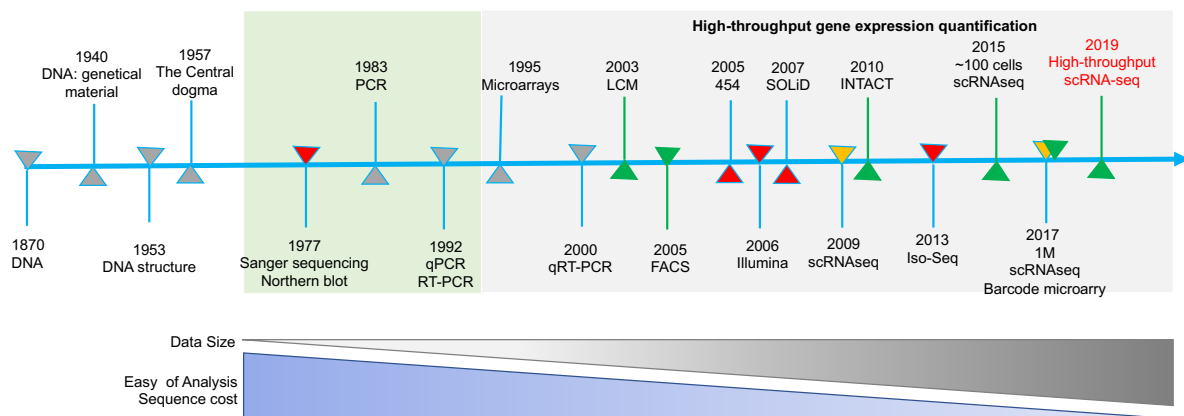
Figure 1. Timeline of gene expression quantification related method discovery. Red arrowheads indicate the introduction of novel sequencing methods. Green arrowheads indicate innovations in the plant filed. Yellow arrowheads indicate innovation in animal fields. Gray arrowheads indicate the milestone of relevant discoveries and innovations.

3.2.1 High-throughput sequencing

The classical sequencing methodologies, such as Maxam and Gilbert, and Sanger sequencing, processed single templates at a time (Alllan and Walter, 1977; Sanger et al., 1977). In contrast, the so-called next-generation sequencing (NGS) methods are high-throughput and can process millions of DNA sequences in parallel (Levy and Myers, 2016). This allows for the quantitative investigation of whole transcriptomes with high sensitivity, high reproducibility, and in a very short time. The first commercial NGS technology was pyrosequencing, which was developed by 454 Life Science (named as 454 sequencing). Pyrosequencing involves bead-based emulsion PCR (emPCR), with hundreds of thousands of unique DNA templates on each bead (Elahi and Ronaghi, 2004). Another method, 454 sequencing detects the light signal released during a flow of enzymatic reactions during each dNTP incorporation (Figure 2C) (Margulies et al., 2005). Ion Torrent (Life Technologies product, also named post-light sequencing), detects pH differences caused by hydrogen release following each dNTP incorporation (Figure 2D) (Rothberg et al., 2011). Besides 454 and Ion Torrent sequencing, oligonucleotide ligation and detection (SOLiD) (Mardis, 2008) was a commonly used option. DNA fragments are sequenced by ligation (SBL). Each fluorometric signal represents a dinucleotide, as it used the two-base-encoded probes. This greatly facilitates the sensitivity of base calling as each base is probed multiple times (Mardis, 2008). However, each ligation signal represents one of several possible dinucleotides, since four fluorometric colors represent 16 two-base-combinations (Figure 2B). This leads to the term 'color-space' (rather than base-space), which must be deconvoluted during data analysis (Goodwin et al., 2016).
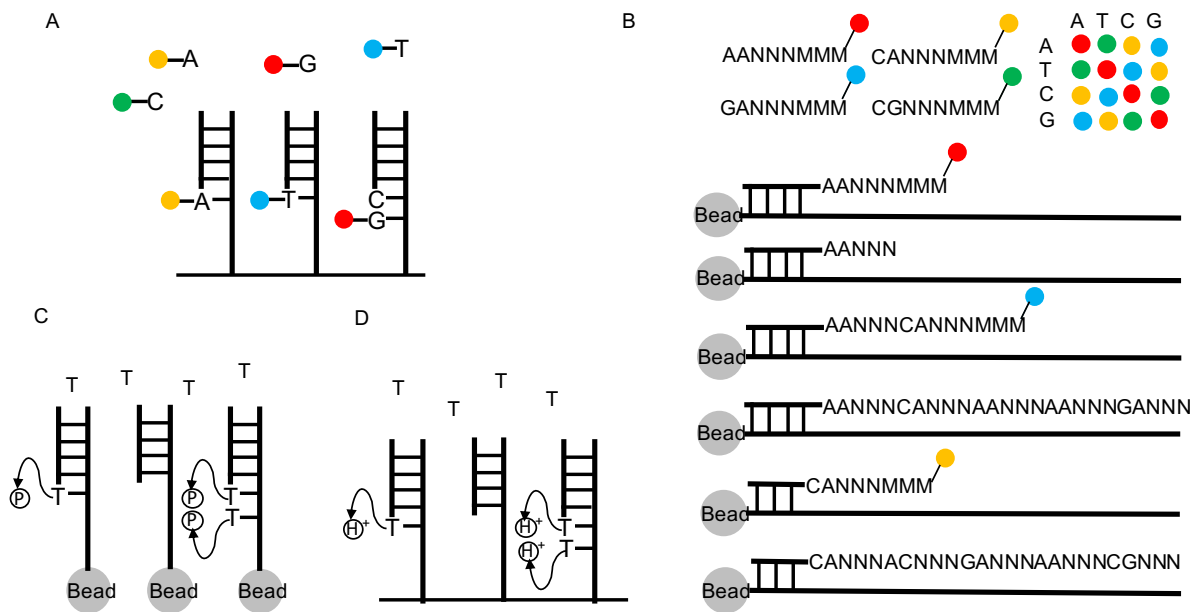
Figure 2. NGS sequencing. (A) Illumina sequencing. All four terminal nucleotides, carrying unique fluorescent labels, and DNA polymerase are added. One complementary nucleotide is added into each template at each time. Signal information is recorded before cleaving of fluorescent and terminating groups and entering into next round base incorporation. (B) SOLiD sequencing. Sequencing primers, DNA ligase and many unique probes (fluorescently labeled based on their first two bases) are added. One complementary probe is ligated to one template each time. After the fluorescence signal is recorded, the three universal bases and fluorophore are cleaved off and they enter into next round. (C) 454 sequencing. One type of natural non-terminating nucleotides is added and one pyrophosphate molecule was released at each nucleotide incorporation. Pyrophosphate was finally converted into light signal to record the nucleotide sequence. (D) Ion Torrent sequencing. Same as 454, one type of natural non-terminating nucleotide is added at each incorporation. Different from 454, H+ was released release in each nucleotide incorporation. Single information coming from pH change was detected and recorded before entering next round nucleotide incorporation.

Currently, Illumina (Solexa) sequencing is by far the most popular NGS technology (Goodwin et al., 2016). It uses flow-cell bound cluster bridge amplification and fluorescent signals to detect each nucleotide addition (Figure 2A) (Heather and Chain, 2016). The primers, complemented to the DNA library fragments adaptors, are densely coated on the surface of the flow cell. This enables the DNA fragments to randomly attach to the surface of the flow cell, creating bridged structures by an active heating and cooling step. Reactants and an isothermal polymerase interact with the bridge fragments to generate double strands, and when the denaturation occurs, the

double stranded DNA becomes single stranded fragments attached to the surface of the flow cell. Repetition of this process leads to clonal clusters of localized identical strands (Figure 3) (Goodwin et al., 2016; Mardis, 2008). The principle of nucleotide addition one by one and detecting the incorporation of each nucleotide is called sequencing-by-synthesis (SBS) (Fuller et al., 2009). Illumina sequencing applies the SBS mechanism and elongates the primer in a stepwise manner. Each of the four DNA bases are attached a different fluorophore in addition to a terminating group (Figure 2A). When the correct base is linked to the template, it is imaged and the nucleotides that have not been incorporated are washed away. The fluorescent branch and the terminating group are cleaved and the cycle can be repeated (Guo et al., 2008).
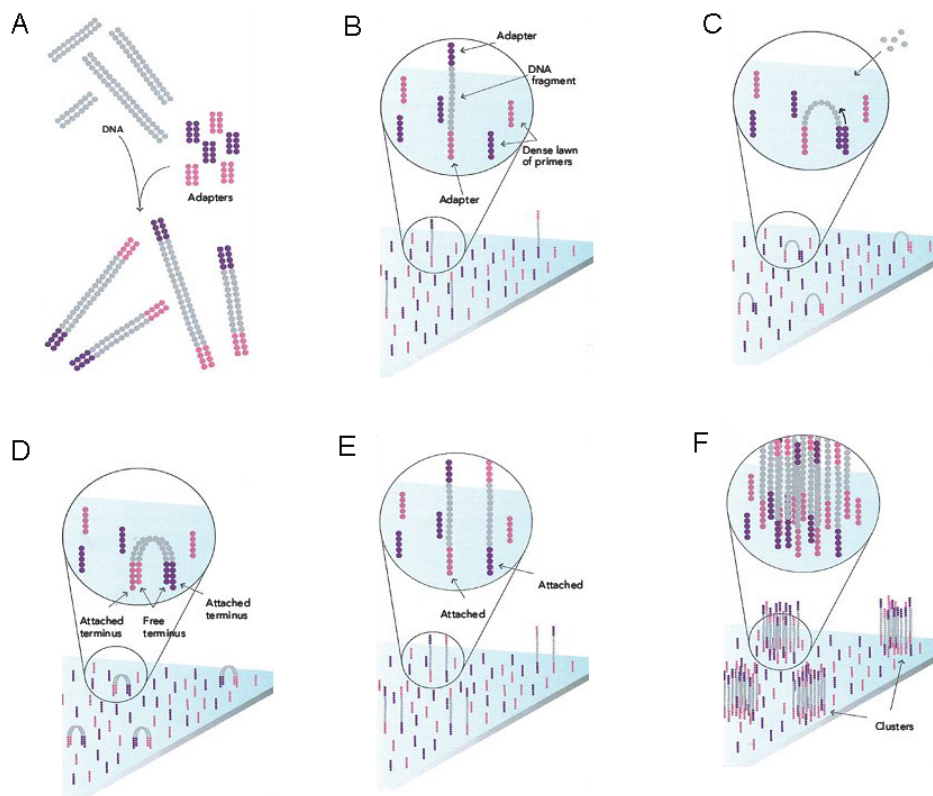
Figure 3. Illumina workflow of bridge amplification generating clusters. (A) Genomic DNA sample preparation; (B) Single-stranded DNA fragments are attached randomly to the surface of flow cell channels; (C) Bridge amplification; (D) Forming double-stranded bridge; (E) Denature forming single-stranded templated for next round amplification; (F) Generating dense clusters. (Image from www.illumina.com)

Illumina sequencing is currently the most cost-effective and has the highest throughput, although 454 sequencing produces longer reads and SOLiD sequencing is more accurate in base calling (Heather and Chain, 2016). Besides that, Illumina dominates the short-read sequencing industry owing to its mature technology, such as a high level of cross-platform compatibility and its wide range of instruments. Illumina instruments range from the low-throughput MiSeq (~20 million reads/run) to the ultra-high-throughput HiSeq X (~6 billion reads/run) (Goodwin et al., 2016).

Library design

The decreasing costs of sequencing and increasing accessibility of NGS machines has allowed researchers to develop a variety of measurement methods based on NGS. For example, the profiling and quantification of genome-wide transcriptomes (Wang et al., 2009), small RNAs (Hagemann-Jensen et al., 2018), TF binding (Johnson et al., 2007; Bartlett et al., 2017), DNA methylation (Li and Tollefsbol, 2011), chromatin structure (Belton et al., 2012), and single nucleotides variants/polymorphisms (Kahvejian et al., 2008; Reuter et al., 2015). A successful -omics sequencing study requires the right choice of library and sequencing methodology (strand-specific or non-strand-specific library; single- or paired- end sequencing), suitable sequencing read length and depth, and the correct number of replicates. For example, single-end sequencing is appropriate for small RNA-Seq (small RNA sequencing) and RNA-Seq used for quantifying gene expression. Paired-end sequencing enables more accurate read alignment and the ability to detect single nucleotide polymorphisms (SNP) and insertion-deletion (indel) variants. Sequencing read-length can be shorter or longer, based on the experimental purpose. Longer reads are better for isoform analysis, while shorter reads are suitable for gene expression quantification. Also sequencing depth (the total number of reads sequenced for the given sample) depends on the goal of the experiment. The more reads, the larger number of genes detected. In most of eukaryotic bulk transcriptomes, five million mapped reads are sufficient to accurately measure medium to highly expressed gene (Sims et al., 2014).

The requirement of full-length transcript sequencing verses 3' end RNA sequencing also needs to be considered. 3' RNA-Seq was designed to quantitatively measure the abundance of mRNA fragments, which produce only one 3' end sequence per transcript (Torres et al., 2008). This overcomes the expression level estimation biases

introduced by transcript length when using full length transcript sequencing. Moreover, yielding the same amount of reads, 3' RNA-Seq has the potential to detect more lower expressed genes than full-length transcript sequencing (Tandonnet and Torres, 2017). Therefore, for model species, such as *Arabidopsis* and maize, 3' RNA-Seq is currently a better choice for quantifying gene expression and identifying differential expression.

Spatially resolved gene expression by bulk RNA-Seq

High-throughput sequencing technologies provide the opportunity to obtain the transcript profiles of any tissue of choice. However, given the complexity of most plant tissues, elucidating the molecular mechanism driving their development requires transcriptome-wide analysis of spatially resolved gene expression patterns to capture expression information on specific target cells. Laser capture microdissection (LCM) (Emmert-Buck et al., 1996), fluorescence-activated cell sorting (FACS) (Birnbaum et al., 2005), tissue specific mRNA purification by translating ribosome affinity purification (TRAP) (Heiman et al., 2014), isolation of nuclei tagged in specific cell types (INTACT) (Moreno-Romero et al., 2017), and spatial transcriptomics (ST) based on barcoded microarrays (Giacomello and Lundeberg, 2018) have been developed to isolate the plant input material of interest from whole tissues so that they can be analyzed at a high-resolution level. Different from animal cells, plant cells are surrounded by a cell wall which complicates the applications of many animal cell isolation technologies to plant tissues. Nonetheless, LCM is commonly used to dissect cell populations from a section of complex tissue (Nakazono, 2003; Scanlon et al., 2009). Tissues are scanned under a microscope and micro-dissected with a laser beam to separate them for further analysis (see Chapter 4.2 and Appendix II). Similarly, a FACS-based method was adapted to isolate plant cells of interest based on reporter lines expressing a fluorescent protein in defined cells of interest (Birnbaum et al., 2005). Tissues are protoplasted into single cells, and GFP-positive cells separated using FACs to study gene expression. This technology has been used to generate detailed gene expression atlases for the *Arabidopsis* root and inflorescence (Birnbaum et al., 2003; Efroni et al., 2015; Yadav et al., 2009). The more recently developed INTACT and TRAP methods eliminate the need for protoplasting. In INTACT, nuclei of cells or tissues of interest are labeled through transgenic expression of a biotinylated nuclear envelop protein, and then affinity purified on streptavidin-coated beads (Moreno-

Romero et al., 2017). Similarly, tissue specific mRNA purification by TRAP combines cell-type-specific transgene expression with affinity purification of translating ribosomes (Tian et al., 2019). Finally, the ST method uses barcoded oligo-deoxythymidine (oligo-dT) primers spotted in clusters on microarrays to capture transcripts from intact tissue sections for cDNA synthesis and sequencing. This technology thus couples transcriptomics to histological imaging to simultaneously quantify and localize gene expression (Giacomello and Lundeberg, 2018).

Each of these technologies have their own strengths and weaknesses. Sample throughput and reliance on fixed tissue are major drawbacks of LCM; however, it is applicable to broader range of species. The time-consuming production of transgenic plant lines (> 6 months to generate stable lines carrying the desired transgene) limits the broader application of FACS, INTACT, and tissue specific TRAP. Whereas the array-based method ST suffers from a limited resolution and is only available to those model species with a reference genome.

Spatially resolved gene expression by single cell RNA-Seq

The gene expression analysis methods described above involve bulk samples, yielding an average gene expression from all target cells. This potentially masks cellular heterogeneity - transcripts that are expressed only in a minority of cells. Measuring the expression profiles of cells individually can reveal such heterogeneity and identify subpopulation expression variability, reflecting for instance developmental time or cell cycle progression, even within a specific tissue (Etzrodt et al., 2014; Qiu et al., 2017; Skelly et al., 2018). In 2009, RNA-Seq was firstly applied to examine the whole-transcriptome of a single cell in mouse blastomeres (Tang et al., 2009). Since then, methods for studying single cells transcriptomes have developed rapidly, and the throughput of single cell sequencing technologies has increased from dozens to millions of cells in animals (Svensson et al., 2018). The increasingly high-throughput nature of single cell RNA sequencing (scRNA-Seq) has been facilitated by the development of droplet technology (Klein et al., 2015; Macosko et al., 2015) and increased automation (Zheng et al., 2017). In brief, a cell is encapsulated within an oil droplet and lysed, and its transcripts reverse transcribed on barcoded beads. Following library production and sequencing, transcripts from individual cells can be identified from the bead-derived barcode and individual transcripts accounted for using

unique molecular identifiers (UMI) (Prakadan et al., 2017). As scRNA-Seq is able to measure transcript accumulation at the cellular level, it brings revolutionary discoveries to many fields such as deconstructing cell type in various tissues (Villani et al., 2017), tracing cell lineage and cell fate commitment in embryonic development and differentiation (Nelms and Walbot, 2019; Semrau et al., 2017; Trapnell et al., 2014), inferring transcription dynamics and regulatory networks (Dixit et al., 2016; Tanay and Regev, 2017), identifying tumor heterogeneity (Venteicher et al., 2017), and detecting variation in stress response (Ricci et al., 2018).

Different from mammalian systems, the complexity of plant cell structure and physiology – including the cell wall, vacuole, high osmotic pressure sensitivities, and high cell size variability and chloroplasts - increases the difficulty in processing tissues for single cell transcriptomic profiling. Nonetheless, low throughput scRNA-Seq has been performed successfully in plants (Efroni et al., 2015). In Chapter 4.1, I describe the first application of high-throughput scRNA-Seq in plants is described. This yielded an expression atlas of the *Arabidopsis* root composed of thousands of independently profiled cells which provides a spatiotemporal perspective of root cell-type differentiation at a resolution not previously achievable (Denyer, Ma et al., 2019)

3.2.2 Computational analysis of -omics data

Currently, NGS has applied broadly in the life science research community, but no single analysis pipeline can be used in all cases. Analysis strategies must be determined based on experimental and sequencing specificities. Given the aims of my dissertation, here I primarily focus on analysis of RNA-Seq data. Here, the major analysis steps include quality control, read alignment (mapping), quantification, normalization, differential expression analysis, functional annotation, and visualization (Conesa et al., 2016). In addition to the above, for species without a high-quality reference genome, a transcript assembly step is required. Furthermore, scRNA-Seq requires identifying variable genes (feature selection), reducing dimension, clustering cells, assigning cell types, building developmental trajectories and finally visualizing the result (Denyer, Ma et al., 2019).

Quality control

The quality of sequencing data directly determines the success of the project. Quality estimation of transcriptome sequencing data involved in analysis of raw reads, read alignment, and reproducibility. For scRNA-Seq, one more parameter, the reads per cell, is added. NGS raw reads are recorded in a *fasta* file, which includes the nucleotide sequence (a sequence read) and the quality of sequence reads (*phred* score). FASTQC (Andrews, 2010) is a popular tool to perform QC analyses on Illumina sequencing reads. Its checks include analysis of *phred* score distribution showing sequence quality, GC content directing sequence bias or contamination, duplicate sequences indicating PCR artifacts, and overrepresented sequences reflecting sequence contamination.

The percentage of mapped reads (alignment will be described later) is an important evaluation parameter, exposing potential contamination of sequencing samples and determining sequencing accuracy. However, the mapping ratio depends on the sequencing type, and the quality and complexity of the reference genome. For example, the mapping ratio in *Arabidopsis* is higher than that of maize, because the maize reference genome quality is lower, and its complexity is higher. For full length transcript sequencing, uniformity of read coverage across the gene body provides an additional valuable measure to assess library quality, as the sequencing reads will show a distribution bias when the RNA quality is low, for instance, due to RNA degradation. RSeQC (Wang et al., 2012) is a popular mapping quality control tool, calculating the reads distribution on the gene body.

RNA-Seq is highly reproducible (Marioni et al., 2008). Variation is typically introduced by biological variation (Hansen et al., 2011). Several biological replicates are necessary to quantify gene expression by RNA-Seq. Increasing the number of replicates can overcome biological variability at a certain level, but also risks increasing the difficulty of statistical analysis and introducing the batch effects during sample library preparation (Liu et al., 2014). Data reproducibility between the replicates at a global gene expression level is typically estimated using Pearson's correlation coefficient analysis. This calculates whether two variables are linearly related. The higher the Pearson's coefficient, the better correlation among replicates. The expected coefficient of Pearson correlation among good replicates is above 0.95 (Mortazavi et al., 2008), depending on the biological variation or heterogeneity of the

material under study. It can be used to detect and correct sample mis-annotation resulting from sample preparation, library construction, or sequencing.

Alignment and quantification

Read alignment (mapping) forms an important step in NGS analysis, such as RNA-Seq, small RNA-Seq, ChiP-Seq (chromatin immunoprecipitation with DNA sequencing), and BS-Seq (Bisulfite sequencing). For RNA-Seq analysis, there are two mapping strategies to identify gene expression, dependent on reference genome availability and quality. Without a good quality reference genome, reads are first assembled into transcripts and subsequently mapped back onto these transcripts to calculate the number of reads per transcript. The availability of a high-quality reference genome allows reads to be mapped directly onto the genome to measure gene expression. Spliced transcript alignment to a reference (STAR) is a leading aligner and accomplishes the alignment step faster and more accurately than other current alternatives. It takes a new alignment algorithm that searches sequential maximum mappable seed in uncompressed suffix arrays followed by seed clustering and stitching procedure (Dobin et al., 2013). Next, HTSeq (Anders et al., 2015) or FeatureCounts (Liao et al., 2014) are applied to assign mapped reads to gene to estimate gene expression.

As scRNA-Seq data is more complicated than bulk RNA-Seq data, scRNA-Seq analysis requires additional analysis: barcode processing. For example, 26 nucleotides of Read1 contain the UMI and cell barcode sequences and 98 nucleotides of Read2 capture the biological information. Tools such as Cell Ranger integrated alignment (same as bulk RNA-Seq analysis, STAR is used as read mapper), barcode assignment and UMI counting. During barcode assignment, Cell Ranger can automatically infer intact barcodes based on all the known listed of barcodes, which allows 1-Hamming-distance away from an observed barcode, considering sequencing error. The final cell barcodes are determined based on the distribution of UMI counts (all top barcodes within the same order of magnitude are potential cell barcodes). For UMI counting, UMIs with 90% base call accuracy were considered valid if they are not homopolymers. When a UMI is 1-Hamming-distance away from another UMI (with more reads) with the same cell barcode, it is corrected to the UMI with more reads. Finally, PCR duplicates (same barcode sequence, same UMI tag, and mapped to the

same gene) are removed. Only confidently mapped (MAP 255), non-PCR duplicates with valid barcodes and UMIs should be used to generate the gene-barcode matrix (Zheng et al., 2017).

Normalization

Normalization is an important step to obtain the signal of interest by reducing unwanted biases generated from capturing efficiency, sequencing depth, dropouts, and other technical effects. In general, there are two types of normalization: within-sample normalization and between-sample normalization (Vallejos et al., 2017). Within-sample normalization aims to remove the gene-specific biases (e.g., gene length), which makes gene expression comparable within one sample (such as RPKM or FPKM, reads/fragments per kilobase per million). In contrast, between-sample normalization is used to adjust sample-specific differences (e.g., sequencing depth and capture efficiency) to enable the comparison of gene expression between samples (such as RPM, reads per million, or quantile normalization). Generally, these simple normalization strategies are based on sequencing depth or upper quartile. scRNA-Seq generates abundant zero-expression values and has a higher level of technical variation than bulk RNA-Seq. Using bulk RNA-Seq normalization may cause overcorrection or introduce artifacts in scRNA-Seq for lowly expressed genes. Seurat is popular tool to analysis single cell data. It employs a global-scaling normalization method that normalizes the gene expression measurements for each cell by the total expression, multiplies this by a scale factor (10,000 by default), and log-transforms the result (Butler et al., 2018; Satija et al., 2015). If UMI or spike-ins are used, normalization can be further refined based on the performance of UMI/spike-ins (Dillies et al., 2013; Evans et al., 2018). In order to reduce the domination of the highly expressed genes and offer them equal weight in downstream analysis, a linear transformation (scaling) was applied before dimensional reduction: shift the expression of each gene, so that the mean expression across cells is zero; scale the expression of each gene, so that the variance across cells is one (Butler et al., 2018; Satija et al., 2015). In summary, normalization and scaling aids to reduce technical noise.

Feature selection and dimension reduction

scRNA-Seq data has a high dimensionality (M x N), which involves thousands of genes (M) and thousands of cells (N). In fact, only select genes are informative and reflect biological variation. Feature selection aids to remove the uninformative genes and identifies the most features to reduce the number of dimensions used in downstream analysis, which can largely speed up the calculations of large-scale scRNA-Seq data. Based on the assumption that the genes with highly variable expression across cells are a consequence of biological effects rather than technical noise, Seurat implements unsupervised feature selection algorithms to identify highly variable genes (HVGs) (Butler et al., 2018; Satija et al., 2015).

After feature selection, principle component analysis (PCA) is used to merge the information from correlated feature genes into one principle component (PC) (Lever et al., 2017). This aids to reduce the dimensions and capture the greatest amount of variance in the data. In principle component analysis, PC1 explains the largest percentage of variance of the data and PC2 explains the second highest percentage of variance of the data, and so on. The PCs are ranked based on the percentage of variance explained by each one, and the lower ranking of the PC contributes to explaining the least variance of the data. Therefore, the lower ranked PCs cannot increase any information on the demonstration of the biological variability of the cells except by increasing the computational load. In Seurat, an 'Elbow plot' can be generated by a heuristic method: a ranking of principle components (Butler et al., 2018; Satija et al., 2015). For example, when we observe an 'elbow' around 10 PCs, it suggests that the majority of true signal is captured in the first 10 PCs. In summary, dimensionality reduction projects the data into a lower dimensional space by optimally preserving some key properties of the original data.

Cell subpopulation identification (clustering cells and assigning cell types)

PCA is often followed by clustering analysis to further identify sets of genes that define cellular subgroups and the biological function and significance of each group. The main goal of clustering is to separate individuals into subsets based on their similarity or distance between the data points. This allows heterogeneity within the population to be identified along with the genes that are responsible for these differences. Approaches for clustering cells can be mainly grouped into two categories based on whether prior information is used. If a set of known markers was used in clustering,

the method is prior information based supervised clustering, which would complicate the input and introduce bias. Alternatively, unsupervised clustering methods can be used for *de novo* identification of cell populations with scRNA-Seq data. The algorithms for unsupervised clustering can be mainly divided into the four types: k-means, hierarchical clustering, density based clustering and graph clustering (Andrews and Hemberg, 2018). K-means is a fast approach that assigns cells to the nearest cluster center, and it requires the predetermined number of clusters. Hierarchical clustering can determine the relationships between clusters, but it generally works slower than k-means. Density-based clustering requires a large number of samples to accurately estimate densities, and it works well for droplet-based datasets. Graph clustering is an extension of density-based clustering, which can deal with millions of cells fast and accurately. Moreover, graphs can easily represent complex nonlinear structure, such as cell populations of different sizes and densities. Seurat identifies the cell clusters mainly based on a graph-based clustering algorithm. Briefly, it embeds cells in a graph structure - for example a K-nearest neighbor (KNN) graph, with edges drawn between cells with similar feature expression patterns (based on the Euclidean distance in PCA space, and refining the edge weights between any two cells based on the shared overlap in their local neighborhoods via Jaccard similarity). It then attempts to partition this graph into highly interconnected communities using modularity optimization techniques (Louvain algorithm) to iteratively group cells together (Butler et al., 2018; Satija et al., 2015).

Once partitioning has been completed, the next step is to identify marker genes that are differentially expressed between different clusters (see differential expression analysis), which aids the assignment of cell types. Based on whether reference cell type information is available, approaches for assigning cell types are mainly classified into two types: (1) known tissue-specific makers, based on directly analyzing correlation with bulk RNA-Seq data captured from reference cells (by reporter-gene driven FACS, for example); and (2) using the cell type specific marker gene's promoter to drive GFP expression and infer their tissue type or using the expression pattern of novel cell type makers, identified by unsupervised clustering, in the existing plant expression atlas to infer the tissue type.

Construction of pseudotemporal trajectories

Cells often exist in a continuum of states, as there are rarely discrete states in many biological systems. scRNA-Seq is able to capture cells at various developmental stages in a single experiment. For example, transitional states defined by the continuously evolving gene expression profile of single cells (Milpied et al., 2018). In such situations, applying algorithms like graph-based clustering to cells transiting through a continuous process like cell differentiation often groups the cells into clusters that fail to reflect their progression through the process and capture the dynamic features of the transition. Trajectory inference (TI) algorithms are able to order cells along the trajectory of a continuously developmental process in a biology system, which allows the identification of cell types at the beginning, intermediate, and end state of the trajectory and reveal the gene expression dynamic across cells (Griffiths et al., 2018). Trajectory topology can be linear, bifurcating, or a tree/graph structure. Therefore, besides uncovering new makers for immediate states and identifying the factors triggering state transitions, developmental trajectories have the potential to reveal branched pathways, and inform regulatory dynamics of differentiation.

Similar to cell clustering analysis, there are also three steps in TI: feature selection (choosing genes that define progress), dimension reduction, and cell ordering. Actually, the genes that are included in the ordering greatly impact the trajectory. Until now, more than 50 TI programs have been developed (Saelens et al., 2019). Based on the extent of the prior information used, TI methods are mainly classified into three types: supervised TI, semi-supervised TI, and unsupervised TI. Supervised TI uses sequential condition labels as input, which increases the complexity of input. For example, scRNA-Seq measurements of cells with a sequence of condition labels corresponding to progression along the process, such as timepoints in a time series (Treutlein et al., 2016). Semi-supervised TI can involve either selecting a set of known makers as features to order cells, or assigning a group of cells into cell states. Prior information can aid in finding the correct trajectory among many likely alternatives but incorrect or noisy prior information could bias the trajectory towards current knowledge. Moreover, prior information is not always easily available. Unsupervised TI discovers the important ordering genes from the data, rather than relying on prior knowledge, by employing as little prior biological information of the interest system as possible. Monocle2, is currently the most popular software, and can be used as either a semi-supervised TI or an unsupervised TI method, considering that ordering genes could

be assigned as known makers or detected automatically (Qiu et al., 2017). Once the genes have been selected to order the cells, Monocle2 incorporates manifold learning and reversed graph embedding (RGE) to reduce the dimensionality of the data, which is robust, fast and powerful (Qiu et al., 2017). Besides that, it could explicitly recover the intrinsic structure from the data, which is different from conventional dimensionality reduction such as PCA or ICA (independent component analysis). In addition, Monocle assumes the trajectory has a tree structure: one end of it is the "root" (the beginning of biological process), and others are the "leaves" (the end of biological process). After projecting the expression data into lower dimensional space, Monocle2 implements an advanced nonlinear reconstruction algorithm, DDRTree, to fit the best tree to the data. Although the Moncle1 is the early version of Moncle2, they employ distinct algorithms and Moncle2 improves branches detection and decreases the sensitive to low quality cells. Monocle1 builds a minimum spanning tree (MST) for cells to search for the longest backbone based on ICA, which is highly complex and require the user to specify the number of branches to search (Trapnell et al., 2014).

Pseudotime is an abstract unit of progress in Monocle2, which tracks changes in expression as a function of progress along the trajectory, instead of as a function of time. Essentially, it is simply the distance between a cell and the start of the trajectory (shortest distance). The trajectory's total length is the total amount of transcriptional change of a cell as it moves from the starting state to the end state (Qiu et al., 2017).

Differential expression analysis

Phenotypical variation results from molecular level diversity. Identifying differential expression of genes between specific conditions or tissues is a great aid to understanding this morphological diversity. It is also one of the most commonly performed tasks for RNA-Seq data analysis. DESeq2 and edgeR are the leading tools for use with fewer replicates (<12) in bulk RNA-Seq analysis (Schurch et al., 2016), which applies the principle of non-zero difference in average expression. However, unlike bulk RNA-Seq data, scRNA-Seq data tends to exhibit an abundance of zero counts and a complicated distribution and huge heterogeneity. Therefore, scRNA-Seq data requires the use of a new differential expression analysis beyond non-zero difference in average expression. Tools such as the Seurat package includes several tests for scRNA-Seq differential expression analysis: bimod, tobit, ttest, poisson, and

negbinom test. Of these, bimod is most popular and accurate (Chen et al., 2018). Bimod combines a discrete/continuous model and likelihood ratio test (LRT) to improve the testing power (McDavid et al., 2013). Discrete/continuous models are designed based on the fact that UMI counts at the single-cell level formed zero and a log-normal distribution. A likelihood ratio test (LRT) can simultaneously test for changes in mean expression (expressed genes) and in the percentage of expressed cells. Finally, it identifies positive and negative markers of a single cluster, compared to all other cells or each other clusters via differential expression analysis (Butler et al., 2018; Satija et al., 2015).

Functional analysis

Characterization of molecular functions or pathways in which DEGs or tissue-specific genes are involved allows us to interpret their biological implication. Gene Ontology (GO) enrichment and pathway enrichment are two major functional analysis methods. Gene Ontology (Ashburner et al., 2000) and Bioconductor (Huber et al., 2015) contain annotation for most model species. For functional unannotated specifies, Blast2GO offers a suitable platform to annotate massive transcriptome datasets, based on a sequence similarity search across a variety of functional annotation database (Conesa et al., 2005). Model organisms usually have their own annotation database, for example, tair (https://www.arabidopsis.org/) for *Arabidopsis*, and maizeGDB (https://www.maizegdb.org/) for *maize*. In addition, there are some other databases such as gramene (http://www.gramene.org/) and phytomoze (https://phytozome.jgi.doe.gov/pz/portal.html), integrating the comparative functional genomics of crops and model plant species.

Visualization

Visualization of data is crucial for the interpretation of results. RNA-Seq data visualization necessities include those to assess the quality of reads (FASTQC), mapping of reads (IGV, Integrative genomics viewer), coverage of reads (RSeQC), and differential expression of genes (DEseq2). In addition, heatmaps benefit the visual comparison of signals on multiple samples. Besides those, scRNA-Seq requires additional visualization tools, such as those for analyzing clustering results. Seurat implements t-SNE (t-Distributed stochastic neighbor embedding) to visualize datasets

at two- or three- dimensions space, without losing information about the relative distance between the plotted cells (local structure). For example, if the diversity of the cells was found to be well represented with ten PCs, then ten dimensions are required to represent the cells. t-SNE will plot the cells on a two- or three- dimensional plot in a way that maintains the ten-dimension relationship between cells, so that cells that are neighbors on a ten-dimension plot remain neighbors on a two- or three-dimension plot.

Regulatory network

A developmental GRN explains the instructions of spatial/temporal expression of regulatory genes or target genes (Li and Davidson, 2009). Accurately deciphering GRNs is central to understanding many developmental processes. Reverse engineering of gene regulatory networks aims at revealing the regulatory interactions among genes from high-throughput sequencing data, via computational algorithms (Hartemink, 2005). Controlling gene transcription by TFs is a key part of gene regulatory and plays vital roles in many fundamental biological processes. There are two typical ways to build TFs-centered gene regulatory networks: Inference of regulation through clustering of co-expression data and/or promoter motif analysis of clustered genes (D'hhaeseleer et al., 2000), and inferring the GRN through high throughput time series gene expression data (Sima et al., 2009). The first method builds a network with undirected edges (without an orientation) between two genes. The basic assumption is that those genes co-expressed or co-regulated, participate in a common regulatory module and share a common biological function, location or process. The second method builds a directed edge (with an orientation) between two genes, indicating a casual control and a regulatory relationship. The basic assumption is that over continuous time, the timings of up-and-down mRNA expression collected from the dynamic mRNA expression profiles, can explain transcriptional regulation. SCODE (Matsumoto et al., 2017), an efficient RGN inference algorithm from scRNA-Seq during differentiation, takes the principle of the second method. Yeast hybridization (Joung et al., 2000), ChIP-Seq (Johnson et al., 2007), and DNA affinity purification sequencing (DAP-Seq) (Bartlett et al., 2017) offer an opportunity to test or verify such GRN linkages.

3.3 Development of the meristem

In contrast to animals, much of plant development occurs post-embryogenesis. New organs are initiated from meristems, specialized niches that maintain a population of undifferentiated stem cells set aside during embryogenesis (Bosca et al., 2011; Zhang et al., 2017). While the shoot apical meristem (SAM) gives rise to all above-ground organs, like the leaves, stems, and flowers, the below-ground root system originates from the root apical meristem (RAM) (Aichinger et al., 2012).

3.3.1 Root apical meristem

Although the function of the RAM is in principle similar to that of the SAM, it is organized differently. Here, the quiescent center (QC) is surrounded by a single layer of tissue-specific stem cells (initials). The QC act as a signaling center, or organizer, and maintains the identity of neighboring stem cells (Jerome et al., 2018). These divide asymmetrically to generate the concentrically arranged tissue files of the root, specifically the stele, cortex, endodermis, and epidermis. The epidermis includes two type of cells: root hair epidermis and non-root hair epidermis. Beneath the QC are the columella initials that give rise to the columella, while the surrounding cells gives rise to the lateral root cap (Iijima et al., 2008). Along the root, there are several zones: the meristematic, elongation, differentiation and mature zones. The meristematic zone is at the root apical tip. Next to the meristematic zone is elongation zone where cells stop division and enter into rapid growth by elongation. Above the elongation region is the differentiation zone where cells undertake their final fate (Figure 4A). The obvious characteristic of the differentiation zone is the appearance of root hairs in the epidermis. This distinct structure of the *Arabidopsis* root provides an ideal tissue for analyzing the promise of scRNA-Seq. Although several root atlases have been generated using reporter lines, they focused primarily on describing either radial or temporal expression profiles. However, scRNA-Seq allows to study them simultaneously. Here in Chapter 4.1, we present a high-resolution scRNA-Seq expression atlas of the *Arabidopsis* root that captures its precise spatiotemporal information.
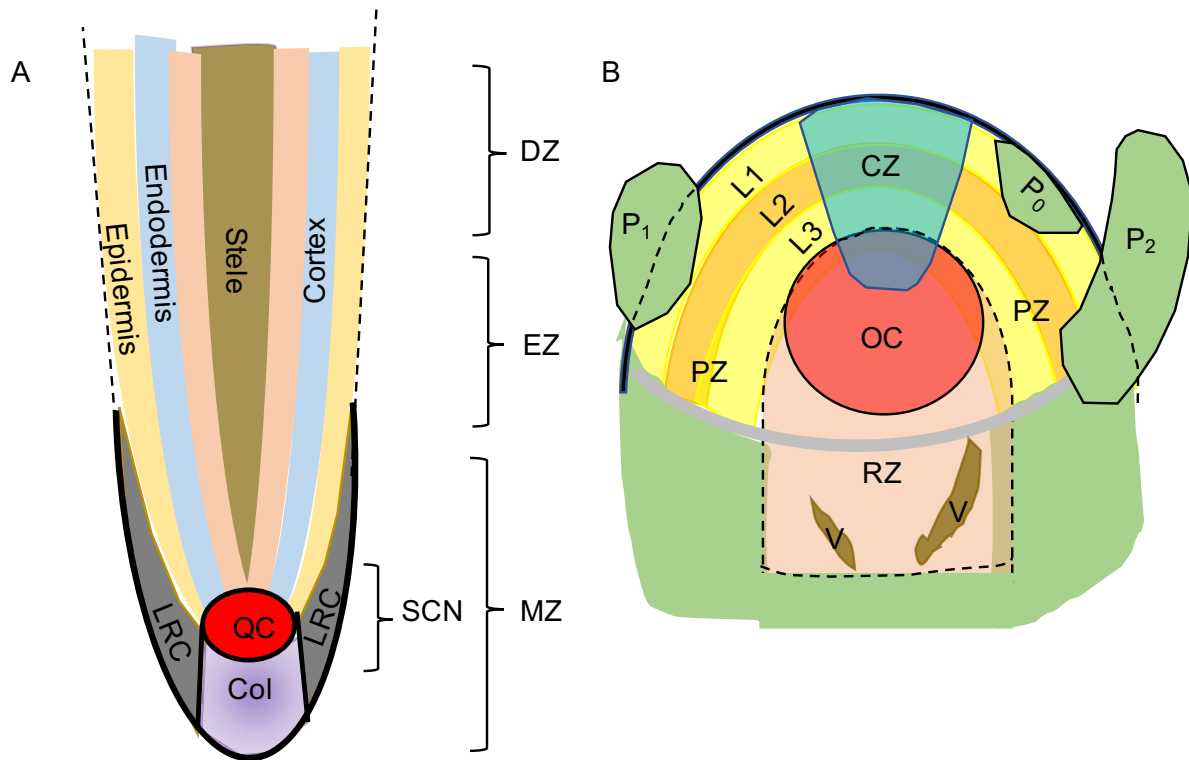
Figure 4. The structure of meristems. A: root meristem. QC: quiescent center; LRC: later root cap; Col: columella; SCN: stem cell niche; MZ: meristem zone; EZ: elongation zone; DZ: differential zone; B: shoot apical meristem. CZ: central zone; OC: organizing center; RZ: rib zone; V: vasculature; L1: Layer1; L2: Layer2; P_0-P_2: different developmental stage of leaf primordia;

3.3.2 Shoot apical meristem

The SAM is composed of different functional domains (Takacs et al., 2012). The central zone (CZ), containing pluripotent stem cells, is located at the summit of the meristem. The peripheral zone (PZ), surrounding the CZ, is located at the lateral flanks of the SAM, and is where leaf primordia initiates. Beneath the CZ is the rib zone (RZ), from which the elongating stem develops (Jerome et al., 2018). Besides the PZ and CZ, organizing center (OC) resides directly underneath CZ and guides the stem cell maintenance in *Arabidopsis* (Pfeiffer et al., 2017) (Figure 4A). Different from *Arabidopsis*, maize has only two clonally separate layers superimposed on the CZ: an outer L1 (layer 1) deriving into the epidermis, and a sub-epidermal L2 (layer 2) (Aichinger et al., 2012). Cell division in the CZ replenishes the stem cell population, renews the meristem, and displaces the PZ. Different from the rapidly dividing PZ,

cells in the CZ divide relative infrequently to reduce accumulation mutations in stem cells (Takacs et al., 2012).

Knauer et al., (2019, in review) generated a high-resolution gene expression atlas of the maize SAM, which revealed distinct gene expression signatures contributing to cell identity and the combinatorial activity of TFs driving the dynamic expression of individual family members defining cell types. In addition, it showed that distinct sets of genes govern the regulation and identity of stem cells in maize versus *Arabidopsis.* The distinct cell fates in the SAM are stably maintained despite the growth of the meristem, so cell fates are continuously defined in a highly dynamic and coordinated manner.

Above all, given the dynamic nature of the stem cell niches and the fact that distinct cell fates are continuously defined in a close spatial and temporal vicinity, meristems are the ideal tissues to study spatiotemporal gene regulation in development.

3.4 The objective of this study:

The distinct structure and abundant of available resources of meristems make them ideal models to study molecular regulation mechanism during plant development. This dissertation includes the following two parts:

Exploration of the high-throughput scRNA-Seq application on the *Arabidopsis* root apex:

High-throughput scRNA-Seq have been widely-used in mammalian research, offering an opportunity to study the molecular heterogeneity of- and developmental progressions within- target tissues. However, the structure of plant cell, including the additional components of cell wall and vacuole increases the difficulty of droplet-based, high throughput scRNA-Seq application on plants. The *Arabidopsis* root, with its distinct spatiotemporal organization provides a perfect material to explore its application. This study aims to explore the high-throughput scRNA-Seq application in plant.

Understand the small RNA spatiotemporal action in maize shoot apex:

MiRNAs play important roles in every aspect of plant development, considering their robust regulation, highly specify and rapid mode of action, and that their mature miRNAs accumulation is regulated both at transcriptional and post-transcriptional level. However, less is known about the spatiotemporal regulation mechanisms of miRNA accumulation and miRNA-mediated regulation mechanisms and regulation modes. The availability of a high resolution of maize shoot apex atlas (Knauer et al., 2019, in review) offers an opportunity to study the small RNAs spatiotemporal action in maize shoot apex. This study aims to reveal insights into the spatiotemporal regulation of miRNAs accumulation, and regulation mechanisms and regulation mode during driving plant development.

3.5 References:

Aichinger, E., Kornet, N., Friedrich, T., and Laux, T. (2012). Plant stem cell niches. Annu Rev Plant Biol *63*, 615-636.

Alllan, M.M., and Walter, G. (1977). A new method for sequencing DNA. Proceedings of the National Academy of Sciences of the United States of America *74*, 560-564.

Alwine, J.C., Kemp, D.J., and STARK, G.R. (1977). Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. Proceedings of the National Academy of Sciences of the United States of America *74*, 5350-5354.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166-169.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Andrews, T.S., and Hemberg, M. (2018). Identifying cell populations with scRNASeq. Mol Aspects Med *59*, 114-122.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T.*, et al.* (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet *25*, 25-29.

Avery, O.T., Macleod, C.M., and Mccarty, M. (1943). Studies on the chemical nature of the substance inducing transformation of penumococcal types: Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. The Journal of Experimental Medicine *79*, 137-159.

Bartlett, A., O'Malley, R.C., Huang, S.C., Galli, M., Nery, J.R., Gallavotti, A., and Ecker, J.R. (2017). Mapping genome-wide transcription-factor binding sites using DAP-seq. Nat Protoc *12*, 1659-1672.

Belton, J.M., McCord, R.P., Gibcus, J.H., Naumova, N., Zhan, Y., and Dekker, J. (2012). Hi-C: a comprehensive technique to capture the conformation of genomes. Methods *58*, 268-276.

Biggin, M.D. (2011). Animal transcription networks as highly connected, quantitative continua. Dev Cell *21*, 611-626.

Birnbaum, K., Jung, J.W., Wang, J.Y., Lambert, G.M., Hirst, J.A., Galbraith, D.W., and Benfey, P.N. (2005). Cell type-specific expression profiling in plants via cell sorting of protoplasts from fluorescent report lines. Nature methods *2*, 615-619.

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. Science *302*, 1956-1960.

Bosca, S., Knauer, S., and Laux, T. (2011). Embryonic development in Arabidopsis thaliana: from the zygote division to the shoot meristem. Front Plant Sci 2, 93.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol 36, 411-420.

Chen, W., Li, Y., Easton, J., Finkelstein, D., Wu, G., and Chen, X. (2018). UMI-count modeling and differential expression analysis for single-cell RNA sequencing. Genome Biol 19, 70.

Chen, X. (2009). Small RNAs and their roles in plant development. Annu Rev Cell Dev Biol 25, 21-44.

Cobb, M. (2017). 60 years ago, Francis Crick changed the logic of biology. PLoS Biol 15, e2003243.

Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21, 3674-3676.

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szczesniak, M.W., Gaffney, D.J., Elo, L.L., Zhang, X., et al. (2016). A survey of best practices for RNA-Seq data analysis. Genome Biol 17, 13.

Crick, F. (1970). Central dogma of molecular biology. Nature 227, 561-563.

D'Ario, M., Griffiths-Jones, S., and Kim, M. (2017). Small RNAs: Big Impact on Plant Development. Trends Plant Sci 22, 1056-1068.

D'hhaeseleer, P., Liang, X., and Somogyi, R. (2000). Genetic network inference: from co-expression clustering to reverse engioneering. Bioinformatics 16, 707-726.

Dahm, R. (2005). Friedrich Miescher and the discovery of DNA. Dev Biol 278, 274-288.

Dillies, M.A., Rau, A., Aubert, J., Hennequet-Antier, C., Jeanmougin, M., Servant, N., Keime, C., Marot, G., Castel, D., Estelle, J., et al. (2013). A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. Brief Bioinform 14, 671-683.

Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C.P., Jerby-Arnon, L., Marjanovic, N.D., Dionne, D., Burks, T., Raychowdhury, R., et al. (2016). Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. Cell 167, 1853-1866 e1817.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-Seq aligner. Bioinformatics 29, 15-21.

Doerks, T., Copley, R.R., Schultz, J., Ponting, C.P., and Bork, P. (2002). Systematic identification of novel protein domain families associated with nuclear functions. Genome Res *12*, 47-56.

Efroni, I., Ip, P.L., Nawy, T., Mello, A., and Birnbaum, K.D. (2015). Quantification of cell identity from single-cell gene expression profiles. Genome Biol *16*, 9.

Elahi, E., and Ronaghi, M. (2004). Pyrosequecing: a tool for DNA sequencing analysis. Methods molecular biology *255*, 211-219.

Emmert-Buck, M.R., Bonner, R.F., Smith, P.D., Chuaqui, R.F., Huang, Z., Goldstein, S.R., Weiss, R.A., and Liotta, L.A. (1996). Laser capture microdissection. Science *274*, 998-1001.

Etzrodt, M., Endele, M., and Schroeder, T. (2014). Quantitative single-cell approaches to stem cell research. Cell Stem Cell *15*, 546-558.

Evans, C., Hardin, J., and Stoebel, D.M. (2018). Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions. Brief Bioinform *19*, 776-792.

Fuller, C.W., Middendorf, L.R., Benner, S.A., Church, G.M., Harris, T., Huang, X., Jovanovich, S.B., Nelson, J.R., Schloss, J.A., Schwartz, D.C.*, et al.* (2009). The challenges of sequencing by synthesis. Nat Biotechnol *27*, 1013-1023.

Giacomello, S., and Lundeberg, J. (2018). Preparation of plant tissue to enable Spatial Transcriptomics profiling using barcoded microarrays. Nat Protoc *13*, 2425-2446.

Goodwin, S., McPherson, J.D., and McCombie, W.R. (2016). Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet *17*, 333-351.

Griffiths, J.A., Scialdone, A., and Marioni, J.C. (2018). Using single-cell genomics to understand developmental processes and cell fate decisions. Mol Syst Biol *14*, e8046.

Guo, J., Xu, N., Li, Z., Zhang, S., and Wu, J. (2008). Four-color DAN sequencing with 3'-O-modified nucleotide reversible terminators and chemically cleavable fluorescent dideoxynucleotides. Proceedings of the National Academy of Sciences of the United States of America *105*, 9145-9150.

Hagemann-Jensen, M., Abdullayev, I., Sandberg, R., and Faridani, O.R. (2018). Small-seq for single-cell small-RNA sequencing. Nat Protoc *13*, 2407-2424.

Hansen, K.D., Wu, Z., Irizarry, R.A., and Leek, J.T. (2011). Sequencing technology does not eliminate biological variability. Nat Biotechnol *29*, 572-573.

Hartemink, A.J. (2005). Reverse engineering gene regulatory networks. Nature biotechnology *23*, 554-555.

Heather, J.M., and Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. Genomics *107*, 1-8.

Heiman, M., Kulicke, R., Fenster, R.J., Greengard, P., and Heintz, N. (2014). Cell type-specific mRNA purification by translating ribosome affinity purification (TRAP). Nat Protoc *9*, 1282-1291.

Huang, Y., Ji, L., Huang, Q., Vassylyev, D.G., Chen, X., and Ma, J.B. (2009). Structural insights into mechanisms of the small RNA methyltransferase HEN1. Nature *461*, 823-827.

Huber, W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B.S., Bravo, H.C., Davis, S., Gatto, L., Girke, T.*, et al.* (2015). Orchestrating high-throughput genomic analysis with Bioconductor. Nat Methods *12*, 115-121.

Iijima, M., Morita, S., and Barlow, P.W. (2008). Structure and function of the root cap. Plant production science *11*, 17-27.

Jerome, E.P., Drapek, C., and Benfey, P.N. (2018). Regulation of division and differentiation of plant stem cells. AnnuRevCell DevBiol *34*, 289-310.

Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007). Genome-wide mapping of in vivo protein-DAN interactions. Science *316*, 1497-1502.

Joung, J.K., Ramm, E.I., and Pabo, C.O. (2000). A bacterial two-hybrie selection system for studying protein-DNA and protein-protein interactions. Proceedings of the National Academy of Sciences of the United States of America *97*, 7382-7387.

Kahvejian, A., Quackenbush, J., and Thompson, J.F. (2008). What would you do if you could sequence everything? Nat Biotechnol *26*, 1125-1133.

Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell *161*, 1187-1201.

Lever, J., Krzywinski, M., and Altman, N. (2017). Principal component analysis. Nature Methods *14*, 641-642.

Levy, S.E., and Myers, R.M. (2016). Advancements in Next-Generation Sequencing. Annu Rev Genomics Hum Genet *17*, 95-115.

Li, E., and Davidson, E.H. (2009). Building developmental gene regulatory networks. Birth Defects Res C Embryo Today *87*, 123-130.

Li, Y., and Tollefsbol, T.O. (2011). DNA methylation detection: bisulfite genomic sequencing analysis. Methods Mol Biol *791*, 11-21.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923-930.

Liu, Y., Zhou, J., and White, K.P. (2014). RNA-Seq differential expression studies: more sequence or more replication? Bioinformatics *30*, 301-304.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M*., et al.* (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell *161*, 1202-1214.

Mardis, E.R. (2008). The impact of next-generation sequencing technology on genetics. Trends Genet *24*, 133-141.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z*., et al.* (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature *437*, 376-380.

Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M., and Gilad, Y. (2008). RNA-Seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res *18*, 1509-1517.

Matsumoto, H., Kiryu, H., Furusawa, C., Ko, M.S.H., Ko, S.B.H., Gouda, N., Hayashi, T., and Nikaido, I. (2017). SCODE: an efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation. Bioinformatics *33*, 2314-2321.

McDavid, A., Finak, G., Chattopadyay, P.K., Dominguez, M., Lamoreaux, L., Ma, S.S., Roederer, M., and Gottardo, R. (2013). Data exploration, quality control and testing in single-cell qPCR-based gene expression experiments. Bioinformatics *29*, 461-467.

Milpied, P., Cervera-Marzal, I., Mollichella, M.L., Tesson, B., Brisou, G., Traverse-Glehen, A., Salles, G., Spinelli, L., and Nadel, B. (2018). Human germinal center transcriptional programs are de-synchronized in B cell lymphoma. Nat Immunol *19*, 1013-1024.

Moreno-Romero, J., Santos-Gonzalez, J., Hennig, L., and Kohler, C. (2017). Applying the INTACT method to purify endosperm nuclei and to generate parental-specific epigenome profiles. Nat Protoc *12*, 238-254.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods *5*, 621-628.

Nakazono, M. (2003). Laser-Capture Microdissection, a Tool for the Global Analysis of Gene Expression in Specific Plant Cell Types: Identification of Genes Expressed Differentially in Epidermal Cells or Vascular Tissues of Maize. The Plant Cell Online *15*, 583-596.

Nelms, B., and Walbot, V. (2019). Defining the developmental program leading to meiosis in maize. Science *6435*, 52-56.

Nick, W.O., and John, J.H. (1993). Reverse transcription-polymerase chain reaction: an overview of the technique and its applications. Biotech Adv *11*, 13-29.

Pfeiffer, A., Wenzl, C., and Lohmann, J.U. (2017). Beyond flexibility: controlling stem cells in an ever changing environment. Curr Opin Plant Biol *35*, 117-123.

Phelps-Durr, T.L., Thomas, J., Vahab, P., and Timmermans, M.C. (2005). Maize rough sheath2 and its Arabidopsis orthologue ASYMMETRIC LEAVES1 interact with HIRA, a predicted histone chaperone, to maintain knox gene silencing and determinacy during organogenesis. Plant Cell *17*, 2886-2898.

Plavskin, Y., Nagashima, A., Perroud, P.F., Hasebe, M., Quatrano, R.S., Atwal, G.S., and Timmermans, M.C. (2016). Ancient trans-Acting siRNAs Confer Robustness and Sensitivity onto the Auxin Response. Dev Cell *36*, 276-289.

Prakadan, S.M., Shalek, A.K., and Weitz, D.A. (2017). Scaling by shrinking: empowering single-cell 'omics' with microfluidic devices. Nat Rev Genet *18*, 345-361.

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.

Reiter, F., Wienerroither, S., and Stark, A. (2017). Combinatorial function of transcription factors and cofactors. Curr Opin Genet Dev *43*, 73-81.

Reuter, J.A., Spacek, D.V., and Snyder, M.P. (2015). High-throughput sequencing technologies. Mol Cell *58*, 586-597.

Ricci, M., Ragonese, F., Gironi, B., Paolantoni, M., Morresi, A., Latterini, L., Fioretti, B., and Sassi, P. (2018). Glioblastoma single-cell microRaman analysis under stress treatments. Sci Rep *8*, 7979.

Rogers, K., and Chen, X. (2013). Biogenesis, turnover, and mode of action of plant microRNAs. Plant Cell *25*, 2383-2399.

Rothberg, J.M., Hinz, W., Rearick, T.M., Schultz, J., Mileski, W., Davey, M., Leamon, J.H., Johnson, K., Milgrew, M.J., Edwards, M.*, et al.* (2011). An integrated semiconductor device enabling non-optical genome sequencing. Nature *475*, 348-352.

Saelens, W., Cannoodt, R., Todorov, H., and Saeys, Y. (2019). A comparison of single-cell trajectory inference methods. Nat Biotechnol *37*, 547-554.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences of the United States of America *74*, 5463-5467.

Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nat Biotechnol *33*, 495-502.

Scanlon, M.J., Ohtsu, K., Timmermans, M.C., and Schnable, P.S. (2009). Laser microdissection-mediated isolation and in vitro transcriptional amplification of plant RNA. Curr Protoc Mol Biol *Chapter 25*, Unit 25A 23.

Schmiedel, J.M., Klemm, S.L., Zheng, Y., Sahay, A., Blüthgen, N., Marks, D.S., and Oudenaarden, A.v. (2015). MicroRNA control of protein expression noise. Science *348*, 128-132.

Schurch, N.J., Schofield, P., Gierlinski, M., Cole, C., Sherstnev, A., Singh, V., Wrobel, N., Gharbi, K., Simpson, G.G., Owen-Hughes, T.*, et al.* (2016). How many biological replicates are needed in an RNA-Seq experiment and which differential expression tool should you use? RNA *22*, 839-851.

Semrau, S., Goldmann, J.E., Soumillon, M., Mikkelsen, T.S., Jaenisch, R., and van Oudenaarden, A. (2017). Dynamics of lineage commitment revealed by single-cell transcriptomics of differentiating embryonic stem cells. Nat Commun *8*, 1096.

Shchena, M., Shalon, D., Davis, R.W., and Brown, P.O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarry. Science *270*, 467-470.

Sima, C., Hua, J., and Jung, S. (2009). Inference of gene regulatory network using time-series data: a survey. Current Genomics *2009*, 416-429.

Sims, D., Sudbery, I., Ilott, N.E., Heger, A., and Ponting, C.P. (2014). Sequencing depth and coverage: key considerations in genomic analyses. Nat Rev Genet *15*, 121-132.

Skelly, D.A., Squiers, G.T., McLellan, M.A., Bolisetty, M.T., Robson, P., Rosenthal, N.A., and Pinto, A.R. (2018). Single-Cell Transcriptional Profiling Reveals Cellular Diversity and Intercommunication in the Mouse Heart. Cell Rep *22*, 600-610.

Skopelitis, D.S., Benkovics, A.H., Husbands, A.Y., and Timmermans, M.C.P. (2017). Boundary Formation through a Direct Threshold-Based Readout of Mobile Small RNA Gradients. Dev Cell *43*, 265-273 e266.

Skopelitis, D.S., Hill, K., Klesen, S., Marco, C.F., von Born, P., Chitwood, D.H., and Timmermans, M.C.P. (2018). Gating of miRNA movement at defined cell-cell interfaces governs their impact as positional signals. Nat Commun *9*, 3107.

Svensson, V., Vento-Tormo, R., and Teichmann, S.A. (2018). Exponential scaling of single-cell RNA-Seq in the past decade. Nat Protoc *13*, 599-604.

Takacs, E.M., Li, J., Du, C., Ponnala, L., Janick-Buckner, D., Yu, J., Muehlbauer, G.J., Schnable, P.S., Timmermans, M.C., Sun, Q.*, et al.* (2012). Ontogeny of the maize shoot apical meristem. Plant Cell *24*, 3219-3234.

Tanay, A., and Regev, A. (2017). Scaling single-cell genomics from phenomenology to mechanism. Nature *541*, 331-338.

Tandonnet, S., and Torres, T.T. (2017). Traditional versus 3' RNA-Seq in a non-model species. Genom Data *11*, 9-16.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A.*, et al.* (2009). mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods *6*, 377-382.

Tian, C., Wang, Y., Yu, H., He, J., Wang, J., Shi, B., Du, Q., Provart, N.J., Meyerowitz, E.M., and Jiao, Y. (2019). A gene expression map of shoot domains reveals regulatory mechanisms. Nat Commun *10*, 141.

Torres, T.T., Metta, M., Ottenwalder, B., and Schlotterer, C. (2008). Gene expression profiling by massively parallel sequencing. Genome Res *18*, 172-177.

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.

Treutlein, B., Lee, Q.Y., Camp, J.G., Mall, M., Koh, W., Shariati, S.A., Sim, S., Neff, N.F., Skotheim, J.M., Wernig, M.*, et al.* (2016). Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-Seq. Nature *534*, 391-395.

Vallejos, C.A., Risso, D., Scialdone, A., Dudoit, S., and Marioni, J.C. (2017). Normalizing single-cell RNA sequencing data: challenges and opportunities. Nat Methods *14*, 565-571.

Venteicher, A.S., Tirosh, I., Hebert, C., Yizhak, K., Neftel, C., Filbin, M.G., and Hovestadt, V. (2017). Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-Seq. Science *355*.

Villani, A.-C., Satija, R., Reynolds, G., Darkizova, S., and Shekhar, K. (2017). Single-cell RNA-Seq reveals types of human blood dendritic cell, monocytes, and progenitors. Science *356*.

Wang, L., Wang, S., and Li, W. (2012). RSeQC: quality control of RNA-Seq experiments. Bioinformatics *28*, 2184-2185.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nature reviews genetics *10*, 57-63.

Watson, J.D., and Crick, F.H.C. (1953). Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid Nature *171*, 732-738.

Xu, L. (2003). Novel as1 and as2 defects in leaf adaxial-abaxial polarity reveal the requirement for ASYMMETRIC LEAVES1 and 2 and ERECTA functions in specifying leaf adaxial identity. Development *130*, 4097-4107.

Yadav, R.K., Girke, T., Pasala, S., Xie, M., and Reddy, G.V. (2009). Gene expression map of the Arabidopssi shoot apical meristem stem cell niche. Proceedings of the National Academy of Sciences of the United States of America *106*, 4941-4946.

Zhang, Z., Tucker, E., Hermann, M., and Laux, T. (2017). A Molecular Framework for the Embryonic Initiation of Shoot Meristem Stem Cells. Dev Cell *40*, 264-277 e264.

Zheng, G.X., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J.*, et al.* (2017). Massively parallel digital transcriptional profiling of single cells. Nat Commun *8*, 14049.

Developmental Cell

4.1 Spatiotemporal Developmental Trajectories in the *Arabidopsis* Root Revealed Using High-Throughput Single-Cell RNA Sequencing

Graphical Abstract

scRNA-Seq of 4,727 Arabidopsis root cells

Cluster arrangement captures developmental trajectories from a meristematic cluster core

Pericycle
Columella
Phloem
QC
Xylem
Meristem
Endodermis
Cortex
Atrichoblast
Trichoblast

trajectory
trajectory

Finely-resolved waves of gene expression underlie developmental trajectories
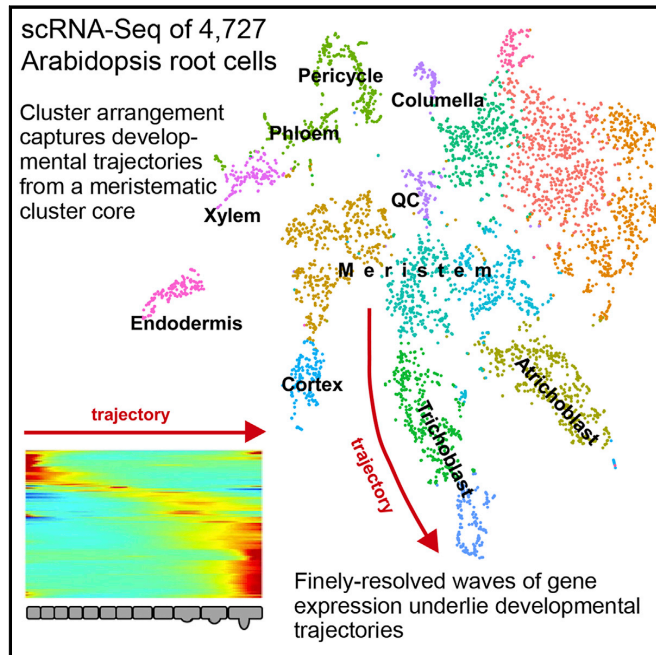
Authors

Tom Denyer, Xiaoli Ma, Simon Klesen, Emanuele Scacchi, Kay Nieselt, Marja C.P. Timmermans

Correspondence

marja.timmermans@ zmbp.uni-tuebingen.de

In Brief

Denyer and Ma et al. generate a single-cell RNA expression atlas of the *Arabidopsis* root that captures spatiotemporal information for all major cell types and uncovers new regulators. Pseudotime-analysis-derived developmental trajectories depict a cascade of developmental progressions between stem cell and final differentiation mirrored by waves of transcription factor expression.

Highlights

- scRNA-seq of *Arabidopsis* root cells captures precise spatiotemporal information

- Defining expression features for cell types identify new developmental regulators

- Cluster arrangement reflects developmental time with a centrally localized niche

- Intricate waves of gene expression finely resolve developmental trajectories

Details see appendix I.

Cell   s

4.2 Genome-wide analysis of miRNA action in plants

Xiaoli Ma[1], Marie Javelle[2], Tom Denyer[1], Antje Feller[1], Steffen Knauer[1], Mike Scanlon[3], Marja C. P. Timmermans[1,2*]

[1] Center for Plant Molecular Biology, University of Tübingen, 72076 Tübingen, Germany ;

[2] Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA ;

[3] Division of Plant Biology, Cornell University, Ithaca, New York 14850, USA

* Corresponding author: marja.timmermans@zmbp.uni-tuebingen.de

This work aimed to understand the spatiotemporal action of miRNA across maize shoot apex. I contributed to the experimental design and performed the data analysis. I wrote the manuscript with the help from Tom Denyer and Marja Timmermans and with the input from co-authors.

4.2.1 Abstract

MiRNAs play important roles in every aspect of plant development. Small RNA regulation confers sensitivity and robustness onto gene regulatory networks, and the morphogen-like readout of small RNA mobility gradients yields sharply delineated domains of target gene expression. However, less is known about how the spatiotemporal patterns of miRNA activity are attained. The availability of a high-resolution gene expression atlas of the maize shoot apical meristem (SAM) offers an opportunity to study this problem. Analysis of miRNA precursor gene expression across twelve functional SAM domains revealed substantial sub-functionalization, with distinct precursor genes marking stem cells versus differentiating organ primordia. Correlation analysis revealed that miRNA abundance, determined by small RNA deep sequencing, and precursor transcript levels across the SAM are highly correlated, with the exception of miRNAs localized to the vasculature and stem cells in the niche. Here, key enzymes involved in the miRNA biology, such as DICER-LIKE1, are minimally expressed. The spatial accumulation of miRNAs thus reflects regulation at both the transcriptional and post-transcriptional level. In addition, we identified miRNA target genes by bioinformatic prediction and experimentally validated them using parallel analysis of RNA ends (PARE); analyzing their regulation across the SAM. Integrating analysis of -omics data, we inferred miRNA regulation mechanism (transcript cleavage or translation repression) and regulation mode (clearance or rheostat) and revealed an unexpected variability in miRNA-mediated transcript cleavage efficiency. Together, our results reveal that the spatiotemporal action of plant miRNAs during development is regulated at multiple levels affecting miRNA transcription, processing/stability, and activity.

4.2.2 Introduction

MiRNA-driven regulation is a key component and a number of highly-conserved miRNAs target transcription factors (TFs) with specific roles in early stage development (D'Ario et al., 2017; Li et al., 2016; Singh et al., 2018; Swarup and Denyer, 2019; Yang et al., 2018). For example, miR156-mediated repression of SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL) transcription factors is required for both the proper leaf initiate rate (Schwarz et al., 2008; Wang et al., 2008). miR160-mediated regulation AUXIN RESPONSE FACTORS (ARF) genes is necessary for normal cotyledon numbers and positions (Liu et al., 2007;

Mallory et al., 2005; Wang et al., 2005). MiR164 regulates organ boundary size and maintain phyllotaxis set-up in meristem by its modulation of the CUC (CUP-SHAPED COTYLEDON) gene (Baker et al., 2005; Mallory et al., 2004; Nikovics et al., 2006; Peaucelle et al., 2007; Sieber et al., 2007). MiR165/166 spatially restrict homeobox-leucine zipper family transcription factors during meristem and leaf development (Chitwood et al., 2009; Juarez et al., 2004; Jung and Park, 2007; Nogueira et al., 2007; Reinhart et al., 2002; Sakaguchi and Watanabe, 2012). MiR172 control APETALA2 (AP2) gene involved flower meristem identity transition (Chen, 2004; D'Ario et al., 2017; Irish and Sussex, 1990). MiR394 target WUSCHEL(WUS) suppressor LEAF CURLING REPRESIVENESS (LCR) in the organizing center (OC) to control meristem size (Knauer et al., 2013; Song et al., 2012). Therefore, miRNAs are able to define cell-type specific gene expression at the early stage according to spatial and temporal cues.

MiRNA-target modules have been coopted for major innovations in plant evolution. MiRNAs bind in a homology-dependent manner to target mRNAs and trigger the degradation through transcript cleavage, and/or their translational inhibition. Mature miRNAs originate from MIR genes, which experiences transcription and processing catalyzed by multiple enzymes. Some key miRNA/target modules, central to proper plant development in *Arabidopsis*, have been found to function by-way-of translational repression besides cleavage. For examples, miR156 and miR172 regulated their targets primarily by translation repression (Aukerman and Sakai, 2003; Chen, 2004; Gandikota et al., 2007; He et al., 2018).

In additional, miRNAs shape and diversify gene expression profiles of different cell types during development by post-transcriptional regulation. Recent finding indicates that miRNA could form highly precise mobile positional signals to generate sharp domains of target gene expression through a morphogen-like readout (Skopelitis et al., 2017). This reflects a non-linear, threshold-based readout, which is a highly sensitive to the relative ratio of small RNA-to-target levels (Skopelitis et al., 2017). When this ratio is above a given threshold, small RNAs can clear target transcripts or protein accumulation. Whereas below the threshold, small RNAs reduce target gene expression noise and show a rheostat behaviour. Thus, there are multiple types of small RNA-mediated regulatory modes in development (Skopelitis et al., 2017).

So far, global study spatiotemporal accumulation and action of miRNAs with a developmental perspective localized tissues are lacking. To address these questions, using the maize (*Zea mays*) shoot apex as a model, we quantified miRNA levels in the shoot apex by miRNA-Seq before determining expression patterns using *in-situ* hybridization. Precursor (pri-miRNA) spatiotemporal accumulation across twelve SAM subdomains was also determined from

previous data (Knauer et al., 2019). These results show substantial sub-functionalization of precursor genes. Integrating analysis of RNA-Seq and degradome-Seq data, we identified the accurate miRNA target genes and systematically assigned the miRNA mediated gene regulation mechanism and regulation mode. Our results also demonstrated the factors related to the different sensitivity of efficacy of miRNA-mediated transcript cleavage. These results provide a foundation for spatiotemporal action of miRNAs during plant development.

4.2.3 Results

Expression patterns of miRNA precursors indicate substantial tissue-specific transcriptional sub-functionalization

To understand small RNA spatiotemporal action during development, we took advantage of a recently described high-resolution maize SAM expression atlas (Knauer et al., 2019). The SAM is a specialized niche located at the tip of the growing plant shoot that orchestrates the balance between stem cell proliferation and organ initiation essential for post-embryonic shoot growth. The SAM provides a perfect context to study miRNA regulation. Cell fates within the growing niche are continuously defined in close a spatial and temporal vicinity, and expression of key cell fate determinants within the SAM is under miRNA control (Fouracre and Poethig, 2016; Petsch et al., 2015). Moreover, miRNA mobility between functional domains of the SAM is dynamically regulated (Skopelitis et al., 2018).

The maize SAM atlas captures genome-wide expression profiles for ten distinct structural and functional domains within the maize vegetative shoot apex: the whole meristem, the stem cell-containing central zone, the incipient leaf (P0) at the meristem periphery, the L1 and L2 lineage layers overlaying these meristematic regions, developing leaf primordia P1, P2 and P3, the internode primordium, and the vasculature (Knauer et al., 2019; Figures 1A). Considering that small RNAs play important roles in adaxial-abaxial leaf polarity (Juarez et al., 2004; Nogueira et al., 2007), we complemented the atlas by generating global gene expression data also for the adaxial and adaxial sides of P2-P3 leaf primordia (Figures 1B and 1C, Figure S1A-D).
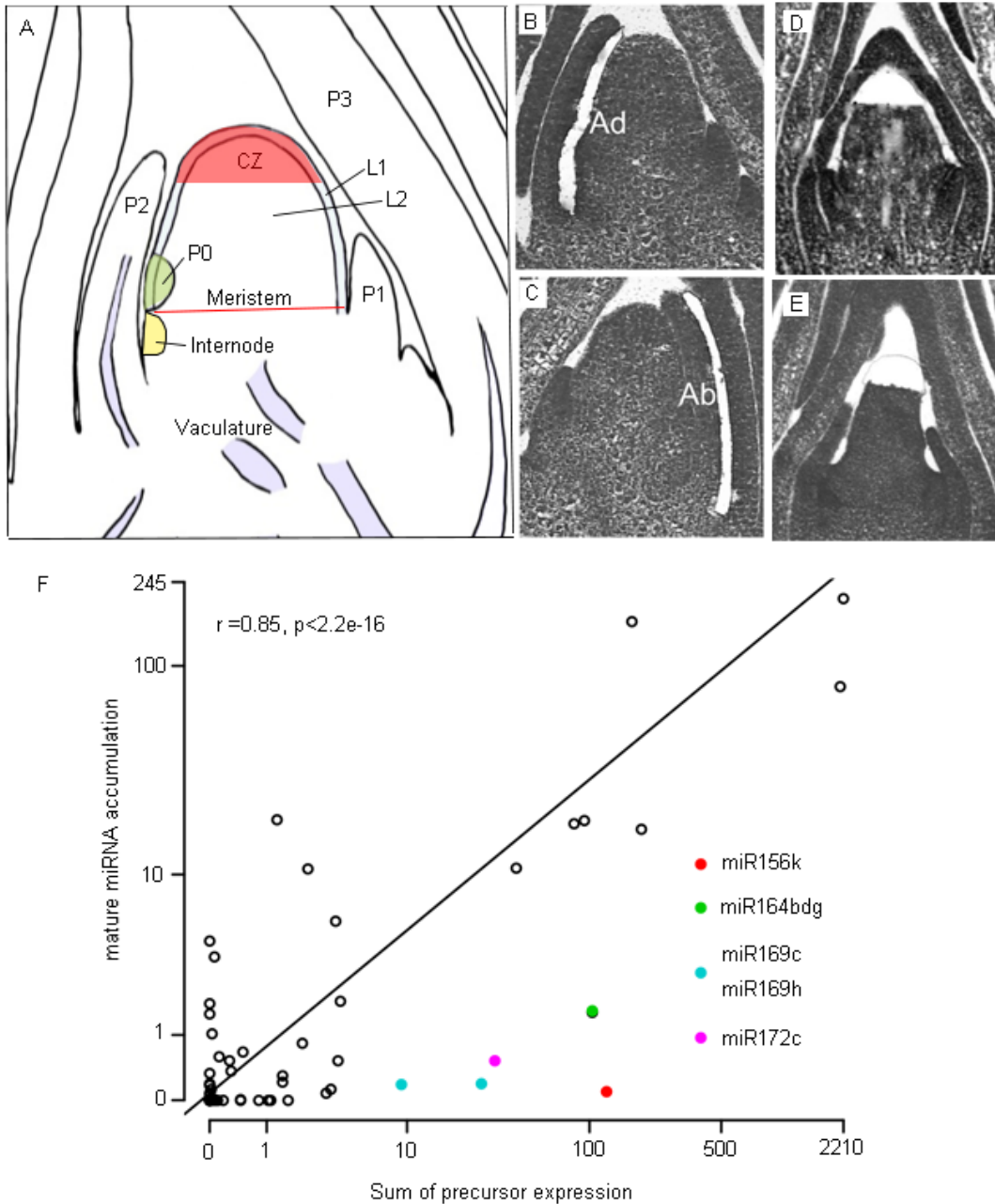
Figure 1. miRNA accumulation and precursor gene expression is highly correlated. (A) Schematic of a longitudinal section of a 14 day-old B73 seedling apex illustrating the ten domains profiled by laser microdissection RNA-Seq (Knauer et al., 2019). (B-E) Sections after laser microdissection of the adaxial (Ad) (B), adaxial (Ab) (C) sides of developing leaf primordia, and the center zone (CZ) in *dcl1-2* (D), and wild type (E). (F) miRNA abundance and precursor expression levels are well correlated (r=0.85). Prominent outliers whose mature miRNA is less abundant than expected, are highlighted.

The 28 confirmed maize miRNA families comprise 141 precursor genes that together are predicted to produce 160 mature miRNAs (Table 1, Dataset S1). Instances where a single precursor is processed into more than one mature miRNA are seen in seven families. Notably, the *MIR395* precursor genes contain sixteen miRNA stem-loops tandemly arrayed within four precursors (Dataset S1). Similar examples of polycistronic miRNAs are also found in the miR166, miR167, miR169, miR396, miR2118 and miR2275 families. 19% of mature miRNAs in maize are polycistronic miRNAs, which is similar percentage to those of *Arabidopsis* (20%) and rice (19%) (Merchan et al., 2009). Interestingly, polycistronic miR166, miR169 and miR395 are conserved in rice and *Arabidopsis*. It suggests a putative conservational origin of polycistronic *MIR* genes in evolution process.

Although most of these miRNAs are broadly conserved across land plant evolution, the families themselves continue to diverge. This is also apparent at the level of the mature miRNA, where single nucleotide polymorphisms are not uncommon. Here, we detected 76 distinct miRNA sequences originating from 28 miRNA families (Table 1).

41 of the 141 precursors were found to be expressed in at least one of the twelve subdomains in the shoot apex at a level ≥1 RPM in at least one of the twelve subdomains analyzed (Figure 1F; Dataset S1). Among them, three precursors are polycistronic *MIR* genes. Consistent with possible roles in development, hierarchical clustering revealed three major clusters describing the leaf primordia, vasculature, and meristem regions. Interestingly, this spatial separation mirrors that observed previously for transcription factors (TFs), which was found to be predictive of tissue identity and the overall expression profiles of domains within the apex (Knauer et al., 2019). Precursors for multiple miRNAs are expressed within each of these apex regions, reinforcing the principle of a prominent contribution of miRNAs in patterning the shoot apex. Combinational input forms diverse gene networks on cell identity. These observed expression profiles hint at defined functions. For instance, *MIR166b, c, d, k* & *n* cluster together, and their expression reflects their recorded role in specifying abaxial identity (Chitwood et al., 2007; Nogueira et al., 2009; Nogueira et al., 2007). Similarly, *MIR168*a, b group together in the SAM. It is consistent with select miRNAs controlling developmental programs.

Table 1. Summary of miRNA family

Family	No. of precursors	No. of miRNAs	No. of unique miRNA[#]
MIR156	12	12	3
MIR159	11	11	5
MIR160	6	6	2
MIR162	1	1	1
MIR164	8	8	4
MIR166*	11	12	3
MIR167*	8	9	2
MIR168	2	2	1
MIR169*	16	17	10
MIR171	11	11	6
MIR172	5	5	2
MIR319	4	4	1
MIR390	2	2	1
MIR393	3	3	1
MIR394	2	2	1
MIR395*	4	16	4
MIR396*	7	8	4
MIR397	1	1	1
MIR398	2	2	1
MIR399	10	10	6
MIR408	1	1	1
MIR482	1	1	1
MIR528	2	2	1
MIR529	1	1	1
MIR827	1	1	1
MIR1432	1	1	1
MIR2118*	5	7	7
MIR2275*	3	4	4
Total	141	160	76

The 28 known miRNA families show variation in complexity. * Precursor genes produced multiple mature miRNAs. [#] The number of distinct miRNA sequences in miRNA family.

For most miRNA families, only a small subset of its members is expressed in the apex, indicating some sub-functionalization among miRNA family members and less redundancy than perhaps expected. The expression patterns of members within single miRNA families reveal a further division of labor, with individual precursors differentially expressed across different domains of the apex (Figure S2). For example, both *MIR394* precursors are expressed in the SAM, but only *MIR394a* is highly expressed in the vasculature, and *MIR394b* is expressed primarily on the adaxial side of leaf primordia. Similarly, precursors for miR159, miR160, miR169, and miR319

show redundant expression in the SAM but are differentially regulated across other regions of the apex. Additionally, for miR156, miR164, miR167, miR168, and miR172, precursor transcripts are primarily derived from just one of the family members (Figure S2). Interestingly, *MIR169c* and *MIR169e,h*, which generate different mature miRNA sequences, show differential accumulation in the meristem and vasculature. Similarly, miR166 precursors producing distinct mature miRNA sequences also show dynamical expression in leaf primordia.

Taken together, these data indicate that the spatiotemporal patterns of miRNA activity result in-part from intricate transcriptional regulation of their precursor genes. This generates tissue-specific patterns of expression resulting in a division of labor among miRNA family members, with distinct genes functioning in the SAM, vasculature, or leaves. This indicates that these members likely play specialized roles in these tissue types.

Expression levels of precursor genes and their mature miRNAs are highly correlated

In maize, mature miRNA accumulation has been reported as being subject to complex tissue- and cell-type- specific post-transcriptional control (Nogueira et al., 2009). RNA-Seq and small RNA sequencing (sRNA-Seq) data provides an opportunity to assess this level of regulation on miRNA accumulation. To investigate this, we performed sRNA-Seq on the apex, comprising the SAM, and up to four leaf primordia. Approximately 90% of the 18- to 26- nucleotide (nt) reads mapped to the maize reference genome, and small RNA size distribution is similar to that previously described in maize (Dotto et al., 2014; Petsch et al., 2015)

28 miRNA families are predicted to produce 76 distinct mature miRNAs. Considering a raw read count cut-off of ≥10, 41 miRNAs, belonging to 25 families, are expressed across the apex. Correlation analysis shows that the relative abundance of a given miRNA is highly correlated (r=0.85) with the cumulative level of expression of the corresponding precursor genes across the meristem, internode, P1 to P3 leaf primordia, and vasculature (representing non-overlapping domains in the apex). Levels of miR166 and miR319 are higher than expected based on expression of their respective precursors. Conversely, for *MIR156k*, *MIR164bdg*, *MIR169c*, *MIR169e,h*

and MIR172c, mature miRNA accumulation is lower than predicted (Figure 1G, Dataset S1). This suggests a certain level of regulation at the level of miRNA processing and/or stability. Interestingly, the latter precursors (except *MIR164b, -d, -g*) are mainly expressed in the vasculature (Figure 2, Figure S2). Genes required for small RNA biogenesis, function, and turnover show some expression variation across the apex (Figure S3, Dataset S2). However, interestingly, transcript levels for *DCL1* are particularly low in the vasculature compared to other regions of the apex, whereas expression of *HEN1 SUPPRESSOR1* (*HESO1*), involved in miRNA turnover, is noticeably higher in the vasculature (Figure S3). However, their expression is lower than expected accumulation is not explained by a miRNA turnover speed (Zhai et al., 2013). It is likely that differences in the activities of these enzymes contribute to the spatial regulation of miRNA accumulation, limiting miRNA levels in vasculature.

Mer	L1	L2	Tip	Vas	Int	P0	P1	P2	P3	Ad	Ab	
27.00	22.00	28.00	10.00	1.20	8.60	100.00	56.00	31.00	22.00	15.00	24.00	MIR167c
170.00	210.00	120.00	33.00	78.00	180.00	1400.00	830.00	400.00	220.00	220.00	330.00	MIR167d
8.20	18.00	1.70	2.70	0.45	1.40	71.00	71.00	23.00	7.90	5.90	7.50	MIR167a,g
0.41	0.46	0.35	0.25	0.30	0.23	1.00	0.31	0.18	0.16	0.14	0.10	MIR2275d
12.00	14.00	8.50	3.60	1.70	120.00	53.00	38.00	23.00	31.00	9.30	180.00	MIR166n
1.00	9.60	3.60	5.20	4.10	7.60	23.00	49.00	43.00	54.00	5.40	180.00	MIR166d
0.08	0.06	0.10	0.04	0.02	0.05	0.82	0.80	0.21	0.02	0.04	1.10	MIR390a
1.40	4.80	0.57	0.84	0.89	2.00	12.00	13.00	12.00	22.00	1.20	26.00	MIR166b
7.70	10.00	3.80	0.60	0.73	45.00	39.00	13.00	5.60	7.70	3.00	31.00	MIR166k,m
7.00	3.10	5.50	0.40	0.39	52.00	10.00	4.50	2.00	1.60	1.10	7.50	MIR166c
3.40	14.00	0.69	0.49	0.16	2.00	11.00	13.00	13.00	18.00	23.00	14.00	MIR171d
5.40	18.00	0.37	7.80	0.45	8.00	8.50	4.40	4.70	8.70	16.00	8.20	MIR171e
20.00	36.00	22.00	34.00	2.40	20.00	29.00	23.00	19.00	47.00	210.00	20.00	MIR394b
130.00	180.00	140.00	230.00	410.00	270.00	380.00	370.00	380.00	330.00	220.00	340.00	MIR159a
4.50	2.30	5.30	0.87	15.00	3.20	7.20	23.00	20.00	22.00	3.80	7.60	MIR160d
0.00	0.15	0.04	0.19	3.00	0.08	0.00	0.02	0.00	0.00	0.16	0.00	MIR1432
0.00	0.13	0.00	0.23	3.70	0.00	0.06	0.00	0.00	0.00	0.00	0.00	MIR393c
2.00	6.00	2.80	2.50	16.00	3.00	3.50	2.20	1.00	1.90	1.50	1.00	MIR169c
0.58	8.20	4.20	3.00	24.00	2.30	2.60	0.83	0.96	3.00	10.00	1.10	MIR172c
5.80	13.00	6.90	9.20	58.00	2.80	12.00	14.00	18.00	24.00	7.40	5.80	MIR156k
0.12	5.70	1.90	2.00	4.60	1.30	0.82	0.26	0.02	0.04	0.77	0.03	MIR160c
0.06	1.20	0.04	0.21	1.60	0.16	0.13	0.00	0.00	0.00	0.08	0.00	MIR319b
0.00	0.09	0.00	0.02	0.62	0.09	0.00	0.02	0.24	1.30	0.16	0.13	MIR827
0.25	2.10	0.46	1.60	1.20	0.15	0.04	0.04	0.00	0.00	0.47	0.07	MIR156d
0.06	3.20	0.12	1.10	1.20	0.25	0.21	0.51	0.00	0.06	0.31	0.00	MIR172e
0.21	1.40	0.15	1.30	0.23	0.14	0.32	0.79	0.00	0.04	0.61	0.00	MIR171j
8.90	34.00	14.00	7.60	14.00	6.50	6.00	25.00	22.00	16.00	4.70	9.40	MIR164b
2.20	2.00	2.00	2.60	2.00	1.50	1.10	0.72	0.93	0.62	0.61	0.70	MIR167b
7.40	7.40	6.90	7.40	2.80	0.92	1.00	0.63	0.67	0.91	1.20	1.40	MIR168a
2.10	1.60	2.60	2.70	0.02	0.25	0.31	0.04	0.00	0.02	0.06	0.00	MIR396a
1.30	0.48	1.60	2.20	0.04	0.03	0.09	0.00	0.00	0.00	0.00	0.00	MIR396e
14.00	12.00	13.00	33.00	25.00	2.10	5.10	4.70	4.50	8.20	12.00	0.86	MIR394a
110.00	92.00	87.00	230.00	51.00	65.00	45.00	30.00	30.00	29.00	26.00	25.00	MIR159f
40.00	28.00	51.00	31.00	57.00	12.00	15.00	11.00	19.00	16.00	9.80	8.00	MIR168b
15.00	7.00	18.00	18.00	21.00	1.80	1.90	0.98	0.12	0.44	0.26	0.09	MIR319c
1.30	0.96	1.30	1.00	0.26	0.52	0.07	0.30	0.24	0.74	0.43	0.28	MIR2118c
3.60	3.40	3.20	2.20	0.55	1.30	1.30	0.94	1.20	1.70	0.91	1.40	MIR169e,h
8.40	9.00	8.60	5.80	6.80	7.00	6.70	4.00	5.10	9.20	4.80	3.90	MIR162
1.80	3.00	1.20	0.59	0.02	2.80	1.20	0.20	0.04	0.12	0.00	0.00	MIR164g
3.60	4.20	1.30	1.80	0.12	1.50	0.61	0.06	0.00	0.04	0.20	0.00	MIR164d
9.20	13.00	14.00	1.60	3.00	5.50	2.80	1.00	0.10	0.22	0.34	0.16	MIR319d

0 0.4 0.8

Figure 2. **variation in precursor gene expression patterns across the maize shoot apex.** Heatmap of 41 expressed precursors (RPM ≥1) are grouped into three clusters via hierarchical clustering analysis. Dash lines distinguish these clusters. Relative expression levels have been normalized and reflected in the color scale. Numbers on the heatmap depict expression value (RPM). Mer - meristem, Int – internode, Vas – vasculature.

Mechanisms appear in place to regulate miRNA accumulation at the post-transcriptional level. Nonetheless, the fact that miRNA abundance and precursor expression is strongly correlated suggests that miRNA accumulation is to a large extent determined at the transcriptional level (Figure 1F). We conclude, therefore, that precursor expression can thus serve as proxy for mature miRNA accumulation.
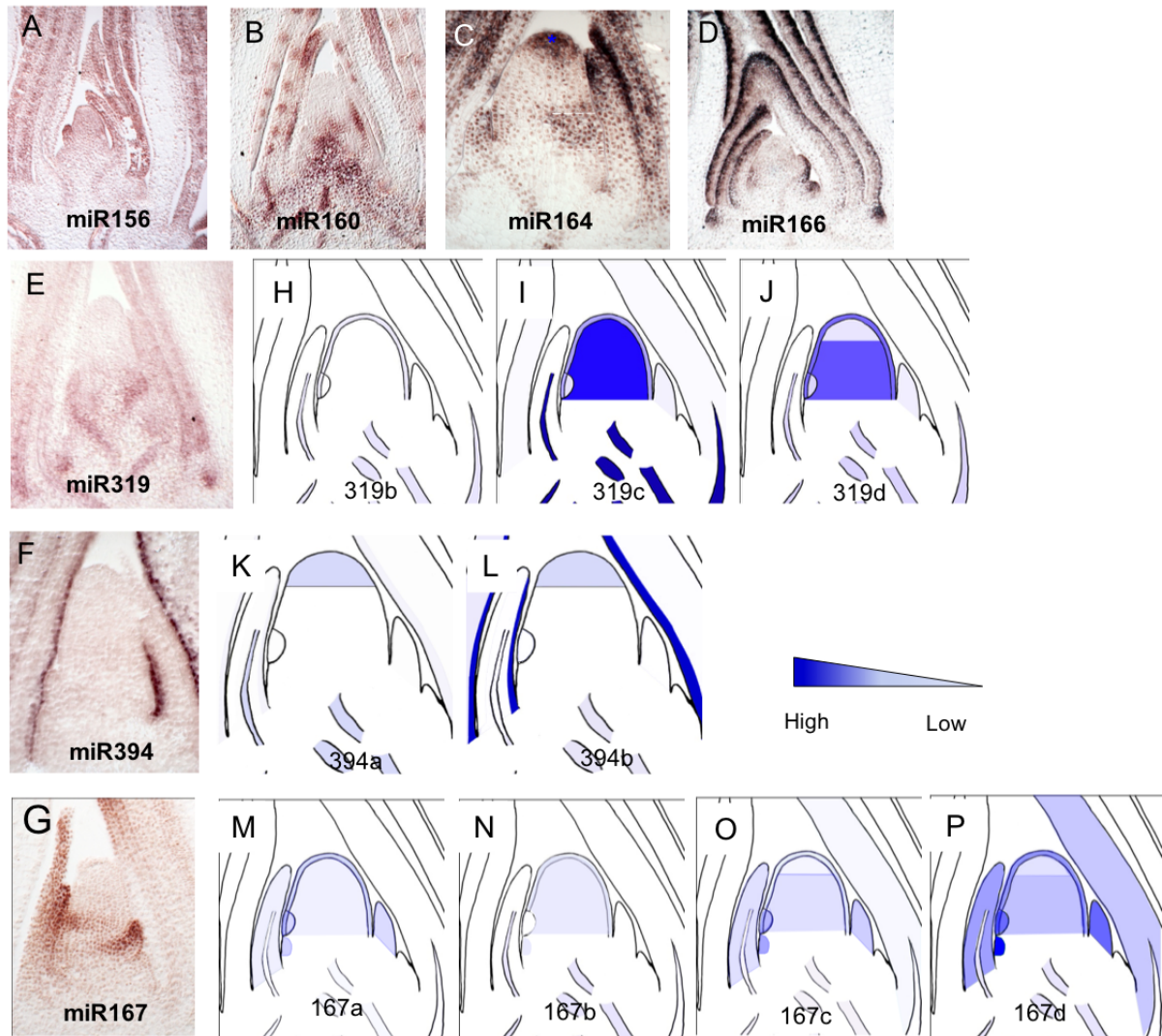
Spatial regulation of miRNA accumulation in the SAM

While a strong indicator of miRNA level, transcriptional regulation of precursor genes may not be a perfect prediction of the patterns of miRNA accumulation. Small RNAs, including miRNAs are able to move from cell to cell via plasmodesmata, and spread systemically through the vasculature (Vaten et al., 2011). The movement of small RNA is a carefully regulated process (Skopelitis et al., 2018). Particularly, small RNA movement between functional domains of the shoot stem cell niche and from the vasculature cambium is restricted. Small RNAs however serve as mobile positional signals within the developing leaf.

To investigate spatiotemporal regulation of miRNA accumulation, we determined patterns of miRNA localization in the shoot apex via *in situ* hybridization, focusing on those small RNAs implicated in development (Figure 3). Most miRNAs were found to accumulate in a pattern across the apex consistent with the expression domains of their precursors (Figure 2 and Figures 3A-E). For example, miR156 and its precursors are broadly expressed in the apex; miR160 and its precursors are mainly accumulated in leaf primordia and the vasculature; and miR164 is found within leaf primordia. miR166 is predominantly expressed in abaxial side of leaf primordia.

While the pattern of miRNA accumulation generally matches that of its precursors, miRNA mobility appears to effect miRNA accumulation. Besides miR166, which is known to move from its source on the abaxial side to generate an accumulate gradient across the adaxial-abaxial axis (Juarez et al., 2004), our data demonstrates the

mobility of miR167, miR319 and miR394. Opposite to miR166, miR394 moves from its source on the adaxial side towards the adaxial, generating an accumulate gradient (Figure 3F). miR319 moves from vasculature to their surrounding cells and miR167 from early leaf primordia to their surrounding cells (Figure 3E, 3G). We therefore summaries that the range of small RNA mobility is in part determined by the abundance of a small RNA.

Figure 3. *in situ* hybridization reveals regulation of miRNA accumulation in the maize shoot apex. (A-G) select miRNAs show distinct expression profiles. (H-J) In contrast to the mature miRNA (F), miR319 precursors show expression also in the SAM. (G-L) Likewise, in contrast to the mature miRNA (G), miR394 precursors accumulate in the CZ and vasculature (K-L). (M-P) Same as above, miR167 precursors also show accumulation in the SAM.

However, there are notable exceptions. For example, while mature miR319 is detectable primarily in the vasculature, its precursors are expressed in both the vasculature and the SAM CZ (Figure 3E and 3H-J). Similarly, miR394 accumulates on the adaxial side of developing leaf primordia in line with the pattern of precursor expression but is undetectable in the CZ despite the equivalent two precursors' expression in this domain of the SAM (Figure 3F and 3K-L). Finally, miR167 precursors are detectable in the CZ, while no mature miRNAs are. Although both of them are accumulated in internode and leaf primordia (Figure 3G and 3M-P).

To further investigate these miRNAs' accumulation regulation mechanisms, we generated RNA-Seq data from the meristematic CZ of a *dcl1-2* mutant line (Figure 1D-E, Dataset S3). miR167, miR319, and miR394 precursors' expression was found to be up-regulated in the mutant compared to wild type (Figure S4). This is consistent with the previous finding that all mature miRNAs' accumulation levels are decreased in *dcl1-2* compared to wild type (Petsch et al., 2015). Interestingly, the transcripts' increasing level among precursors is different in *dcl-1* relative to wildtype. This indicates that DCL1 processing ratio is different among precursors, and the relative expression level of precursor in *dcl1-2*/wildtype serves as a proxy for miRNA processing efficiency. This result reinforces the principle that mature small RNAs accumulation is regulated at spatial, transcriptional, and post-transcriptional levels in the SAM.

The efficiency of miRNA-directed transcript cleavage

To identify all targets, including conserved and non-conversed (novel) targets, we performed target prediction analysis by using the TargetFinder tool and validated them by degradome analysis. Altogether, 114 target genes were identified (Dataset S4), nearly all of which are conserved when compared to those recorded for *Arabidopsis*. This highlights the extensive conservation of not only miRNAs, but also their targets (Chorostecki et al., 2017).

The fact that potential target genes might not be expressed, weakly expressed, or translational repressed, means it is likely that other targets could not be detected by degradome sequencing. Given the observed conservation of targets, an additional 48 target genes were detected via phylogenetic analysis, 39 of which are not expressed

in the maize shoot apex (Dataset S4). For the other nine expressed target genes, seven of them belong to five expressed miRNA families: miR159, miR166, miR167, miR393 and miR396 (three) (Figure S5). miR166 target is highly accumulated in the CZ of the meristem, while miR166 does not accumulate in there (Figure S5B). miR159 and its target are co-expressed in the same domain, while polymorphisms in complementary binding sites are different from its other targets. Same to miR393, different polymorphisms in complementary binding sites are detected compared to their other targets (Figure S7). Three expressed miR396 targets have the same complementary binding sites and expression pattern as the other targets (Figure S5E, Figure S7), and the probability is that they are regulated at a translation level.
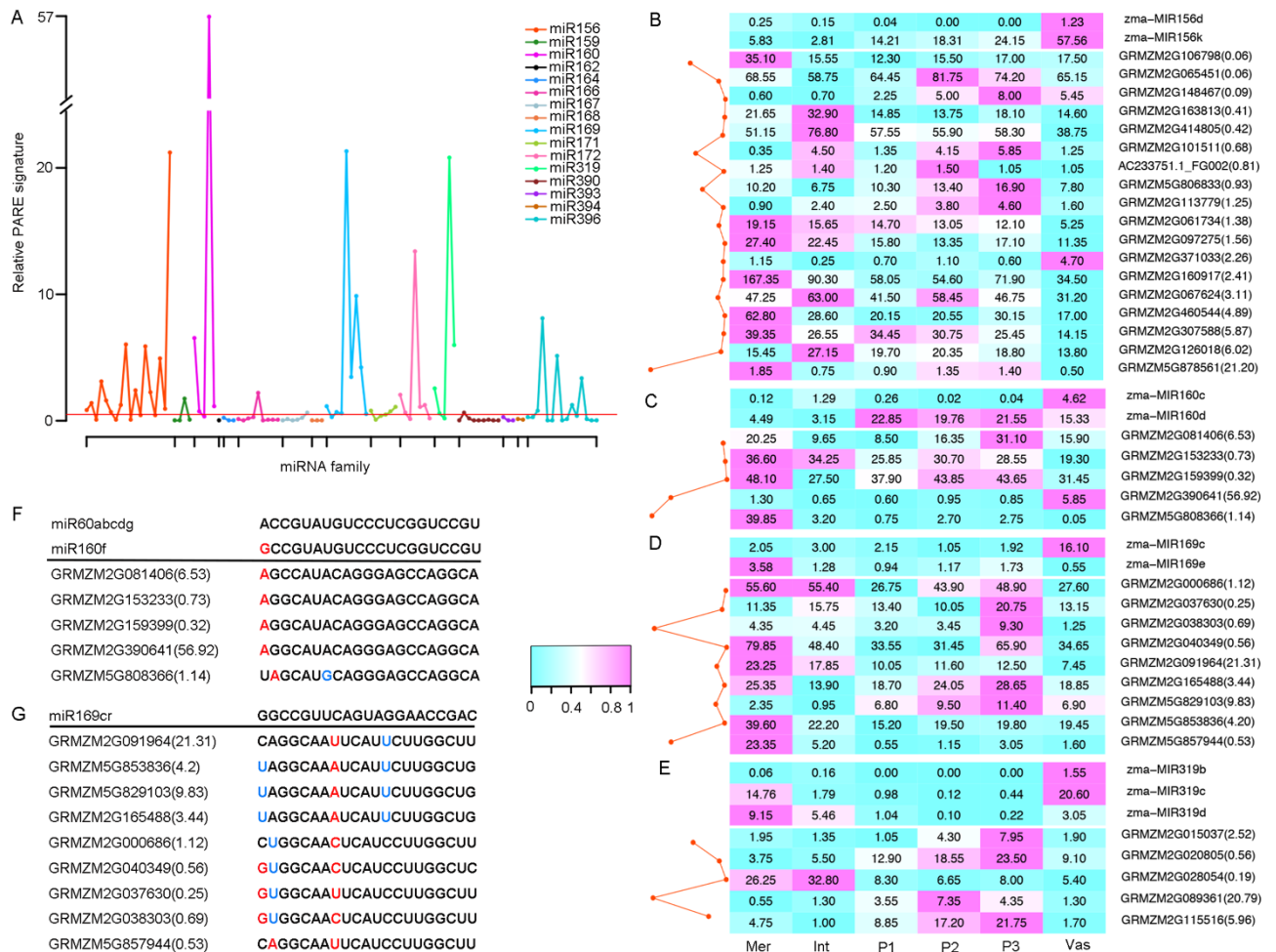
Figure 4. The sensitivity of miRNA-mediated transcript cleavage. (A) For 105 expressed targets (≥1 RPM in any one subdomain), the ratio of cleaved transcript to un-cleaved transcript was calculated by relative PARE signature. The PARE signature was divided by the targets' average expression in a miRNA accumulation domain. Below the red line (the relative

degradome signature is 0.5), the relative PARE signature of targets is low. (B-E) Heatmaps of miR156 (B), miR160 (C), miR169 (D), and miR319 (E) accumulation, and their targets' expression level in select apex subdomains: Mer - meristem, Int – internode, Vas - vasculature, P1, P2, and P3. Left side of heatmap reflects the relative accumulation of miRNA to targets. Bracketed numbers indicate the relative PARE signature. Red line (left of heatmap) is the ratio of miRNA expression to target expression (in miRNA accumulation domain). (F-G) Complementary binding sites between select miR160 (F) and miR169 (G) and their targets. The blue bases identify a G:U bulge. Red indicated a mismatch.

Sequence divergence during evolution can give rise to novel miRNA-target modules. Indeed, most miRNA families generate multiple mature miRNA variants, and the differences in complementarity affect the range of possible targets. For example, four miR159 variants do not always share the same targets (Dataset S4). Similarly, miR166, miR169, miR171, and miR396 variants regulate the different targets. Interestingly, one of three non-conserved targets are regulated by two miR166 variants, which is distinct from novel targets identified for miR166 in rice (Salvador-Guirao et al., 2018). This reflects the evolvability of miRNAs and their targets.

To globally estimate the sensitivity of efficacy of miRNA-directed transcript cleavage, we calculated the ratio of cleaved transcripts reflected by the PARE signatures, to the un-cleaved target mRNA levels (Figure 4A). The results show these ratios vary substantially even among targets within individual miRNA families, as well as targets across miRNA families. This indicates that miRNA-directed transcript cleavage efficacy differs between targets of the same miRNA. Relative, lower PARE signatures are indicative of either low miRNA accumulation, or miRNA mediated translational regulation. For example, miR166 targets are regulated at translational level and have low cleavage signatures.

Four targets coming from miR156, miR160, miR169 and miR319 show the highest sensitivity of cleavage efficacy. Interestingly, through analyzing the ratio of miRNA to targets, we discovered that the highly efficient cleavage of targets of miR156 and miR319 have the highest ratio of relative miRNA accumulation to targets (Figure 4B). However, miRNA160 and miR169 failed the above rule. A miR160 target gene (GRMZM5G808366) was found to be have the highest ratio of miRNA to target, but its cleavage efficacy is relatively low. By further analyzing the binding sites of miRNAs,

we uncovered that binding sites' complementarity polymorphisms are the second factor affecting miRNA efficacy besides the ratio of miRNA to targets. The low efficacy of targets gene (GRMZM5G808366) is because of more single nucleotide polymorphisms (SNPs) in miRNA complementarity binding site than other targets (Figure 4F). Similarly, SNPs in miRNA::target complementary binding site play a major role in determining the cleavage efficacy of targets gene (GRMZM2G0919164) regulated by miR169 (Figure 4G).

Aside from sequence complementarity polymorphisms, and the ratio of miRNA to target affecting the cleavage efficacy of target genes regulated by the same miRNA, there are several other factors that can affect the cleavage efficacy of target genes regulated by different miRNA families. For example, RNA-binding proteins or target accessibility (Li et al., 2014a). AGO1 mediates both target cleavage and translational repression (Iwakawa and Tomari, 2015) and the different binding efficacy of miRNA to AGO protein affects the efficacy of target cleavage .

**Complexity of the mechanisms of miRNA-mediated gene regulation across the meristem**

To understand the action of small RNAs on their target genes, it is necessary to know the regulatory mechanism (transcript cleavage or translational repression) and the regulatory mode (clearance or rheostat) that a miRNA employs in the regulation of its targets. To determine this, we built a model to infer the regulation mechanism and regulation mode by integrating analysis of the miRNA accumulation level, un-cleaved targets' mRNA accumulation level in miRNA accumulation domains, and the PARE signatures reflecting the target cleavage level (Table S1). A regulation mechanism model such as this relies on the principle that translational repression has only a subtle effect on mRNA accumulation. If the relative PARE signature is low, this indicates that a given miRNA does not cleave many transcripts. Conversely, if the miRNA-target level ratio is high (≥1). It suggests that there is enough miRNA to regulate the targets. If a target's expression in miRNA accumulating domains is high (≥10), this suggests that the targets are predominantly regulated by translation inhibition (Table S1). A regulation mode model relies on the principle that clearance would clear out the targets on a mRNA or protein level. If the PARE signature is high (≥0.5) and target's

accumulation in miRNA accumulating domains is low (<10), this suggests that targets are clear out by miRNA and it is regulated by miRNA at clearance mode (Table S1).

Based on our model, we systematically estimated the regulation mechanism and mode of miRNA/target modules. Nine families were found to regulate their targets at transcript cleavage level and seven miRNA families were found to regulate their targets at both a transcriptional cleavage and translational repression level (Dataset S5, Table S2). The regulation mechanism of miRNA families is highly consistent with the reported *Arabidopsis* miRNA regulation mechanisms (Table S2). miR156 and miR159 has been found to regulate their targets at both a transcript cleavage and translational inhibition level (Aukerman and Sakai, 2003; Gandikota et al., 2007; He et al., 2018; Li et al., 2014b; Li et al., 2013; Schwab et al., 2005). We predict that miR166, miR167, and miR394 predominantly regulate their targets through translational repression. Further, for the miR166 family, where miRNA accumulation is abundant, target expression also appears high, suggesting that miR166 targets are predominantly regulated by translational repression. Reinforcing this, a recent study showed that targets were cleared by miR166/miR165 at a protein level (Skopelitis et al., 2017).

To study further the link between the action of small RNAs on their target genes, and the contribution of miRNAs to development in the SAM, we analyzed the spatial correlation between the sum expression of precursors for each family, and the sum expression of corresponding targets genes. Typically, for translational repression, target mRNA was found to be only slightly changed following dynamic miRNA accumulation. For example, miR156 and miR172 accumulate in the vasculature, where target expression is reduced far less than in other domains (Figure 6SA). The same principle applies to abaxially-accumulated miR166 (Figure 6SA). Here, target expression decreases little in the abaxial domain, compared to the adaxial domain. With regards the other three dynamically expressed miRNA families: miR167 is predominantly expressed in leaf primordia, miR171 is mainly accumulated in L1, distinguishing it from L2 and miR394 distinguishes the adaxial and abaxial side. Different from other miRNAs, miR159 is broadly expressed in the shoot apex (Figure 6SB).

Taken together, we summarize that these miRNAs provide a major contribution to target-protein accumulation regulation in the SAM, and/or fine-tune target accumulation in a particular domain in which the miRNA accumulates. For the targets which are regulated on a transcript cleavage level, the dynamically expressed miRNAs spatially restrict their target mRNA accumulation in different cell types (Figure 5A). For example, meristem versus leaf primordia (miR160), vasculature (miR169), L1 versus L2 (miR164). Non-dynamically expressed miRNAs (such as miR162, miR390, miR393 and miR396) limit their targets broadly in all tissues (Figure 5B).
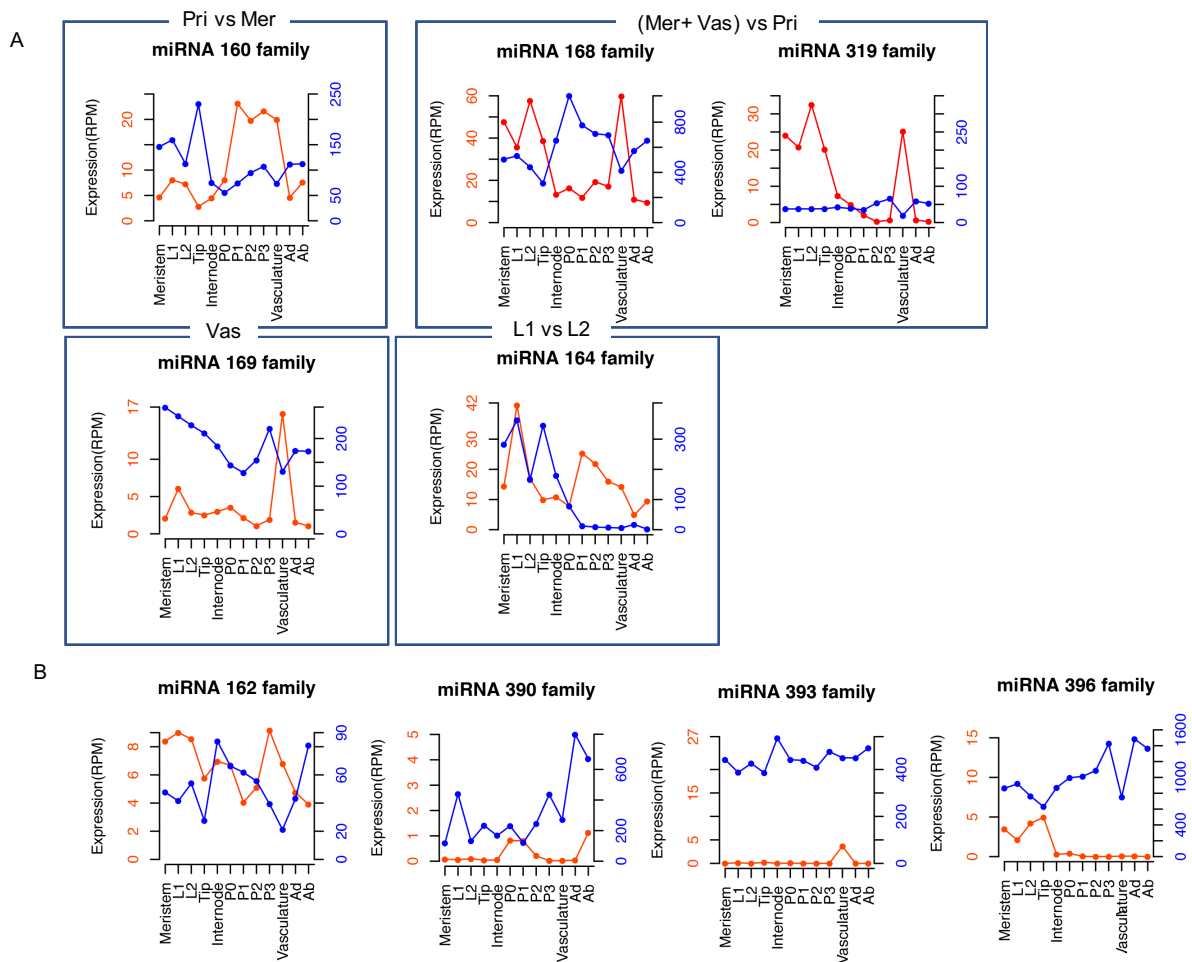
Figure 5: Nine miRNA families were found to regulate targets solely on a transcript-cleavage level in the maize apex. Expression of pri-miRNA and targets across 12 subdomains. (A) Dynamically expressed miRNAs spatially restrict target accumulation: miRNA160, miR168, miR319, miR169, and miR164 families. (B) Non-dynamically expressed miRNAs and targets: miR162, miR390, miR393, and miR396. Left axis (red): pri-miRNA expression level (sum expression of all expressed precursors in same family); right axis (blue): target gene expression level (sum expression of all targets to the same miRNA family). On the

top of box annotate miRNA mainly accumulation domain verse non-accumulation domain. Mer - meristem, Vas – vasculature, Pri – Primordia.

4.2.4 Discussion

The availability of high-resolution maize shoot apex transcriptome atlas data (Knauer et al., 2019) offered an opportunity to investigate the spatial temporal action of miRNAs during development. Substantial sub-functionalization of precursor genes reflects the concepts that miRNAs are regulated at transcriptional level. The inconsistency of miRNAs' and their precursors' expression level suggests that miRNAs' accumulation are also regulated at post-transcriptional level. Dynamic expression of enzymes participated in miRNA biology across shoot apex point to *DCL1* and *HESO1* as the two major factors affecting miRNAs spatiotemporally accumulation. Besides the above, the mobility of miRNAs is also revealed as a contributor to miRNA accumulation. These complex regulation of miRANs' spatiotemporal accumulation confers miRNAs robustness to drive development.

Targets are the central to understand miRNA function. miRNAs can obtain new function by acquiring new targets in the evolution process (Baldrich et al., 2018). Here, novel targets are revealed in maize by computational prediction and degradome-Seq validation. Actually, miRNA targets are highly conserved between *Arabidopsis* and maize. In *Arabidopsis*, SPL9 is more sensitive than SPL13 to miR156/miR157-directed transcript cleavage (He et al., 2018). Consistent with this, our result reveals the distinct sensitivity of efficacy of miRNAs-mediated transcriptional cleavage. These different efficacies of miRNA-mediated cleavages are related to the nature of polymorphisms in miRNA/targets complementary binding sites and miRNA/target ratio. In animals, miRNA polymorphisms are associated with disease, and single nucleotide polymorphisms in miRNA/targets binding sites may result in the escape of a target from inhibition or degradation by a miRNA (Chin et al., 2008; de Almeida et al., 2018; Gebert and MacRae, 2019; Gong et al., 2012). In addition, target accessibility is also determined by the secondary structure of target genes (Li et al., 2014a; Zheng et al., 2017), RNA-binding proteins (Li et al., 2014a), and AGO proteins (Iwakawa and Tomari, 2015; Jeong et al., 2013). These are also suggested factors affecting the different miRNA families mediated cleavage efficacy.

Here, we systematically reveal more miRNAs mediated regulation mechanisms by integrating RNA-seq and degradome-Seq. Besides the known miRNA regulation mechanisms (Gandikota et al., 2007; He et al., 2018; Li et al., 2014b; Schwab et al., 2005), here we report three more miRNAs who mediated their targets' expression through translational regulation, pointing out the underestimation of the translational regulation in plants. Additionally, we also reveal the miRNAs regulation mode (clearance and rheostat), which offers the resource for further genetic analysis of the biological phenotype variation. Finally, we reveal the spatiotemporal interaction of miRNA/target modules across maize shoot apex, which form sharp boundaries of expression within tissues and determine the fate of cell types. The strong or slight anti-correlation of spatiotemporal transcriptional accumulation of miRNA and their targets was related to the regulation mechanism and mode.

In conclusion, this study provides a foundation for further investigate mechanism of miRNA regulation and offers new insights to miRNA dynamical action in plant development.

4.2.5 Materials and methods:

Plant materials

All analyses were performed on 14 day-old seedlings grown under 16 h 24°C light and 8 h 20°C dark cycles. The *dcl1-2* allele (Petsch et al., 2015) was introgressed for 4 generations into B73 prior to analysis.

Laser microdissection and RNA-Seq library construction and sequencing

Cells of interest were collected by laser microdissection as described previously (Knauer et al. 2019) from at least six independent apices per biological replicate. Adaxial and abaxial tissues were collected from P2 and P3 leaf primordia of 14-day-old B73 seedling apices. For the dcl1-2 to wild-type comparisons, the center zone and P2-P3 leaf primordia were dissected from 14-day-old seedling apices of mutant and wild-type siblings. RNA was extracted and linearly amplified as described in Knauer et al. (2019), and single-end RNA-Seq libraries constructed using standard Illumina protocols (Illumina). Libraries for the adaxial and abaxial sides of leaf primordia were

sequenced (100bp) on the Illumina HighSeq2000 platform, whereas the *dcl1-2* and wild-type sibling libraries were sequenced (68 bp) on the Illumina HiSeq2500 platform.

Annotation of pri-miRNAs and genes with functions in miRNA biology

Considering that some miRNA genes were not well predicted and wrongly annotated among the Filtered Gene Set (version FGSv5b), genomic coordinates for all pri-miRNAs were manual curation based on available transcriptome data for B73 and Mo17 apices (Li et al., 2013a). Pre-miRNA annotations in miRbase (version 2) were blasted back to B73 RefGen_v2, and the location of the pri-miRNA defined based on the distribution of RNA-Seq reads at the locus. For MIR genes for which transcriptome data was not available, the gene model was annotated to cover the pre-miRNA hairpin +/- 100 nucleotides. Curated precursor locations were used to determine transcript accumulation across tissues. Maize genes with functions in small RNA biology were identified based on information from published work, or through identification of maize homologs of known Arabidopsis genes using standard homology searches (Blastp 2.26++) or the paralog search tool in BioMart (http://www.gramene.org) (Tello-Ruiz et al., 2017). Expression values for genes with paralogous functions were summed together, and relative expression across the twelve SAM domains determined.

Gene expression analysis

RNA-Seq data for the adaxial and abaxial tissue samples was processed and analyzed as described in Knauer et al. (2019). Trimmed reads were aligned to the B73 RefGen_V2 using GSNAP, and uniquely mapped reads allowing ≤2 mismatches every 36 bp and less than 5 bases for every 75 bp as tails were used for subsequent analyses. pri-miRNA expression values were calculated for all twelve B73 shoot apex samples based on the annotated precursor models. Target gene expression values were calculated similarly, or collected from Knauer et al. (2019). Heatmap analysis used the R package gplots heatmap.2 and clustering method was set default hierarchical clustering.

For the dcl1-2 to wild-type comparisons, sequence reads (single-end, 68 bp) were trimmed using Trimmomatic version 0.36 (Bolger et al., 2014), and aligned to the B73 RefGen_v2 reference genome with TopHat version 2.1.1(Kim et al., 2013). Gene

(based on version FGSv5b and the corrected precursor annotation) expression values were calculated on uniquely mapped reads using HTSeq version 0.9 (Anders et al., 2015), and DEseq2 (Love et al., 2014) was used to calculate differentially expression (absolute log2FC ≥1 and q <0.05) on genes with expression levels ≥1 RPM in at least one library.

Small RNA data analysis

Total RNA was extracted from shoot apices comprising the SAM and up to four leaf primordia using TRIzol reagent (Invitrogen) followed by treatment with DNase I (Promega). Small RN-Seq libries were prepared from 1.2 μg total RNA using TruSeq Small RNA sample preparation kit (Illumina). Libraries were quantified with the KAPA Illumina library Quantification Kit (KAPABIOSYSTEMS) and sequenced on the Illumina HiSeq2000 platform. Sequence reads were trimmed using Cutadapt version 1.13 (Martin, 2011) and trimmed reads, 18-26 nucleotides in length, aligned to the maize B73 RefGen_v2 genome (release 5a.57) using Bowtie version 1.1.2 (Langmead et al., 2009), allowing no mismatches and a maximum of 20 alignments per read. Reads matching known structural RNAs (rRNAs, tRNAs, sn-RNAs and sno-RNAs) from Rfam database (http://www.sanger.ac.uk/software/Rfam) and GeneBank noncoding RNA database (http://blast.ncbi.nlm.nih.gov/) were removed from further analysis. The remaining sequences were annotated using bedtools version 2.25.0 (Quinlan and Hall, 2010) to known miRNAs in miRbasev2.

Small RNA Target identification

Potential targets were predicted using Target Finder (https://github.com/carringtonlab/TargetFinder), allowing a maximum score of 5. Predicted targets were validated by PARE data generated previously (Dotto et al., 2014) from B73 seedling apices. PARE tag abundance was calculated for the large (31nt, W_L) and small (5nt, W_s) window, and cleavage sites filtered to retain only those for which W_s/W_L ≥0.75 and W_s ≥ 4. Homologous analysis of predicted targets using MapMan annotations of the maize filtered gene set v5b.60 (Usadel et al., 2009), aids to annotate *Arabidopsis* homologous genes and allowed us to identify the potential target genes without PARE signature.

Correlation analysis

Total expression across the meristem, internode, P1, P2, P3 and vasculature (non-overlapping tissues in the atlas) was calculated for all precursors associated with a give mature miRNA. Total precursor expression and miRNA abundance were normalized at log-scale, and the Pearson-correlation calculated in R.

RT-PCR and *in situ* hybridization

For RT-PCR, 4 ug RQ1 DNAse (Promega) treated a RNA was converted into cDNA using the Supterscript III First-Strand synthesis System (Invitrogen) with random hexamer primers according to manufacturer's protocol. *In situ* hybridizations were performed on apices of 14 day-old B73 seedling according to (Javelle and Timmermans, 2012) using the following probe concentrations and hybridization:

probe	sequence	label	hybridization temperature	amount pmol/slide
miR156	GTGCTCACTCTCTTCTGTCA	5'+3'-DIG	50°C	10
miR160	TGGCATACAGGGAGCCAGGCA	5'+3'-DIG	55°C	10
miR164	TGCACGTGCCCTGCTTCTCCA	5'-DIG	57°C	0.5
miR166	GGGAATGAAGCCTGGTCCGA	5'-DIG	50°C	10
miR167	TAGATCATGCTGGCAGCTTCA	5'-DIG	55°C	10
miR319	GGGAGCACCCTTCAGTCCAA	5'+3'-DIG	50°C	10
miR394	GGAGGTGGACAGAATGCCAA	5'+3'-DIG	55°C	5

4.2.6 References:

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166-169.

Aukerman, M.J., and Sakai, H. (2003). Regulation of flowering time and floral organ identity by a MicroRNA and its APETALA2-like target genes. Plant Cell *15*, 2730-2741.

Baker, C.C., Sieber, P., Wellmer, F., and Meyerowitz, E.M. (2005). The early extra petals1 mutant uncovers a role for microRNA miR164c in regulating petal number in Arabidopsis. Curr Biol *15*, 303-315.

Baldrich, P., Beric, A., and Meyers, B.C. (2018). Despacito: the slow evolutionary changes in plant microRNAs. Curr Opin Plant Biol *42*, 16-22.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114-2120.

Chen, X. (2004). A MicroRNA as a Translational Repressor of APETALA2 in Arabidopsis Flower Development. Science *303*, 2022-2025.

Chin, L.J., Ratner, E., Leng, S., Zhai, R., Nallur, S., Babar, I., Muller, R.U., Straka, E., Su, L., Burki, E.A.*, et al.* (2008). A SNP in a let-7 microRNA complementary site in the KRAS 3' untranslated region increases non-small cell lung cancer risk. Cancer Res *68*, 8535-8540.

Chitwood, D.H., Guo, M., Nogueira, F.T., and Timmermans, M.C. (2007). Establishing leaf polarity: the role of small RNAs and positional signals in the shoot apex. Development *134*, 813-823.

Chitwood, D.H., Nogueira, F.T., Howell, M.D., Montgomery, T.A., Carrington, J.C., and Timmermans, M.C. (2009). Pattern formation via small RNA mobility. Genes Dev *23*, 549-554.

Chorostecki, U., Moro, B., Rojas, A.M.L., Debernardi, J.M., Schapire, A.L., Notredame, C., and Palatnik, J.F. (2017). Evolutionary Footprints Reveal Insights into Plant MicroRNA Biogenesis. Plant Cell *29*, 1248-1261.

D'Ario, M., Griffiths-Jones, S., and Kim, M. (2017). Small RNAs: Big Impact on Plant Development. Trends Plant Sci *22*, 1056-1068.

de Almeida, R.C., Chagas, V.S., Castro, M.A.A., and Petzl-Erler, M.L. (2018). Integrative Analysis Identifies Genetic Variants Associated With Autoimmune Diseases Affecting Putative MicroRNA Binding Sites. Front Genet *9*, 139.

Dotto, M.C., Petsch, K.A., Aukerman, M.J., Beatty, M., Hammell, M., and Timmermans, M.C. (2014). Genome-wide analysis of leafbladeless1-regulated and phased small RNAs underscores the importance of the TAS3 ta-siRNA pathway to maize development. PLoS Genet *10*, e1004826.

Fouracre, J.P., and Poethig, R.S. (2016). The role of small RNAs in vegetative shoot development. Curr Opin Plant Biol *29*, 64-72.

Gandikota, M., Birkenbihl, R.P., Hohmann, S., Cardon, G.H., Saedler, H., and Huijser, P. (2007). The miRNA156/157 recognition element in the 3' UTR of the Arabidopsis SBP box gene SPL3 prevents early flowering by translational inhibition in seedlings. Plant J *49*, 683-693.

Gebert, L.F.R., and MacRae, I.J. (2019). Regulation of microRNA function in animals. Nat Rev Mol Cell Biol *20*, 21-37.

Gong, J., Tong, Y., Zhang, H.M., Wang, K., Hu, T., Shan, G., Sun, J., and Guo, A.Y. (2012). Genome-wide identification of SNPs in microRNA genes and the SNP effects on microRNA target binding and biogenesis. Hum Mutat *33*, 254-263.

He, J., Xu, M., Willmann, M.R., McCormick, K., Hu, T., Yang, L., Starker, C.G., Voytas, D.F., Meyers, B.C., and Poethig, R.S. (2018). Threshold-dependent repression of SPL gene expression by miR156/miR157 controls vegetative phase change in Arabidopsis thaliana. PLoS Genet *14*, e1007337.

Irish, V.F., and Sussex, M. (1990). Function of the apetala-1 Gene during Arabídopsis Floral Development. Plant Cell *2*, 741-753.

Iwakawa, H.O., and Tomari, Y. (2015). The Functions of MicroRNAs: mRNA Decay and Translational Repression. Trends Cell Biol *25*, 651-665.

Javelle, M., and Timmermans, M.C. (2012). In situ localization of small RNAs in plants by using LNA probes. Nat Protoc *7*, 533-541.

Jeong, D.H., Thatcher, S.R., Brown, R.S., Zhai, J., Park, S., Rymarquis, L.A., Meyers, B.C., and Green, P.J. (2013). Comprehensive investigation of microRNAs enhanced by analysis of sequence variants, expression patterns, ARGONAUTE loading, and target cleavage. Plant Physiol *162*, 1225-1245.

Juarez, M.T., Kui, J.S., Thomas, J., Heller, B.A., and Timmermans, M.C. (2004). microRNA-mediated repression of rolled leaf1 specifies maize leaf polarity. Nature *428*, 81-84.

Jung, J.H., and Park, C.M. (2007). MIR166/165 genes exhibit dynamic expression patterns in regulating shoot apical meristem and floral development in Arabidopsis. Planta *225*, 1327-1338.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene

fusions. Genome Biol *14*, 36-49.

Knauer, S., Holt, A.L., Rubio-Somoza, I., Tucker, E.J., Hinze, A., Pisch, M., Javelle, M., Timmermans, M.C., Tucker, M.R., and Laux, T. (2013). A protodermal miR394 signal defines a region of stem cell competence in the Arabidopsis shoot meristem. Dev Cell *24*, 125-132.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

Li, J., Reichel, M., Li, Y., and Millar, A.A. (2014a). The functional scope of plant microRNA-mediated silencing. Trends Plant Sci *19*, 750-756.

Li, J., Reichel, M., and Millar, A.A. (2014b). Determinants beyond both complementarity and cleavage govern microR159 efficacy in Arabidopsis. PLoS Genet *10*, e1004232.

Li, L., Petsch, K., Shimizu, R., Liu, S., Xu, W.W., Ying, K., Yu, J., Scanlon, M.J., Schnable, P.S., Timmermans, M.C.*, et al.* (2013a). Mendelian and non-Mendelian regulation of gene expression in maize. PLoS Genet *9*, e1003202.

Li, S., Liu, L., Zhuang, X., Yu, Y., Liu, X., Cui, X., Ji, L., Pan, Z., Cao, X., Mo, B.*, et al.* (2013b). MicroRNAs inhibit the translation of target mRNAs on the endoplasmic reticulum in Arabidopsis. Cell *153*, 562-574.

Li, X., Shahid, M.Q., Wu, J., Wang, L., Liu, X., and Lu, Y. (2016). Comparative Small RNA Analysis of Pollen Development in Autotetraploid and Diploid Rice. Int J Mol Sci *17*, 499.

Liu, P.P., Montgomery, T.A., Fahlgren, N., Kasschau, K.D., Nonogaki, H., and Carrington, J.C. (2007). Repression of AUXIN RESPONSE FACTOR10 by microRNA160 is critical for seed germination and post-germination stages. Plant J *52*, 133-146.

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.

Mallory, A.C., Bartel, D.P., and Bartel, B. (2005). MicroRNA-directed regulation of Arabidopsis AUXIN RESPONSE FACTOR17 is essential for proper development and modulates expression of early auxin response genes. Plant Cell *17*, 1360-1375.

Mallory, A.C., Dugas, D.V., Bartel, D.P., and Bartel, B. (2004). MicroRNA regulation of NAC-domain targets is required for proper formation and separation of adjacent embryonic, vegetative, and floral organs. Curr Biol *14*, 1035-1046.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal *17*, 10-12.

Merchan, F., Boualem, A., Crespi, M., and Frugier, F. (2009). Plant polycistronic precursors containing non-homologous microRNAs target transcripts encoding functionally related proteins. Genome Biol *10*, R136.

Nikovics, K., Blein, T., Peaucelle, A., Ishida, T., Morin, H., Aida, M., and Laufs, P. (2006). The balance between the MIR164A and CUC2 genes controls leaf margin serration in Arabidopsis. Plant Cell *18*, 2929-2945.

Nogueira, F.T., Chitwood, D.H., Madi, S., Ohtsu, K., Schnable, P.S., Scanlon, M.J., and Timmermans, M.C. (2009). Regulation of small RNA accumulation in the maize shoot apex. PLoS Genet *5*, e1000320.

Nogueira, F.T.S., Madi, S., Chitwood, D.H., Juarez, M.T., and Timmermans, M.C.P. (2007). Two small regulatory RNAs establish opposing fates of a developmental axis. Genes & Development *21*, 750-755.

Peaucelle, A., Morin, H., Traas, J., and Laufs, P. (2007). Plants expressing a miR164-resistant CUC2 gene reveal the importance of post-meristematic maintenance of phyllotaxy in Arabidopsis. Development *134*, 1045-1050.

Petsch, K., Manzotti, P.S., Tam, O.H., Meeley, R., Hammell, M., Consonni, G., and Timmermans, M.C. (2015). Novel DICER-LIKE1 siRNAs Bypass the Requirement for DICER-LIKE4 in Maize Development. Plant Cell *27*, 2163-2177.

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

Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B., and Bartel, D.P. (2002). MicroRNAs in plants. Genes Dev *16*, 1616-1626.

Sakaguchi, J., and Watanabe, Y. (2012). miR165/166 and the development of land plants. Development, Growth & Differentiation *54*, 93-99.

Salvador-Guirao, R., Hsing, Y.I., and San Segundo, B. (2018). The Polycistronic miR166k-166h Positively Regulates Rice Immunity via Post-transcriptional Control of EIN2. Front Plant Sci *9*, 337.

Schwab, R., Palatnik, J.F., Riester, M., Schommer, C., Schmid, M., and Weigel, D. (2005). Specific effects of microRNAs on the plant transcriptome. Dev Cell *8*, 517-527.

Schwarz, S., Grande, A.V., Bujdoso, N., Saedler, H., and Huijser, P. (2008). The microRNA regulated SBP-box genes SPL9 and SPL15 control shoot maturation in Arabidopsis. Plant Mol Biol *67*, 183-195.

Sieber, P., Wellmer, F., Gheyselinck, J., Riechmann, J.L., and Meyerowitz, E.M. (2007). Redundancy and specialization among plant microRNAs: role of the MIR164 family in developmental robustness. Development *134*, 1051-1060.

Singh, A., Gautam, V., Singh, S., Sarkar Das, S., Verma, S., Mishra, V., Mukherjee, S., and Sarkar, A.K. (2018). Plant small RNAs: advancement in the understanding of biogenesis and role in plant development. Planta *248*, 545-558.

Skopelitis, D.S., Benkovics, A.H., Husbands, A.Y., and Timmermans, M.C.P. (2017). Boundary Formation through a Direct Threshold-Based Readout of Mobile Small RNA Gradients. Dev Cell *43*, 265-273 e266.

Skopelitis, D.S., Hill, K., Klesen, S., Marco, C.F., von Born, P., Chitwood, D.H., and Timmermans, M.C.P. (2018). Gating of miRNA movement at defined cell-cell interfaces governs their impact as positional signals. Nat Commun *9*, 3107.

Song, J.B., Huang, S.Q., Dalmay, T., and Yang, Z.M. (2012). Regulation of leaf morphology by microRNA394 and its target LEAF CURLING RESPONSIVENESS. Plant Cell Physiol *53*, 1283-1294.

Swarup, R., and Denyer, T. (2019). miRNAs in Plant Development. Annual Plant Review *Accepted, pending publication.*

Tello-Ruiz, M.K., Naithani, S., Stein, J.C., Gupta, P., Campbell, M., Olson, A., Wei, S., Preece, J., Geniza, M.J., and Jiao, Y. (2017). Gramene 2018: unifying comparative genomics and pathway resources for plant research. Nucleic Acids Research *46*, D1181-D1189.

Usadel, B., Obayashi, T., Mutwil, M., Giorgi, F.M., Bassel, G.W., Tanimoto, M., Chow, A., Steinhauser, D., Persson, S., and Provart, N.J. (2009). Co-expression tools for plant biology: opportunities for hypothesis generation and caveats. Plant Cell Environ *32*, 1633-1651.

Vaten, A., Dettmer, J., Wu, S., Stierhof, Y.D., Miyashima, S., Yadav, S.R., Roberts, C.J., Campilho, A., Bulone, V., Lichtenberger, R*., et al.* (2011). Callose biosynthesis regulates symplastic trafficking during root development. Dev Cell *21*, 1144-1155.

Wang, J.W., Schwab, R., Czech, B., Mica, E., and Weigel, D. (2008). Dual effects of miR156-targeted SPL genes and CYP78A5/KLUH on plastochron length and organ size in Arabidopsis thaliana. Plant Cell *20*, 1231-1243.

Wang, J.W., Wang, L.J., Mao, Y.B., Cai, W.J., Xue, H.W., and Chen, X.Y. (2005). Control of root cap formation by MicroRNA-targeted auxin response factors in Arabidopsis. Plant Cell *17*, 2204-2216.

Yang, T., Wang, Y., Teotia, S., Zhang, Z., and Tang, G. (2018). The Making of Leaves: How Small RNA Networks Modulate Leaf Development. Front Plant Sci *9*, 824.

Zhai, J., Zhao, Y., Simon, S.A., Huang, S., Petsch, K., Arikit, S., Pillay, M., Ji, L., Xie, M., Cao, X*., et al.* (2013). Plant microRNAs display differential 3' truncation and tailing modifications that are ARGONAUTE1 dependent and conserved across species. Plant Cell *25*, 2417-2428.

Zheng, Z., Reichel, M., Deveson, I., Wong, G., Li, J., and Millar, A.A. (2017). Target RNA Secondary Structure Is a Major Determinant of miR159 Efficacy. Plant Physiol *174*, 1764-1778.

4.2.7 Supplementary figures:

Figure S1. Precision of LCM of Ab and Ad. (A-B) *in-situ* expression pattern for maker genes *IG1* and *ARF3a*, matching their transcripts levels in the Ad and Ab domain of RNA-Seq libraries (C).

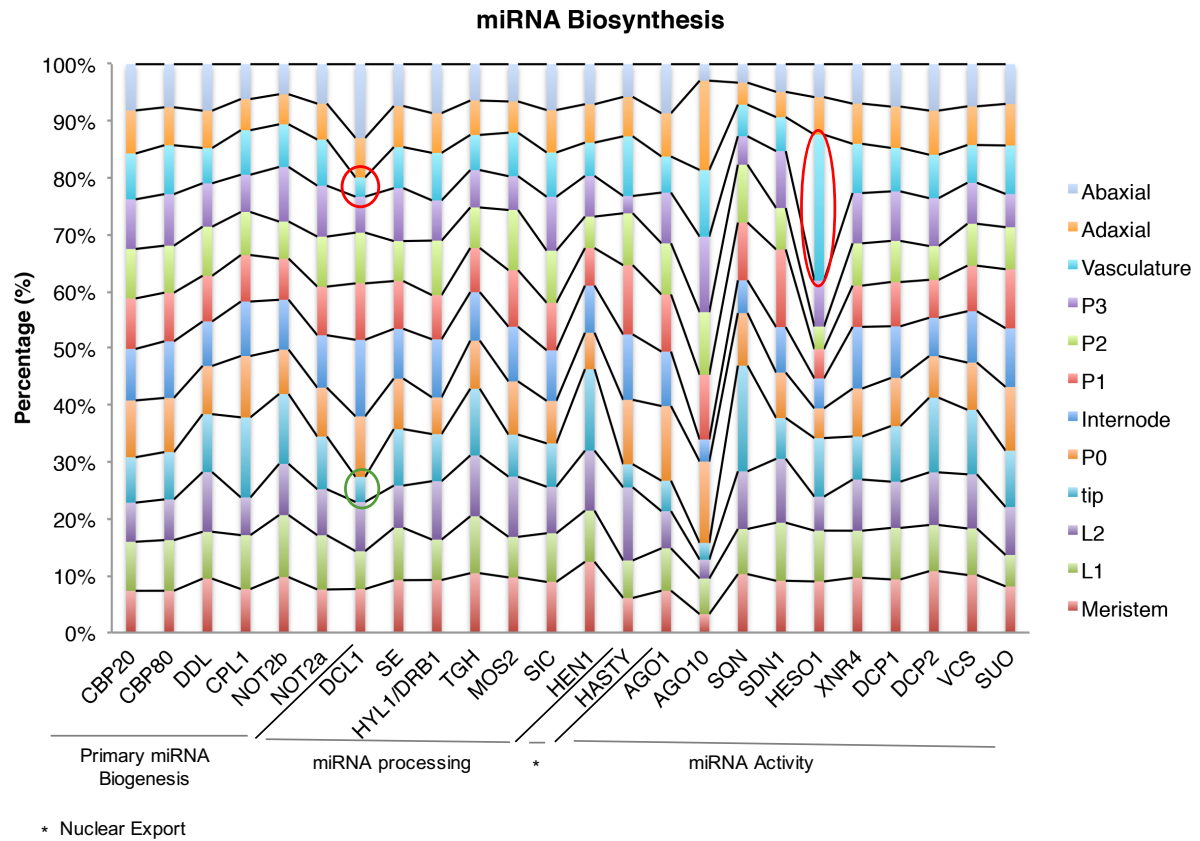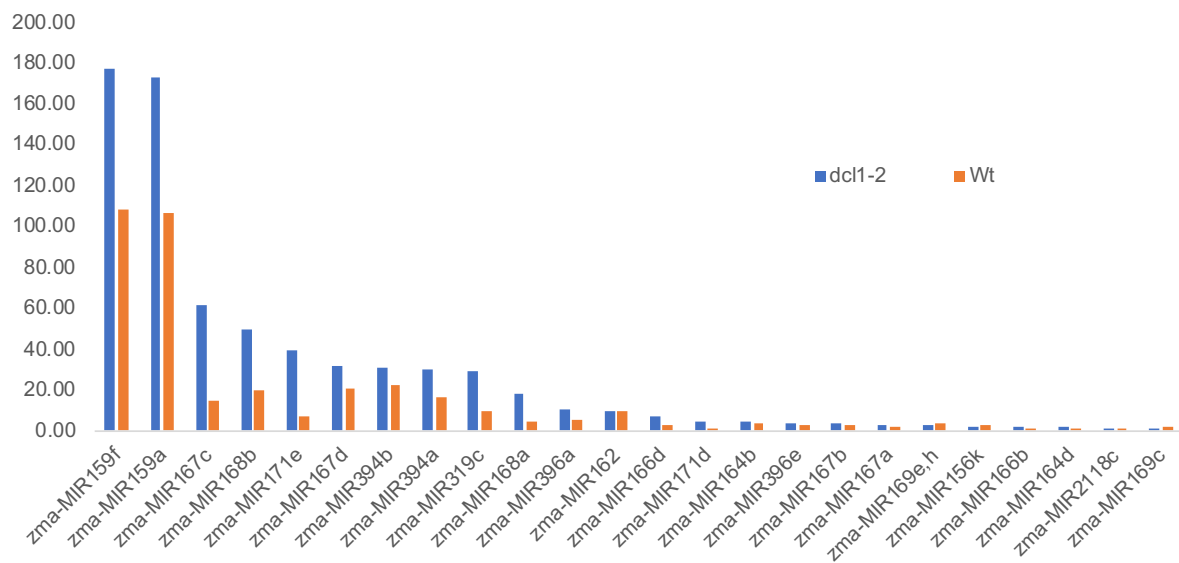| Mer	L1	L2	Tip	Int	P0	P1	P2	P3	Vas	Ad	Ab	
0.25	2.10	0.46	1.60	0.15	0.04	0.04	0.00	0.00	1.20	0.47	0.07	MIR156d
5.80	13.00	6.90	9.20	2.80	12.00	14.00	18.00	24.00	58.00	7.40	5.80	MIR156k
130.00	180.00	140.00	230.00	270.00	380.00	370.00	380.00	330.00	410.00	220.00	340.00	MIR159a
110.00	92.00	87.00	230.00	65.00	45.00	30.00	30.00	29.00	51.00	26.00	25.00	MIR159f
0.12	5.70	1.90	2.00	1.30	0.82	0.26	0.02	0.04	4.60	0.77	0.03	MIR160c
4.50	2.30	5.30	0.87	3.20	7.20	23.00	20.00	22.00	15.00	3.80	7.60	MIR160d
8.90	34.00	14.00	7.60	6.50	6.00	25.00	22.00	16.00	14.00	4.70	9.40	MIR164b
3.60	4.20	1.30	1.80	1.50	0.61	0.06	0.00	0.04	0.12	0.20	0.00	MIR164d
1.80	3.00	1.20	0.59	2.80	1.20	0.20	0.04	0.12	0.02	0.00	0.00	MIR164g
1.40	4.80	0.57	0.84	2.00	12.00	13.00	12.00	22.00	0.89	1.20	26.00	MIR166b
7.00	3.10	5.50	0.40	52.00	10.00	4.50	2.00	1.60	0.39	1.10	7.50	MIR166c
1.00	9.60	3.60	5.20	7.60	23.00	49.00	43.00	54.00	4.10	5.40	180.00	MIR166d
7.70	10.00	3.80	0.60	45.00	39.00	13.00	5.60	7.70	0.73	3.00	31.00	MIR166k,m
12.00	14.00	8.50	3.60	120.00	53.00	38.00	23.00	31.00	1.70	9.30	180.00	MIR166n
8.20	18.00	1.70	2.70	1.40	71.00	71.00	23.00	7.90	0.45	5.90	7.50	MIR167a,g
2.20	2.00	2.00	2.60	1.50	1.10	0.72	0.93	0.62	2.00	0.61	0.70	MIR167b
27.00	22.00	28.00	10.00	8.60	100.00	56.00	31.00	22.00	1.20	15.00	24.00	MIR167c
170.00	210.00	120.00	33.00	180.00	1400.00	830.00	400.00	220.00	78.00	220.00	330.00	MIR167d
7.40	7.40	6.90	7.40	0.92	1.00	0.63	0.67	0.91	2.80	1.20	1.40	MIR168a
40.00	28.00	51.00	31.00	12.00	15.00	11.00	19.00	16.00	57.00	9.80	8.00	MIR168b
2.00	6.00	2.80	2.50	3.00	3.50	2.20	1.00	1.90	16.00	1.50	1.00	MIR169c
3.60	3.40	3.20	2.20	1.30	1.30	0.94	1.20	1.70	0.55	0.91	1.40	MIR169e,h
3.40	14.00	0.69	0.49	2.00	11.00	13.00	13.00	18.00	0.16	23.00	14.00	MIR171d
5.40	18.00	0.37	7.80	8.00	8.50	4.40	4.70	8.70	0.45	16.00	8.20	MIR171e
0.21	1.40	0.15	1.30	0.14	0.32	0.79	0.00	0.04	0.23	0.61	0.00	MIR171j
0.58	8.20	4.20	3.00	2.30	2.60	0.83	0.96	3.00	24.00	10.00	1.10	MIR172c
0.06	3.20	0.12	1.10	0.25	0.21	0.51	0.00	0.06	1.20	0.31	0.00	MIR172e
0.06	1.20	0.04	0.21	0.16	0.13	0.00	0.00	0.00	1.60	0.08	0.00	MIR319b
15.00	7.00	18.00	18.00	1.80	1.90	0.98	0.12	0.44	21.00	0.26	0.09	MIR319c
9.20	13.00	14.00	1.60	5.50	2.80	1.00	0.10	0.22	3.00	0.34	0.16	MIR319d
14.00	12.00	13.00	33.00	2.10	5.10	4.70	4.50	8.20	25.00	12.00	0.86	MIR394a
20.00	36.00	22.00	34.00	20.00	29.00	23.00	19.00	47.00	2.40	210.00	20.00	MIR394b
2.10	1.60	2.60	2.70	0.25	0.31	0.04	0.00	0.02	0.02	0.06	0.00	MIR396a
1.30	0.48	1.60	2.20	0.03	0.09	0.00	0.00	0.00	0.04	0.00	0.00	MIR396e

Figure S2. Precursor gene expression patterns in shoot apex. Heatmap of 41 expressed precursors (RPM ≥1) (corresponding to Figure 2) are shown based on family order. Relative expression levels have been normalized and reflected in color scale. Numbers on the heatmap depict expression value (RPM). Mer - meristem, Int – internode, Vas – vasculature.
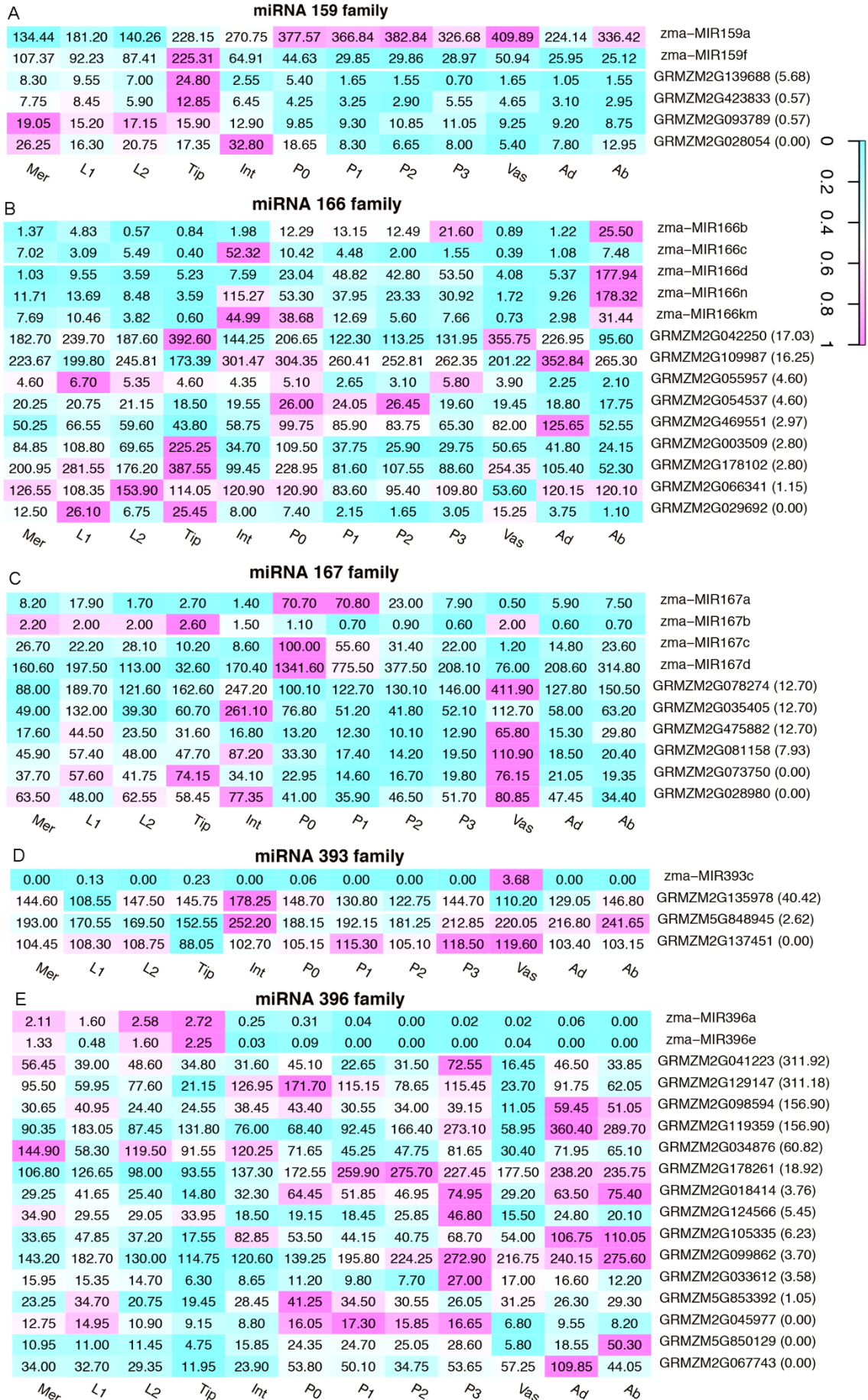
Figure S3. 24 enzymes participating in miRNA biology show differential expression across twelve maize shoot apex subdomains. Expression level across subdomains of the same enzymes is summed together, and the percentage share of expression for each subdomain is plotted. DCL1 is lowest expressed in the vasculature and tip and while HESO1 is highest accumulated in the vasculature (indicated by red circles). Enzymes are separated into four functional processes: primary miRNA biogenesis, miRNA processing, nuclear export and miRNA activity.

Figure S4. Expression level of precursor genes in mutant (*dcl1-2*) and wildtype. Here only include precursor gene are expressed in dcl1-2 (RPM≥1).

A — miRNA 159 family

	Mer	L1	L2	Tip	Int	P0	P1	P2	P3	Vas	Ad	Ab	
	134.44	181.20	140.26	228.15	270.75	377.57	366.84	382.84	326.68	409.89	224.14	336.42	zma−MIR159a
	107.37	92.23	87.41	225.31	64.91	44.63	29.85	29.86	28.97	50.94	25.95	25.12	zma−MIR159f
	8.30	9.55	7.00	24.80	2.55	5.40	1.65	1.55	0.70	1.65	1.05	1.55	GRMZM2G139688 (5.68)
	7.75	8.45	5.90	12.85	6.45	4.25	3.25	2.90	5.55	4.65	3.10	2.95	GRMZM2G423833 (0.57)
	19.05	15.20	17.15	15.90	12.90	9.85	9.30	10.85	11.05	9.25	9.20	8.75	GRMZM2G093789 (0.57)
	26.25	16.30	20.75	17.35	32.80	18.65	8.30	6.65	8.00	5.40	7.80	12.95	GRMZM2G028054 (0.00)

B — miRNA 166 family

	Mer	L1	L2	Tip	Int	P0	P1	P2	P3	Vas	Ad	Ab	
	1.37	4.83	0.57	0.84	1.98	12.29	13.15	12.49	21.60	0.89	1.22	25.50	zma−MIR166b
	7.02	3.09	5.49	0.40	52.32	10.42	4.48	2.00	1.55	0.39	1.08	7.48	zma−MIR166c
	1.03	9.55	3.59	5.23	7.59	23.04	48.82	42.80	53.50	4.08	5.37	177.94	zma−MIR166d
	11.71	13.69	8.48	3.59	115.27	53.30	37.95	23.33	30.92	1.72	9.26	178.32	zma−MIR166n
	7.69	10.46	3.82	0.60	44.99	38.68	12.69	5.60	7.66	0.73	2.98	31.44	zma−MIR166km
	182.70	239.70	187.60	392.60	144.25	206.65	122.30	113.25	131.95	355.75	226.95	95.60	GRMZM2G042250 (17.03)
	223.67	199.80	245.81	173.39	301.47	304.35	260.41	252.81	262.35	201.22	352.84	265.30	GRMZM2G109987 (16.25)
	4.60	6.70	5.35	4.60	4.35	5.10	2.65	3.10	5.80	3.90	2.25	2.10	GRMZM2G055957 (4.60)
	20.25	20.75	21.15	18.50	19.55	26.00	24.05	26.45	19.60	19.45	18.80	17.75	GRMZM2G054537 (4.60)
	50.25	66.55	59.60	43.80	58.75	99.75	85.90	83.75	65.30	82.00	125.65	52.55	GRMZM2G469551 (2.97)
	84.85	108.80	69.65	225.25	34.70	109.50	37.75	25.90	29.75	50.65	41.80	24.15	GRMZM2G003509 (2.80)
	200.95	281.55	176.20	387.55	99.45	228.95	81.60	107.55	88.60	254.35	105.40	52.30	GRMZM2G178102 (2.80)
	126.55	108.35	153.90	114.05	120.90	120.90	83.60	95.40	109.80	53.60	120.15	120.10	GRMZM2G066341 (1.15)
	12.50	26.10	6.75	25.45	8.00	7.40	2.15	1.65	3.05	15.25	3.75	1.10	GRMZM2G029692 (0.00)

C — miRNA 167 family

	Mer	L1	L2	Tip	Int	P0	P1	P2	P3	Vas	Ad	Ab	
	8.20	17.90	1.70	2.70	1.40	70.70	70.80	23.00	7.90	0.50	5.90	7.50	zma−MIR167a
	2.20	2.00	2.00	2.60	1.50	1.10	0.70	0.90	0.60	2.00	0.60	0.70	zma−MIR167b
	26.70	22.20	28.10	10.20	8.60	100.00	55.60	31.40	22.00	1.20	14.80	23.60	zma−MIR167c
	160.60	197.50	113.00	32.60	170.40	1341.60	775.50	377.50	208.10	76.00	208.60	314.80	zma−MIR167d
	88.00	189.70	121.60	162.60	247.20	100.10	122.70	130.10	146.00	411.90	127.80	150.50	GRMZM2G078274 (12.70)
	49.00	132.00	39.30	60.70	261.10	76.80	51.20	41.80	52.10	112.70	58.00	63.20	GRMZM2G035405 (12.70)
	17.60	44.50	23.50	31.60	16.80	13.20	12.30	10.10	12.90	65.80	15.30	29.80	GRMZM2G475882 (12.70)
	45.90	57.40	48.00	47.70	87.20	33.30	17.40	14.20	19.50	110.90	18.50	20.40	GRMZM2G081158 (7.93)
	37.70	57.60	41.75	74.15	34.10	22.95	14.60	16.70	19.80	76.15	21.05	19.35	GRMZM2G073750 (0.00)
	63.50	48.00	62.55	58.45	77.35	41.00	35.90	46.50	51.70	80.85	47.45	34.40	GRMZM2G028980 (0.00)

D — miRNA 393 family

	Mer	L1	L2	Tip	Int	P0	P1	P2	P3	Vas	Ad	Ab	
	0.00	0.13	0.00	0.23	0.00	0.06	0.00	0.00	0.00	3.68	0.00	0.00	zma−MIR393c
	144.60	108.55	147.50	145.75	178.25	148.70	130.80	122.75	144.70	110.20	129.05	146.80	GRMZM2G135978 (40.42)
	193.00	170.55	169.50	152.55	252.20	188.15	192.15	181.25	212.85	220.05	216.80	241.65	GRMZM5G848945 (2.62)
	104.45	108.30	108.75	88.05	102.70	105.15	115.30	105.10	118.50	119.60	103.40	103.15	GRMZM2G137451 (0.00)

E — miRNA 396 family

	Mer	L1	L2	Tip	Int	P0	P1	P2	P3	Vas	Ad	Ab	
	2.11	1.60	2.58	2.72	0.25	0.31	0.04	0.00	0.02	0.02	0.06	0.00	zma−MIR396a
	1.33	0.48	1.60	2.25	0.03	0.09	0.00	0.00	0.00	0.04	0.00	0.00	zma−MIR396e
	56.45	39.00	48.60	34.80	31.60	45.10	22.65	31.50	72.55	16.45	46.50	33.85	GRMZM2G041223 (311.92)
	95.50	59.95	77.60	21.15	126.95	171.70	115.15	78.65	115.45	23.70	91.75	62.05	GRMZM2G129147 (311.18)
	30.65	40.95	24.40	24.55	38.45	43.40	30.55	34.00	39.15	11.05	59.45	51.05	GRMZM2G098594 (156.90)
	90.35	183.05	87.45	131.80	76.00	68.40	92.45	166.40	273.10	58.95	360.40	289.70	GRMZM2G119359 (156.90)
	144.90	58.30	119.50	91.55	120.25	71.65	45.25	47.75	81.65	30.40	71.95	65.10	GRMZM2G034876 (60.82)
	106.80	126.65	98.00	93.55	137.30	172.55	259.90	275.70	227.45	177.50	238.20	235.75	GRMZM2G178261 (18.92)
	29.25	41.65	25.40	14.80	32.30	64.45	51.85	46.95	74.95	29.20	63.50	75.40	GRMZM2G018414 (3.76)
	34.90	29.55	29.05	33.95	18.50	19.15	18.45	25.85	46.80	15.50	24.80	20.10	GRMZM2G124566 (5.45)
	33.65	47.85	37.20	17.55	82.85	53.50	44.15	40.75	68.70	54.00	106.75	110.05	GRMZM2G105335 (6.23)
	143.20	182.70	130.00	114.75	120.60	139.25	195.80	224.25	272.90	216.75	240.15	275.60	GRMZM2G099862 (3.70)
	15.95	15.35	14.70	6.30	8.65	11.20	9.80	7.70	27.00	17.00	16.60	12.20	GRMZM2G033612 (3.58)
	23.25	34.70	20.75	19.45	28.45	41.25	34.50	30.55	26.05	31.25	26.30	29.30	GRMZM5G853392 (1.05)
	12.75	14.95	10.90	9.15	8.80	16.05	17.30	15.85	16.65	6.80	9.55	8.20	GRMZM2G045977 (0.00)
	10.95	11.00	11.45	4.75	15.85	24.35	24.70	25.05	28.60	5.80	18.55	50.30	GRMZM5G850129 (0.00)
	34.00	32.70	29.35	11.95	23.90	53.80	50.10	34.75	53.65	57.25	109.85	44.05	GRMZM2G067743 (0.00)

Color scale: 0, 0.2, 0.4, 0.6, 0.8, 1

Figure S5. Five miRNA families and their target genes expression patterns in shoot apex. Heatmap of expressed precursors (RPM ≥1) and their miRNA targets are shown in (A) miR169, (B) miR166, (C)miR167, (D)miR393, (E) miR396. Relative expression levels have been normalized and reflected in color scale. Numbers on the heatmap depict expression value (RPM). Bracketed numbers indicate the PARE signature. Mer - meristem, Int – internode, Vas – vasculature.
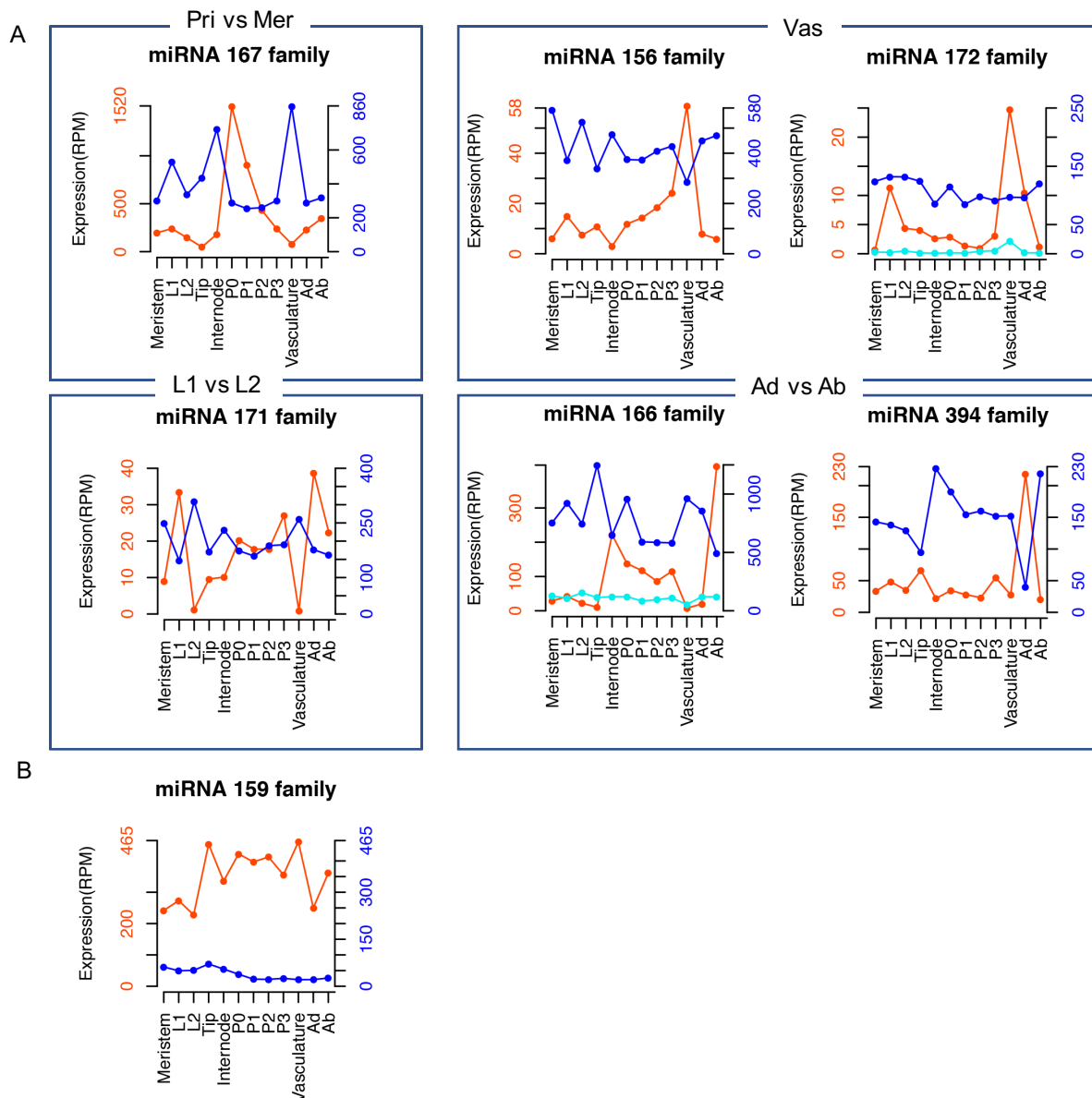
Figure S6. Seven miRNA families were found to regulate targets on both a transcript-cleavage, and translational repression level in the maize apex. Expression of pri-miRNA and targets across 12 subdomains (corresponding to Figure 5). (A) Dynamically expressed miRNAs spatially restrict target accumulation: miR167, miR156, miR172, miR166, and miR394. (B) Non-dynamically expressed miRNAs and targets: miR159. Left axis (red): pri-

miRNA expression level (sum expression of all expressed precursors in same family); right axis (blue): target gene expression level (sum expression of all targets to the same miRNA family). Mer - meristem, Vas – vasculature, Pri – Primordia.

miR156k	CACGAGCGAGAGAAGACAGU
miR156abcdefghil	CACGAGUGAGAGAAGACAGU
GRMZM2G067624(97.1)	AUGCUCUCUCUCUUCUGUCA
GRMZM2G126018(83.12)	GUGCUCUCUCUCUUCUGUCA
GRMZM2G307588(83.12)	GUGCUCUCUCUCUUCUGUCA
GRMZM2G160917(83.12)	GUGCUCUCUCUCUUCUGUCA
GRMZM2G460544(83.12)	GUGCUCUCUCUCUUCUGUCA
GRMZM2G097275(17.76)	GUGCUCUCUCUCUUCUGUCA
GRMZM2G414805(16.38)	GUGCUCUCUCUCUUCUGUCA
GRMZM2G371033(10.60)	GUGCUCUCUCUCUUCUGUCA
GRMZM5G878561(10.60)	GUGCUCUCUCUCUUCUGUCA
GRMZM5G806833(7.27)	GUGCUCUCUCUCUUCUGUCA
GRMZM2G061734(7.27)	GUGCUCUCUCUCUUCUGUCA
GRMZM2G163813(5.93)	GUGCUCUCUCUCUUCUGUCA
GRMZM2G065451(3.98)	GUGCUCUCUCUCUUCUGUCA
GRMZM2G113779(2.00)	AUGCUCUCUCUCUUCUGUCA
GRMZM2G106798(1.03)	GUGCUCUCUCUCUUCUGUCA
GRMZM2G101511(0.85)	GUGCUCUCUCUCUUCUGUCA
AC233751.1_FG002(0.85)	GUGCUCUCUCUCUUCUGUCA
GRMZM2G148467(0.52)	GUGCUCUCUCUCUUCUGUCA

miR159af	GUCUCGAGGGAAGUUAGGUUU
GRMZM2G139688(5.68)	UGGAGCUCCCUUCACUCCAAG
GRMZM2G423833(0.57)	CAGAGAUCCCUUCGAUCCAAA
GRMZM2G093789(0.57)	CAGAGAUUCCUUCGAUCCAAA
GRMZM2G028054(0.00)	UGGAGCCCCCUUCAGUCCAAA

miR60abcdg	ACCGUAUGUCCCUCGGUCCGU
miR160f	GCCGUAUGUCCCUCGGUCCGU
GRMZM2G081406(117.38)	AGCCAUACAGGGAGCCAGGCA
GRMZM2G153233(18.98)	AGGCAUACAGGGAGCCAGGCA
GRMZM2G159399(18.98)	AGGCAUACAGGGAGCCAGGCA
GRMZM2G390641(12.63)	AGGCAUACAGGGAGCCAGGCA
GRMZM5G808366(1.78)	UAGCAUGCAGGGAGCCAGGCA

miR162	ACCUACGUCUCCAAA-UAGCU
GRMZM2G040762(0.40)	UGGAUGCAGAGGUUUUAUCGA

miR164abcdg	ACGUGCACGGGACGAAGAGGU
GRMZM2G114850(0.40)	AGCAGGUGCCCUGCUUCUCCA
GRMZM2G139700(0.43)	AGCUCGUGCCCUGCUUCUCCA
GRMZM2G393433(2.32)	AGCUCGUGCCCUGCUUCUCCA
GRMZM2G063522(0.00)	AGCAAGUGCCCUGCUUCUCCA

miR166bcd	CCCCUUACUUCGGACCAGGCU
GRMZM2G042250**(17.03)**	CUGGGAUGAAGCCUGGUCCGG
GRMZM2G109987**(16.25)**	CUGGGAUGAAGCCUGGUCCGG
GRMZM2G469551**(2.97)**	CUGGGAUGAAGCCUGGUCCGG
GRMZM2G003509**(2.80)**	UUGGGAUGAAGCCUGGUCCGG
GRMZM2G178102**(2.80)**	UUGGGAUGAAGCCUGGUCCGG
GRMZM2G066341**(1.15)**	GGGAUGUGAAGUCUGGUCCAG
AC187157.4_FG005**(0.00)**	CCGGGAUGAAGCCUGGUCCGG

miR166m	CUCCUUACUUCGGACCAGGCU
GRMZM2G042250**(17.03)**	CUGGGAUGAAGCCUGGUCCGG
GRMZM2G109987**(16.25)**	CUGGGAUGAAGCCUGGUCCGG
GRMZM2G469551**(2.97)**	CUGGGAUGAAGCCUGGUCCGG
GRMZM2G003509**(2.80)**	UUGGGAUGAAGCCUGGUCCGG
GRMZM2G178102**(2.80)**	UUGGGAUGAAGCCUGGUCCGG
AC187157.4_FG005**(0.00)**	CCGGGAUGAAGCCUGGUCCGG

miR166nk	UCCCUAACUUCGGACCAGGCU
GRMZM2G042250**(17.03)**	UGGGAU-GAAGCCUGGUCCGG
GRMZM2G109987**(16.25)**	UGGGAU-GAAGCCUGGUCCGG
GRMZM2G469551**(2.97)**	UGGGAU-GAAGCCUGGUCCGG
GRMZM2G003509**(2.80)**	UGGGAU-GAAGCCUGGUCCGG
GRMZM2G178102**(2.80)**	UGGGAU-GAAGCCUGGUCCGG
AC187157.4_FG005**(0.00)**	CGGGAU-GAAGCCUGGUCCGG
GRMZM2G029692**(0.00)**	ACUGAUUGCAGCCUGGUCCGG
miR166nk	UCCCUA-ACUUCGGACCAGGCU
GRMZM2G066341**(1.15)**	GGGGAUGUGAAGUCUGGUCCAG

miR167abcd	AUCUAGUACGACCGUCGAA-GU
GRMZM2G078274**(12.70)**	UAGAUCAGGCUGGCAGCUUGUA
GRMZM2G035405**(12.70)**	GAGAUCAGGCUGGCAGCUUGUA
GRMZM2G475882**(7.93)**	UAGAUCAGGCUGGCAGCUUGUA
GRMZM2G081158**(7.93)**	AAGAUCAGGCUGGCAGCUUGUA
GRMZM2G073750**(0.00)**	AAGAUCAGGCUGGCAGCUUGUA
GRMZM2G028980**(0.00)**	GAGAUCAGGCUGGCAGCUUGUA

miR168ab	CAGGGCUAGACGUGGUUCGCU
GRMZM2G441583**(2.75)**	UUCCCGAGCUGCACCAAGCCC
GRMZM2G039455**(2.33)**	UUCCCGAGCUGCACCAAGCCC
GRMZM2G361518**(0.00)**	CUCCCGAGCUGCGCCAAGCAA

```
miR169cr                     GGCCGUUCAGUAGGAACCGAC
GRMZM2G091964(158.78)        CAGGCAAUUCAUUCUUGGCUU
GRMZM5G853836(81.78)         UAGGCAAAUCAUUCUUGGCUG
GRMZM5G829103(67.82)         UAGGCAAAUCAUUCUUGGCUG
GRMZM2G165488(64.80)         UAGGCAAAUCAUUCUUGGCUG
GRMZM2G000686(30.97)         CUGGCAACUCAUCCUUGGCUU
GRMZM2G040349(19.50)         GUGGCAACUCAUCCUUGGCUC
GRMZM2G037630(3.35)          GUGGCAAUUCAUCCUUGGCUU
GRMZM2G038303(0.87)          GUGGCAACUCAUCCUUGGCUU
GRMZM5G857944(0.85)          CAGGCAAUUCAUCCUUGGCUU


mmiR169fh                    AUCCGUUCAGUAGGAACCGAU
GRMZM2G091964(158.78)        CAGGCAAUUCAUUCUUGGCUU
GRMZM5G853836(81.78)         UAGGCAAAUCAUUCUUGGCUG
GRMZM5G829103(67.82)         UAGGCAAAUCAUUCUUGGCUG
GRMZM2G165488(64.80)         UAGGCAAAUCAUUCUUGGCUG
GRMZM2G000686(30.97)         CUGGCAACUCAUCCUUGGCUU
GRMZM2G037630(3.35)          GUGGCAAUUCAUCCUUGGCUU
GRMZM5G857944(0.85)          CAGGCAAUUCAUCCUUGGCUU
miR169fh                     AU-CCGUUCAGUAGGAACCGAU
GRMZM2G040349(19.50)         UGUGGCAACUCAUCCUUGGCUC
GRMZM2G038303(0.87)          UGUGGCAACUCAUCCUUGGCUU


miR171deij                   CUAUAACCGUGCCGAGUUAGU
GRMZM2G037792(32.15)         GAUAUUGGCGCGGCUCAAUCA
GRMZM5G825321(31.25)         GAUAUUGGCGCGGCUCAAUCA
GRMZM2G098800(12.48)         GAUAUUGGCGCGGCUCAAUCA
GRMZM2G110579(12.47)         GAUAUUGGCGCGGCUCAAUCA
GRMZM2G079470(7.52)          GAUAUUGGCGCGGCUCAAUUA
GRMZM2G051785(0.60)          GAUAUUGGCGCGGCUCAAUCA
GRMZM2G060265(0.00)          GAUACUGGCGCGGCUCAACUA


172c                         UACGUCGUAGUAGUU-CUAAGA
GRMZM5G862109(37.63)         CUGCAGCAUCAUCAG-GAUUCU
GRMZM2G176175(37.40)         CUGCAGCAUCAUCAG-GAUUCU
GRMZM2G076602(10.87)         CUGCAGCAUCAUCAC-GAUUCC
GRMZM2G174784(5.45)          CUGCAGCAUCAUCAC-GAUUCC
GRMZM2G160730(4.32)          CUGCAGCAUCAUCAG-GAUUCC
GRMZM5G879527(3.93)          GUGUGGCAUCAUCAAGUAUUCA
GRMZM2G700665(3.40)          CUGCAGCAUCAUCAG-GAUUCU


miR319bcd                    CCCUCGUGGGAAGUCAGGUU
GRMZM2G089361(19.23)         AGGGGGACCCUUCAGUCCAA
GRMZM2G115516(19.23)         AGGGGGACCCUUCAGUCCAA
GRMZM2G015037(4.85)          AGGGGACCCCUUCAGUCCAG
GRMZM2G020805(3.57)          AGGGGGGCCCUUCAGUCCAA
GRMZM2G028054(3.03)          UGGAGCCCCCUUCAGUCCAA
```

miR393c	CUAGUUACGCUAGGGAAACCU
GRMZM2G135978(40.42)	GA-CAAUGCGAUCCCUUUGGA
GRMZM5G848945(2.62)	GA-CAAUGCGAUCCCUUUGGA
GRMZM2G137451(0.00)	ACACAAUGCGAUCCCUUUGGA

miR394ab	CCUCCACCUGUCUUACGGUU
GRMZM2G064954(2.55)	GGAGGUGGACAGAAUGCCAA
GRMZM2G119650(1.88)	GGAGGUGGACAGAAUGCCAA

miR396abg	GUCAAGUUCUUUCG-ACACCUU
GRMZM2G041223(311.92)	CCGUUCAAGAAAGCCUGUGGAA
GRMZM2G129147(311.18)	CCGUUCAAGAAAGCCUGUGGAA
GRMZM2G098594(156.90)	CCGUUCAAGAAAGCCUGUGGAA
GRMZM2G119359(156.90)	CCGUUCAAGAAAGCCUGUGGAA
GRMZM2G045977(0.00)	CCGUUCAAGAAAGCCUGUGGAA
GRMZM2G034876(60.82)	CCGUUCAAGAAAGCCUGUGGAA
GRMZM2G067743(0.00)	CCGUUCAAGAAAGCCUGUGGAA
GRMZM2G178261(18.92)	UCGUUCAAGAAAGCAUGUGGAA
GRMZM2G018414(3.76)	CCGUUCAAGAAAGCCUGUGGAA
GRMZM2G124566(5.45)	CCGUUCAAGAAAGCCUGUGGAA
GRMZM2G105335(6.23)	CCGUUCAAGAAAGCCUGUGGAA
GRMZM2G099862(3.70)	CCGUUCAAGAAAGCAUGUGGAA
GRMZM2G033612(3.58)	CCGUUCAAGAAAGCCUGUGGAA
GRMZM5G853392(1.05)	UCGUUCAAGAAAGCAUGUGGAA
GRMZM5G850129(0.00)	CCGUUCAAGAAAGCCUGUGGAA
GRMZM2G045977(0.00)	CCGUUCAAGAAAGCCUGUGGAA
GRMZM2G067743(0.00)	CCGUUCAAGAAAGCCUGUGGAA

miR396cd	GUCAAGUUCUUUCGGACACCU
GRMZM2G041223(311.92)	CCGUUCAAGAAAGCCUGUGGA
GRMZM2G129147(311.18)	CCGUUCAAGAAAGCCUGUGGA
GRMZM2G098594(156.90)	CCGUUCAAGAAAGCCUGUGGA
GRMZM2G119359(156.90)	CCGUUCAAGAAAGCCUGUGGA
GRMZM2G045977(0.00)	CCGUUCAAGAAAGCCUGUGGA
GRMZM2G034876(60.82)	CCGUUCAAGAAAGCCUGUGGA
GRMZM2G067743(0.00)	CCGUUCAAGAAAGCCUGUGGA
GRMZM2G178261(18.92)	UCGUUCAAGAAAGCAUGUGGA
GRMZM2G018414(3.76)	CCGUUCAAGAAAGCCUGUGGA
GRMZM2G124566(5.45)	CCGUUCAAGAAAGCCUGUGGA
GRMZM2G105335(6.23)	CCGUUCAAGAAAGCCUGUGGA
GRMZM2G099862(3.70)	CCGUUCAAGAAAGCAUGUGGA
GRMZM2G033612(3.85)	CCGUUCAAGAAAGCCUGUGGA
GRMZM5G853392(1.05)	UCGUUCAAGAAAGCAUGUGGA
GRMZM5G850129(0.00)	CCGUUCAAGAAAGCCUGUGGA
GRMZM2G045977(0.00)	CCGUUCAAGAAAGCCUGUGGA
GRMZM2G067743(0.00)	CCGUUCAAGAAAGCCUGUGGA

miR396ef	UUCAAGUUCUUUCG-ACACCUU
GRMZM2G018414(3.76)	CCGUUCAAGAAAGCCUGUGGAA
GRMZM2G033612(3.58)	CCGUUCAAGAAAGCCUGUGGAA
GRMZM2G034876(60.82)	CCGUUCAAGAAAGCCUGUGGAA
GRMZM2G041223(311.92)	CCGUUCAAGAAAGCCUGUGGAA
GRMZM2G099862(3.70)	CCGUUCAAGAAAGCAUGUGGAA
GRMZM2G105335(6.23)	CCGUUCAAGAAAGCCUGUGGAA
GRMZM2G124566(5.45)	CCGUUCAAGAAAGCCUGUGGAA
GRMZM2G129147(311.18)	CCGUUCAAGAAAGCCUGUGGAA
GRMZM2G178261(18.92)	UCGUUCAAGAAAGCAUGUGGAA
GRMZM5G853392(1.05)	UCGUUCAAGAAAGCAUGUGGAA
GRMZM2G098594(156.90)	CCGUUCAAGAAAGCCUGUGGAA
GRMZM2G119359(156.90)	CCGUUCAAGAAAGCCUGUGGAA
GRMZM5G850129(0.00)	CCGUUCAAGAAAGCCUGUGGAA
GRMZM2G045977(0.00)	CCGUUCAAGAAAGCCUGUGGAA
GRMZM2G067743(0.00)	CCGUUCAAGAAAGCCUGUGGAA
GRMZM5G850129(0.00)	CCGUUCAAGAAAGCCUGUGGAA
GRMZM2G067743(0.00)	CCGUUCAAGAAAGCCUGUGGAA

GRMZM2G058522(16.77)	CGGGGGUCGCCUGAGAUCACA
GRMZM2G169890(16.77)	CGGGGGUCGCCUGAGAUCACA

Figure S7. Complementary binding sites between sixteen select miRNA families and their targets. These sixteen families corresponding to figure 5 and figure S6. The blue bases identify a G:U bulge and mismatch. Bracketed numbers (red text) indicate the PARE signature.

Supplementary datasets:

Dataset S1. RNA-Seq data revealing pri-miRNA and mature miRNA expression in the maize shoot apex.

Dataset S2. RNA-Seq data revealing the expression of enzymes involved in miRNA biology in the maize shoot apex.

Dataset S3. RNA-Seq data revealing gene expression in CZ of *dcl1-2* and wildtype

Dataset S4. RNA-Seq data revealing miRNA-target gene expression in the maize shoot apex.

Dataset S5. The regulatory mechanism of miRNAs.

Supplementary Tables:

Table S1. A model used to infer the regulation mechanism and regulation mode of miRNAs by integrating degradome-Seq data and RNA-Seq data.

RP	RMT	TE	Mechanism	Mode
High (≥0.5)	High (≥1)	High (≥10)	Cleavage/translational	?
		Low (<10)	Cleavage	clearance
	Low (<1)	High (≥10)	cleavage	rheostat
		N.A.	N.A.	N.A.
Low (<0.5)	High (≥1)	High (≥10)	Translational/cleavage	?
		Low (<10)	Promoter/cleavage	?
	Low (<1)	High (≥10)	Cleavage/translational	rheostat
		N.A.	N.A.	N.A.

RP: Relative PARE (PARE/targets); RMT: Ratio of miRNA to target; TE: Target expression in miRNA accumulation domain; Mechanism: regulation mechanism (transcript cleavage or translational repression); Mode: regulation mode (clearance or rheostat).

Table S2. Regulatory mechanisms of select miRNAs in maize shoot apex

Family	Maize	Arabidopsis	Ref
miR156	both	both	Schwab et al; 2005; German et al., 2008; Gandikota et al., 2007; Brodersen et al., 2008; He et al., 2018
miR159	Both	both	Achard et al., 2004; Schwab et al; 2005; German et al., 2008; Li et al., 2014; Li et al., 2013;
miR160	cleavage	N.A.	German et al., 2008;
miR162	cleavage	N.A.	German et al., 2008; Li et al., 2013;
miR164	both	both	Schwab et al; 2005; German et al., 2008; Li et al., 2013;
miR166	both	N.A.	German et al., 2008;
miR167	both	N.A.	German et al., 2008;
miR168	cleavage	N.A.	German et al., 2008;
miR169	cleavage	N.A.	German et al., 2008;
miR171	both	both	Llave et al., 2002; Kasschau et al., 2003; German et al., 2008; Brodersen et al., 2008;
miR172	-	both	Kasschau et al., 2003; Schwab et al; 2005; German et al., 2008; Aukerman et al., 2003; Chen, 2004; Li et al., 2013;
miR319	cleavage	cleavage	Schwab et al; 2005; German et al., 2008;
miR390	cleavage	cleavage	Marin et al., 2010; petsch et al., 2015
miR393	cleavage	N.A.	German et al., 2008;
miR394	both	N.A.	German et al., 2008;
miR396	cleavage	N.A.	German et al., 2008;

"-": no evidence. "N.A.": have PARE signature signal but not sure the regulation mechanism.

Chapter 5: General conclusion and discussion:

Plant meristems are highly organized stem cell niches characterized by the presence of distinct cell types at various developmental stages. Meristems are therefore offer ideal models for the study molecular regulatory mechanisms during development via generation of high-resolution, spatiotemporal expression atlases. In the first part of this dissertation, I describe the first gene expression atlas of the *Arabidopsis* root generated by scRNA-Seq that reveals new developmental regulators and precise differentiation trajectories. In the second part: I describe using high resolution spatiotemporal expression data of the maize shoot apex to reveal the regulation mechanisms of miRNA spatiotemporal accumulation and action during development.

I. **A cellular resolution gene expression atlas of the *Arabidopsis* root generated by high-throughput single cell RNA sequencing technology reveals new developmental regulators and precise differentiation trajectories.**

Previous scRNA-Seq studies were limited to the analysis of just hundreds of cells (Efroni and Birnbaum, 2016; Efroni et al., 2016)**.** However, development of combinatorial barcoding and increased automation allowed scRNA-Seq to profile the thousands of cells in one experiment (Klein et al., 2015; Macosko et al., 2015; Zheng et al., 2017). In this study, we first used high-throughput scRNA-Seq to build a gene expression atlas of the *Arabidopsis* root at cellular resolution. Here, scRNA-Seq proved to be highly sensitive and reproducible. Protoplasting-time, cell buoyancy, robustness and size, and tissue organization are all factors that can limit scRNA-Seq application. It is still worth to know how broad the application of this technology can be, by applying it to many plants and tissues.

Genes induced by protoplasting were removed from all analysis to reduce technical noise. Among these induced genes are some key makers, including QC marker gene *WOX5* (Forzani et al., 2014) and phloem companion cell marker, SUC2 (Birmbaum et al., 2003).Curiously, *WOX5* was also detected in some cells of an endodermal cluster, suggesting cell type would be wrongly-annotated if we relied only on *WOX5* to assign. In this analysis, supervised (known maker genes expression in cluster) and un-supervised (novel identified cluster maker genes expression in root) methods was

used to assign cell types. The thousands of individual cell transcriptomes represent all major root cell types. Notably, the protoplast induced genes revealed a strong association with lateral root cap (LRC) cells, and therefore, the LCR cluster lost its identity after removing the protoplast induced genes, and therefore, LRC cluster lose its identity after removing the protoplast induced genes. However, in other recently published plant scRNA-Seq data, they did not remove such genes and LRC were identified based on a supervised method (Jean-Baptiste et al., 2019; Ryu et al., 2019; Shulse et al., 2019; Zhang et al., 2019). It could be that genes induced by protoplasting would probably not affect cell clustering but cell type calling. However, it is still worth to know how much impact of protoplast-induced genes on cell clustering and cell type calling in future. Also, these genes can complicate further analysis such as the construction of pseudotime trajectories and the production of GRNs. In this regard, we have found highly-induced genes to dominate these networks, introducing significant, potential, misleading bias. Besides the protoplast induced genes, sequencing depth and cell number are different in each of these recent studies. Next, it is also interesting to estimate how much effect on clustering by different sequencing depth, cell number treatment, which would offer useful suggestion for scRNA-Seq study.

Meristem cells were enriched in our single cell root atlas data. This is likely due to the fact that meristematic cells are smaller and as such the technology may preferentially capture them. Sub-clustering analysis was able to distinguish cell types within the original meristem cluster. For example, QC and columella cells was separated from their shared original cluster by the sub-clustering analysis (Chapter 4.1 Figure 2E). Previous study (using FACS to cell sort followed by RNA-Seq) failed to separate meristematic cells to such a degree (Birnbaum et al., 2003; Brady et al., 2007). Here, they captured either single cell types without distinguishing developmental stages or single developmental stages without distinguishing tissue types. They would finally obtained the average expression of makers genes from multiple cell types/states (Brady et al., 2007). Here, the identified cell-type specific expression genes would offer an abundant resource for further genetic study.

Pseudotime analysis offered a high-resolution developmental trajectory map of root cell development, and revealed regulators driving the transition from different developmental states along the root, on a single cell type level, e.g. trichoblast and

cortex. GRN analysis of the cascade of TFs the trichoblast trajectory revealed the feedback regulation towards meristem and suggested the major nodes coordinating root cell transition.

Here, scRNA-Seq captured all-types of cells in the RAM, which offered a good opportunity to study the molecular regulatory mechanisms behind of meristem development. Going forward, it should be possible to use this single cell data to address the gene networks defining and maintaining the quiescent center and stem cells niche. Further, it might be applied to answering how do stem cell initials define the identity of cell types, and which signals and gene networks regulate the developmental transition of individual cell types.

II. Using maize shoot apex high resolution spatiotemporal expression data to reveal the regulation mechanism of miRNA spatiotemporal accumulation and action during development.

miRNAs are critical for plant normal development. The available of high-resolution maize shoot apex transcriptome atlas data (Knauer et al., 2019) offered an opportunity to investigate the spatial temporal action of miRNAs during development. Comprehensive investigation miRNAs across the maize shoot apex was enhanced by analysis of their stability, mobility, efficacy, and activity. The complex regulation of miRNA accumulation confers miRNAs the robustness to drive plant development. Distinct miRNAs are identified shaping the sharp tissue boundary and determining the cell types fates during maize shoot apex development. Besides mature miRNAs, many miRNA* were detected in the shoot apex. Next, it would be interesting to investigate how miRNA* function in the maize shoot apex, or the reason why remain there. In addition, different isoforms of miRNAs are detected during data analysis. For example, zma-miR162 produces fourteen 20-nt (TCGATAAACCTCTGCATCCA) and 1113 21-nt (TCGATAAACCTCTGCATCCAG) isoforms. Multiple distinct miRNA-like RNAs arising from a single miRNA precursor were identified in *Arabidopsis* (Zhang et al., 2010). Novel miRNAs were discovered in maize ear and inflorescence (Liu et al. 2014). Therefore, it is also worth to identifying miRNA isoforms, novel miRNAs, and miRNA-like RNAs to study the mechanism of those isoforms' biogenesis, distribution and function in maize shoot apical meristem development.

Future prospect

The technologies used to study the transcriptome are rapidly developing. Long-read (third-generation) sequencing enables us to directly read nucleotide sequence at single molecule level. For example, Pacific Bioscience's (PacBio) single molecule real-time sequencing (SAMRT) and Oxford Nanopore's platform can be used to generate full-length cDNA sequences within targeted genes or across an entire transcriptome (Clarke et al., 2009; Eid et al., 2009). Long-read sequencing offers an opportunity to characterize transcript isoforms (Iso-Seq analysis), which increases the accuracy of transcript quantification with isoform-level resolution (Sharon et al., 2013). Recently, combing the scRNA-Seq with Iso-Seq analysis could identify more accurate transcript accumulation and reveal the cellular alternative splicing (Gupta et al., 2018). If we can combine the Iso-Seq and high throughput scRNA-Seq, it will aid to build a high resolution and highly-accurate transcriptomic atlas of plant cell types, and precisely assess the variety and range of cell types, cell states, and their transitions in the target tissues.

Our ability to study the miRNA spatiotemporal accumulation regulation at a cellular level could be aided by integrating the high throughput scRNA-Seq and Iso-Seq. Currently, single cell miRNA profiling only is limits to only a few cells (Faridani et al., 2016; Henriksen et al., 2018; Wang et al., 2019). Inevitably, the barcode-combining technologies will also be used in single cell miRNA profiling, and the single cell protein profiling will likely follow. These will be great step forward in the analysis of miRNA regulation mechanisms and regulation modes on a cellular level.

In future, single cell profiling will not only to create a high-resolution, temporal, and spatial transcriptome maps of the plant tissues, but will also allow us to decipher high resolution transcriptional and post-transcriptional gene regulation networks on a cellar level. It will benefit us greatly to understand how plant cells are organized at molecular level and will allow as to answer what cell-specific regulatory mechanisms are applied to change their states during development, and in response to biotic and abiotic stress.

References:

Birnbaum, K., Shasha, D.E., Wang, J., 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., Wang, J., 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.

Clarke, J., Wu, H.C., Jayasinghe, L., Patel, A., Reid, S., and Bayley, H. (2009). Continuous base identification for single-molecule nanopore DNA sequencing. Nat Nanotechnol *4*, 265-270.

Efroni, I., and Birnbaum, K.D. (2016). The potential of single-cell profiling in plants. Genome Biol *17*, 65.

Efroni, I., Mello, A., Nawy, T., Ip, P.L., Rahni, R., DelRose, N., Powers, A., Satija, R., and Birnbaum, K.D. (2016). Root Regeneration Triggers an Embryo-like Sequence Guided by Hormonal Interactions. Cell *165*, 1721-1733.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B.*, et al.* (2009). Real-time DNA sequencing from single polymerase molecules. Science *323*, 133-138.

Faridani, O.R., Abdullayev, I., Hagemann-Jensen, M., Schell, J.P., Lanner, F., and Sandberg, R. (2016). Single-cell sequencing of the small-RNA transcriptome. Nat Biotechnol *34*, 1264-1266.

Forzani, C., Aichinger, E., Sornay, E., Willemsen, V., Laux, T., Dewitte, W., and Murray, J.A. (2014). WOX5 suppresses CYCLIN D activity to establish quiescence at the center of the root stem cell niche. Curr Biol *24*, 1939-1944.

Gupta, I., Collier, P.G., Haase, B., Mahfouz, A., Joglekar, A., Floyd, T., Koopmans, F., Barres, B., Smit, A.B., Sloan, S.A.*, et al.* (2018). Single-cell isoform RNA sequencing characterizes isoforms in thousands of cerebellar cells. Nat Biotechnol.

Henriksen, T.I., Heywood, S.E., Hansen, N.S., Pedersen, B.K., Scheele, C.C., and Nielsen, S. (2018). Single Cell Analysis Identifies the miRNA Expression Profile of a Subpopulation of Muscle Precursor Cells Unique to Humans With Type 2 Diabetes. Front Physiol *9*, 883.

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.

Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell *161*, 1187-1201.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M.*, et al.* (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell *161*, 1202-1214.

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.

Sharon, D., Tilgner, H., Grubert, F., and Snyder, M. (2013). A single-molecule long-read survey of the human transcriptome. Nat Biotechnol *31*, 1009-1014.

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.*, et al.* (2019). High-Throughput Single-Cell Transcriptome Profiling of Plant Cell Types. Cell Rep *27*, 2241-2247 e2244.

Wang, N., Zheng, J., Chen, Z., Liu, Y., Dura, B., Kwak, M., Xavier-Ferrucio, J., Lu, Y.C., Zhang, M., Roden, C.*, et al.* (2019). Single-cell microRNA-mRNA co-sequencing reveals non-genetic heterogeneity and mechanisms of microRNA regulation. Nat Commun *10*, 95.

Zhang, T.Q., Xu, Z.G., Shang, G.D., and Wang, J.W. (2019). A Single-Cell RNA Sequencing Profiles the Developmental Landscape of Arabidopsis Root. Mol Plant *12*, 648-660.

Zhang, W., Gao, S., Zhou, X., Xia, J., Chellappan, P., Zhou, X., Zhang, X., and Jin, H. (2010). Multiple distinct small RNAs originate from the same microRNA precursors. Genome Biol *11*, 81-99.

Zheng, G.X., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J.*, et al.* (2017). Massively parallel digital transcriptional profiling of single cells. Nat Commun *8*, 14049.

Appendices

Appendix I. Spatiotemporal developmental trajectories in the *Arabidopsis* root revealed using high-throughput single cell RNA sequencing

Tom Denyer*[1], Xiaoli Ma*[1], Simon Klesen[1], Emanuele Scacchi[1], Kay Nieselt[2], and Marja C. P. Timmermans#[1]

This study was published in Developmental Cell. T.D., X.M., and M.C.P.T. designed the project and experiments. T.D. generated the scRNA-Seq and bulk RNA-seq libraries and, together with S.K., generated and analyzed reporter lines. X.M. performed bioinformatics analyses. X.M. and E.S. carried out the GRN analysis. K.N. provided statistical and bioinformatic support. T.D. and M.C.P.T. wrote the manuscript.

Appendix II. A high-resolution gene expression atlas links dedicated meristem genes to key architectural traits

Steffen Knauer[1,2,10], Marie Javelle[2,8,10] , Lin Li[3], Xianran Li[4], Xiaoli Ma[1], Kokulapalan Wimalanathan[5,9], Sunita Kumari[2], Robyn Johnston[6], Samuel Leiboff[6], Robert Meeley[7], Patrick S. Schnable[4], Doreen Ware[2], Carolyn Lawrence-Dill[4,5], Jianming Yu[4], Gary J. Muehlbauer[3], Michael J. Scanlon[6] and Marja C.P. Timmermans[1,2]*

This study has been submitted to Genome Research (under revision). It reports on a high-resolution gene expression atlas of the maize shoot apex and links spatiotemporal resolved genes to key architectural traits. I contributed to data analysis and offered input for writing.

Appendix III. Nonrandom domain organization of the *Arabidopsis* genome at the nuclear periphery.

Xiuli Bi,[1,5] Ying-Juan Cheng,[2,3,5] Bo Hu,[1] Xiaoli Ma,[1] Rui Wu,[4] Jia-Wei Wang,[2] and Chang Liu[1]

This study was published in Genome Research. It demonstrates that chromatin within the nuclear space has a defined spatial organization with silenced protein-coding genes and transposable elements preferentially located at the nuclear periphery. It further provides evidence for a spatial separation of DNA methylation pathways, such that transposons silenced via CHH methylation are located at the periphery, whereas interior positioned transposons are show signatures of distinct methylation pathways. I contributed to the latter methylation data analysis.

Appendix I. Spatiotemporal developmental trajectories in the *Arabidopsis* root revealed using high-throughput single cell RNA sequencing

Tom Denyer[*,1], Xiaoli Ma[*,1], Simon Klesen[1], Emanuele Scacchi[1], Kay Nieselt[2], and Marja C. P. Timmermans[#,1]

[1] Center for Plant Molecular Biology, University of Tübingen, Auf der Morgenstelle 32, 72076 Tübingen, Germany

[2] Center for Bioinformatics, University of Tübingen, Tübingen, Germany

*These authors contributed equally

Corresponding author: marja.timmermans@zmbp.uni-tuebingen.de

Abstract

High-throughput single cell RNA-sequencing (scRNA-Seq) is becoming a cornerstone of developmental research, providing unprecedented power in understanding dynamic processes. Here, we present a high-resolution scRNA-Seq expression atlas of the *Arabidopsis* root, composed of thousands of independently profiled cells. This atlas provides detailed spatiotemporal information, identifying defining expression features for all major cell types, including the scarce cells of the quiescent center. These reveal new developmental regulators and downstream genes that translate cell fate into distinctive cell shapes and functions. Developmental trajectories derived from pseudotime analysis depict a finely resolved cascade of cell progressions from the niche through differentiation that are supported by mirroring expression waves of highly interconnected transcription factors. This study demonstrates the power of applying scRNA-Seq to plants, and provides a unique spatiotemporal perspective of root cell differentiation.

Introduction

In recent years, high-throughput single cell transcriptomics has developed to a point of becoming a fundamental, widely-used method in mammalian research (Potter *et al.,* 2018). Thousands of cells can be profiled simultaneously and analyzed accurately, revealing unique insights into developmental progressions, transcriptional pathways, and the molecular heterogeneity of tissues. The increasingly high-throughput nature of single cell RNA-sequencing (scRNA-Seq) has been facilitated by the development of droplet technology (Macosko *et al*., 2015; Klein *et al*., 2015) and increased automation (Zheng *et al*., 2017). In brief, a cell is encapsulated within an oil droplet, lysed, and its transcripts reverse transcribed on barcoded beads. Following library production and sequencing, transcripts from individual cells can be identified from the bead-derived barcode and individual transcripts accounted for using Unique Molecular Identifiers (UMIs) (Prakadan *et al.,* 2017). However, while commonly used in animal systems, additional technical demands such as the necessity to break down cell walls (with subsequent transcriptional effects), high osmotic pressure sensitivities, and high cell size variability, present potential challenges when applying this technology to plants.

The *Arabidopsis* root provides an ideal tissue for analyzing the promise of scRNA-Seq. The transcriptomes of key cell types have been well profiled, and the root shows a strict spatiotemporal organization. Radially, the root is organized in concentric rings of endodermis, cortex, and epidermis that surround a central stele, comprising the pericycle, phloem and xylem (Figure S1A). These cell types originate from a specialized stem cell niche in which initials, surrounding the quiescent center (QC), divide in a predictable manner, giving rise to long cell files that capture their developmental trajectory along the length of the root (Figure S1B). Several gene expression atlases of the *Arabidopsis* root have been produced (Birnbaum *et al.,* 2003; Brady *et al.,* 2007a; Li *et al.*, 2016). These, however, have focused primarily on describing either radial or temporal expression profiles, and typically relied on reporter lines to assess select cell types. ScRNA-Seq on the other hand, allows the simultaneous, unbiased sampling of every type of cell at every developmental stage, in one experiment.

Here, we present a high-resolution scRNA-Seq expression atlas of the *Arabidopsis* root that captures its precise spatiotemporal information, revealing new regulators, and defining features for all major cell types. We show how QC cells and meristematic cells are distinguished, and resolve intricate developmental trajectories that cells undergo during their transition from stem cell through differentiation. The precise waves of gene expression characterizing this process are mirrored by similar expression changes of highly interconnected transcription factors (TFs). Our atlas offers a unique spatiotemporal perspective of root cell-type differentiation at a resolution not previously achievable.

Results

Single cell RNA sequencing is highly sensitive and highly reproducible

4,727 *Arabidopsis* root cells from two biological replicates were isolated and profiled using droplet-based scRNA-Seq. At ~87k reads per cell, the median number of genes and transcripts detected per cell was 4,276 and 14,758, respectively (Figure S1C; Table S1). In total, transcripts for 16,975 genes were detected (RPM ≥1), which, after correction for read depth, represents ~90% of genes detected by bulk RNA-seq of protoplasted root tissue. Further, the global gene expression profiles of pooled scRNA-

Seq and bulk RNA-seq are highly correlated (r = 0.9; Figure S1D), indicating that plant scRNA-Seq is highly sensitive. This methodology is also highly reproducible, as demonstrated by the facts that ~96% of genes expressed (RPM ≥1) in one scRNA-Seq replicate are detectable in the second, and that expression across the two replicates is highly correlated (r=0.99; Figure S1E).

Clusters comprise the major cell types in the root

To identify distinct cell populations based on gene expression profiles, an unbiased, graph-based clustering was performed on the 4,727 single cell transcriptomes using the Seurat software package (Sajita *et al.,* 2015; Butler *et al*., 2018) (Figure 1A). Genes induced by protoplasting (≥ 2-fold, q <0.05) were identified by standard RNA-seq and dismissed prior to analysis (Figure S1F; Table S1). 15 distinct clusters were identified, each containing between 81 and 596 cells. These clusters harbored similar numbers of cells from each replicate and their gene expression profiles were highly correlated across the replicates (r between 0.95 and 1; Table S2), highlighting again the impressive reproducibility of this technique.

In order to attribute cell identities to these clusters, expression of cell-type specific marker genes, either well established, or identified from a curated collection of root transcriptomic datasets (Table S2; Efroni *et al.,* 2015), was compared across clusters. This allowed cell identities to be confidently assigned to eight of the 15 clusters in the cluster-cloud (Figure 1B; Table S2). Expression of key root development genes amongst these markers, such as *PLT1, SCR, SHR, APL, COBL9*, and *GL2,* shows high specificity to particular clusters (Figure S2). Cluster identities were confirmed with a complementary approach, whereby transcription profiles of differentially expressed (DE) genes governing the clusters were harvested from microarray datasets (Brady *et al.,* 2007a), and analyzed for tissue specificity (Figure S1G). Together, these approaches revealed that, with the exception of lateral root cap cells (for which limited marker data is available), all known major tissue types in the root were captured and are represented by identifiable clusters.
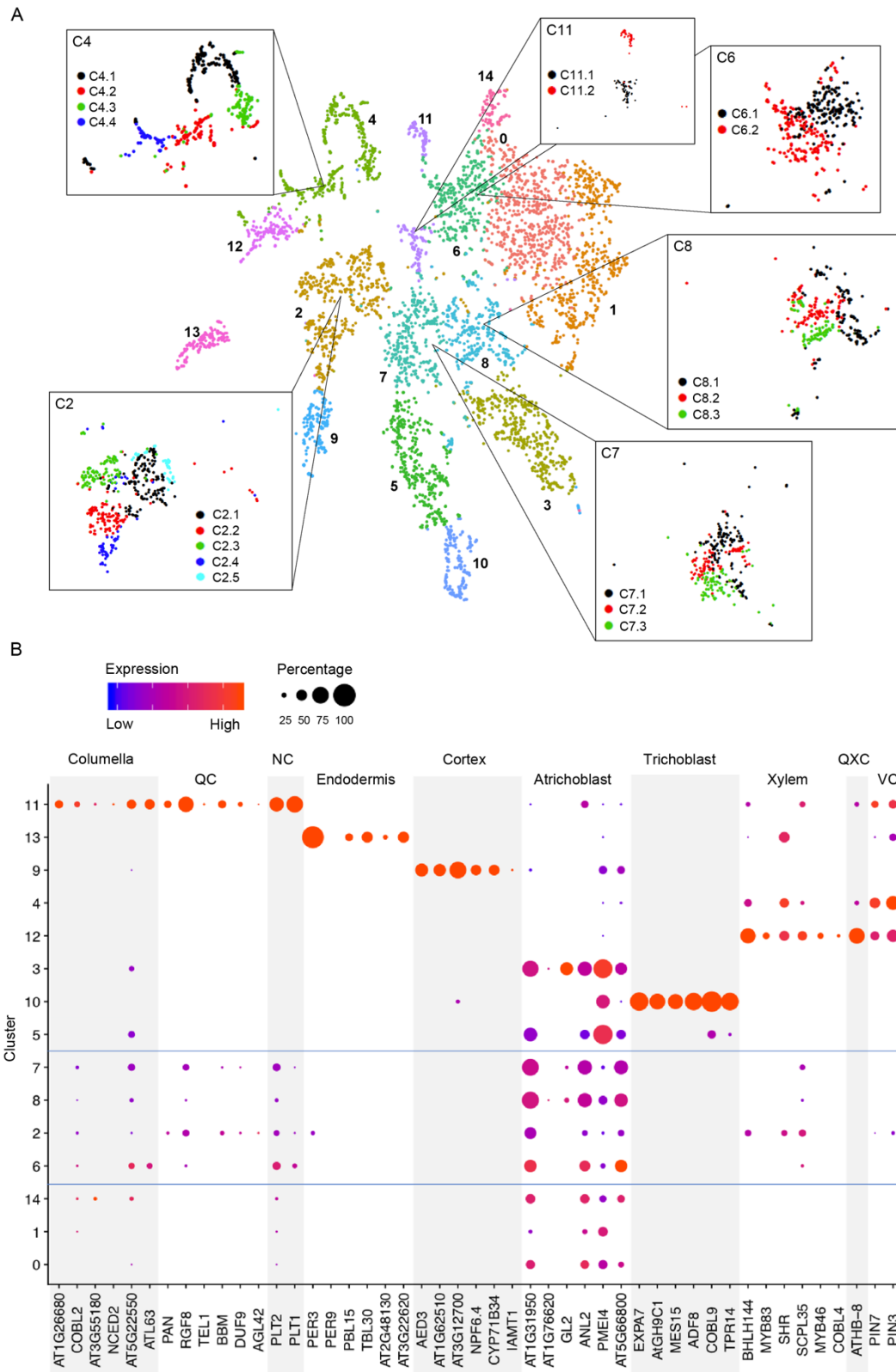
Figure 1. Sequenced single cells cluster by identity. (A) t-SNE plot of 4,727 *Arabidopsis* root cells shows these group into 15 clusters with additional sub-clusters. (B) Expression of known cell-type marker genes across cells reveals the identity of clusters. Dot diameter:

proportion of cluster cells expressing a given gene; Color: mean expression across cells in that cluster. NC: niche and columella; QXC: QC, xylem and columella; VC: vasculature and columella. See Table S2 for details of all marker genes assessed.

Clusters 9 and 13 comprise cortex and endodermal cells, respectively (Figure 1B; Figure S1G; Table S2). The identity of the endodermal cluster was further validated by the localized accumulation of *GFP* transcripts in one of the replicates generated from the *pMIR166A:erGFP* reporter line (see Methods; Figure 2A; Figure S3A). Additionally, when cells were re-clustered incorporating scRNA-Seq data of *shortroot* mutants (*shr-3*), which lack a defined endodermis (Helariutta *et al.,* 2000; Figure S3A), otherwise well-dispersed *shr-3* cells were absent from a cluster comprising endodermal cells of both wild-type replicates (Figure 2B; Figure S3). This cluster analysis also shows that *shr* cells, while present in all other clusters, localise on the outskirts of some (Figure 2B). This points to subtler effects of SHR on cell types other than the endodermis; although some of this phenomenon may also be attributable to the fact that *shr-3* is in the Ler. background. Irrespective, this observation nicely highlights the potential of applying scRNA-Seq to identify hidden phenotypic changes, whether stemming from natural variation or mutations.

Clusters 10 and 3 comprise trichoblast and atrichoblast cells, respectively (Figure 1B; Figure S1G; Table S2). Cluster 5 also contains trichoblast cells (Figure S1G). Although cells in this cluster show low expression of a number of atrichoblast marker genes, crucially, the trichoblast marker *COBL9* is expressed in this cluster whereas the atrichoblast marker, *GL2*, is not (Figure 1B; Figure S2). The co-expression of atrichoblast marker genes hints at a degree of commonality between this subset of trichoblasts and their epidermal counterparts, perhaps reflecting a distinction in developmental stage to the trichoblast cells contained in clusters 10.

Cluster 4 comprises stele cells while a neighboring cluster (12) comprises maturing xylem cells (Figure 1B; Figure S1G; Table S2). Consistent with the tissue complexity of the stele, subclustering reveals cell heterogeneity within cluster 4. Particularly, phloem and pericycle cells are separated into two discrete subclusters (Figure S4), as highlighted by the highly subcluster-specific expression of genes such as *APL* (C4.2), *LBD29,* and *TIP2-3* (C4.1) (Figure 2D; Bonke *et al.,* 2003; Porco *et al.,* 2016; Gattolin *et al.,* 2009).

Finally, cluster 11 comprises both columella and QC cells (Figure 1B; Figure S1G; Table S2), which can be separated into two subclusters. Subcluster 11.2 contains columella cells that express marker genes such as *COBL2*, *NCED2*, and *ATL63* (Figure 2E; Brady *et al.,* 2007b; Efroni *et al.,* 2015). In contrast, transcripts for the QC-expressed genes *AGL42*, *BBM*, and *TEL1* are largely limited to cells in subcluster 11.1 (Figure 2E; Nawy *et al.,* 2005; Efroni *et al.,* 2015). Given the small number of QC cells per root, this cluster may well contain other, transcriptionally similar cells, perhaps the adjacent initials in the niche. However, importantly, the fact that QC cells are captured illustrates well the possibilities of this methodology for studying rare cell types or to elucidate transcriptional subtleties affecting small numbers of cells within a tissue.

Meristematic cells cluster independently of tissue identity

The identity of cells in the remaining clusters is less obvious. Overall gene expression in cells within clusters 0, 1 and 14 is comparatively low (Figure S5A), likely masking their identity at this level of sequencing resolution. However, expression values extracted from a longitudinal microarray dataset (Brady *et al.,* 2007a) for the top DE genes defining these clusters suggest that they comprise mature cells of mixed identity (Figure S5B). In contrast, cells in the final four clusters (2, 6, 7 and 8) show markedly meristematic-based expression profiles (Figure S5B). Notable histone and cytokinesis-linked genes, such as *KNOLLE*, *ENODL14*, and *ENODL15*, are amongst the most prominently differentially expressed genes for these clusters (Figure S2; Table S2; Lauber *et al.,* 1997; Adrian *et al.,* 2015). Subclustering revealed some cell type identities, albeit that they are generally less distinct than those of the clusters described above. For example, subcluster 2.4 shows a distinct cortex identity (Figure S4). Curiously, this subcluster is positioned adjacent to the main cortex cell cluster. A comparable pattern is seen in clusters 7 and 8 with trichoblast cell identity apparent in those subclusters (7.3 and 8.3) closest to the adjacent defined trichoblast clusters (Figure S4).
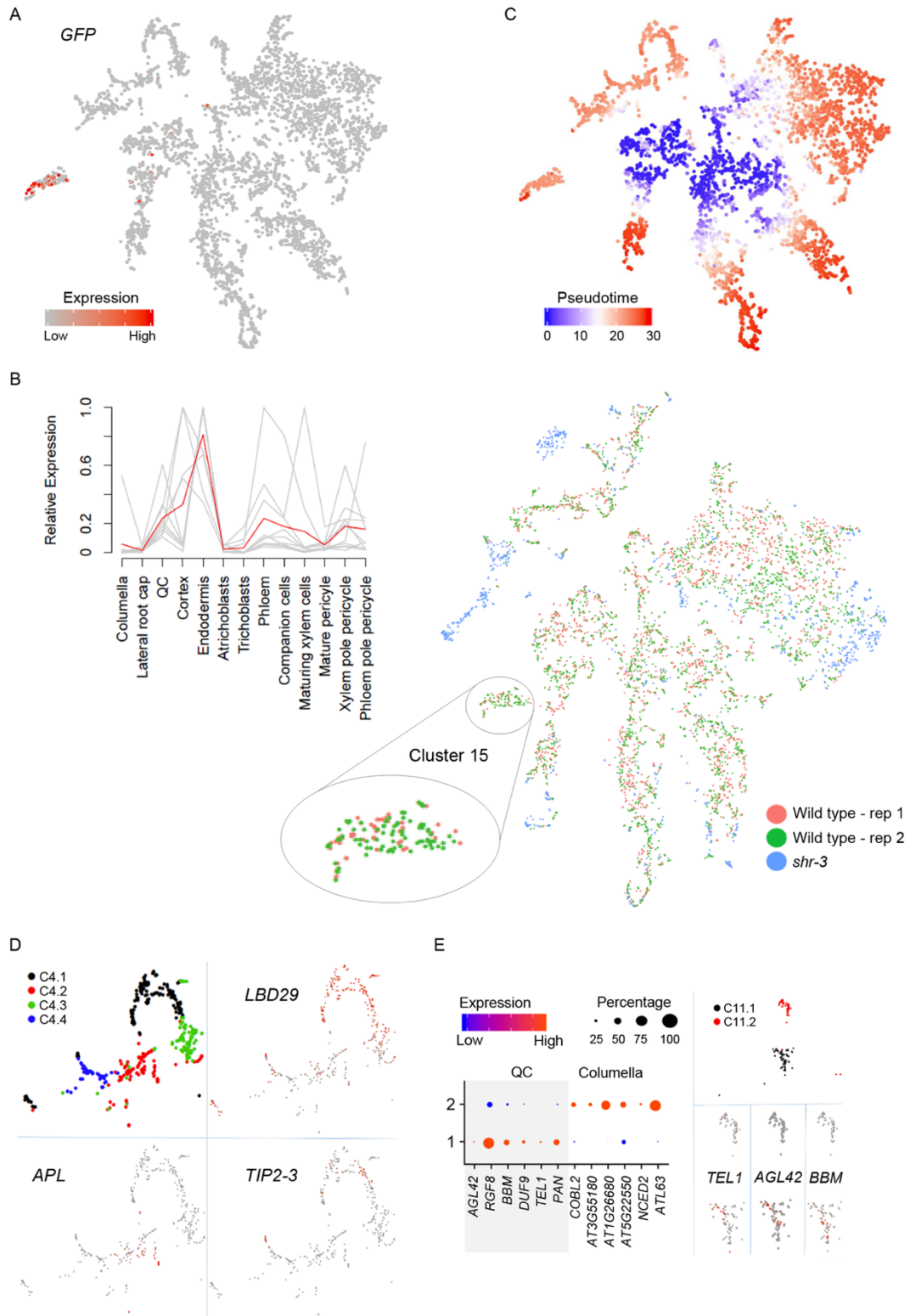
Figure 2. Cell identity and developmental stage are reflected in the (sub)clustering. (A) t-SNE visualisation of the cluster cloud (both replicates) shows *GFP* transcripts localise specifically to *pMIR166:erGFP* cells in the endodermal cluster. (B) Wild-type and *shr-3* cells were combined and clustered. (left) The expression profiles for the top 10 DE genes from

cluster 15, taken from a microarray root atlas (Brady *et al.,* 2007a), reveal an endodermal identity. Red line, mean expression profile; grey lines, individual expression profiles. (right) t-SNE visualization of the wild-type and *shr-3* cell cluster cloud shows endodermal cluster 15 lacks *shr-3* cells. (C) Pseudotime analysis of all wild-type cells reveals cells in the central clusters are earlier in the pseudotime trajectory, consistent with their meristematic identity. (D) t-SNE visualization of cluster 4 subclusters. Expression profiles of selected genes reveals pericycle and phloem identities for C4.1 and C4.2, respectively. (E) Expression of known cell-type marker genes across cells of cluster 11 reveals the identity of subclusters. Dot diameter: proportion of cluster cells expressing a given gene; Colour: mean expression across cells in that cluster. t-SNE visualization of expression profiles of selected QC marker genes.

It is interesting to note that when comprehending all the clusters together, the meristematic clusters are closely localized in the center of the cluster-cloud with the subcluster containing QC cells (11.1) at the heart of this. Meanwhile, those clusters with distinct, mature cell identities span out from the meristematic clusters (Figure 1A; Movie S1), suggesting an overall cluster arrangement that reflects developmental time. Subclustering of the meristematic clusters refines this idea, showing a degree of closeness of mature and developing cells of the same eventual fate. This notion is further supported by pseudotime analysis across all cells which reveals that genes differentially expressed in cells of the central clusters describe the beginning of cell fate progressions (Figure 2C).

Likewise, the cluster-cloud reveals an organization that captures the "lineage" relationships between cell and tissue types. For instance, the trichoblast and atrichoblast clusters, as well as the xylem and vasculature, and the cortex and endodermis clusters, are positioned next to each other within the cluster cloud. The position of the columella in a cluster with QC cells indicates a higher degree of transcriptional accord between these cell types than between these cell types individually, and others. This is reflected in the fact that key developmental regulators, such as *PLT2*, *PLT3*, and *PIN4* are coexpressed in the columella and QC (Galinha *et al.*, 2007; Feraru and Friml, 2008). This way of contemplating clusters, along with pseudotime visualization, thus offers a valuable director for early comprehension of developmental trajectories, particularly in the absence of *a priori* knowledge, such as a reference atlas.

Novel marker genes define cluster identity in an unbiased manner

Given that detailed reference datasets are available only for select tissue and organ types in very few plant species, we developed an unbiased approach to assign cell type identities to scRNA-Seq generated cell clusters. Genes differentially expressed in a given cluster compared to all other clusters (q <0.01; Average Log FC ≥0.25) were identified using 'biomod' on Seurat (McDavid *et al.,* 2013). Differentially expressed genes were further narrowed down by applying the criteria that cluster-specific marker genes must be expressed in ≥10% of cells within the cluster (PCT1), and ≤10% of cells across all other clusters (PCT2). Applying these criteria, we uncovered expected marker genes alongside hundreds of novel genes diagnostic for a given developmental stage or cell type that encompass every cluster (Table S3).

The top two cluster-specific genes (based on average Log FC) for each cluster are expressed across a substantial proportion of cells specifically within one cluster, with the exception of genes for clusters 0 and 1, which show substantial co-expression in cluster 14 (Figure 3A). In addition, GO overrepresentation analysis on cluster-specific gene sets reveal GO terms appropriate to their biology (Table S3). For example, the meristematic clusters 2 and 8 show an abundance of marker genes implicated in processes related to cell proliferation and DNA replication, respectively. Further, markers for the root-hair-cell cluster 10 are enriched in trichoblast differentiation and maturation terms, for the QC- and columella-containing cluster 11, in root development and starch biosynthesis, and for cluster 12, in xylem development and secondary cell wall biogenesis. Finally, genes required for the formation and suberisation of the Casparian strip are among the markers for cluster 13, which comprises endodermal cells. However, a notable outcome of this analysis is the number of marker genes for which a root function has yet to be assigned. This illustrates the potential of scRNA-Seq for identifying new developmental regulators.

To further validate this strategy for marker gene calling and for assigning cell identity to clusters without other reference, we assessed the spatiotemporal patterns of expression for select genes using transcriptional *promoter:3xVenus-NLS* reporter lines. Prioritizing by a balance of high log fold change, high PCT1, low PCT2, and a lack of prior biological information relating to cell type specificity and root development, we selected ten genes from across clusters. Expression for eight of the ten genes

tested localized to specific cell types and/or root zones in line with predictions. Specific expression in the cortex (*AT1G62510*), and maturing trichoblasts (*MES15*) was observed for marker genes for clusters 9 and 10, respectively (Figures 3B and 3C), while genes selected from cluster 4 revealed highly specific phloem (*PME32*) and pericycle (*ATL75*) expression (Figures 3D and 3E). *MLP34* is expressed in the atrichoblasts, as expected for a marker for cluster 3 (Figure 3F). However, expression is also seen in cells of the lateral root cap (Figure 3F), a cell type to which a cluster could not be assigned. *MLP34* shows expression in some cells in cluster 1 (Figure S2), indicating that this cluster may in-fact contain cells of the lateral root cap, although further analysis is needed to confirm this. Finally, expression of genes selected from the meristematic cluster 2 was found to localize to the meristematic cortex and endodermis (*AT3G22120*), the meristematic cortex (*AT1G62500*), or the meristematic vasculature (*PIP2-8*) (Figures 3G, 3H and 3I).

Given the common occurrence of *cis*-regulatory motifs in introns of genes, the fact that promoter fusions for eight out of the ten marker genes tested confirm predictions is notable. This unbiased approach for assigning identity to cell clusters could prove invaluable when no reference data is available. Moreover, our results reveal a level of sensitivity beyond that of assigning whole cluster identity. This is typified by *PME32* and *ATL75* whose promoter fusions show expression in the phloem and pericycle, respectively, in accordance with their expression being predominant in cells of subclusters 4.2 and 4.1, respectively (Figure S2). Furthermore, the meristematic vasculature marker, *PIP2-8*, is primarily expressed in cells of subcluster 2.3, positioned adjacent to the vasculature cluster, while the meristematic cortex & endodermis marker, *AT3G22120,* is mainly expressed in cells of subcluster 2.2 positioned near these mature cell types (Figure S2). The expression patterns observed for the latter reporters thus further reinforce the hypothesis that meristematic subclusters share expression features with the mature cell types they are closest to. Moreover, a gene's expression profile across the cluster cloud is a confident predictor of its localized expression *in planta*.
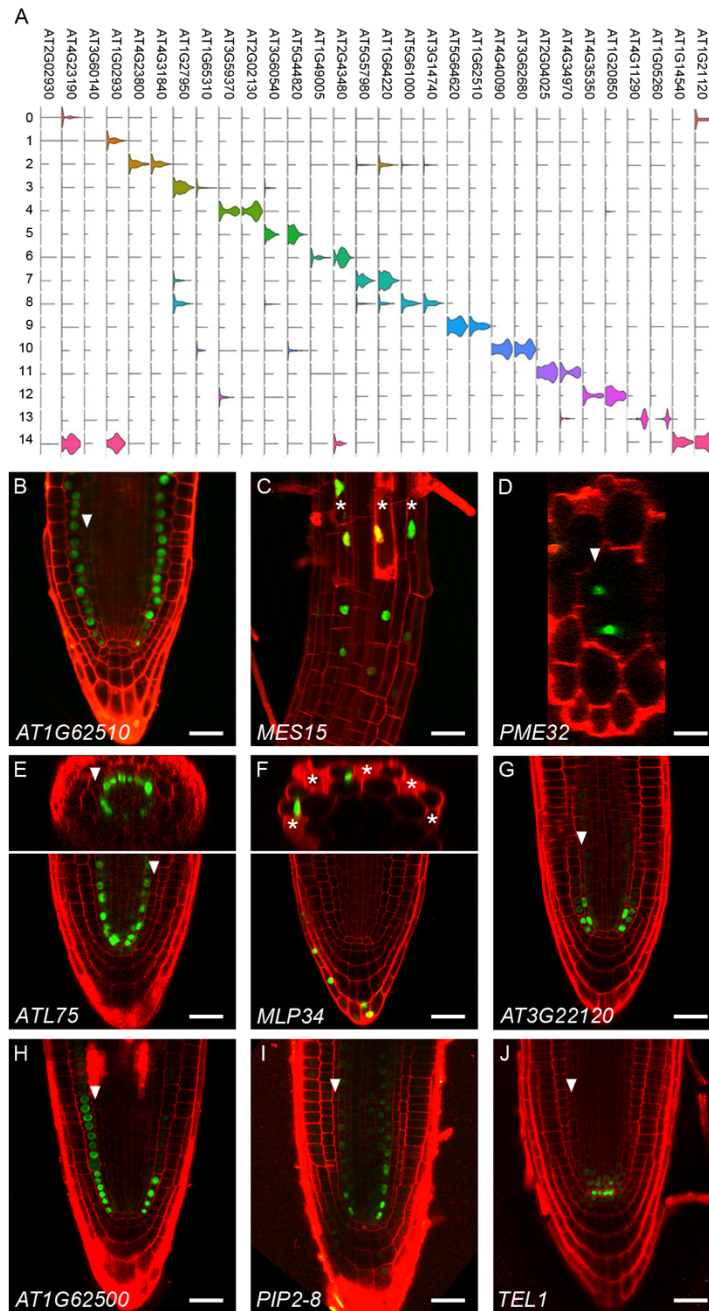
Figure 3. An unbiased, marker-gene-selection method is validated with transcriptional fusions. (A) Cluster specificity revealed by violin plots depicting expression level (length) and proportion of cells expressing (width) for the top two marker genes by differential expression (x-axis) across each cluster (y-axis). (B-J) Spatiotemporal expression patterns for promoter fusions for the following new marker genes reveals the predicted cell type specificities: (B) *AT1G26510*, cortex; (C) *MES15*, differentiating trichoblasts; (D) *PME32*, phloem; (E) *ATL75*, pericycle; (F) *MLP34*, atrichoblasts and lateral root cap, (G) *AT3G22120*, meristematic cortex and endodermis; (H) *AT1G62500*, meristematic cortex; (I) *PIP2-8*, meristematic vasculature;

(J) *TEL1*, QC. White arrowheads indicate endodermal cells; asterisks indicate trichoblasts. Scale bars = 20 μm.

QC marker genes identified from scRNA-Seq data

The fact that QC cells are captured offers a unique opportunity to study this rare cell type, a point quite pertinent given that RNA-seq analysis shows the established marker *WOX5* to be induced upon protoplasting (Table S1). Within subcluster 11.1, 36 cells express at least half of 15 proposed QC genes (Figure 4A; Table S2; Efroni *et al.,* 2015; Nawy *et al.*, 2005), which is in line with the sampling depth of scRNA sequencing and the relative low expression of most QC genes. The high number of QC cells captured likely reflects a bias in the methodology towards capturing small cells (see Methods), which may also account for an over-representation of meristematic cells. Reinforcing our QC cell calling, it is of note that genes marking initial cells directly neighboring the QC, such as *AT3G22120, AT1G62500,* and *PIP2-8* (Figures 3G, 3H, and 3I), show no (33 cells), or negligible (3 cells) expression in the QC cells. Additionally, cells expressing such genes cluster away from the QC, in localized regions of meristematic cluster 2, adjacent to their mature-cell counterparts (Figure S2).

Transcriptomic comparison between the QC cells and undifferentiated cells of the meristem (cells in clusters 2, 6, 7, and 8), identified 254 genes preferentially expressed in the QC (Table S3). While meristematic cells are distinguished by expression of genes involved in cell division and DNA replication, cells of the QC are not. Instead, transcription is an enriched GO term, as is auxin biosynthesis, which is fitting given the role of auxin in QC specification (Sabatini *et al.*, 1999; Galinha *et al.*, 2007). Further, unexpectedly, genes with functions in glucosinolate biogenesis and callose deposition are overrepresented among those genes differentially expressed in the QC (Table S3). This finding in particular is intriguing. Both processes are characteristic of a defence response, which seems curious given the QC's internal location, insulated from external stimuli. Their prominence instead points towards a biology of QC cells not previously appreciated. The recent finding that 3-hydroxypropylglucosinolate acts as a reversible inhibitor of root growth (Malinovsky *et al.*, 2017), is in this regard intriguing. Likewise, that small RNAs are prevented from moving in and out of the QC (Skopelitis

et al., 2018) points to unique regulation of cell-cell communication via plasmodesmata in the QC.

Among the genes differentially expressed between the QC and meristematic cells of the root, 47 show a particularly strong expression bias to the QC cells (Log FC ≥0.25, PCT1 ≥10; PCT2 ≤10) (Table S3). Many of these genes are also expressed in mature cell types, predominantly columella cells, further reinforcing a certain shared-biology not present in the apical root meristem. However, ten genes clearly mark the QC cells (Figures 4B and 4C). Reporter lines for one them (*TEL1*) revealed high expression in the QC cells, and minimal expression elsewhere (Figure 3J). This expression contrasts strongly with that of genes marking various cell type initials (Figures 3G, 3H, and 3I). Amongst the eleven QC markers are several genes not previously described in connection to the QC, including *AT3G18900* and *SOT17* (Figure 4C; Table S3). Intriguingly, the latter functions in the glucosinolate pathway (Klein *et al.*, 2006). This reinforces the prospect of there being a new layer to QC function, and highlights the potential for this technology in identifying novel developmental regulators and downstream genes that translate cell fate into distinctive cell shapes and functions.

The limited number of QC-specific genes, in comparison to the number of specific markers identified for cell types such as the endodermis, xylem, and trichoblasts, would indicate that QC identity reflects the integrative outcome of multiple overlapping expression signatures. This idea is supported by the position of the QC at the intersection of an auxin maximum and SCARECROW activity (Shimotohno *et al*., 2018). However, an alternative, non-mutually exclusive interpretation of this finding is that QC identity reflects a "subtractive" expression signature. In this scenario, the absence of expression of drivers of tissue identity and differentiation forms a key feature of QC identity. The facts that WOX5 acts as a transcriptional repressor *(*Forzani *et al*., 2014; *Pi et al., 2015), and that genes, such as AT3G22120, AT1G62500,* and *PIP2-8* are mostly undetectable in the QC, *supports this idea.*
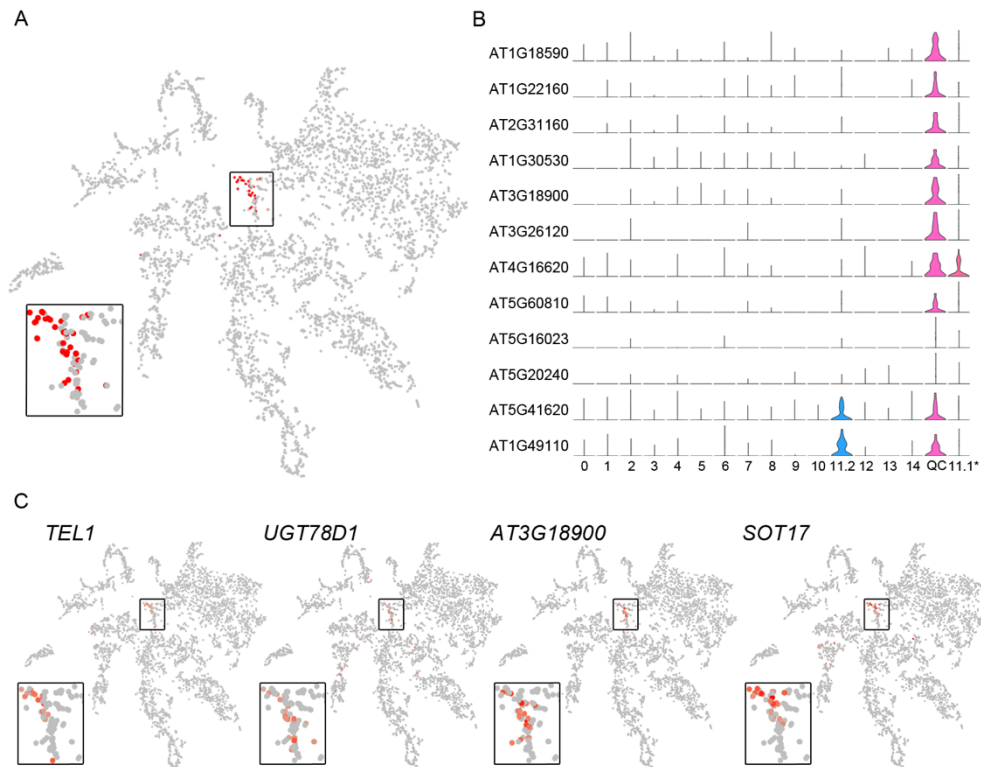
Figure 4. scRNA-Seq data of QC cells identifies novel QC marker genes. (A) t-SNE visualisation indicating the position of 36 QC cells at the centre of the cluster cloud. (B) QC cell specificity revealed by violin plots depicting expression level (length) and proportion of cells expressing (width) for 10 QC cell marker genes (y-axis) across each cluster (x-axis). Violin plots for two genes showing expression in the QC cells and the columella cluster (11.2) are also shown. *C11.1, subcluster 11.1 with QC cells removed. (C) t-SNE visualisation of expression profiles of selected QC marker genes. Cluster 11.1 is magnified in each case.

Cell differentiation reflects finely resolved waves of gene expression

One of the most exciting benefits of scRNA-Seq is that it allows for the simultaneous, unbiased analysis of every type of cell at every developmental stage, in one experiment. This is broadly illustrated by pseudotime analysis across all cells, which shows how central clusters are defined by expression of genes at the beginning of cell fate progressions, whereas mature cell types are peripheral in the cluster cloud (Figure 2C). However, although the cluster cloud represents a coarse landscape of developmental cell states, it does not reveal how individual cells traverse these states. To resolve the progressions that cells undergo during their transition from stem cell to mature trichoblast, we performed pseudotime analysis on cells of clusters 7, 5, and

10. This reveals a linear ordering of cells that reflects the cluster and subcluster arrangement (Figures 5A and 5B). In addition, we identified 3,657 highly dispersed, differentially expressed genes that fall into 8 distinct gene clusters (*a-h*), and depict successive waves of gene expression across pseudotime (Figure 5C; Figure S6A; Table S4).

Reporter lines generated for representative genes of select clusters precisely reflect their pseudotime profiles (Figures 5D to 5H). *AT5G62330* and *PAE2* (cluster *d*) show a distinct peak of expression near the center of the pseudotime trajectory. Reporter lines for both genes capture this expression dynamic. *pAT5G62330:3xVenus-NLS* expression initiates in the distal meristem and persists into the elongation zone, whereas *PAE2* is expressed slightly later and shows strong expression particularly in the early elongation zone (Figures 5D and 5E). Expression of *AT1G05320* (cluster *f*) overlaps with that of *AT5G62330* and *PAE2* in elongating cells, but expression persists into the maturation zone and differentiated trichoblasts (Figure 5F). Finally, reporter activity for *MES15* and *EXT12* (cluster *g*), whose expression initiates late in the pseudotime trajectory, is first detected in cells exiting the elongation zone, with *MES15* expression starting slightly earlier with respect to the first visible protrusion of the emerging hair (Figures 5G and 5H), in accordance with its slightly earlier pseudotime projection.

A

B

C

Cluster

Meristematic ——————→ Mature   Enriched GO terms

a
- DNA replication
- DNA packaging
- Chromatin assembly
- Cell proliferation

b
- DNA confirmation change
- Response to cytokinin
c

d
- Cell wall organisation

e

f
- Cell growth
- Cell wall biogenesis
- Developmental growth

g
- Vesicle-mediated transport
- Regulation of cytoskeleton organisation

- Root hair elongation, Unidimensional cell growth
- Trichoblast maturation

h

C7.1 C7.2 C7.3 C5 C10

D *AT5G62330*

E *PAE2*

F *AT1G05320*

G *MES15*

H *EXT12*

I Meristematic ——————→ Mature
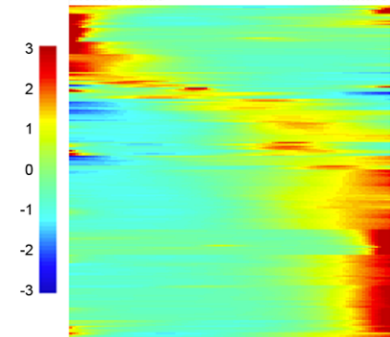
Figure 5. Trichoblast development is guided by progressive waves of gene expression.
(A) Pseudotime reconstruction of trichoblast development reveals a linear ordering of cells, reflecting cluster and subcluster arrangement. (B) Density distribution of cells across pseudotime. (C) Expression heatmap of 3,657 highly dynamically expressed genes ordered across pseudotime reveals trichoblast differentiation reflected in multiple progressive waves of gene expression. Significantly enriched GO terms for clusters are labelled. Lower bar, cell density distribution across clusters. (D-H) Spatiotemporal expression patterns for promoter fusions for the following genes reveals the pseudotime-predicted temporal localization: (D) *AT5G62330*; (E) *PAE2*; (F) *AT1G05320*; (G) *MES15*; (H) *EXT12*. White arrowheads indicate the start of expression. Expression dynamics for single genes plotted across trichoblast pseudotime are under the corresponding root images. (I) Expression heatmap of 230 transcription factors (TFs) extracted from Figure 4C shows similar waves of TF expression. See Table S4 for full data.

The pseudotime trajectory thus reflects with great precision the temporal expression changes of individual genes along the length of the root (Figure 5C), providing a refined view of the changes a cell undergoes during its transition from stem cell through to full differentiation. Genes predominantly expressed at the beginning of the developmental trajectory show an overrepresentation of DNA replication, cell proliferation, and ribosomal functions, as is expected for meristematic cells (clusters *a* and *b*; Table S4). At the other end of the trajectory, cluster *g* captures expression of genes involved in unidimensional growth, and root hair elongation and maturation. A previous gene regulatory network for epidermal cell differentiation had identified 154 core root hair genes from which a temporal progression could be deduced (Bruex *et al.,* 2012). ScRNA-Seq provides a new dimension of temporal resolution. Of the subset (98) of core root hair genes that show a dynamic pattern of expression across the pseudotime, the vast majority (84) are expressed late in the trajectory (cluster *g*; Table S4). In contrast, early cell fate determinants, including *GL2, TRY*, and *WER*, are expressed in cluster *a* (meristem). Our analysis thus reveals stepwise temporal progressions more dynamic than previously appreciated that connect early cell fate decisions to morphological and cellular changes.

These stepwise progressions are primarily captured by the central gene expression clusters. Clusters *d* and *e* reveal cell growth and cell wall biogenesis amongst their enriched GO terms, while those of cluster *f* indicate a burst of cell morphogenesis

activity with an overrepresentation of genes involved in cytoskeleton reorganization, vesicle trafficking and a plethora of transport processes. Cell expansion and cell reorganization are processes known to occur during root hair development (Balcerowicz *et al.*, 2015), and the pseudotime analysis identifies specific genes that could drive processes such as these during root hair differentiation (Table S4). Also found in cluster *d* is an abundance of flavonoid-associated genes, suggesting that this signaling is occurring during cell elongation, downstream of *GL2* and *WER*, but upstream of many auxin/ethylene responsive genes which are overwhelmingly found in the later gene clusters (*f* and *g*).

A highly interconnected TF gene regulatory network coordinates cell differentiation

Expression profiles of the 239 TFs amongst the dynamically expressed genes mirror their waves of expression (Figure 5I; Table S4), pointing towards a causative relationship and an intricate regulation of cell fate progressions. Several of these are known to regulate specific stages of root hair development, including TRY, WER, and many bHLH TFs (notably, RSL2, RSL4, and LRL1). The expression clusters in which they reside accurately reflect their biological roles (Schellmann *et al.*, 2002; Diet *et al.,* 2006; Bruex *et al.,* 2012; Balcerowicz *et al.,* 2015). Likewise, TMO7, previously reported as a central cell-to-cell communicator and regulator of root meristem activity (Lu *et al.*, 2018), is found early on in the trajectory, while the central clusters contain TCP TFs connected to the exit of cell proliferation (Nicolas and Cubas, 2016), and ARR1, which regulates root meristem size (Dello Ioio *et al.*, 2008).

To elucidate further the genetic coordination along the trichoblast differentiation process, we inferred a Gene Regulatory Network (GRN) using a pipeline integrating the transformation of linear ordinary differential equations (ODEs) and linear regression (SCODE) (Matsumoto *et al.*, 2017; Table S4; see Methods). Incorporating TF expression dynamics across pseudotime, the resultant network reveals key players in this process and their regulatory interactions. What is immediately clear from the network is the presence of major, highly-connected central regulators along the whole trajectory (Figure 6A), the majority of which have not previously been implicated in root development. Filtering the network down to its 25 core components (see Methods), we see several distinct passages of feedback regulation along each step of the

trichoblast differentiation trajectory (Figure 6B). Notable amongst these is considerable negative feedback from TFs at the end of the trajectory back to major nodes in the meristem, especially LRL3 (Figure 6B). This suggests that while positive and negative feedback loops in the meristem (also distinct) might maintain meristematic identity, the progression to differentiation gives rise to dominant components that influence the meristematic master-players.

ATHB-20, and to a lesser extent, ATHB-23, stand out as positive regulators of the core network components in the neighboring, upstream, meristematic cluster, including the aforementioned TMO7. While it has been linked to ABA sensitivity in the root (Barrero *et al.,* 2010), no defined role has yet been assigned to ATHB-20 in root development; yet a connection to a major player such as TMO7 suggests this to be the case. Pertinent to this, it is important to note that, with the exception of those in the late developmental cluster, the 25 core genes displayed in the network are not specific to trichoblast cell development (Figure S2; Table S4). This would indicate that they fulfil a similar role in other tissue layers, possibly coordinating growth and differentiation across the root. This notion is supported by pseudotime analysis of the cortex, which reveals an equally dynamic gene expression cascade mirrored by waves of transcription factor expression (Figures S6B and S6C; Table S4). With the notable exception of TMO7, many of the core nodes identified in the trichoblast network, including ATHB-20, ATHB-23, CBF5, GATA4 and GATA2, are also present in equivalent positions along cortex pseudotime.

While negative feedback from differentiating root hair cells towards the meristem is notable in the simplified network, forward regulation of cell differentiation can be found as a culmination of many, weaker interactions from the meristematic and central nodes, in the extended network (Figure 6A; Table S4). The well-established regulators of trichoblast maturation, expressed at the end of the trajectory, are regulated by a combination of additive and opposing effects of broadly expressed (for example, ATHB-20, ATHB-23), and locally restricted (such as GL2 and WER) TFs. Network configurations in which tissue specific expression reflects the combinatorial output of many broadly expressed and locally restricted TFs are emerging as a general feature underlying development (Sparks *et al*., 2016; Reiter *et al*., 2017; Niwa, 2018; Barolo and Posakony, 2018), buffering the system and providing robustness.

Our network serves to particularly highlight the complex wiring between components mediating the transition of cells from the stem cell niche to the root maturation zone. Altogether, the pseudotime analysis of scRNA-Seq data indicates that these transitions are more gradual than previous data suggested. We see evidence of intricate dynamic transcriptional regulation, particularly across cells of the elongation zone. Termination of meristem activity and the initiation of the differentiation process appear coordinated across tissues, while the trichoblast maturation process relies on tissue-specific TFs whose activity comes about from the combinatorial regulation of dozens of temporally upstream players.
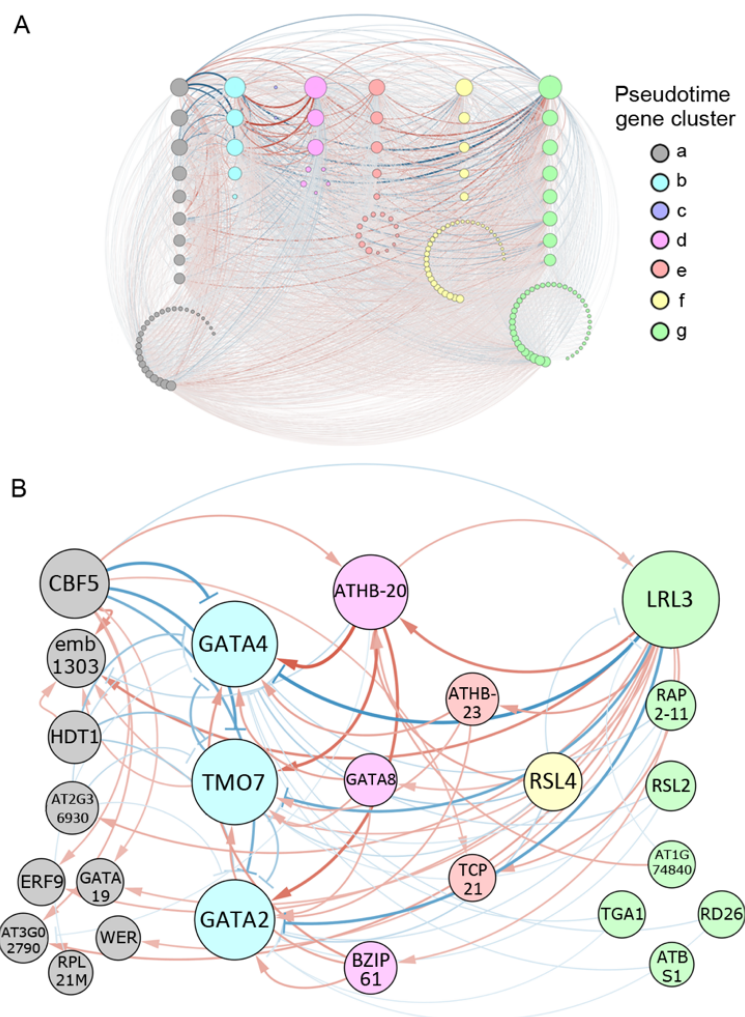
Figure 6. A Gene Regulatory Network (GRN) predicts key regulators during trichoblast cell differentiation. (A) A GRN built of 229 TFs expressed dynamically across trichoblast pseudotime with a parameter cut-off of 0.1 (Table S4). (B) The same network with a parameter cut-off of 1.5. Node size is equivalent to the number of predicted connections. Edge colour

represents activation (red) or repression (blue). Edge width represents the strength of the predicted connection.

Discussion

While scRNA-Seq in plants has previously been limited to analysis of cells in their hundreds, the data presented here shows that profiling of developmentally complex tissues using high-throughput scRNA-Seq of thousands of cells offers an unparalleled view of the spatiotemporal expression dynamics cells undergo between exiting the niche and their final differentiation.

scRNA-Seq of the *Arabidopsis* root proved to be sensitive and highly reproducible. Select genes induced by protoplasting as well as other plant physiology-based challenges are not prohibitive, but must be considered. For instance, the technology has an inherent bias towards smaller cells, such as those from meristems. Mature cells, while captured in sufficient numbers for analysis, are underrepresented in the atlas as their identities appear to become partially clouded during the protoplasting and cell capture processes. Likewise, epidermal cells are captured more readily than cells of the central stele. These trends, however, do not affect the types of analyses presented here.

The nearly five thousand root cells profiled form a high-resolution atlas that captures all major cell types and developmental stages, including the scarce cells of the QC and niche. With this atlas, we can predict with precision the spatiotemporal patterns of gene expression within the root, as demonstrated by our reporter lines. Further, the atlas offers new insights into cellular processes characterizing distinct cell and tissue types. An unbiased approach to marker gene calling identified expression features defining each cluster and subcluster. These negate the need for *a priori* knowledge in assigning cluster identities and identify novel developmental regulators and downstream genes that give cells their distinctive forms and functions. As an example, *SOT17* shows a QC-specific pattern of expression, and genes implicated in glucosinolate biosynthesis and callose deposition more generally show preferential expression in QC cells. Although a connection to cell-specific defence responses cannot formally be excluded, the latter finding more likely relates to plasmodesmata-

mediated signaling, given the recent finding that small RNAs are blocked from moving in and out of the QC (Skopelitis *et al.*, 2018).

It is remarkable just how well the arrangement of cell clusters and subclusters described in our atlas reflects developmental time. In the very center are QC cells with a transcriptome distinctive from that of the surrounding initials and meristematic cells. Differentiating cells are on the periphery of the cluster cloud, adjacent to relevant meristematic subclusters. This cluster arrangement is likely a general feature of tissues that capture developmental trajectories, as a similarly developmentally-informed cluster orientation has been recorded in a study of mouse spermatogenesis (Lukassen *et al.,* 2018). In the case of the *Arabidopsis* root, it indicates that expression signatures linked to cell fate are not as strong as those defining stem cell or meristem identity.

The progressions that cells undergo during differentiation are, however, far more dynamic than captured in just the cluster arrangement. As illustrated by the successive waves of gene expression revealed by pseudotime analysis, cells transitioning from the niche through differentiation follow finely resolved developmental trajectories, with progression steps beyond the commonly described meristematic, elongation and maturation zones. Our scRNA-Seq data offers the required resolution to distinguish the stepwise temporal progressions connecting early cell fate decisions to morphological and cellular changes. While the beginning and end of the trichoblast developmental trajectory have been described (Breux *et al.*, 2012; Balcerowicz *et al.*, 2015), our data not only adds to this, but reveals additional gene expression dynamics occurring particularly in cells in the elongation zone. The number of progression steps observed, compared to the number of cells along the root, from the meristem to maturation, implies that the distinct progressions are characteristic of few, perhaps even individual, cells along the root elongation zone.

Recent studies elaborate the idea that cell fate reflects the output of intricate gene regulatory networks in which numerous TFs control gene expression in a combinatorial manner (Sparks *et al.*, 2016; Reiter *et al.*, 2017; Niwa, 2018; Barolo and Posakony, 2018). This notion is reinforced by the pseudotime-derived trichoblast gene regulatory network, which shows that expression of major regulators of trichoblast maturation is linked to many, weak interactions from meristematic and central nodes.

Interestingly, the network also implies a high degree of feedback regulation towards the meristem, not only from the elongation zone, but also from genes in the maturation zone, such as LRL3. Normal root development requires that growth and differentiation be coordinated across tissue layers. In line with this, the central highly-integrated nodes predicted to coordinate the transitions between zones show similar pseudotime profiles in both the cortex and trichoblast lineages.

An additional key application for the scRNA-Seq technology will be the profiling of mutants to more precisely define the cellular processes, cell types, or developmental stage affected. Cluster analysis of *shr* and wild type scRNA-Seq data revealed an expected absence of endodermal cells in the mutant (Helariutta *et al.,* 2000), but also points to more extensive SHR-dependent cell fate changes. The *shr* scRNA-Seq data thus nicely exemplifies the enticing prospect this technology offers to discern phenotypes not easily recognized by standard RNA-sequencing, physiological, or even reporter-based approaches, whether stemming from natural variation, mutations, localized stress responses, or plant-microbe interactions.

In summary, the atlas of the *Arabidopsis* root described here provides a unique spatiotemporal perspective of root cell type differentiation, increasing the number of discernible developmental domains along the length of the root and pointing to countless candidate developmental regulators that orchestrate this process. ScRNA-Seq will rapidly become a central technique in the plant sciences as it already is in mammalian studies, providing previously unobservable developmental insights.

Material and methods

Experimental model and subject details

All analyses were performed in the *Arabidopsis* Columbia (Col-0) ecotype with the exception of the *shr-3* mutant line, which is in the Ler. ecotype. The *pMIR166A:erGFP* reporter and *shr-3* mutant have been described previously (Miyashima *et al*., 2011; Helariutta *et al*., 2000, respectively). Plants were grown at 22°C on 1% agar plates containing 0.5x Murashige and Skoog (MS) medium (Duchefa Biochem.).

Protoplast isolation

Seedlings were grown vertically on nylon mesh on agar plates. Roots of 6-day-old seedlings were cut approximately one centimeter from their tip, broadly diced with a scalpel blade, and treated with 7 ml protoplasting solution optimized for scRNA-Seq from a protocol in Birnbaum *et al.*, 2003. Immediately before use, 1.5% Cellulase R-10 and 0.1% Pectolyase (Duchefa Biochem.) were added to fresh protoplast buffer (0.1 M KCl, 0.02 M $MgCl_2$, 0.02 M $CaCl_2$, 0.1% BSA (Sigma Aldrich), 0.08 M MES, and 0.6 M Mannitol, adjusted to pH 5.5 with 0.1M Tris HCl), and mixed thoroughly. Root tissues were protoplasted for 2 hours at 20°C on an orbital shaker set at 200 revolutions/minute. The mixture was subsequently filtered through a 100 µm nylon filter and rinsed with 1-5 ml of root protoplast buffer. Protoplasts were then centrifuged for 10 minutes (500 g – 4°C), the supernatant gently removed, and the pellet resuspended in 10 ml root protoplast buffer containing 0.4 M Mannitol and no $CaCl_2$. This wash procedure was repeated once more, the protoplasts centrifuged as before, and resuspended in ~500 µl or less protoplast buffer without $CaCl_2$ and with 0.4 M Mannitol. Protoplasts were validated under a light microscope, and if necessary any excess debris or un-protoplasted tissues removed with an additional washing step. Cells were filtered through a 40 µm cell strainer (Flowmi Bel Art SP Scienceware), quantified using a haemocytometer, and adjusted to a density of approximately 800-900 cells per µl.

Bulk RNA-seq library preparation and sequencing

RNA was extracted using the Spectrum Plant RNA Extraction Kit (Sigma) from protoplasted and equivalent un-protoplasted root tissue collected at completion of the protoplasting procedure. RNA samples were quantified by Nanodrop, and quality assured based on Agilent RNA Bioanalyzer chip traces. mRNA was enriched by oligo-dT pull-down using the NEBNext Poly(A) mRNA Magnetic Isolation Module, and RNA libraries constructed using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina with NEB Multiplex oligos. Final library size and quality was checked on a DNA High Sensitivity Bioanalyzer chip (Agilent), and libraries were quantified using the NEBNext Library Quantification Kit for Illumina.

Single Cell RNA-seq library preparation and sequencing

Single cell RNA-seq libraries were prepared from fresh protoplasts according to the 10x Genomics Single Cell 3' Reagent Kit v2 protocol. For each replicate, 12,200 cells were loaded in the 10x Genomics Chromium single cell microfluidics device with the aim of capturing 7,000 cells. 11 cycles were used for cDNA amplification, as well as for final PCR amplification of the adapter-ligated libraries. Final library size and quality was checked on a DNA High Sensitivity Bioanalyzer chip (Agilent), and libraries were quantified using the NEBNext Library Quantification Kit for Illumina. ScRNA-Seq library sequencing was performed on the NextSeq (Illumina) platform, using the sequencing parameters 26,8,0,98.

Generation and confocal imaging of reporter lines

To verify select marker genes in vivo, *promoter:3xYFP-NLS* reporter lines were generated for the following genes: *AT1G62510, MES15, PME32, TEL1, AT3G22120, ATL75, MLP34, AT1G62500, PIP2-8, AT6G53980, AT3G15357, AT5G62330, AT1G57590, AT1G05320* and EXT*12.* Promoter fragments between approx. 1.2 - 3.5 kb were amplified using PCR primers containing *KpnI* and *XmaI* restriction sites, and introduced by classical cloning into binary vector JM164 (Mathieu *et al*., 2009) to generate transcriptional fusions to a nuclear-localized triple Venus tag. All reporter constructs were transformed into the Col-0 background, and multiple independent events per construct (n ≥3) analyzed. Roots of 7-day-old seedlings were mounted in 10 µg/mL Propidium Iodide (PI) (Sigma-Aldrich) and imaged using a Zeiss LSM880 laser-scanning confocal microscope. Excitation for YFP was at 514 nm and images were acquired at 517 - 571 nm. For PI, the excitation wavelength was 561nm, and images were collected at 589 - 718nm.

Bulk RNA-seq analysis

Sequence reads (pair-end, 75 bp) were trimmed using Trimmomatic (version 0.36; Bolger *et al.,* 2014), and aligned to the *Arabidopsis* TAIR10 reference genome with STAR (Dobin *et al*., 2013). Gene expression values were calculated on uniquely mapped reads using HTSeq (version 0.7.2; Anders *et al*., 2014), and DEseq2 (Love *et al.,* 2014) was used to calculate differentially expression (absolute log2FC ≥1 and q <0.05) on genes with expression levels ≥1 RPM in either replicate. For correlation analysis of gene expression between protoplasted and un-protoplasted root tissues,

the Log2 (mean RPM+1) expression values were calculated for each gene and the Pearson-correlation coefficient determined in R.

Generation of single cell expression matrices

Cell Ranger 2.0.2 (10X Genomics) was used to process scRNA-Seq data. Cell Ranger Count aligned the sequencing reads to the *Arabidopsis* TAIR10 reference genome using STAR (Dobin *et al.*, 2013). For the mapping of *GFP*-derived transcripts, the sequence and gene structure for GFP were added to the reference fastq and gtf files, respectively. Aligned sequence reads with a valid cell barcode and UMI that mapped to exons (Ensembl GTFs TAIR10.37) were used to generate gene expression matrices from which PCR duplicates were removed. Valid cell barcodes were defined based on UMI distribution with the cutoff: cell read count > 5% of 99th percentile of 7000 cells (Zheng *et al.,* 2017). The output files for the two replicates were aggregated into one gene-cell expression matrix using Cell Ranger aggr with the mapped read depth normalization option.

Dimensionality reduction, t-SNE visualization, and cell clustering analysis

The Seurat R package (version 2.3.4) (Sajita *et al.,* 2015; Butler *et al*., 2018) was used for dimensionality reduction analysis. Highly variable genes were identified across the single cells, after controlling for the relationship between average expression and dispersion. Genes were placed into 20 bins based on their average expression, and genes with an average expression value <0.011 removed. Within each bin, a z-score of log transformed dispersion measure (variance/mean) was calculated. A z-score cut-off of 1 was applied to identify the highly variable genes. PCA was then performed using the variable genes as input. 50 PCs were selected as input for a graph-based approach to cluster cells by cell type (Villani *et al.,* 2017) and used as input for t-distributed stochastic neighbor embedding (t-SNE; Van Der Maaten and Hinton, 2008) for reduction to two- or three-dimensional visualization. A resolution value of 0.8 was used in all clustering analyses. Additionally, we used a random forest classifier (Breiman, 2001; Butler *et al,* 2018) to examine cluster distinctness and merged any clusters where the out-of-bag error (OOBE) of the classifier was >10%.

Identification of differentially expressed genes and cluster-specific marker genes

Genes differentially expressed across clusters or subclusters were identified by comparing average expression values in cells of a given cluster to that of cells in all other clusters using the Seurat package likelihood ratio test (Bimod). The following cutoffs were applied: average expression difference ≥ 0.25 natural Log and q <0.01. Cluster-specific marker genes were selected from among the differentially expressed genes based on the criteria that marker genes must be expressed in ≥10% of cells within the cluster (PCT1), and ≤10% of cells across all other clusters (PCT2).

Identification of cluster identities

The top 10 DE genes (q <0.01) by fold change were identified for each cluster and expression profiles harvested from a cell-type specific and longitudinal microarray dataset (Brady *et al.*, 2007a). In the case that microarray data was not available, the next best DE gene was selected. Average normalized expression for 10 DE genes across cell types and developmental stages was calculated and visualized in R. In a complementary approach, marker genes for key cell types were identified from Efroni *et al.* (2015), which integrates root expression data from multiple independent studies. Genes with high normalized expression in a particular cell type (spec. score ≥0.6 as detailed in Efroni *et al.,* 2015) were filtered for specificity by applying a <0.2 spec. score cutoff for all other cell types. This latter criterion was not applied to phloem/phloem companion cells, and phloem/protophloem comparisons, as these cell types show considerable co-expression of most genes. See Supplementary Table S2 for the list of marker genes. Expression of these genes was extracted from the combined single cell expression matrix and visualized using Seurat's SplitDotPlot GG function. Genes with well-defined expression patterns were considered similarly.

Correlation analysis

For correlation analysis of merged single cell and bulk RNA sequence data, Log2 (mean RPM+1) expression values for each gene from two replicates of pooled single cell and bulk RNA sequencing were quantile normalized and the Pearson-correlation coefficient calculated in R. For correlation analysis between single cell RNA-seq

replicates, the single cell data replicates were simulated as bulk RNA sequencing data, the Log2 (mean RPM+1) expression values calculated for each gene, and the Pearson-correlation coefficient between replicates calculated in R. For correlation analysis between the single cell replicates across individual clusters, the average expression of cells within a cluster was calculated for each replicate using the Seurat command AverageExpression(object, use.raw=T). The Pearson-correlation coefficient between the replicates was then determined for each cluster using Seurat CellPlot.

Single cell developmental trajectory analysis

Pseudotime trajectory analysis was performed using the Monocle2 R package (version 2.8.0) algorithm (Trapnell *et al.,* 2014) on genes with a mean expression value ≥ 0.011, and dispersion empirical value larger than the dispersion fit value. Cells were ordered along the trajectory and visualized in a reduced dimensional space. Significantly changed genes along the pseudotime were identified using the differential GeneTest function of Monocle2 with q-value < 0.01. Genes dynamically expressed along the pseudotime were clustered using the 'plot_pseudotime_heatmap' function with the default parameters. Transcript factors were annotated based on information from AtTFDB (https://agris-knowledgebase.org/AtTFDB/). Gene description information was downloaded from (https://www.arabidopsis.org).

GO enrichment analysis

Gene ontology (GO) biological process enrichment analyses (http://pantherdb.org) were performed on cluster-grouped differently expressed genes along the pseudotime (average expression ≥ 0.011) via fisher exact test (q <0.01, Fold enrichment >1).

Gene regulatory network analysis

The gene expression levels of transcription factors (Table S4) without genes with dual-polar expression (cluster h) were normalized using the Monocle2 R package (version 2.8.0) genesmoothcurve function (Trapnell *et al.,* 2014). The pseudotime of each cell assigned by Monocle2 was normalized from 0 to 1. Gene regulatory network inference was calculated on dynamic TFs using SCODE (Matsumoto *et.al*., 2017) with

parameter z setting as 4, averaging 50x results to obtain reliable relationships. Gene regulatory inference was filtered using various cutoffs on parameter value, the results visualized using Cytoscape (Shannon *et al.,* 2003), and the network topological parameters obtained with NetworkAnalyzer (Assenov *et al*., 2008).

Data and software availability

All high-throughput sequencing data, both raw and processed files, have been deposited in NCBI's Gene Expression Omnibus and are accessible upon publication under accession number GSE123818.

Acknowledgements

References

Adrian, J., Chang, J., Ballenger, C., Bargmann, B., Alassimone, J., Davies, K., Lau, O., Matos, J., Hachez, C., Lancot, A., Vatén, A., Birnbaum, K., and Bergmann D. (2015). Transcriptome dynamics of the stomatal lineage: birth, amplification and termination of a self-renewing population. Dev. Cell 33, 107-118.

Anders, S., Pyl, P., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 15, 166-169.

Assenov, Y., Ramírez, F., Schelhorn, S., Lengauger, T., and Albrecht, M. (2008). Computing topological parameters of biological networks. Bioinformatics 24, 282-284.

Balcerowicz, D., Schoenaers, S., and Vissenberg, K. (2015). Cell Fate Determination and the switch from diffuse growth to planar polarity in Arabidopsis root epidermal cells. Front. Plant Sci. 6, doi: 10.3389/fpls.2015.01163.

Barolo, S., and Posakony, J. (2002). Three habits of highly effective signaling pathways: principles of transcriptional control by developmental cell signaling. Genes and Dev. 16, 1167-1181.

Barrero, J., Millar, A., Griffiths, J., Czechowski, T., Scheible, W., Udvardi, M., Reid, J., Ross, J., Jacobsen, J., and Gubler, F. (2010). Gene expression profiling identifies two regulatory genes controlling dormancy and ABA sensitivity in *Arabidopsis* seeds. Plant Journal 61, 611-622.

Birnbaum, K., Shasha, D., Wang, J., Jung, J., Lambert, G., Gailbraith, D., and Benfey, P. (2003). A gene expression map of the *Arabidopsis* root. Science 302, 1956-1960.

Bolger, A., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 1, 2114-2120.

Bonke, M., Thitamadee, S., Mähönen, A., Hauser, M., and Helariutta, Y. (2003). APL regulates vascular tissue identity in *Arabidopsis*. Nature 13, 181-186.

Brady, S., Orlando, D., Lee, J., Wang, J., Koch, J., Dinney, J., Mace, D., Ohler, U., and Benfey P. (2007a). A high-resolution root spatiotemporal map reveals dominant expression patterns. Science 2, 801-806.

Brady, S., Song, S., Dhugga, K., Rafalski, A., and Benfey, P. (2007b). Combining expression and comparative evolutionary analysis. The COBRA gene family. Plant Phys. 143, 172-187.

Breiman, L. (2001). Random forests. Mach. Lear. 45, 5-32.

Bruex, A., Kainkaryam, R., Wieckowski, Y., Kang, Y., Bernhardt, C., Xia, Y., Zheng, X., Wang, J., Lee, M., Benfey, P., Woolf, P., and Schiefelbein, J. (2012). A gene regulatory network for root epidermis cell differentiation in *Arabidopsis*. PLoS Genet. 8, doi: 10.1126/science.aah4573

Butler, A., Hoffman, P., Smibert, P., Papalexi, E. and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. 36, 411-420.

Dello Ioio, R., Nakamura, K., Moubayidin, L., Perilli, S., Taniguchi, M., Morita, M., Aoyama, T., Costantino, P., and Sabatini, S. (2008). A genetic framework for the control of cell division and differentiation in the root meristem. Science 322, 1380-1384.

Diet, A., Link, B., Seifert, G., Schellenberg, B., Wagner, U., Pauly, M., Reiter, W-D., and Ringli, C. (2006). The *Arabidopsis* root hair cell wall formation mutant *lrx1* is suppressed by mutations in the *RHM1*gGene encoding a UDP-L-Rhamnose synthase. Plant Cell 18, 1630-1641.

Dobin, A., Davis, C., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 1, 15-21.

Efroni, I., Ip, P-L., Nawy, T., Mello, A., and Birnbaum, K. (2015). Quantification of cell identity from single-cell gene expression profiles. Genome Biol. 16, doi: 10.1186/s13059-015-0580-x.

Feraru, E., and Friml, J. (2008). PIN polar targeting. Plant Phys. 147, 1553-1559.

Forzani, C., Aichinger, E., Sornay, E., Willemsen, V., Laux, T., Dewitte, W., and Murray, A. (2014). WOX5 suppresses *CYCLIN D* activity to establish quiescence at the center of the root stem cell niche. Curr. Biol. 24, 1939-1944.

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.

Gattolin, S., Sorieul, M., Hunter, P., Khonsari, R., and Frigerio, L. (2009). In vivo imaging of the tonoplast intrinsic protein family in *Arabidopsis* roots. BMC Plant Biol. 9, doi:10.1186/1471-2229-9-133

Helariutta, Y., Fukaki, H., Wysocka-Diller, J., Nakajima, K., Jung, J., Sena, G., Hauser, M-T., and Benfey, P. (2000). The SHORT-ROOT gene controls radial patterning of the *Arabidopsis* root through radial signalling. Cell 101, 555-567.

Klein, M., Reichelt, M., Gershenzon, J., and Papenbrock, J. (2006). The three desulfoglucosinolate sulfotransferase proteins in *Arabidopsis* have different substrate specificities and are differentially expressed. FEBS J. 273, 122-136.

Klein, A., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D., and Kirschner, M. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell 161, 1187-1201.

Lauber, M. Waizenegger, I., Steinmann, T., Schwarz, H., Mayer, U., Hwang, I., Lukowitz, W., and Jürgens, G. (1997). The *Arabidopsis* KNOLLE protein is a cytokinesis-specific syntaxin. Cell Biol. 139, 1485-1493.

Li, S., Yamada, M., Han, X., Ohler, U., and Benfey, P. (2016). High-resolution expression map of the *Arabidopsis* root reveals alternative splicing and lincRNA regulation. Dev. Cell 39, 508-522.

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

Lu, K-J., De Rybel, B., van Mourik, H., and Weijers, D. (2018). Regulation of intercellular TARGET OF MONOPTEROS 7 protein transport in the Arabidopsis root. Development 145, doi:10.1242/dev.152892.

Lukassen, S., Bosch, E., Ekici, A., and Winterpacht, A. (2018). Characterization of germ cell differentiation in the male mouse through single-cell RNA sequencing. Scientific Reports 8, doi:10.1038/s41598-018-24725-0.

van der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. Machine Learning Research 9, 2579-2605.

Macosko, E., Basu, A., Satjia, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A., Kamitaki, N., Martersteck, E., Trombetta, J., Weitz, D., Sanes, J., Shalek, A., Regev, A., and McCarroll, S. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell 161, 1202-1214.

Malinovsky, F., Thomsen, M., Nintemann, S., Jagd, L., Bourgine, B., Burrow, M., and Kliebenstein, D. (2017). An evolutionarily young defense metabolite influences the root growth of plants via the ancient TOR signaling pathway. eLife 6. e29353.

Mathieu, J., Yant, L.J., Mürdter, F., Küttner, F., and Schmid, M. (2009). Repression of flowering by the miR172 target *SMZ*. PLoS biology 7, e1000148.

Matsumoto, H., Kiryy, H., Furusawa, C., Ko, M., Ko, S., Gouda, N., Hayashi, T., and Nikaido, I. (2017). SCODE: an efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation. Bioinformatics 33, 2314-2321.

McDavid, A., Finak, G., Chattopadyay, P., Dominguez, M., Lamoreaux, L., Ma, S., Roederer, M., and Gottardo, R. (2013). Data exploration, quality control and testing in single-cell qPCR-based gene expression experiments. Bioinformatics 29, 461-467.

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.

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

Nicolas, M., and Cubas, P. (2016). TCP factors: new kids on the signaling block. Curr. Opin. Plant Biol. 33, 33-41.

Niwa, H. (2018). The principles that govern transcription factor network functions in stem cells. Development 145, doi:10.1242/dev.157420.

Pi, L., Aichinger, E., van der Graaff, E., Llavata-Peris, C., Weijers, D., Hennig, L., Groot, E., and Laux, T. (2015). Organizer-derived WOX5 signal maintains root columella stem cells through chromatin-mediated repression of *CDF4* expression. Dev. Cell 33, 576-588.

Porco, S., Larrieu, A., Du, Y., Gaudinier, A., Goh, T., Swarup, K., Swarup, R., Kuempers, B., Bishopp, A., Lavenus, J., Casimiro, I., Hill, K., Benkova, E., Fukaki, H., Brady, S., Scheres, B., Péret, B., and Bennett, M. (2016). Lateral root emergence in *Arabidopsis* is dependent on transcription factor LBD29 regulation of auxin influx carrier LAX3. Development 143, 3340-3349.

Potter, S. (2018). Single-cell RNA sequencing for the study of development, physiology and disease. Nat. Rev. Nephrology 14, 479-492.

Prakadan, S., Shalek, A., and Weitz, D. (2017). Scaling by shrinking: empowering single-cell 'omics' with microfluidic devices. Nat. Rev. Genet. 18, 345-361.

Reiter, F., Wienerroither, S., and Stark, A. (2017). Combinatorial function of transcription factors and cofactors. Curr. Opin. Genet. Dev. 43, 73-81.

Sabatini, S., Beis, D., Wolkenfelt, H., Murfett, J., Guilfoyle, T., Malamy, J., Benfey, P., Leyser, O., Bechtold, N., Weisbeek, P., and Scheres, B. (1999). An auxin-dependent distal organizer of pattern and polarity in the Arabidopsis root. Cell 99, 463-472.

Satija, R., Farrell, J., Gennert, D., Schier, A., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol, 33, 495-502.

Schellmann, S., Schnittget, A., Kirik, V., Wada, T., Okada, K., Beermann, A., Thumfahrt, J., Juergens, G., and Huelskamp, M. (2002). TRIPTYCHON and CAPRICE mediate lateral inhibition during trichome and root hair patterning in *Arabidopsis*. EMBO 21, 5036-5046.

Shimotohno, A., Heidstra, R., Blilou, I., and Scheres, B. (2018). Root stem cell niche organizer specification by molecular convergence of PLETHORA and SCARECROW transcription factor modules. Genes & Dev. 32, 1085-1100.

Skopelitis, D., Hill, K., Klesen, S., Marco, C., von Born, P., Chitwood, D., and Timmermans, M. (2018). Gating of miRNA movement at defined cell-cell interfaces governs their impact as positional signals. Nat. Commun. 9, doi: 10.1038/s41467-018-05571-0.

Shannon, P., Markiel, A., Ozier, O., Baliga, N., Wang, J., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 12, 2498-2505.

Sparks, E., Drapek, C., Gaudinier, A., Li, S., Ansariola, M., Shen, N., Hennacy, J., Zhang, J., Turco, G., Petricka, J., Foret, J., Hartemink, A., Gordan, R., Megraw, M.,

Brady, S., and Benfey, P. (2016). Establishment of expression in the SHORTROOT-SCARECROW transcriptional cascade through opposing activities of both activators and repressors. Dev. Cell 39, 585-596.

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

Villani, A., Satija, R., Reynolds, G., Sarkizova, S., Shekhar, K., Fletcher, J., Griesbeck, M., Butler, A., Zheng, S., Lazo, S., Jardine, L., Dixon, D., Stephenson, E., Nilsson, E., Grundberg, I., McDonald, D., Filby, A., Li, W., De Jager, P., Rozenblatt-Rosen, O., Lane, A., Haniffa, M., Regev, A., and Hacohen, N. (2017). Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science 356, 4573.

Zheng, G., Terry, J., Belgrader, P., Ryvkin, P., Bent, Z., Wilson, R., Ziraldo, S., Wheeler, T., McDermott, G., Zhu, J., Gregory, M., Shuga, J., Montesclaros, L., Underwood, J., Masquelier, D., Nishimura, S., Schnall-Levin, M., Wyatt, P., Hindson, C., Bharadwaj, R., Wong, A., Ness, K., Beppu, L., Deeg, H., McFarland, C., Loeb, K., Valente, W., Ericson, N., Stevens, E., Radich, J., Mikkelsen, T., Hindson, B., and Bielas, J. (2017). Massively parallel digital transcriptional profiling of single cells. Nat. Comm. 8, 14049.

Supplementary data
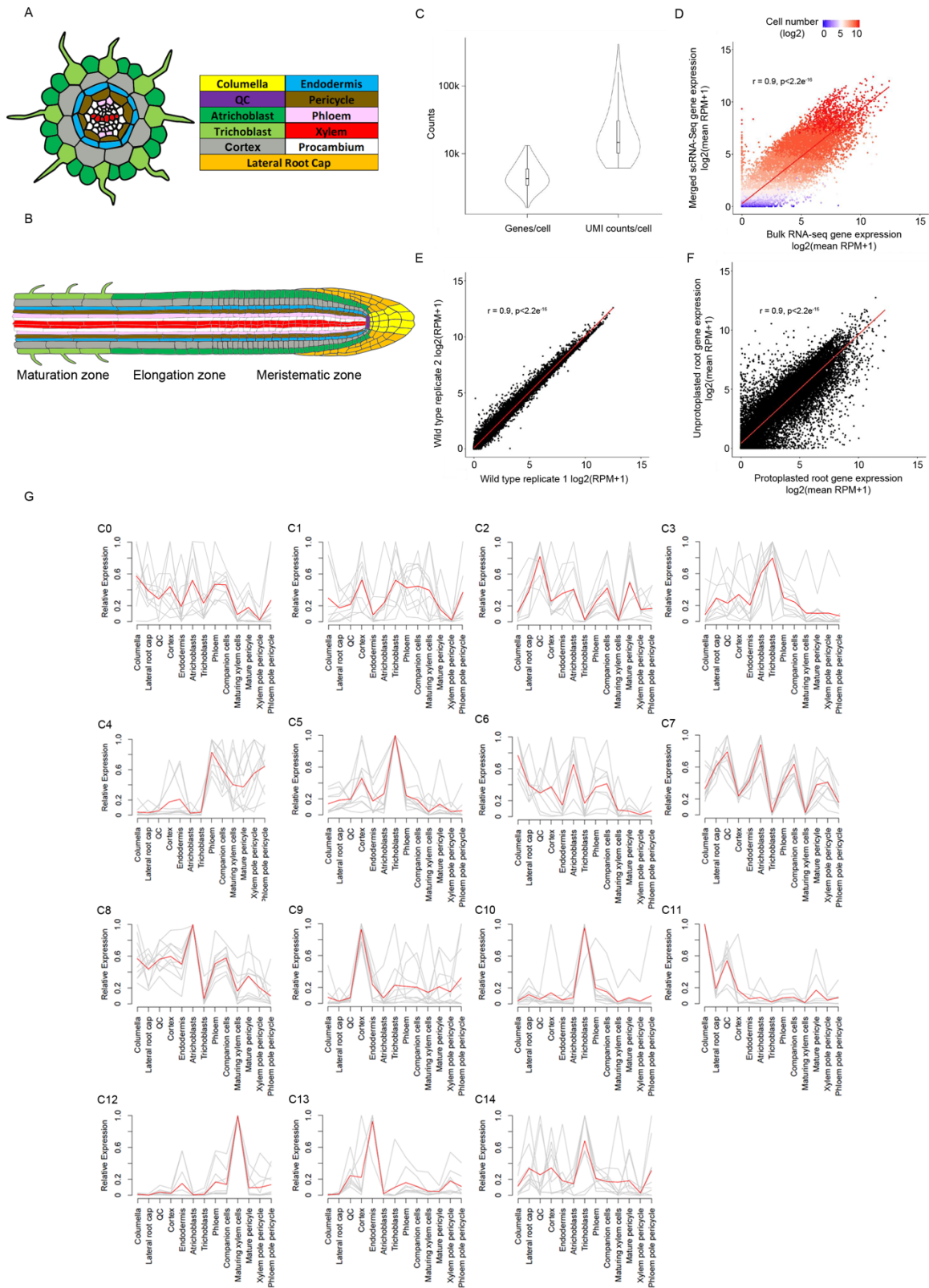
Figure S1. ScRNA-Seq of the *Arabidopsis* root reveals distinct cell-type specific clusters; Related to Figure 1. (A) Transverse section of a mature root illustrating its radial

organisation. (B) Longitudinal cross-section of a mature root illustrating an organisation in three distinct developmental zones. Figures adapted from Root illustrations – figshare collection (2018). (C) Violin plots showing the distribution of genes and transcripts (UMIs) detected per cell. (D) Correlation between merged single cell and bulk-tissue RNA-seq measurement of gene expression from *Arabidopsis* root protoplasts. For each gene, the quartile normalisation of the log2-transformed mean RPM + 1 values from the merged scRNA-Seq and bulk RNA-seq are plotted against each other. Colour represents cell numbers (log2, +1) from the scRNA-Seq data. (E) Correlation between single cell RNA-seq replicates of *Arabidopsis* Col-0 (Rep. 1) and *pMIR166A:erGFP* (Rep 2) root protoplasts. For each gene, log2-transformed RPM +1 values are plotted against each other. (F) Correlation between bulk RNA-seq measurements of gene expression from protoplasted and un-protoplasted *Arabidopsis* root tissue. r = Pearson's correlation coefficient. 3,545 genes were induced upon protoplasting (see Table S1). (G) The expression profiles for the top 10 DE genes defining clusters, taken from a microarray root atlas (Brady *et al*., 2007a), reveal discreet cell identities. Red line, mean expression profile; grey lines, individual expression profiles.
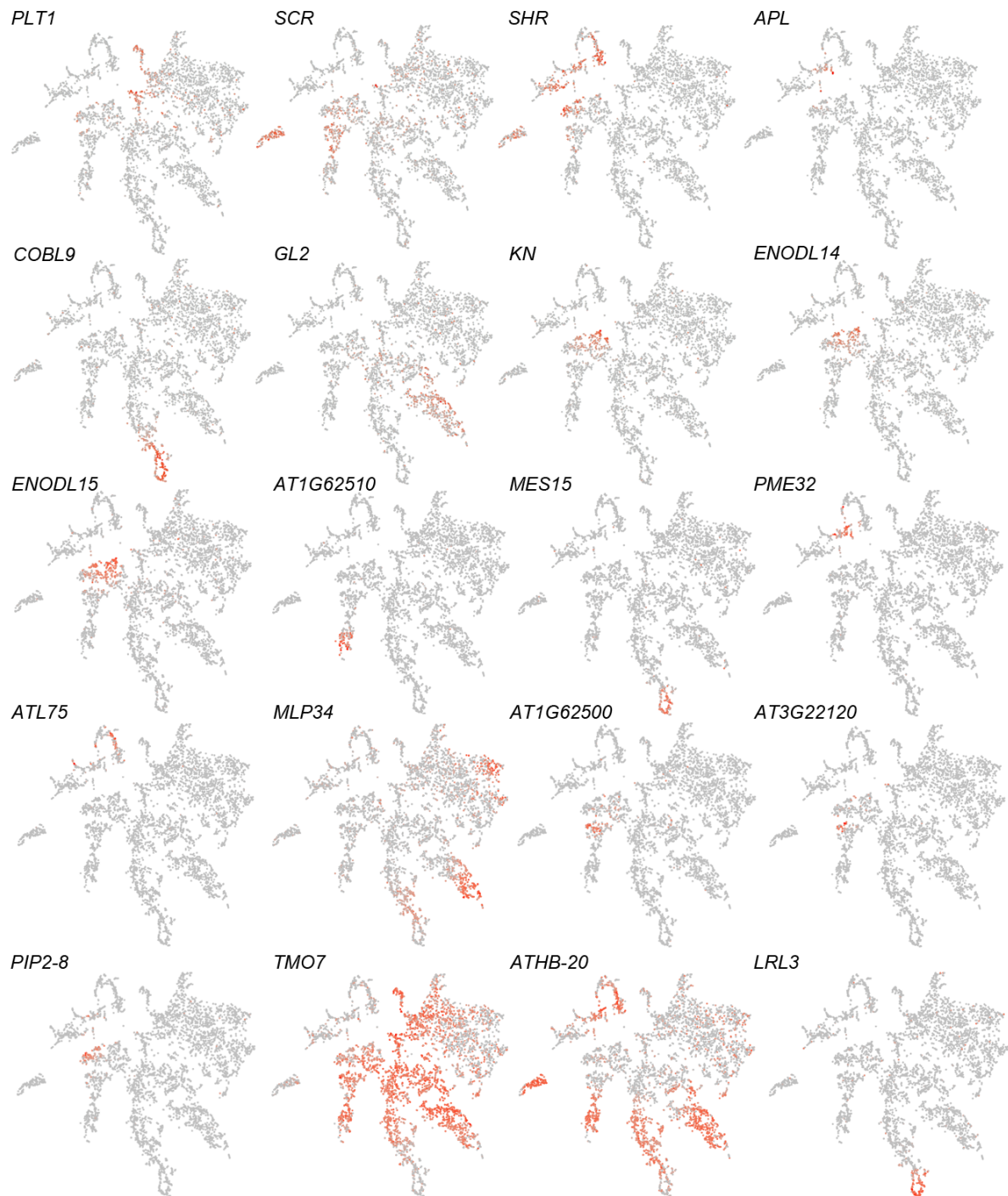
Figure S2. Key *Arabidopsis* genes reveal distinct expression profiles across clusters. Related to Figure 1. t-SNE visualizations of cell clusters revealing defined expression profiles of key *Arabidopsis* development genes.
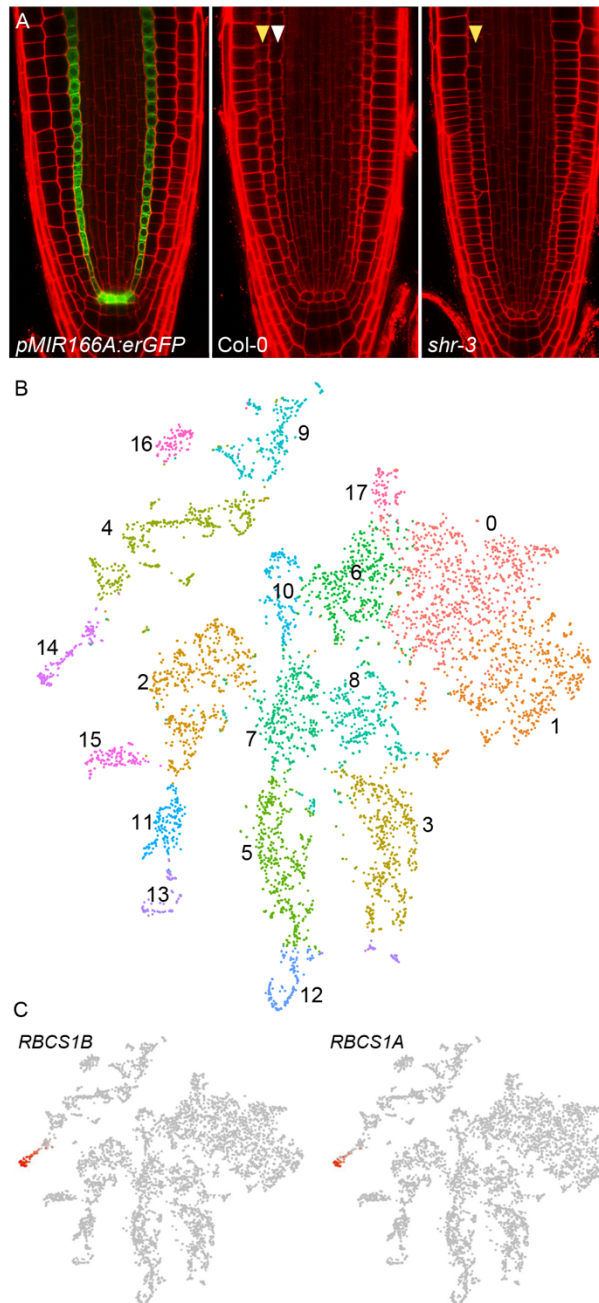
Figure S3. A localised GFP marker and mutant line validate cluster calling; Related to Figure 2.
(A) Left - *pMIR166A:erGFP* expression is limited to the endodermal cell layer and the QC cells. Centre and right - wild-type (Col-0) and *SHR* mutant (*shr-3*). Yellow arrowheads indicate cortex cell-layers, white arrowhead indicates endodermal cell layer in Col-0, missing in *shr-3*. 6-day old root tips. (B) t-SNE visualisation (right) of wild-type and *shr-3* cells combined reveals 18 distinct clusters. (C) t-SNE visualisation of the cluster cloud (wild-type replicates) reveals expression of select rubisco subunit genes localised to a leaf-cell cluster.
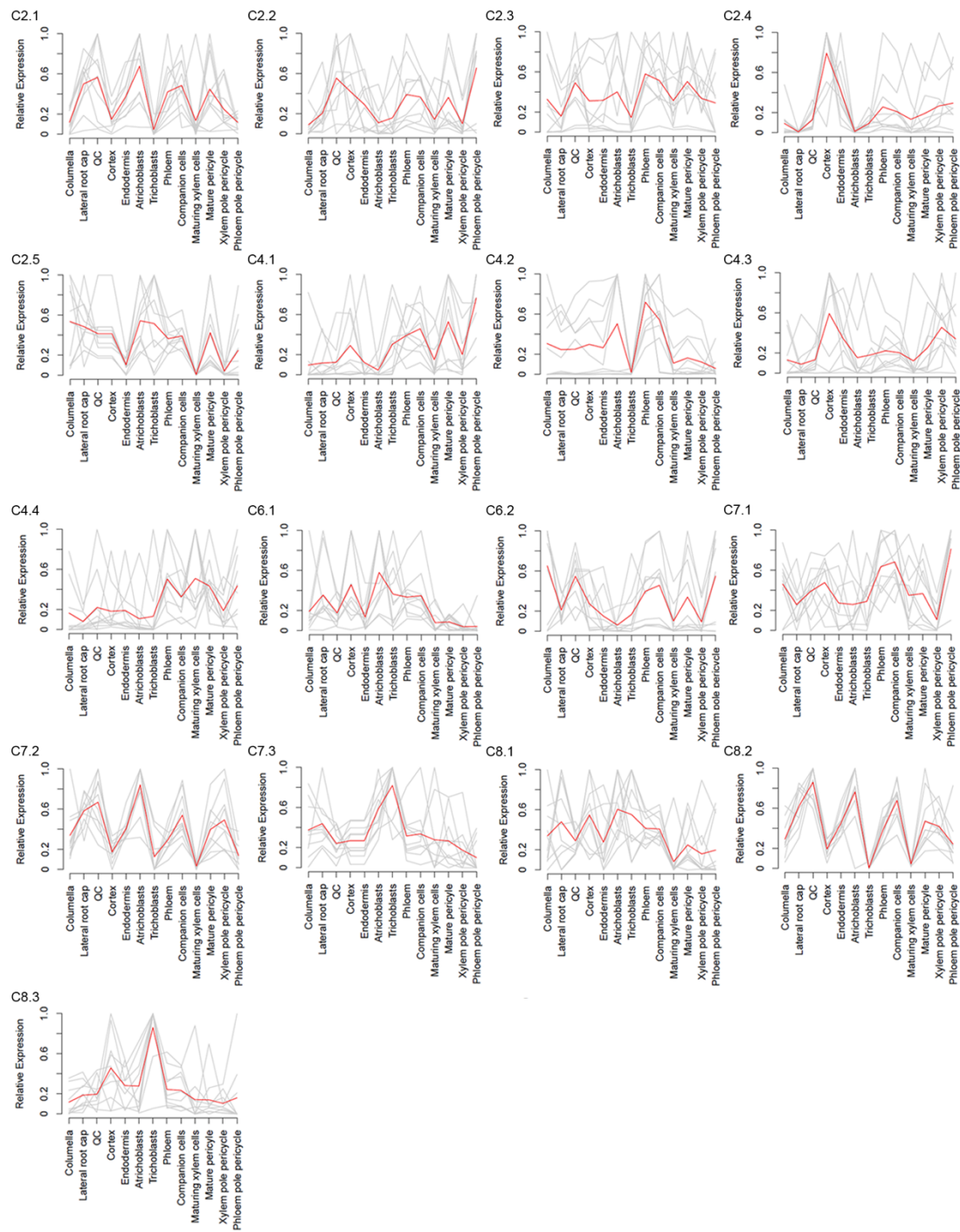
Figure S4. Discreet cell identities can be found within select subclusters; Related to Figure 1. The expression profiles for the top ten DE genes defining subclusters, taken from a microarray root atlas (Brady *et al.,* 2007a), reveal discreet cell identities. Red line, mean expression profile; grey lines, individual expression profiles.
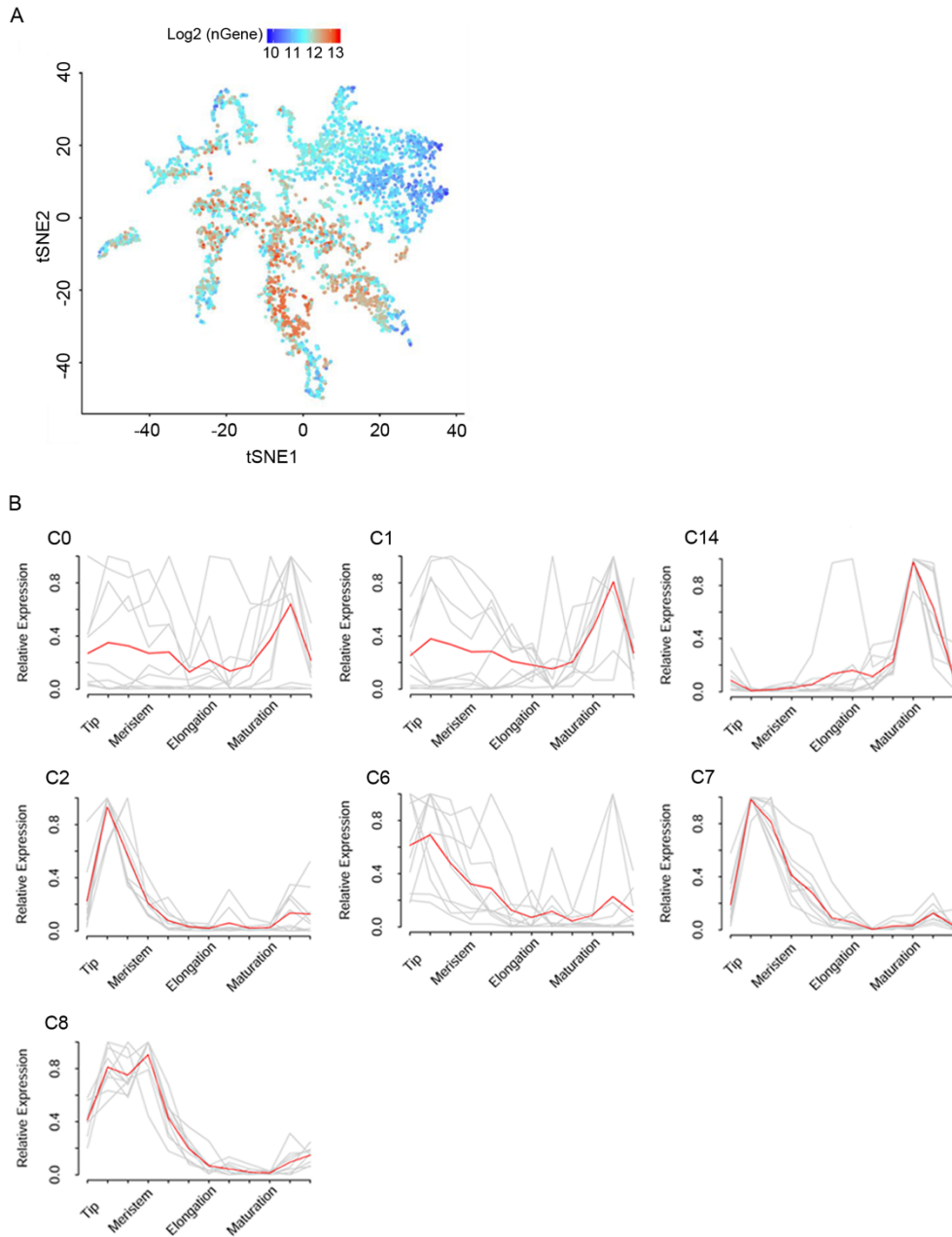
Figure S5. Developmental stage-specific clusters can be identified. Related to Figure 1. (A) t-SNE visualization of UMI counts across all clusters. (B) The expression profiles for the top 10 DE genes defining subclusters, taken from a longitudinal microarray root atlas (Brady *et al.,* 2007a), reveal meristematic- and differentiated-cell identities. Red line, mean expression profile; grey lines, individual expression profiles.
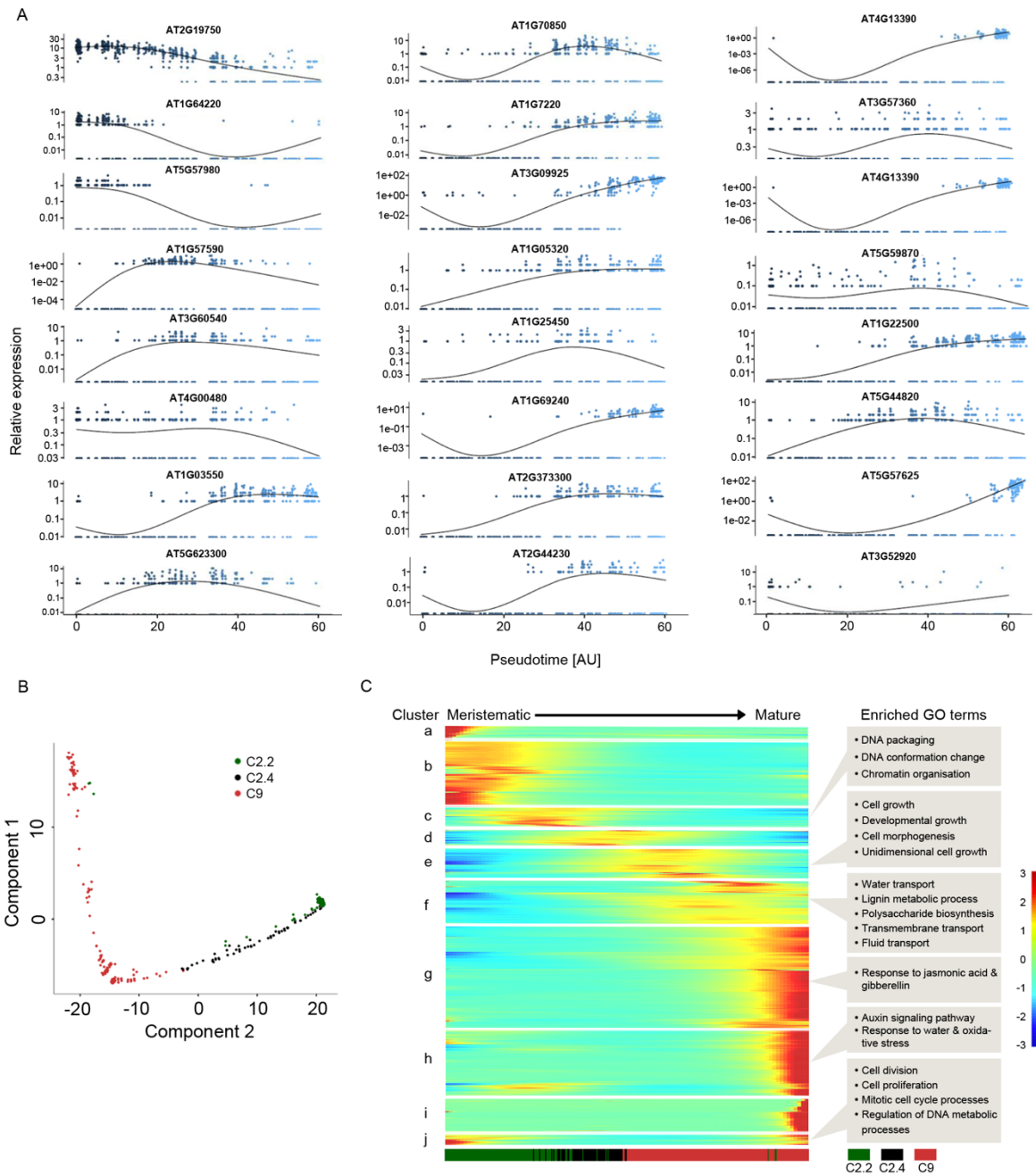
Figure S6. Cortex development is guided by distinct waves of gene expression; Related to Figure 5. (A) Expression dynamics for select single genes across pseudotime. Blue-scale, pseudotime value. (B) Pseudotime reconstruction of cortex development reveals a linear ordering of cells, reflecting cluster and subcluster arrangement. (C). Expression heatmap of highly dynamic genes ordered across pseudotime reveals cortex differentiation reflected in multiple distinct waves of gene expression. Significantly enriched GO terms for clusters are labelled. Lower bar, cell density distribution across clusters. See Table S4 for full data.

Table S1 to 5 (separate file)

Movie S1. ScRNA-Seq of the *Arabidopsis* root reveals distinct clusters; Related to Figure 1.

Data and software availability

All high-throughput sequencing data, both raw and processed files, have been deposited in NCBI's Gene Expression Omnibus and are accessible under accession number GEO: GSE123818.

Appendix II: A high-resolution gene expression atlas links dedicated meristem genes to key architectural traits

Steffen Knauer[1,2,10], Marie Javelle[2,8,10] , Lin Li[3], Xianran Li[4], Xiaoli Ma[1], Kokulapalan Wimalanathan[5,9], Sunita Kumari[2], Robyn Johnston[6], Samuel Leiboff[6], Robert Meeley[7], Patrick S. Schnable[4], Doreen Ware[2], Carolyn Lawrence-Dill[4,5], Jianming Yu[4], Gary J. Muehlbauer[3], Michael J. Scanlon[6] and Marja C.P. Timmermans[1,2]*

[1] Center for Plant Molecular Biology, University of Tuebingen, 72076 Tuebingen, Germany

[2] Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA

[3] Department of Agronomy and Plant Genetics, University of Minnesota, Saint Paul, MN 55108, USA

[4] Department of Agronomy, Iowa State University, Ames, IA 50011, USA

[5] Interdepartmental Bioinformatics and Computational Biology Program, Iowa State University, Ames, IA 50011, USA

[6] Plant Biology Section, School of Intergrated Plant Science, Cornell University, Ithaca, NY 14853, USA

[7] DuPont Pioneer, Agricultural Biotechnology, Johnston, Iowa 50131, USA

[8] Current address: Biogemma, CRC, route d'Ennezat, 63720 Chappes, France

[9] Current address: Sainsbury Laboratory, Norwich Research Park, Norwich NR4 7UH, UK

[10] These Authors contributed equally to this work

* Corresponding author: marja.timmermans@zmbp.uni-tuebingen.de

Abstract

The shoot apical meristem (SAM) orchestrates the balance between stem cell proliferation and organ initiation essential for post-embryonic shoot growth. Meristems show a striking diversity in shape and size. How this morphological diversity relates to variation in plant architecture, and the molecular circuitries driving it are unclear. By generating a high-resolution gene expression atlas of the vegetative maize shoot apex, we show here that distinct sets of genes govern the regulation and identity of stem cells in maize versus *Arabidopsis.* Cell identities in the maize SAM reflect the combinatorial activity of transcription factors (TFs) that drive the preferential, differential expression of individual members within gene families functioning in a plethora of cellular processes. Sub-functionalization thus emerges as a fundamental feature underlying cell identity. Moreover, we show that adult plant characters are to a remarkable degree regulated by gene circuitries acting in the SAM, with natural variation modulating agronomically-important architectural traits enriched specifically near dynamically expressed SAM genes, and the TFs that regulate them. Besides unique mechanisms of maize stem cell regulation, our atlas thus identifies key new targets for crop improvement.

Introduction

The SAM positioned at the plant's growing shoot tip harbors a population of pluripotent stem cells, which serve as a persistent source of cells for postembryonic growth and organogenesis. A striking aspect of meristems is the tremendous diversity in morphology seen across plant species (Steeves and Sussex 1989). How this diversity relates to variation in overall plant architecture is unclear. SAM morphology does not seem to follow phylogeny (Steeves and Sussex 1989). This implies that the architectural diversity of the angiosperms is elaborated post-meristematically, and that the main function of the SAM is to balance stem cell proliferation with organogenesis. Contrary to this concept, quantitative variation in SAM structure in maize is correlated with adult morphological traits such as node number and flowering time (Leiboff et al. 2015). This suggests that variations in adult plant architecture may be determined in part by regulatory mechanisms acting in the SAM, and that such regulatory networks form targets for selection in the improvement of agronomically-important traits.

The nature of these regulatory networks remains unclear. Much of our understanding of SAM function originates from studies in *Arabidopsis*. These illustrate that gene expression within the growing meristem is precisely coordinated in a highly spatial and temporal manner. Mobile signals, mechanical inputs, and environmental cues, all provide positional information to specify cell fates within the dynamic stem cell niche (Besnard et al. 2011; Pfeiffer et al. 2017). These inputs in part converge onto a negative feedback loop involving *WUSCHEL* (*WUS*) and *CLAVATA* (*CLV*) signaling that maintains stem cell number in the central zone (CZ) at the SAM tip. Additionally, these inputs, through their effects on auxin polar transport and signaling, link proliferation in the meristem to organ initiation in the peripheral zone (PZ) (Besnard et al. 2011; Pfeiffer et al. 2017).

Many of the recognized regulators of meristem function predate the origin of the angiosperms, and their functions at the molecular level are conserved (Plackett et al. 2015). However, whereas the roles for meristem regulators such as *WUS* and *CLV1/3* appear conserved across eudicots, substantial diversification in these regulatory pathways between monocot and eudicot lineages has been noted (Nardmann and Werr 2007). This raises the question as to whether diversity in SAM morphology is reflected at the level of molecular circuitries.

To address this question and assess a link between SAM function and adult plant architecture, we generated a high-resolution gene expression atlas of the vegetative maize shoot apex. This shows that the molecular circuitries underlying distinct cell fates within the SAM are largely divergent from *Arabidopsis*, illustrating fundamental differences in stem cell regulation in both species. Cell identities in the maize SAM reflect the combinatorial activity of TFs that show subtle quantitative expression differences across the apex. These drive the preferential, differential expression of individual members within gene families functioning in a plethora of cellular processes. Allelic diversity present near these dynamically expressed genes, and the TFs that regulate them, links the molecular circuitries acting in the SAM to post-meristematic morphological traits. Besides unique mechanisms of maize stem cell regulation, the atlas thus identifies novel targets for selection in the improvement of agronomically-important traits.

Results

Tissue-specific genes contributing to cell identity

To identify gene expression signatures associated with meristem function in maize, we used laser microdissection to isolate cells from the following distinct structural and functional domains within the shoot apex of 14 day-old B73 seedlings: the entire meristem, the stem cell comprising meristem tip (hereafter referred to as Tip), the incipient leaf (P0) at the meristem periphery, the L1 and L2 lineage layers overlaying these meristem regions, developing leaf primordia at P1, P2 and P3, as well as the internode primordium and vasculature (Figure 1A–C; Figure S1). Collectively, transcripts of 19,278 genes, or about half of the 39,656 annotated maize genes, are detectable at levels ≥ 2 RPM in at least one of the 10 domains sampled (Table S2). The number of genes expressed in discrete shoot apical domains varies little, and also their genome-wide expression profiles are highly correlated (Figure 1D, E). This indicates that the differential expression of a relatively small subset of genes underlies the specialized functions of distinct domains within the shoot apex.
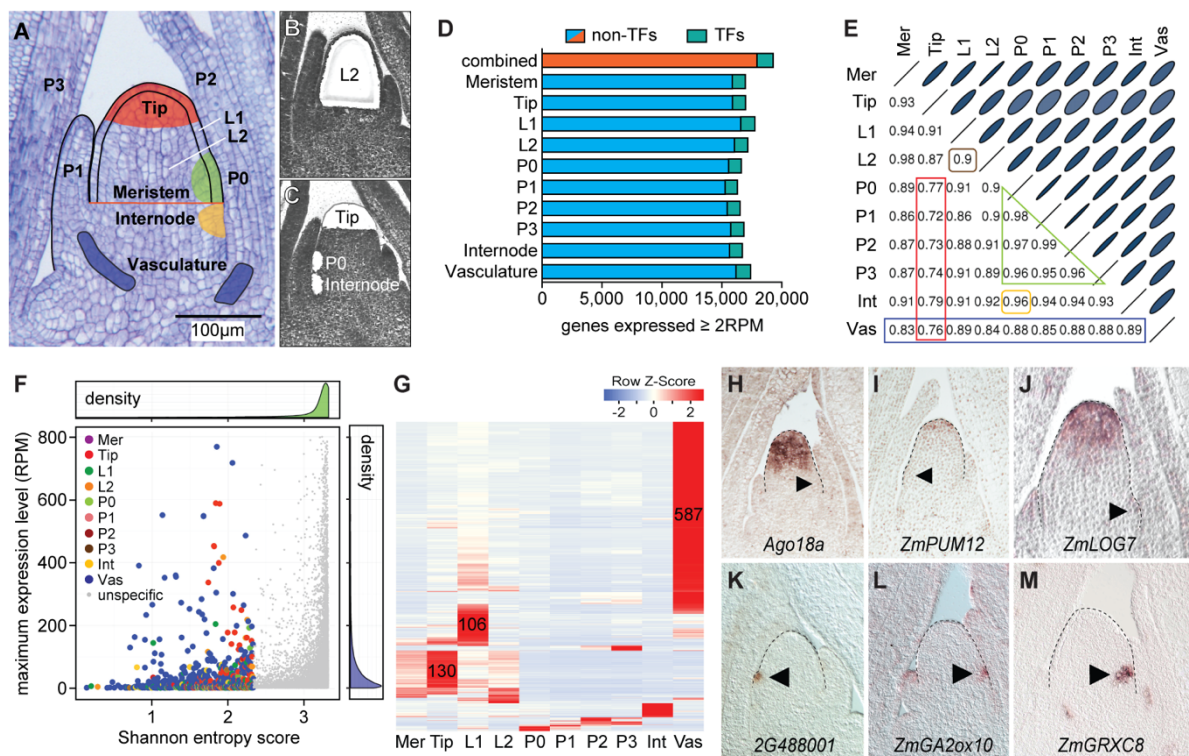
Figure 1. Tissue-specific genes contributing to cell identity. (*A*) Longitudinal section of a 14 day-old B73 seedling apex. The 10 domains/tissues captured are illustrated. (*B,C*) Sections after laser microdissection of the L2 (*B*), or Tip, P0, and internode (*C*). (*D*) The number of genes expressed ≥2 RPM varies only slightly across tissues. (*E*) Correlation analysis identifies the Tip as the most distinctive cell type (red), followed by the vasculature (blue). Overall

expression in the two clonal layers is highly correlated (brown). Expression profiles of leaf primordia of successive plastochron (P) stages are also highly correlated (green), and closely match expression in the internode (yellow). Mer, meristem; Int, internode; Vas, vasculature. (*F*) Density plots show that most genes have high SE scores and low expression values. Genes with SE score < 2.33 (colored) show tissue-specific expression. For visual simplification, only genes with expression values ≤ 800 RPM are shown. (*G*) Heatmap of cell type specific genes shows most mark the vasculature, Tip, or L1. Primordium stage-specific transcripts were also identified. (*H-M*) *In situ* hybridization verifying specificity for select Tip- (*H-J*) and P0-specific genes (*K-M*). Meristem shape is outlined. Arrowheads, P0.

To identify such genes, we first used Shannon entropy (SE) (Schug et al. 2005) to define genes showing domain-specific patterns of expression. Of the 964 genes with a SE score < 2.33, which considering the partial overlap of the regions captured were defined as domain specific, most mark the vasculature (587), Tip (130), or L1 (106) (Figure 1F, G; Table S3). Particularly the Tip-specific genes are noteworthy (Figure 1F–J), as genes underlying maize stem cell identity have remained elusive. Additionally, genes specific to individual leaf stages, including the P0, had not been noted previously (Figure 1K–M). These gene sets provide a powerful resource to infer tissue-specific enhancer elements, to modulate spatial patterns of gene expression, or to assign spatiotemporal origins to single cell transcriptomic data (Denyer et al. 2019).

Genes with signaling-associated functions are overrepresented among the domain-specific genes, pointing to particular signaling pathways underlying cell fate decisions (Table S4). For example, key cytokinin-synthesis genes show a Tip-specific pattern of expression, whereas gibberellic acid (GA) associated genes are among the primordium-specific genes, and genes involved in auxin, abscisic acid (ABA), jasmonic acid (JA) and ethylene signaling predominate the vasculature (Tabs S3, S4). Such an organization of hormone activities will increase the possibilities for specific spatial interactions needed to coordinate the many cell fate decisions within the growing shoot. In addition, of the 48 genes encoding CLE signaling peptides (Goad et al. 2017), 17 are expressed in the apex, of which 4 show specificity for the vasculature, internode, or Tip (Figure S2A). The peptide derived from the single Tip-specific *CLE* gene is orthologous to rice *FCP1* (Je et al. 2016). Although previously reported to be

primordium-derived (Je et al. 2016), *ZmFCP1* is in fact specific to the Tip (Figure S2A,B), and could act through a *CLE-WOX* regulatory module within the SAM itself.

Besides genes connected to cell-cell signaling, transcripts for a surprisingly large number of TFs accumulate in a tissue- or domain-specific manner. Whereas 7% of all expressed genes encode TFs (Figure 1D), consistent with a role in driving the primary molecular changes underlying cell identity, 117 of the 964 domain-specific genes represent TFs (>12%, Table S3). Expression of all 12 tissue-specific Dof TFs is limited to the vasculature, suggestive of a role as master regulators of vascular development. However, most TF families show some evidence of sub-functionalization with individual members showing specificity for distinct cell or tissue types. Conversely, each tissue expresses a diverse set of TFs, pointing to combinatorial inputs from TFs on cell identity.

Dynamic expression of individual gene family members defines cell identity

While SE identifies genes with a near on/off state in expression, more genes likely contribute in a quantitative manner to distinguish cell identities. We therefore next analyzed differentially expressed genes (DEGs) between the vasculature, Tip, and P0, the latter as a representative of the closely related leaf primordia (Figure 2A; Table S5). Genes preferentially expressed in the vasculature largely overlap with the vascular-specific genes, and show an overrepresentation for genes involved in signaling as well as cell wall homeostasis (Table S6). In the pairwise comparison between the Tip and P0, genes preferentially expressed in the P0 are enriched for functions connected to actively dividing cells and auxin signaling (Figure 2B; Table S6), known features of primordium initiation. Additionally, consistent with TFs driving developmental programs, differences are seen for select TF families. For instance, the NAC, GeBP, and ABI3/VP1 families are overrepresented among Tip-enriched genes, whereas YABBY TFs mark the P0 (Figure 2B).

Perhaps the most striking feature stemming from this analysis, however, is the fact that few functional categories or pathways are enriched among the DEGs (Figure 2B; Table S6). DEGs from all three pair-wise comparisons are annotated to function in numerous, widely diverse metabolic and cellular processes. Although these processes are typically represented by multi-gene families, single or highly select subsets of members within these families show a differential or tissue-specific pattern of expression (Figure 2C–F). Besides the *CLE* example mentioned above, individual *LONELY GUY* (*LOG*), *ISOPENTENYL TRANSFERASE* (*IPT*), *PIN-FORMED* (*PIN*), and *YUCCA* genes involved in hormone metabolism/signaling are DE across these tissues (Figure 2C, D; Figure S2C). Likewise, individual family members for cell wall modifying enzymes and redox regulation (e.g., expansins, pectin methylesterases, and glutaredoxins) known to act downstream of TFs and hormone signaling in meristem homeostasis and organogenesis (Schippers et al. 2016; Tognetti et al. 2017), are DE (Figure 2E; Figure S2D,E). Key features of cell identity are thus regulated across tissues by a comparatively small but functionally highly diverse set of DEGs.
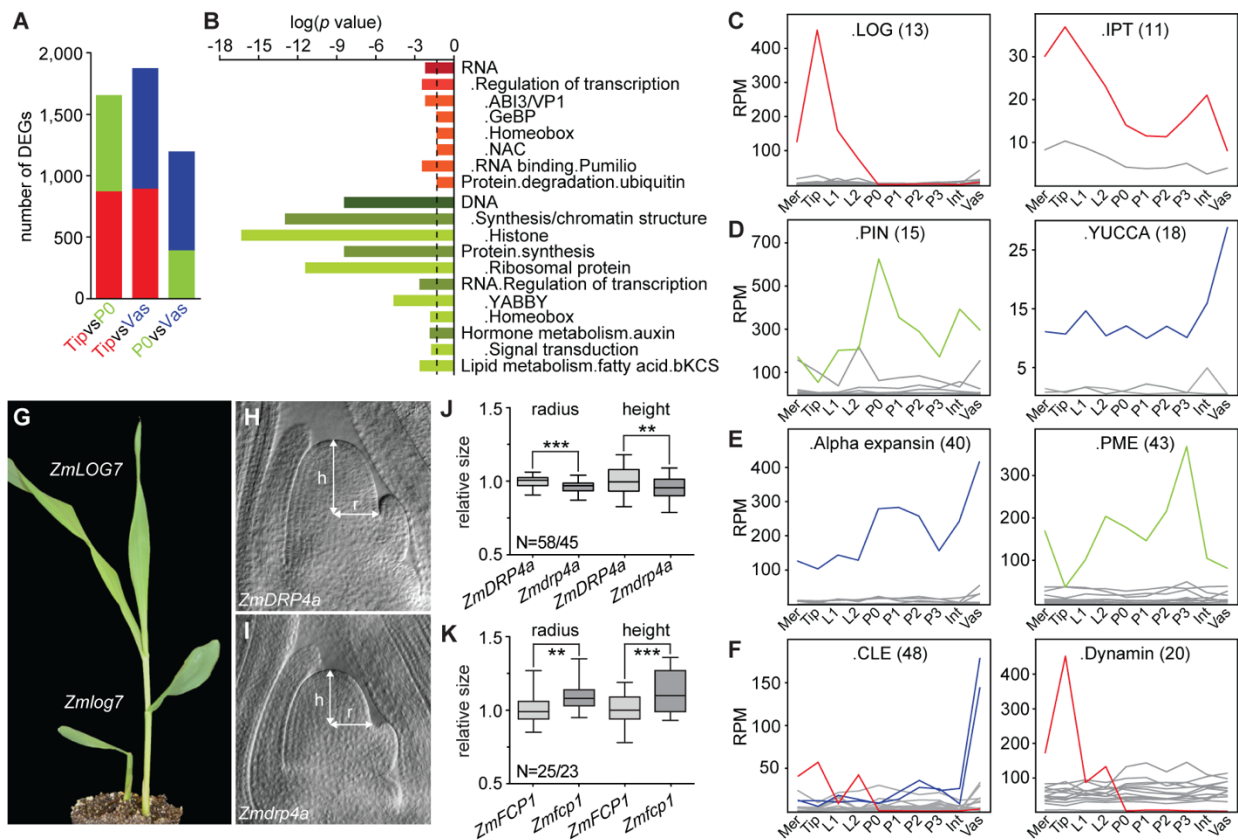
Figure 2. Dynamic expression of individual gene family members determines cell identity. (*A*) Pair-wise differential gene expression analyses between Tip, P0 and vasculature show ~10% of all expressed genes are DE ($q < 0.01$). Colors represent preferential expression within respective domains. (*B*) Enrichment analyses identified few functional categories over-

represented among Tip vs P0 DEGs. Dashed line, significance threshold ($p = 0.05$). (*C-F*) Individual genes within gene families show abundant and differential expression across the apex. Examples are shown for gene families functioning in cytokinin (*C*), auxin (*D*), cell wall (*E*), or other signaling processes (*F*). Number of family members is shown in parenthesis. Only genes expressed ≥ 2 RPM are illustrated. Mer, meristem; Int, internode; Vas, vasculature. (*G*) *Zmlog7* mutants display meristem termination phenotypes. (*H,I*) Cleared shoot apices of wildtype (*H*) and the small meristem mutant *Zmdrp4a* (*I*) . (*J,K*) Box-and-whisker plots of SAM size measurements of *Zmdrp4a* (*J*) or *Zmfcp1* (*K*) and their respective wildtype siblings. N, number of wildtype/mutant apices measured. **, $p < 0.01$; ***, $p < 0.001$ according to Student's t-test.

Intriguingly, the expression level of such DEGs across all 10 tissues sampled often far exceeds that of the remaining more uniformly expressed family members (Figure 2C–F; Figure S2C–E). This predicts a more limited degree of redundancy. Indeed, putative loss-of-function alleles available for 3 DEGs with strong preferential expression in the Tip, each show meristem phenotypes. Mutation of *ZmLOG7* conditions a meristem termination phenotype (Figure 2G). This is contrary to *Arabidopsis* and rice, where *log* mutations have little or no effect on vegetative meristem size (Kurakawa et al. 2007; Tokunaga et al. 2012). In addition, mutations in the dynamin family member *ZmDRP4a* cause a reduction in SAM radius, whereas *Zmfcp1* mutants show a significant increase in SAM size (Figure 2H–K; (Je et al. 2016)), consistent with the presence of a local CLE-WOX module regulating stem cell number.

Thus, the cellular mechanisms linking patterns of TF activity to the differentiation of distinct cell types are highly complex, with DEGs predicted to function in a wide array of metabolic and cellular processes. Inputs from discrete signaling components and combinatorially acting TFs, many of which are expressed in a tissue-specific manner, drive strong, differential expression of select, often individual, genes within gene families to confer distinctive properties onto functional domains within the apex. Sub-functionalization among members of functionally diverse gene families thus emerges as a key feature underlying cell fate differentiation.

Molecular signatures underlying functional SAM domains

Besides the classically defined peripheral and central zones, studies in *Arabidopsis* revealed the presence of an organizing center (OC) positioned immediately below the

CZ. The OC provides positional information required to specify stem cell fate, and balances activities in the central and peripheral zones of the SAM (Pfeiffer et al. 2017). Cells in the OC are characterized by expression of the *WUS* TF (Mayer et al. 1998). Interestingly, whereas *WUS* orthologs in other eudicot species share this pattern of expression (Galli and Gallavotti 2016), *ZmWUS1* expression within the vegetative SAM is not conserved (Nardmann and Werr 2007). Likewise the *CLV1* ortholog *thick tassel dwarf1* is expressed in leaf primordia (Nardmann and Werr 2007), and *ZmFCP1* rather than the maize *CLV3* ortholog specifically marks the SAM tip (Figure S2A, B). This indicates substantial diversification in the WUS-CLV signaling pathway between monocot and eudicot species, and leaves open the identity of an OC in the morphologically distinct maize SAM. To address this question and to identify molecular signatures that distinguish the functional SAM domains, we clustered genes based on their transcript profiles across the meristem, Tip, P0, and P1-P3 leaf primordia (Figure 3A). Inclusion of the latter enhanced the ability to identify SAM specific signatures and allowed to follow the transition from indeterminate stem cell through differentiation.
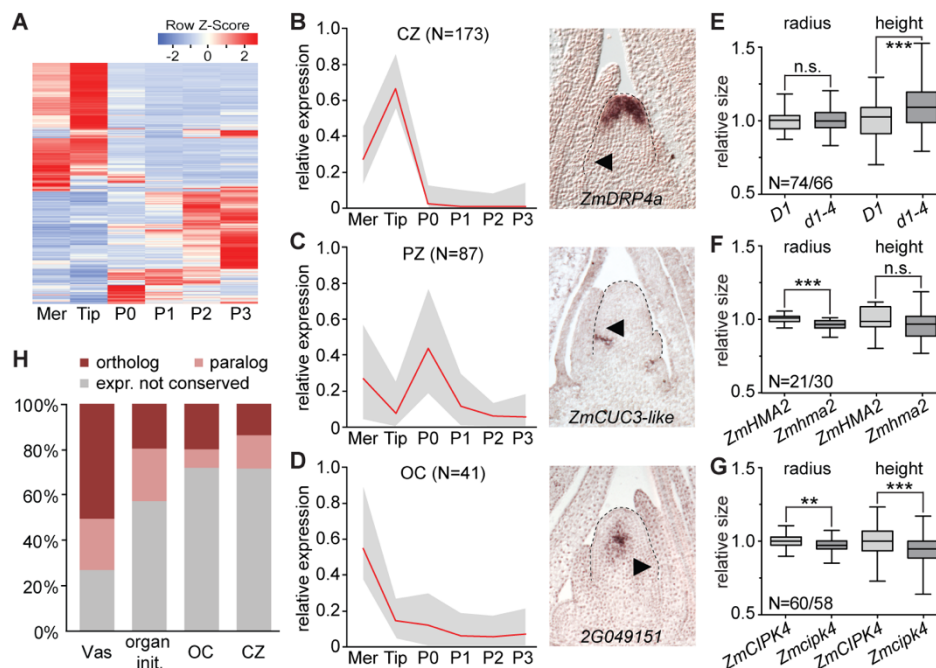
Figure 3. Divergent gene sets define functional SAM domains in maize versus *Arabidopsis*. (*A*) Expression heatmap of genes dynamically expressed during the transition from stem cell (Tip) to P3 organ primordium. Genes of Cluster 1 whose expression changes minimally are excluded. Mer, meristem. (*B-D*) Left: composite expression profiles of gene clusters marking the CZ (*B*), PZ (*C*) or OC (*D*). Right: *in situ* hybridization patterns of select

cluster members. Arrowheads, P0; N, number of genes in respective clusters; Red line, mean expression; Grey profile; range between highest and lowest values. (*E-G*) Mutations in the OC genes *D1* (*E*), *ZmHMA2* (*F*), and *ZmCIPK4* (*G*) effect meristem height, radius, or both. N, number of measured wildtype/mutant individuals. n.s., not significant; **, $p < 0.01$; ***, $p < 0.001$ according to Student's t-test. (*H*) Percentages of maize genes expressed specifically in vasculature, during organ initiation, or in the OC or CZ with a similarly expressed *Arabidopsis* ortholog (dark red) or related paralog (light red). Maize genes without an identifiable *Arabidopsis* ortholog or near paralog are not shown.

Among the 46 clusters obtained, two large gene clusters show increased or decreased expression with leaf ontogeny (Figure S3A, B), providing numerous markers for comparative analyses of leaf development. Additionally, few clusters show expression profiles expected for meristem core genes, with high relative expression in both meristem and Tip, and minimal expression during organogenesis (Figure S3C; Table S7). These clusters point to the presence of regulatory circuits promoting general meristem identity that presumably interact with more localized determinants to specify regional identities within the SAM.

Importantly, several clusters show expression profiles consistent with expectations for the functional domains of the SAM. Two CZ-clusters with 173 genes (Figure 3B) comprise mostly Tip specific genes, but also genes expressed in both the Tip and vasculature, possibly reflecting a general stem cell function, as well as genes that specifically mark the L1 (e.g., *ZmWOX9b*, *ZmWOX9c*) or L2 (e.g., *ZmFCP1)*, predicting layer specific contributions to the CZ (Figure S3D). In addition to P0 specific genes, the four PZ-clusters with 87 genes total (Figure 3C), include genes connected to leaf initiation (e.g., *ZmWOX3a, Arf3b, ZmGA2ox*, and *fused leaves1*), boundary formation (e.g. *ZmCUC3-like*), and axillary meristem formation (e.g., *barren inflorescence2*; Table S7). RNA *in situ* hybridization verified that selected genes in these clusters indeed show the predicted domain specific patterns of expression (Figs. 1H–M, 3B, C).

Genes expressed at a position equivalent to that of the *Arabidopsis* OC are predicted to show high expression in the meristem overall, and minimal expression in the Tip and leaf primordia. This profile is seen in three clusters, comprising 41 genes total (Figure 3D). *In situ* hybridization shows that transcripts for a gene of unknown function

representative of these clusters localize to a small group of cells at the SAM center below the CZ, recapitulating the canonical OC expression pattern of *WUS* (Mayer et al. 1998) (Figure 3D). Further, in line with a role as a signaling center critical for balancing stem cell maintenance and organogenesis, genes encoding the GA 3-oxidase *dwarf plant1* (*D1*), *ZmBAS1*, a LRR receptor kinase, as well as proteins related to calcium-, redox-, and sugar-based signaling are included in these clusters (Table S7). Moreover, three of the six genes with available loss-of-function alleles display quantitative effects on meristem height (*d1*), radius (*Zmhma2*) or both (*Zmcipk4*) (Figure 3E–G).

Genes in these clusters thus identify a domain in the maize SAM equivalent in position and function to the *Arabidopsis* OC. Noteworthy, ten TFs, each representing a distinct family, are among the OC genes. However, neither *ZmWUS1* nor any other WOX member shows this expression signature. In maize, the activity of a central signaling center required to balance cell fates within the shoot stem cell niche can thus be separated from WOX activity, pointing to divergence in the molecular networks underlying meristem function in maize compared to dicot species.

Divergent molecular circuitries underlie SAM function in maize and Arabidopsis

To examine the extent of such divergence and to assess the degree to which morphological diversity between the maize and *Arabidopsis* SAM is reflected at the level of molecular circuitry, we asked whether *Arabidopsis* homologs of the maize SAM domain specific genes show analogous patterns of expression. Notably, for about one third of maize domain specific genes, an *Arabidopsis* ortholog could not be identified. This percentage is surprising given that the origin of the layered meristem predates the divergence of monocot and dicot lineages (Gifford 1954), and that merely 19% of all expressed maize genes lack an identifiable *Arabidopsis* ortholog (Table S8). Moreover, whereas ~73% of maize vasculature specific genes have an *Arabidopsis* ortholog or close paralog with a vascular enriched pattern of expression, only 28% of remaining maize OC and CZ genes have a close *Arabidopsis* relative whose pattern of expression is conserved (Figure 3H). Also genes marking the incipient primordium (organ initiation clusters, Table S7) show more extensive expression conservation (Figure 3H; Table S8).

Aside from *D1,* the signaling-related genes connected to the maize OC lack homologs in *Arabidopsis* with a similar pattern of expression, and only two of the 10 TFs marking the maize OC have related genes enriched in the OC of *Arabidopsis* (Table S8). Within the CZ, particularly genes involved in transcriptional and post-transcriptional regulation, such as TFs, *AGO5* and *PUMILIO* (*PUM*) genes, show a conserved pattern of expression (Table S8). AGO and Pumilio proteins are also required for germline and stem cell maintenance in animals (Siomi and Kuramochi-Miyagawa 2009; Slaidina and Lehmann 2014), pointing to possible fundamental features of stem cell regulation. Thus, while select processes are shared, highly divergent sets of genes define the OC and CZ in maize and *Arabidopsis*. This not only indicates crucial differences in stem cell regulation but also that the stem cell state itself can reflect the differential activity of a range of processes, a notion supported by the finding that quantitative expression level changes in hundreds of genes involved in wide array of metabolic and cellular functions distinguish the CZ from leaf primordia (Figure 2A; Table S5).

Complex TF signatures drive cell identities

The expression divergence between maize and *Arabidopsis* raises the question how cell fates within the maize SAM are specified. Many TFs show tissue-specific or differential patterns of expression, predicting a causal relationship to the dynamic expression changes characterizing individual cell and tissue types. To address this, we performed a principal component analysis (PCA) considering expression values of all TFs across the 10 domains under study. This identified 3 principal components that distinguish meristematic tissues (PC1), vasculature (PC2) and internode (PC3), respectively (Figure 4A; Table S9). Interestingly, this spatial separation based merely on the expression profiles of TFs, which represent ~7% of all expressed genes, mirrors the overall trends observed in correlation analysis (Figure 1E). This supports the idea that defined sets of TFs underlie the cell- and domain-specific expression of select target genes within diverse gene families to specify cell fate.

Further predicting causative relationships between specific TF families and tissue identity, the mean expression values of most TF families are strongly correlated with a given PC (Figure 4B,C; Table S9). 11 of the 55 maize TF families (http://grassius.org/) are positively correlated to vascular identity (PC2; Figure 4C), including the earlier mentioned Dof TFs, type-B ARRs known to promote vascular identity (Yokoyama et

al. 2007), as well as TUBBY and G2-like TFs of which several show vascular specific expression (Table S3). The mean expression values for 23 TF families are highly correlated with PC1, of which 17 are positively-correlated (Figure 4C). Among them the GeBP, FARL, NAC, and Homeobox TFs, which have a demonstrated link to meristem function (Vollbrecht et al. 2000; Aida and Tasaka 2006; Chevalier et al. 2008; Aguilar-Martínez et al. 2014). Mean expression for the remaining 6 families is negatively-correlated with PC1, suggesting they act as repressors of meristem identity and/or promote organogenesis. Indeed, YABBY, GRF, and OFP TFs either directly or indirectly repress KNOX gene function (Dai et al. 2007; Hay and Tsiantis 2010; Kuijt et al. 2014; Wang et al. 2016).

Of the TF families positively-correlated to PC1, the GeBP, HB, ABI3/VP1 and NAC TF families are overrepresented among Tip enriched genes in DE and SE analyses (Figure 2B; Table S4). In addition, 81 out of 583 PC1-correlated TFs are included in the meristem core or subdomain clusters (Table S7). However, most TFs are expressed more broadly, and correlation to PC1 reflects the additive effect of more subtle quantitative expression differences from multiple TF family members across the vegetative apex. Interestingly, the collective quantitative expression differences of all PC1-correlated TFs offer a basis for generating meristem- and CZ-specific patterns of expression. Positive PC1-correlated TFs show highest cumulative expression in the meristem and the CZ particularly, whereas expression of TFs negatively-correlated to PC1 is lowest in these tissues. When combined, the opposing effects of these TFs could conceivably bring about tissue specificity (Figure 4D). Accordingly, the spatially restricted expression of meristem- and CZ-specific genes is predicted to reflect the combinatorial activities of multiple meristem-promoting and -repressing TFs.
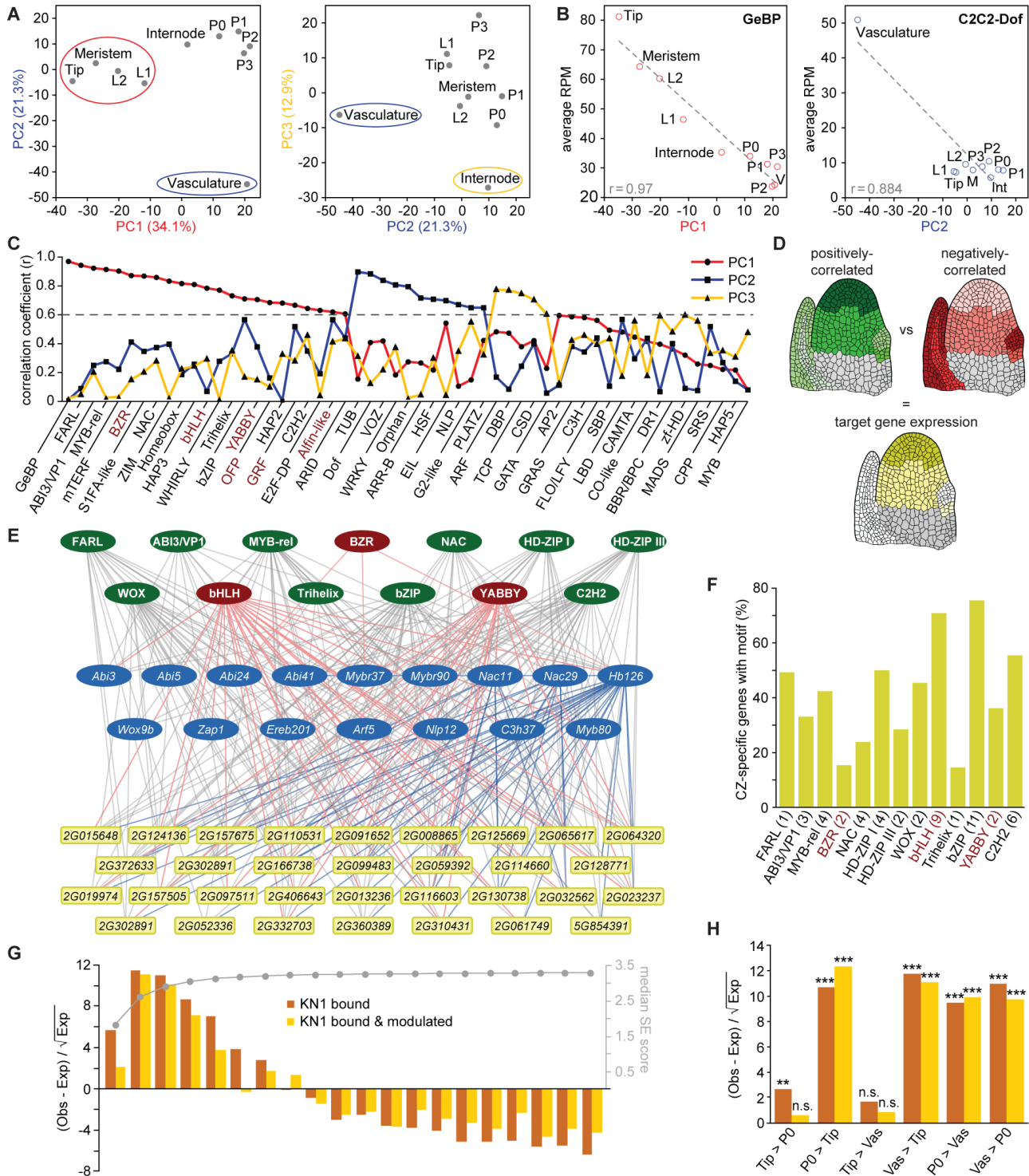
Figure 4. Combinatorial effects of multiple TFs distinguish cell identities. (*A*) PCA showing TF expression across domains mirrors trends of genome-wide expression profiles with PC1 (red) and PC2 (blue) separating Tip and vasculature, respectively. PC3 (yellow) distinguishes the internode, and leaf primordia group together. (*B*) Mean TF family expression across domains correlates with given PCs. Examples are shown for PC1 (left) and PC2 (right). Correlation

coefficients (r) are indicated. V, vasculature; M, meristem; Int, internode. (*C*) Overview of correlation coefficients (r) for each TF family to each major PC. A threshold of 0.6 (dashed line) was used as correlation cut-off. Mean expression of TF families shown in red is negatively-correlated to meristem identity. (*D*) Diagrams illustrating combined expression values for all TFs positively- (green) and negatively-correlated (red) to PC1 together can govern cell type specificity (yellow). Grey, not examined. (*E*) Visualization of a GRN for select CZ-specific genes (yellow and blue) with combinatorial and hierarchical interactions from TF families positively- (green) and negatively-correlated (red) to PC1, as well as individual CZ-specific TFs (blue). (*F*) Percentage of CZ-specific genes with binding sites for PC1-correlated TF families in their promoter. The number of tested motifs is given in parentheses. TF families shown in red are negatively-correlated to PC1. (*G*) Distribution of KN1 bound (dark orange) and KN1 bound and modulated (light orange) targets among apex-expressed genes grouped into 20 bins based on SE score (right y-axis) shows KN1 primarily targets dynamically-expressed genes. Obs, observed; Exp, expected. (*H*) DEGs targeted by KN1 are preferentially expressed in organ primordia and vasculature. n.s., not significant; **, $p < 0.01$; ***, $p < 0.001$ based on Chi-square test with Yates' continuity correction.

Combinatorial effects of diverse TFs promote stem cell fate

To test this hypothesis and to assess a contribution of PC1-correlated TFs to cell fate specification, we constructed a gene regulatory network (GRN) based on the occurrences of TF binding motifs within the promoters of CZ-specific genes. Indicative of functional regulatory interactions *in vivo* (Reece-Hoyes et al. 2013; Sparks et al. 2016), *cis*-regulatory elements for all 13 PC1-correlated TF families with available binding position weight matrices are highly enriched in proximal promoters of CZ genes relative to whole genome incidence (Table S10).

The GRN reveals a highly interconnected arrangement of possible transcriptional regulatory interactions. Each of the 13 PC1-correlated TF families can target a substantial number of CZ-specific genes (Figure 4E, F; Figure S4A). For instance, both FARL and HD-ZIPI DNA-binding motifs are present in the promoters of nearly half the CZ-specific genes (Figure 4F). Interestingly, both TF families mediate transcriptional responses downstream of light signaling (Harris et al. 2011; Siddiqui et al. 2016), pointing to mechanisms allowing for plasticity in the specification of functional SAM domains in response to environmental cues. Conversely, the promoters of CZ-specific genes contain *cis*-regulatory motifs for on average 5 distinct

PC1-correlated TF families, with some promoters containing binding sites for as many as 10 of the 13 families analyzed (Table S11). This reinforces the idea that domain- or tissue-specific expression reflects the combinatorial actions of multiple TFs. Further, the combination of TFs target sites in individual promoters varies considerably (Table S11), suggesting that tissue specificity can be achieved in many ways.

Noteworthy, most promoters include binding sites for both TFs positively- and negatively-correlated to PC1 (Figure 4E, F; Figure S4A), consistent with the idea that the spatially restricted expression of CZ genes reflects the combinatorial activities of both meristem-promoting and -repressing TFs. The GRN further shows hierarchical transcriptional regulation, with more broadly expressed PC1-correlated TFs converging on the promoters of CZ-specific TFs (Figure 4e). Binding sites for both more broadly expressed and tissue-specific TFs are present in promoters of other CZ genes, generating a network configuration that would reinforce tissue specificity of cell fate determinants (Barolo and Posakony 2002; Niwa 2018). Given the overrepresentation of TFs among tissue-specific genes, such regulatory relationships appear a general feature underlying cell identities.

KN1 promotes meristem fate by repressing organogenesis and differentiation

The above GRN points to substantial redundancy among TFs in meristem regulation. Nonetheless, mutations in *kn1*, which is part of the meristem core cluster (Figure S3C), can show a highly penetrant meristem termination phenotype (Vollbrecht et al. 2000). Interestingly, KN1 targets (Bolduc et al. 2012) are specifically enriched among dynamically expressed genes (Figure 4G, H), indicating a role for KN1 in generating differential patterns of expression across the apex. KN1 particularly targets genes preferentially expressed in the vasculature and P0, where it is not itself expressed. KN1 thus seems to regulate meristem activity primarily by repressing differentiation rather than promoting meristem identity. Supporting this notion, genes predicted to drive organogenesis, e.g. auxin signaling, cytokinin turnover, and cell wall remodeling, are overrepresented among KN1 targets (Figure S4B; Table S12). Moreover, consistent with being a central hub in the CZ GRN, KN1 binds a substantial number (~23%) of PC1-correlated TFs. However, only TFs negatively-correlated to PC1 are enriched among the KN1 targets, whereas those positively-correlated are depleted (Table S13). Thus, KN1 is a master regulator of meristem activity that mediates

indeterminacy by selectively targeting key transcriptional regulators and signaling pathways that promote organogenesis and differentiation.

Dynamically-expressed meristem genes modulate important architectural traits

A combinatorial quantitative contribution from diverse TFs to cell fate specification confounds mutational analyses. We therefore took advantage of data from genome-wide association studies (GWAS) to address function, and tested whether potential natural variation present at these loci contributes to the quantitative variation in plant morphology present among maize varieties. Individual traits measured in more than a dozen GWAS studies (see Table S14 for details) were broadly classified as being architectural or non-architectural in nature. Collectively, ~60% of SNPs associated with traits in either broad category are located within 10 kb of apex-expressed genes (Table S14). Strikingly, given the stringent criteria applied (see Methods), PC1-correlated TFs are significantly enriched among expressed genes associated with plant architectural traits (Figure 5A). In contrast, natural variation linked to disease and metabolic traits maps preferentially near TFs underlying PC3. A breakdown into individual architectural traits shows that different PC1-correlated TF families control distinct morphological features (Figure 5B). For instance, natural variation near ABI3/VP1 and YABBY TFs is linked to diversity in plant height, whereas allelic diversity at Myb-rel and OFP TFs is strongly associated with leaf morphology traits, and at Trihelix and GRF TFs with node number.

Plant morphological diversity is thus determined in part by TFs connected to cell fate decisions in the SAM. In addition, genes identified by SE to have a tissue-specific or dynamic pattern of expression are uniquely enriched near architectural Trait Associated SNPs (TASs), whereas constitutively expressed genes are strongly depleted (Figure 5A; Figure S5). This highlights an unexpected importance for polymorphisms near dynamically expressed genes in shaping morphological variation, and identifies such genes as key targets during breeding selection. Among the dynamically expressed genes, members of the meristem core and subdomain clusters are significantly enriched again only near architectural TASs, affecting a variety of morphology traits (Figure 5A, C). However, the subset of genes marking the CZ is depleted near SNPs underlying variation in the majority of adult plant characters.

Instead, natural variation near genes marking cells of the OC or PZ strongly influences the overall morphology of the plant (Figure 5C). Adult plant characters, including key agronomic traits, such as leaf angle, leaf shape, plant height, flowering time, and inflorescence morphology, are thus to a remarkable degree regulated by gene circuitries acting in the SAM.
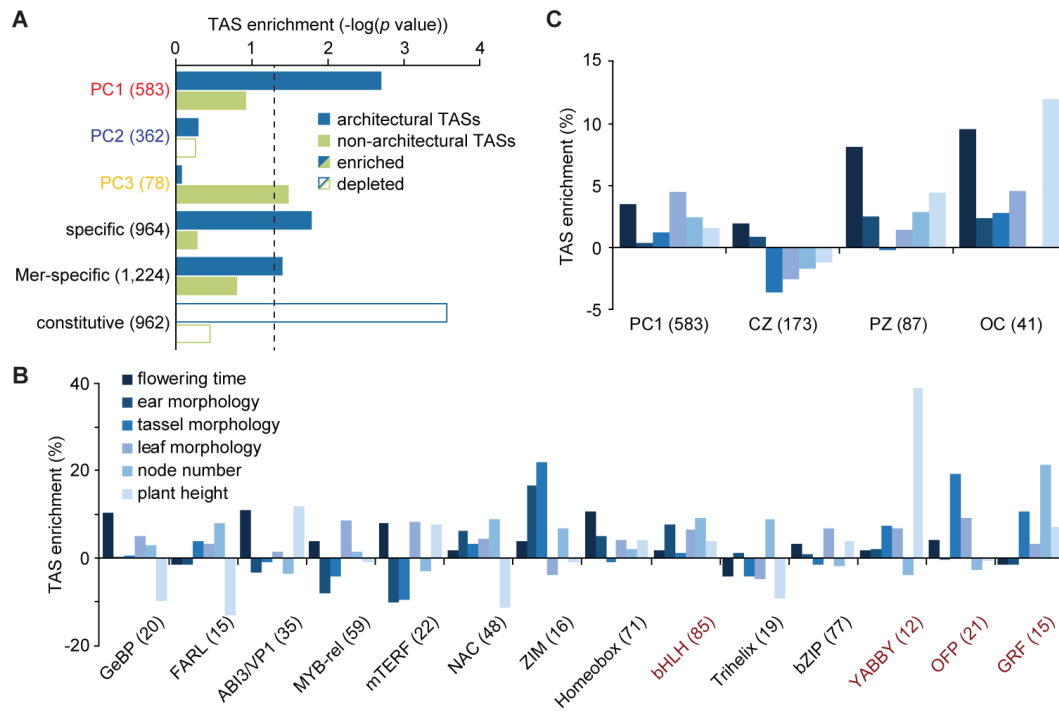
Figure 5. Dynamically-expressed meristem genes modulate important architectural traits. (*A*) Significance plot illustrating that natural variation underlying plant architectural traits is significantly enriched specifically near PC1-correlated TFs, dynamically expressed genes, and meristem-specific genes. In contrast, architectural TASs are strongly depleted near constitutively expressed genes. Dashed line, significance threshold (*p* = 0.05) based on Chi-square test with Yates' continuity correction. Input gene numbers are given in parentheses. (*B*) Enrichments of TASs near individual TF families either positively- or negatively-correlated (red) to PC1 shows different TF families shaping distinct morphological features. (*C*) Enrichments of TASs influencing individual architectural traits near PC1-correlated TFs and genes marking the PZ and OC. TASs for most architectural traits are depleted near CZ-specific genes.

Discussion

The SAMs of most angiosperms share a basic functional organization. Meristems are nonetheless, characterized by striking diversity in shape and size (Steeves and

Sussex 1989), as nicely exemplified by the highly divergent morphologies of the maize and *Arabidopsis* SAM. Our study shows that this morphometric diversity is reflected at the level of molecular circuitry. The high-resolution gene expression atlas identified molecular signatures defining critical domains of the maize vegetative apex. This reveals that distinct sets of genes underlie the regulation and identity of stem cells in maize versus *Arabidopsis*. Nevertheless, expression of *ZmFCP1* specifically in the CZ of the maize SAM predicts the presence of a locally restricted CLE/WOX module to balance stem cell number and stably anchor the CZ to the growing shoot tip. In *Arabidopsis*, two opposing signaling centers provide relevant positional cues; mobile WUS from the OC promotes stem cell identity in distal cells, and epidermal-derived miR394 anchors the CZ to the SAM tip (Yadav et al. 2011; Knauer et al. 2013; Daum et al. 2014). In maize, the OC is not defined by WOX expression, and miR394 is absent from the SAM (Javelle and Timmermans, unpublished data), suggesting an alternative mechanism to maintain a region of stem cell competence at the tip. An intriguing hypothesis is that in the vegetative SAM of maize the sources of these signals are displaced. Likely candidates to promote stem cell identity are *ZmWOX9b* and *ZmWOX9c*, which are both expressed in the L1 of the CZ. Consistent with this idea, *ZmFCP1* is primarily expressed in subepidermal layers, where it overlaps with its receptors FEA2 (Je et al. 2018) and FEA3 (Je et al. 2016). WOX9 protein could represent a steady, but inwards-directed stem cell-promoting factor that in conjugation with hormones and other signals originating from the OC provide the positional information required to stabilize stem cell activity in the growing niche.

The cellular mechanisms linking positional inputs to patterns of TF activity and the differentiation of distinct cell and tissue types are highly complex. Cell identities are distinguished by the differential expression of hundreds of genes involved in wide array of metabolic and cellular processes. A surprising number of these DEGs are expressed at levels far exceeding those of other family members. Thus, reminiscent of divergent paralogs driving morphological innovation (Panchy et al. 2016), sub-functionalization of gene family members is a key feature underlying the differentiation of distinct cell types. These findings further identify single DEGs as prime targets via which to shape plant morphology and manipulate developmental traits critical to crop improvement.

The CZ GRN further predicts that genes effecting cell identity are targeted by both activating and repressing TFs and that more broadly expressed TFs act in a combinatorial and hierarchical manner with cell type-specific TFs to define their spatially restricted patterns of expression. Network configurations in which target gene expression reflects the combinatorial additive and opposing effects of more general and locally restricted TFs is emerging as a general feature underlying developmental patterning (Barolo and Posakony 2002; Sparks et al. 2016; Reiter et al. 2017; Niwa 2018). These complex network architectures buffer gene expression by reducing the impact of mutations in individual TFs and, moreover, allow cells to discriminate true signaling inputs from background gene expression fluctuations to provide robustness (Sokolik et al. 2015). These features are particularly relevant to plants given their sessile nature and the need to maintain stable developmental programs under highly variable conditions.

Nevertheless, this mechanism of robustness remains inevitably accompanied by vulnerabilities, as perturbations in highly connected nodes, such as TFs at the top of a transcription cascade, can lead to a collapse of the entire network (Barabasi and Oltvai 2004). For instance, misregulation of KN1, which targets nearly half the TFs negatively-correlated with meristem fate, leads to severe developmental defects (Smith et al. 1992; Vollbrecht et al. 2000). Curiously, expression of very few targets is altered in such mutants (Bolduc et al. 2012). Perhaps, KN1 functions as a pioneer factor that facilitates binding of other TFs to regulate target gene expression in this context dependent manner (Reiter et al. 2017). However, given that KN1 primarily targets genes that function in organ primordia, an alternative, non-mutually exclusive view for the contribution of KN1 to SAM function is that it generates a state of default repression (Barolo and Posakony 2002). In this scenario, KN1 safeguards cells in the meristem from erroneously activating the differentiation program.

Although morphological diversity between plant species is thought to elaborate primarily post-meristematically (Steeves and Sussex 1989), our data shows that maize plant architecture is to a remarkable degree regulated by molecular circuitries acting in the vegetative SAM. Both the high degree of TF connectivity and the broad spectrum of cellular processes underlying cell identity would allow a degree of circuitry evolvability. This is measured as quantitative variation, specifically in morphological

traits. Our findings highlight a distinctive contribution from allelic diversity near dynamically expressed genes to phenotypic variation. Particularly polymorphisms near PC1-correlated TFs connected to cell fate decisions in the SAM, as well as genes expressed in the organogenic PZ or in the OC, which orchestrates the balance between stem cell maintenance and organogenesis, are associated with morphological diversity. Besides defining genes governing the identity and function of critical domains within the maize SAM, our gene expression atlas thus identifies key targets for selection in the improvement of agronomically-important traits.

Methods

Plant materials

All analyses were performed on 14 day-old B73 seedlings grown under 16 h 24°C light and 8 h 20°C dark cycles. Mutant alleles for *ZmLOG7* (mu1030680, mu1052820), *dwarf plant1* (*d1-4*), *ZmCIPK4* (mu1046464); *ZmHMA2* (mu1076495), *GRMZM2G050234* (mu1037811, non-phenotypic), *GRMZM2G416817* (mu1070579, non-phenotypic) and *GRMZM2G168807* (mu1038844, non-phenotypic) were obtained from the Maize Genetics Cooperation Stock Center. DuPont Pioneer kindly screened for exon insertion alleles for *ZmFCP1* and *ZmDRP4a*. Transposon insertion alleles were introgressed for 3-4 generations into B73 (*Zmfcp1*) or T43 (all other mutations) prior to genetic and phenotypic analysis. See Table S15 for gene IDs.

Laser microdissection and RNAseq library construction

Hand-dissected apices of 14 day-old B73 seedlings were fixed in acetone, embedded into paraffin, and sections of 8 µm spread on 1.0 PEN membrane Slides (Zeiss), as described (Scanlon et al. 2009). After deparaffinization in xylene, cells of interest were captured into AdhesiveCap 500 tubes (Zeiss) using the PALM Micro-Beam system. To minimize variation, tissue samples were captured from sections from at least 10 individual apices for each of two biological replicates. RNA was extracted using the PicoPure RNA isolation kit (Arcturus), treated with DNaseI (Qiagen), and amplified with the TargetAmp 2-Round aRNA Amplification Kit 2.0 (Epicentre Biotechnologies). Single-end RNAseq libraries were constructed using standard Illumina protocols (Illumina) and sequenced (100 bp) on the Illumina HighSeq2000 platform.

Gene expression analyses

The nucleotides of each raw read were scanned for low quality bases. Bases with a PHRED quality value <15 (out of 40), i.e. with error rates $\leq$ 3%, were removed by Data2Bio's trimming pipeline. Trimmed reads were aligned to the B73 RefGen_V3 using GSNAP (Wu and Nacu 2010), and uniquely mapped reads allowing ≤2 mismatches every 36 bp and less than 5 bases for every 75 bp as tails were used for subsequent analyses. Read counts per gene were computed using B73 gene annotation version FGSv5b. Library metrics are listed in Table S1. Gene expression levels were normalized to RPM rather than RPKM, as linearly amplified RNA captures the 3' 400-500 nucleotides of transcripts (Scanlon et al. 2009). Relatedness across tissue samples was determined based on the expression values of all genes expressed in each pairwise comparison using Pearson's correlation in R. Subsequent analysis were performed on genes with a mean expression value ≥ 2 RPM in at least one of the 10 tissues sampled. Differential gene expression was determined using the DESeq (Anders and Huber 2010) package in R with default parameters and a BH-corrected p value < 0.01 cut-off (Benjamini and Hochberg 1995). Cell type specific genes were identified by Shannon entropy (SE) (Schug et al. 2005). The SE density distribution for all expressed genes fits a Chi-square distribution (p < 2.2E-16, Pearson's Chi-square normality test). SE scores <2.33 were considered as domain-specific to account for overlaps among some of the domains analyzed. For cluster analysis, genes showing a ≥ 2-fold expression change in any 2-way comparison between meristem, Tip, P0 – P3 were identified using Cluster Affinity Search Technique. Cluster analysis on the dynamically expressed genes was conducted in the MultiExperiment Viewer (MeV) software package using Spearman Rank Correlation as a distance metric and a threshold parameter of 0.8 according to (Ben-Dor and Yakhini 1999). Heatmaps were generated in R using the heatmap.2 function in the gplots package (Warnes et al. 2016).

Enrichment analysis

MapMan annotations of the maize filtered gene set (v5b.60 (Usadel et al. 2009)) were used for functional enrichment analyses using hypergeometric distribution-based enrichment testing with GOseq (Young et al. 2010). TFs and hormone related genes

were manually annotated based on information from Grassius (http://grassius.org) (Yilmaz et al. 2009) and published work, or through identification of maize homologs of known *Arabidopsis* genes using the paralog search tool in BioMart (http://www.gramene.org) (Tello-Ruiz et al. 2017). Enrichments were determined using Fisher's exact test and reported as BH-adjusted *p* values (Benjamini and Hochberg 1995). Pathway analyses were performed with the ClueGO plug-in in Cytoscape (Shannon et al. 2003; Bindea et al. 2009) with standard settings on the GO/MolecularFunction, GO/BiologicalProcess, and KEGG databases.

PCA and TF correlation analyses

An expression matrix of all expressed TFs was compiled and after standardization, PCA was conducted using the Prcomp function in R. Under the assumption that TFs belonging to the same family are functionally interchangeable, the contributions of TF families to each principal component was estimated based on the correlations between principal component variables and the average expression values for each TF family using the Corrgram package in R.

***cis*-regulatory motif enrichment**

To determine enrichment of *cis*-regulatory elements for PC1-correlated TFs within the promoters of Tip-specific genes, we used the computational prediction pipeline described in (Eveland et al. 2014) that leverages the Search Tool for Occurrences of Regulatory Motifs (STORM) from the Comprehensive Regulatory Element Analysis and Detection (CREAD) suite of tools (Smith et al. 2006). Enrichment scores for 51 distinct position weight matrices (PWMs) for 13 PC1-correlated TF families obtained from (Franco-Zorrilla et al. 2014; Mathelier et al. 2015), were calculated based on their occurrence within promoter regions spanning 1 kb upstream to 500 bp downstream of the transcription start site in CZ specific genes over the complete B73 filtered gene set (v5b.60; Table S10). The Gene Regulatory Network was constructed in Cytoscape (Shannon et al. 2003).

KN1 target enrichment

Bound, and bound and modulated KN1 targets were obtained from (Bolduc et al. 2012), of which 2960 and 574 are expressed within the apex, respectively. Target enrichments were calculated as divergence from expectation using the following formula, . Significance was calculated using Chi-square test with Yates' continuity correction in R.

Trait associated SNP (TAS) enrichment

Genes located within 10 kb of SNPs associated with architectural traits and non-architectural traits were identified in R (script available on request). See Table S14 for a complete list of trait associated SNPs and their classifications. Enrichments over the occurrence of all expressed genes near trait associated SNPs were calculated as divergence from expectation. Stringency was applied by removing possible redundancies among trait associations by: 1) considering genes containing multiple SNPs associated with a given trait only once, and 2) counting genes associated with multiple architectural, respectively, non-architectural traits only once. Significance was calculated using Chi-square test with Yates' continuity correction in R.

RT-PCR and *in situ* hybridization

For RT-PCR, 4 ug RQ1 DNAse (Promega) treated aRNA was converted into cDNA using the Superscript III First-Strand Synthesis System (Invitrogen) with random hexamer primers according to manufacturer's protocol. *In situ* hybridizations were performed on apices of 14 day-old B73 seedlings according to (Javelle and Timmermans 2012). Gene specific primers used in these analyses are listed in Table S15.

Meristem size measurements

Hand dissected apices of mutant and non-mutant siblings were vacuum infiltrated for 2x 15 min in FAA (10% formaldehyde, 5% acetic acid, 45% ethanol solution) and fixed overnight in fresh FAA on a shaker at 4°C. Dehydration was performed on a shaker at 4°C for 1 h in 70%, 85%, 95%, and 100% ethanol, respectively, followed by 1 h in 100% ethanol at room temperature (RT). Apices were cleared overnight at RT on a shaker with ethanol:methyl salicylate (1:1), followed by 1 day in 100% methyl salicylate at RT,

changing the solution once. Images were acquired with Nomarski optics on a Leica DMRB transmitted light microscope connected to a MicroPublisher 5.0 RTV camera (QImaging). Meristem height and radius were measured at the height of the P1 cleft from near-median longitudinal optical sections. Values were determined from two independently introgressed lines, and normalized to the height and radius of wildtype siblings. Significance was determined by unpaired two-tailed Student's t-test in GraphPad Prism.

Arabidopsis expression analysis

Arabidopsis orthologs and paralogs of domain specific maize genes were identified using BioMart v.07 Plant Genes 56 (http://www.gramene.org) (Tello-Ruiz et al. 2017). Expression profiles for the following *Arabidopsis* domains AtHB8 (xylem), S17shoot (phloem), WUS (OC), CLV3 (CZ), FIL (organ primordia), HDG4 (meristematic L2), HMG (meristematic L1), KAN1 (outer PZ) and LAS (organ boundary), were obtained from (Yadav et al. 2014). To accommodate the partial overlap between domains, the following comparatively lax criteria were applied to determine tissue specificity: vasculature, relative AtHB8 or S17shoot expression > 25% and expression in all other domains < 20%; organ initiation, exclude S17shoot and AtHB8 expression, relative FIL or LAS expression > 25%, expression in KAN, CLV3 and WUS < 20%; OC, exclude S17shoot and AtHB8 expression, relative WUS expression > 25%, expression in FIL, KAN, LAS, HMG and HGD4 < 20%; CZ, exclude S17shoot and AtHB8 expression, relative CLV3 expression > 25%, expression in FIL, KAN, LAS < 20% (see Figure S3E; Table S8).

Data access

All high-throughput sequencing data, both raw and processed files, have been deposited in NCBI's Gene Expression Omnibus and are accessible upon publication under accession number SRP101301. Mean expression values for most tissues are also visible as a track on the MaizeGDB Genome Browser (http://www.maizegdb.org) under 'Expression and Transcripts – Shoot Apex Atlas'. The authors declare that all other data supporting the findings of this study are available within the manuscript and its Supplemental files or are available from the corresponding author upon request.

Acknowledgements

DISCLOSURE DECLARATION

The authors declare no competing interests.

References

Aguilar-Martínez JA, Uchida N, Townsley B, West DA, Yanez A, Lynn N, Kimura S, Sinha N. 2014. Transcriptional, post-transcriptional and post-translational regulation of STM gene expression in Arabidopsis determine gene function in the shoot apex. *Plant Physiology* **114**: 248625.

Aida M, Tasaka M. 2006. Genetic control of shoot organ boundaries. *Current Opinion in Plant Biology* **9**: 72-77.

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biology* **11**: R106.

Barabasi A-L, Oltvai ZN. 2004. Network biology: understanding the cell's functional organization. *Nature Reviews Genetics* **5**: 101.

Barolo S, Posakony JW. 2002. Three habits of highly effective signaling pathways: principles of transcriptional control by developmental cell signaling. *Genes & Development* **16**: 1167-1181.

Ben-Dor A, Yakhini Z. 1999. Clustering gene expression patterns. In *Proceedings of the third annual international conference on Computational molecular biology*, pp. 33-42. ACM.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society* **1**: 289-300.

Besnard F, Vernoux T, Hamant O. 2011. Organogenesis from stem cells in planta: multiple feedback loops integrating molecular and mechanical signals. *Cellular and Molecular Life Sciences* **68**: 2885-2906.

Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman W-H, Pagès F, Trajanoski Z, Galon J. 2009. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**: 1091-1093.

Bolduc N, Yilmaz A, Mejia-Guerra MK, Morohashi K, O'connor D, Grotewold E, Hake S. 2012. Unraveling the KNOTTED1 regulatory network in maize meristems. *Genes & Development* **26**: 1685-1690.

Chevalier F, Perazza D, Laporte F, Le Hénanff G, Hornitschek P, Bonneville J-M, Herzog M, Vachon G. 2008. GeBP and GeBP-like proteins are noncanonical leucine-zipper transcription factors that regulate cytokinin response in Arabidopsis. *Plant Physiology* **146**: 1142-1154.

Dai M, Hu Y, Zhao Y, Liu H, Zhou D-X. 2007. A WUSCHEL-LIKE HOMEOBOX gene represses a YABBY gene expression required for rice leaf development. *Plant Physiology* **144**: 380-390.

Daum G, Medzihradszky A, Suzaki T, Lohmann JU. 2014. A mechanistic framework for noncell autonomous stem cell induction in Arabidopsis. *Proceedings of the National Academy of Sciences* **111**: 14619-14624.

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

Eveland AL, Goldshmidt A, Pautler M, Morohashi K, Liseron-Monfils C, Lewis MW, Kumari S, Hiraga S, Yang F, Unger-Wallace E. 2014. Regulatory modules controlling maize inflorescence architecture. *Genome Research* **24**: 431-443.

Franco-Zorrilla JM, López-Vidriero I, Carrasco JL, Godoy M, Vera P, Solano R. 2014. DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proceedings of the National Academy of Sciences* **111**: 2367-2372.

Galli M, Gallavotti A. 2016. Expanding the regulatory network for meristem size in plants. *Trends in Genetics* **32**: 372-383.

Gifford EM. 1954. The shoot apex in angiosperms. *The Botanical Review* **20**: 477-529.

Goad DM, Zhu C, Kellogg EA. 2017. Comprehensive identification and clustering of CLV3/ESR related (CLE) genes in plants finds groups with potentially shared function. *New Phytologist* **216**: 605-616.

Harris JC, Hrmova M, Lopato S, Langridge P. 2011. Modulation of plant growth by HD-Zip class I and II transcription factors in response to environmental stimuli. *New Phytologist* **190**: 823-837.

Hay A, Tsiantis M. 2010. KNOX genes: versatile regulators of plant development and diversity. *Development* **137**: 3153-3165.

Javelle M, Timmermans MC. 2012. In situ localization of small RNAs in plants by using LNA probes. *Nature Protocols* **7**: 533.

Je BI, Gruel J, Lee YK, Bommert P, Arevalo ED, Eveland AL, Wu Q, Goldshmidt A, Meeley R, Bartlett M. 2016. Signaling from maize organ primordia via FASCIATED EAR3 regulates stem cell proliferation and yield traits. *Nature Genetics* **48**: 785.

Je BI, Xu F, Wu Q, Liu L, Meeley R, Gallagher JP, Corcilius L, Payne RJ, Bartlett ME, Jackson D. 2018. The CLAVATA receptor FASCIATED EAR2 responds to distinct CLE peptides by signaling through two downstream effectors. *Elife* **7**: e35673.

Knauer S, Holt AL, Rubio-Somoza I, Tucker EJ, Hinze A, Pisch M, Javelle M, Timmermans MC, Tucker MR, Laux T. 2013. A protodermal miR394 signal defines a region of stem cell competence in the Arabidopsis shoot meristem. *Developmental Cell* **24**: 125-132.

Kuijt SJ, Greco R, Agalou A, Shao J, CJ't Hoen C, Övernäs E, Osnato M, Curiale S, Meynard D, van Gulik R. 2014. Interaction between the GRF and KNOX families of transcription factors. *Plant Physiology* **113**: 222836.

Kurakawa T, Ueda N, Maekawa M, Kobayashi K, Kojima M, Nagato Y, Sakakibara H, Kyozuka J. 2007. Direct control of shoot meristem activity by a cytokinin-activating enzyme. *Nature* **445**: 652.

Leiboff S, Li X, Hu H-C, Todt N, Yang J, Li X, Yu X, Muehlbauer GJ, Timmermans MC, Yu J. 2015. Genetic control of morphometric diversity in the maize shoot apical meristem. *Nature Communications* **6**: 8974.

Mathelier A, Fornes O, Arenillas DJ, Chen C-y, Denay G, Lee J, Shi W, Shyr C, Tan G, Worsley-Hunt R. 2015. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Research* **44**: D110-D115.

Mayer KF, Schoof H, Haecker A, Lenhard M, Jürgens G, Laux T. 1998. Role of WUSCHEL in regulating stem cell fate in the Arabidopsis shoot meristem. *Cell* **95**: 805-815.

Nardmann J, Werr W. 2007. The evolution of plant regulatory networks: what Arabidopsis cannot say for itself. *Current Opinion in Plant Biology* **10**: 653-659.

Niwa H. 2018. The principles that govern transcription factor network functions in stem cells. *Development* **145**: dev157420.

Panchy N, Lehti-Shiu M, Shiu S-H. 2016. Evolution of gene duplication in plants. *Plant Physiology* **171**: 2294-2316.

Pfeiffer A, Wenzl C, Lohmann JU. 2017. Beyond flexibility: controlling stem cells in an ever changing environment. *Current Opinion in Plant Biology* **35**: 117-123.

Plackett AR, Di Stilio VS, Langdale JA. 2015. Ferns: the missing link in shoot evolution and development. *Frontiers in Plant Science* **6**: 972.

Reece-Hoyes JS, Pons C, Diallo A, Mori A, Shrestha S, Kadreppa S, Nelson J, DiPrima S, Dricot A, Lajoie BR. 2013. Extensive rewiring and complex evolutionary dynamics in a C. elegans multiparameter transcription factor network. *Molecular Cell* **51**: 116-127.

Reiter F, Wienerroither S, Stark A. 2017. Combinatorial function of transcription factors and cofactors. *Current Opinion in Genetics & Development* **43**: 73-81.

Scanlon MJ, Ohtsu K, Timmermans MC, Schnable PS. 2009. Laser microdissection-mediated isolation and in vitro transcriptional amplification of plant RNA. *Current Protocols in Molecular Biology* **87**: 25A. 23.21-25A. 23.15.

Schippers JH, Foyer CH, van Dongen JT. 2016. Redox regulation in shoot growth, SAM maintenance and flowering. *Current Opinion in Plant Biology* **29**: 121-128.

Schug J, Schuller W-P, Kappen C, Salbaum JM, Bucan M, Stoeckert CJ. 2005. Promoter features related to tissue specificity as measured by Shannon entropy. *Genome Biology* **6**: R33.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* **13**: 2498-2504.

Siddiqui H, Khan S, Rhodes BM, Devlin PF. 2016. FHY3 and FAR1 act downstream of light stable phytochromes. *Frontiers in Plant Science* **7**: 175.

Siomi MC, Kuramochi-Miyagawa S. 2009. RNA silencing in germlines—exquisite collaboration of Argonaute proteins with small RNAs for germline survival. *Current Opinion in Cell Biology* **21**: 426-434.

Slaidina M, Lehmann R. 2014. Translational control in germline stem cell development. *Journal of Cell Biology* **207**: 13-21.

Smith AD, Sumazin P, Xuan Z, Zhang MQ. 2006. DNA motifs in human and mouse proximal promoters predict tissue-specific expression. *Proceedings of the National Academy of Sciences* **103**: 6275-6280.

Smith LG, Greene B, Veit B, Hake S. 1992. A dominant mutation in the maize homeobox gene, Knotted-1, causes its ectopic expression in leaf cells with altered fates. *Development* **116**: 21-30.

Sokolik C, Liu Y, Bauer D, McPherson J, Broeker M, Heimberg G, Qi LS, Sivak DA, Thomson M. 2015. Transcription factor competition allows embryonic stem cells to distinguish authentic signals from noise. *Cell Systems* **1**: 117-129.

Sparks EE, Drapek C, Gaudinier A, Li S, Ansariola M, Shen N, Hennacy JH, Zhang J, Turco G, Petricka JJ. 2016. Establishment of expression in the SHORTROOT-SCARECROW transcriptional cascade through opposing activities of both activators and repressors. *Developmental Cell* **39**: 585-596.

Steeves TA, Sussex IM. 1989. *Patterns in plant development*. Cambridge University Press.

Tello-Ruiz MK, Naithani S, Stein JC, Gupta P, Campbell M, Olson A, Wei S, Preece J, Geniza MJ, Jiao Y. 2017. Gramene 2018: unifying comparative genomics and pathway resources for plant research. *Nucleic Acids Research* **46**: D1181-D1189.

Tognetti VB, Bielach A, Hrtyan M. 2017. Redox regulation at the site of primary growth: auxin, cytokinin and ROS crosstalk. *Plant, Cell & Environment* **40**: 2586-2605.

Tokunaga H, Kojima M, Kuroha T, Ishida T, Sugimoto K, Kiba T, Sakakibara H. 2012. Arabidopsis lonely guy (LOG) multiple mutants reveal a central role of the LOG-dependent pathway in cytokinin activation. *The Plant Journal* **69**: 355-365.

Usadel B, Poree F, Nagel A, Lohse M, CZEDIK-EYSENBERG A, Stitt M. 2009. A guide to using MapMan to visualize and compare Omics data in plants: a case study in the crop species, Maize. *Plant, Cell & Environment* **32**: 1211-1229.

Vollbrecht E, Reiser L, Hake S. 2000. Shoot meristem size is dependent on inbred background and presence of the maize homeobox gene, knotted1. *Development* **127**: 3161-3172.

Wang S, Chang Y, Ellis B. 2016. Overview of OVATE FAMILY PROTEINS, a novel class of plant-specific growth regulators. *Frontiers in Plant Science* **7**: 417.

Warnes GR, Bolker B, Bonebakker L, Gentleman R, Liaw WHA, Lumley T, Maechler M, Magnusson A, Moeller S, Schwartz M. 2016. gplots: various R programming tools for plotting data. R package version 3.0. 1. *The Comprehensive R Archive Network*.

Wu TD, Nacu S. 2010. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* **26**: 873-881.

Yadav RK, Perales M, Gruel J, Girke T, Jönsson H, Reddy GV. 2011. WUSCHEL protein movement mediates stem cell homeostasis in the Arabidopsis shoot apex. *Genes & Development* **25**: 2025-2030.

Yadav RK, Tavakkoli M, Xie M, Girke T, Reddy GV. 2014. A high-resolution gene expression map of the Arabidopsis shoot meristem stem cell niche. *Development* **141**: 2735-2744.

Yilmaz A, Nishiyama MY, Fuentes BG, Souza GM, Janies D, Gray J, Grotewold E. 2009. GRASSIUS: a platform for comparative regulatory genomics across the grasses. *Plant Physiology* **149**: 171-180.

Yokoyama A, Yamashino T, Amano Y-I, Tajima Y, Imamura A, Sakakibara H, Mizuno T. 2007. Type-B ARR transcription factors, ARR10 and ARR12, are implicated in cytokinin-mediated regulation of protoxylem differentiation in roots of Arabidopsis thaliana. *Plant and Cell Physiology* **48**: 84-96.

Young MD, Wakefield MJ, Smyth GK, Oshlack A. 2010. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology* **11**: R14.
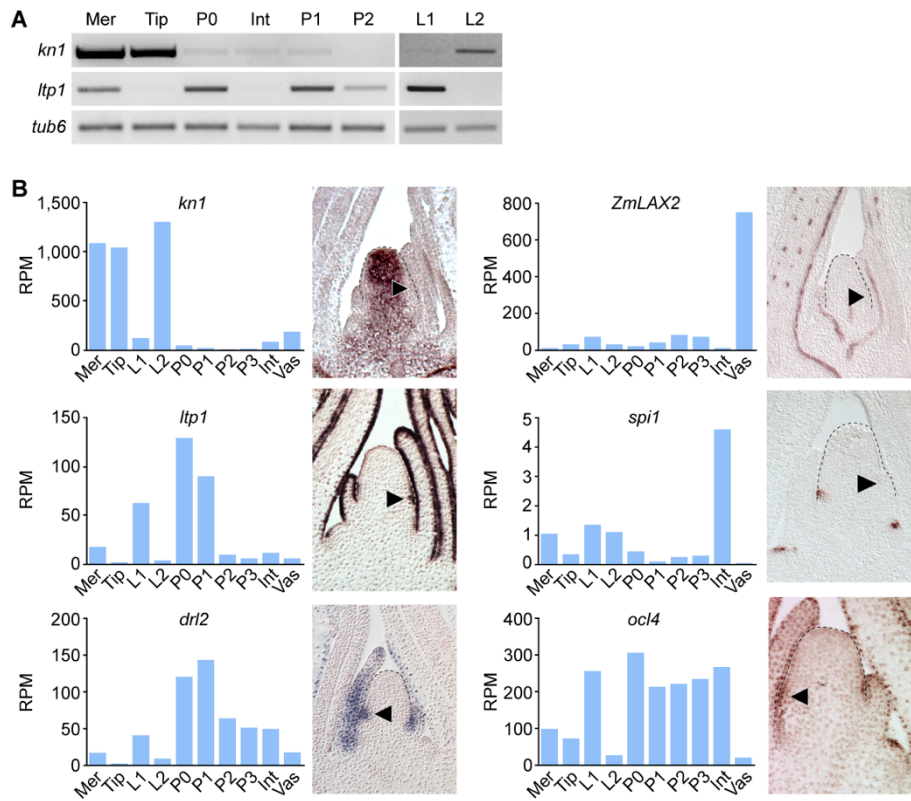
Figure S1. Precision of LCM. (*A*) RT-PCR for marker genes *kn1* and *ltp1* on amplified RNA from laser captured domains reveals a high precision of laser microdissection. Consistent with their known expression domains, *kn1* transcripts are present at high levels in the L2 of the meristem and Tip and barely detectable in the L1, internode, and developing leaf primordia, whereas *ltp1* transcripts are detec Table Specifically in the epidermis of incipient and developing leaf primordia. Mer, meristem; Int, internode. (*B*) Transcript levels for the marker genes *kn1, ltp1, outer cell layer4* (*ocl4*), *drooping leaf ortholog2* (*drl2*), *sparse inflorescence1* (*spi1*), and *like auxin resistant2* (*ZmLAX2*) across the domain RNAseq libraries reflect their previously reported *in situ* expression patterns, further confirming the precision of microdissections. See Table S15 for gene IDs and references.
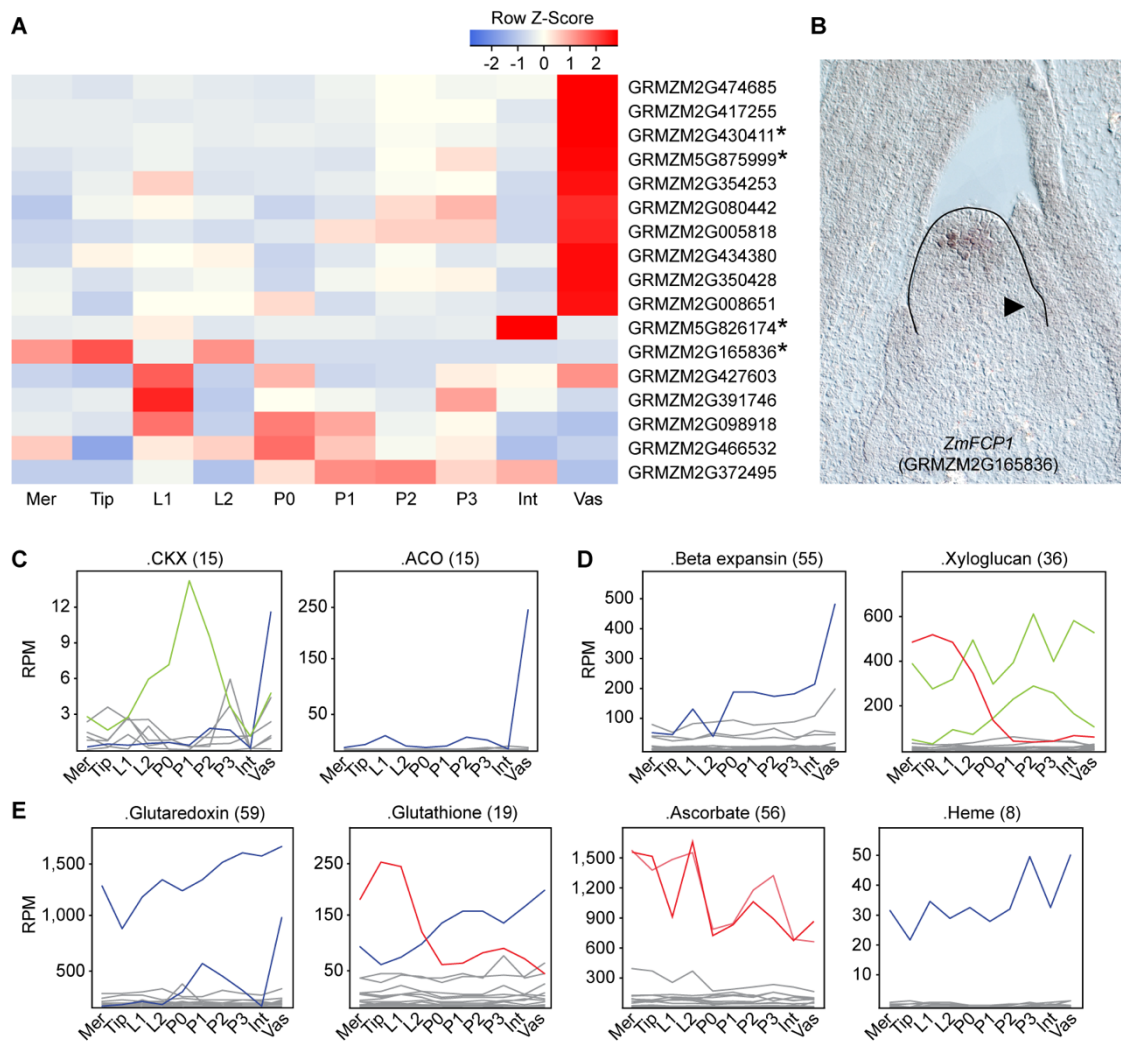
Figure S2. Predominant differential expression of individual gene family members distinguishes cell identities. (*A*) Heatmap displaying expression profiles of expressed *CLE* genes. Asterisk, cell type specific *CLEs*. (*B*) *In situ* hybridization shows *FCP1* transcripts accumulate in the L2 of the Tip. Arrowhead, P0. (*C-E*) Additional examples of gene families in which individual or select members show abundant and differential expression across the apex. Examples are shown for gene families with functions related to hormone signaling (*C*), cell wall remodeling (*D*), or redox regulation (*E*). Number of members in the family is shown in parenthesis. Only genes expressed at levels ≥ 2 RPM are shown. Mer, meristem; Int, internode; Vas, vasculature.
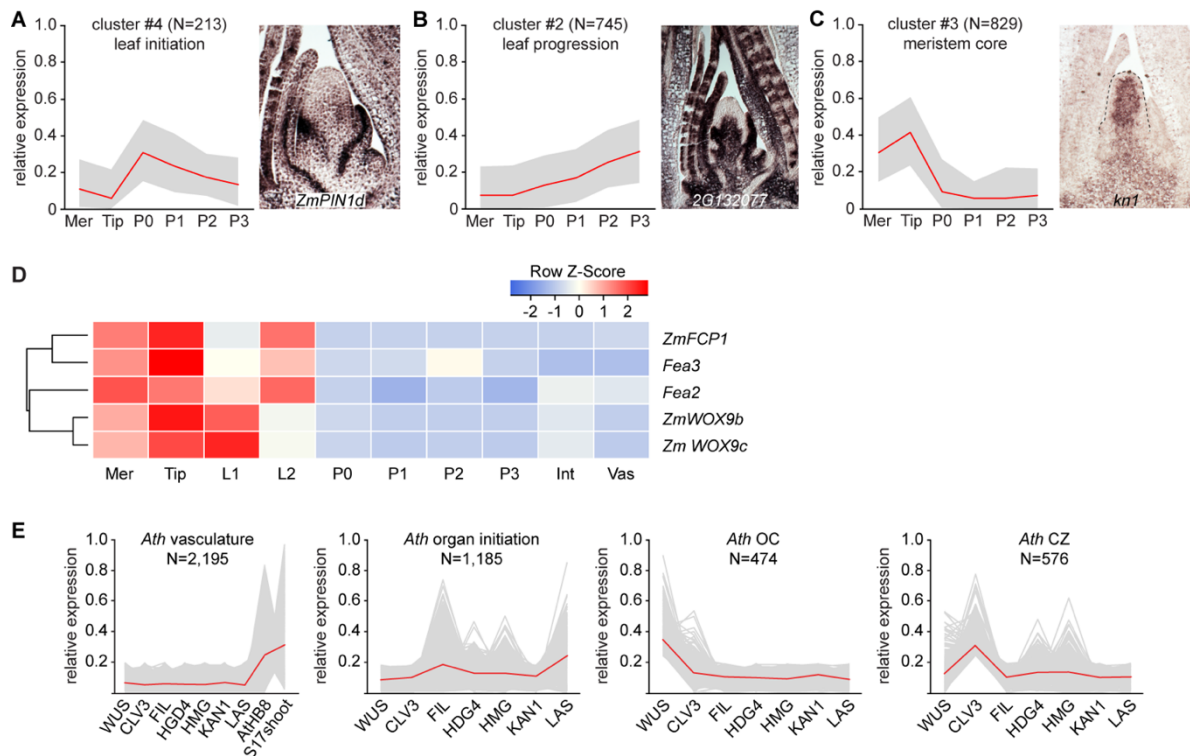
Figure S3. CAST clustering and transcriptional conservation. (*A-C*) Left: main gene clusters with expression profiles characteristic for leaf initiation (*A*), leaf progression (*B*) or meristem core identity (*C*). Right: representative *in situ* hybridization images. Red line, mean expression values; Grey profile, range between highest and lowest expression values. (*D*) Heatmap displaying expression profiles of key genes in the proposed SAM CLE-WOX signaling module. *ZmFCP1* is expressed specifically in subepidermal layers of the CZ, expression overlaps with that of receptors *FEA2* and *FEA3*. In contrast, expression of *ZmWOX9b* and *ZmWOX9c* specifically marks the L1 of the CZ. Mer, meristem; Int, internode; Vas, vasculature. (*E*) Expression profiles of all domain-enriched *Arabidopsis* genes considered for comparative analyses. Red line, mean expression values. Grey lines, individual expression profiles; N, Number of genes; *Ath*, *Arabidopsis thaliana*.

A

B

log(*p* value)

Hormone metabolism
.Auxin
..Signal transduction
..Response
.BR
..Synthesis/degradation
.GA
.ABA
..Signal transduction
..Cytokinin
RNA
.Regulation of transcription
..AP2/EREBP
..ARF
..Homeobox
..Dof
..bHLH
..SRS
..GARP
..OFP
..GATA
..YABBY
..MYB
..TCP
..zf-HD
..MADS
..SBP
..WRKY
Cell wall
.Modification
..Beta expansin
..Alpha expansin
Signalling
.Receptor kinases
..LRR III
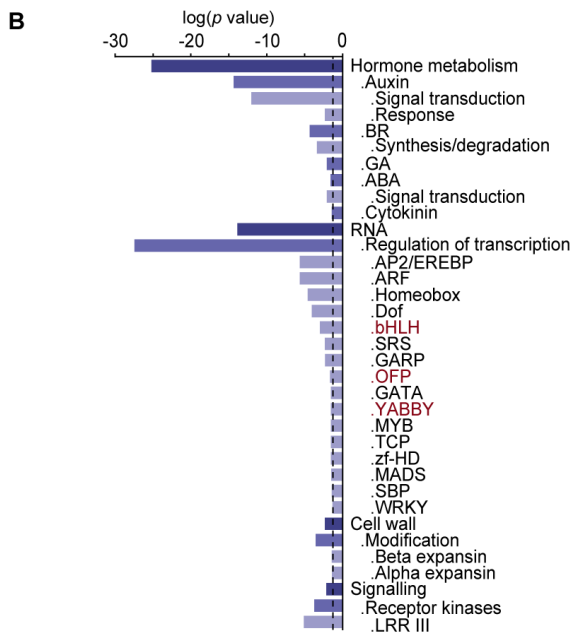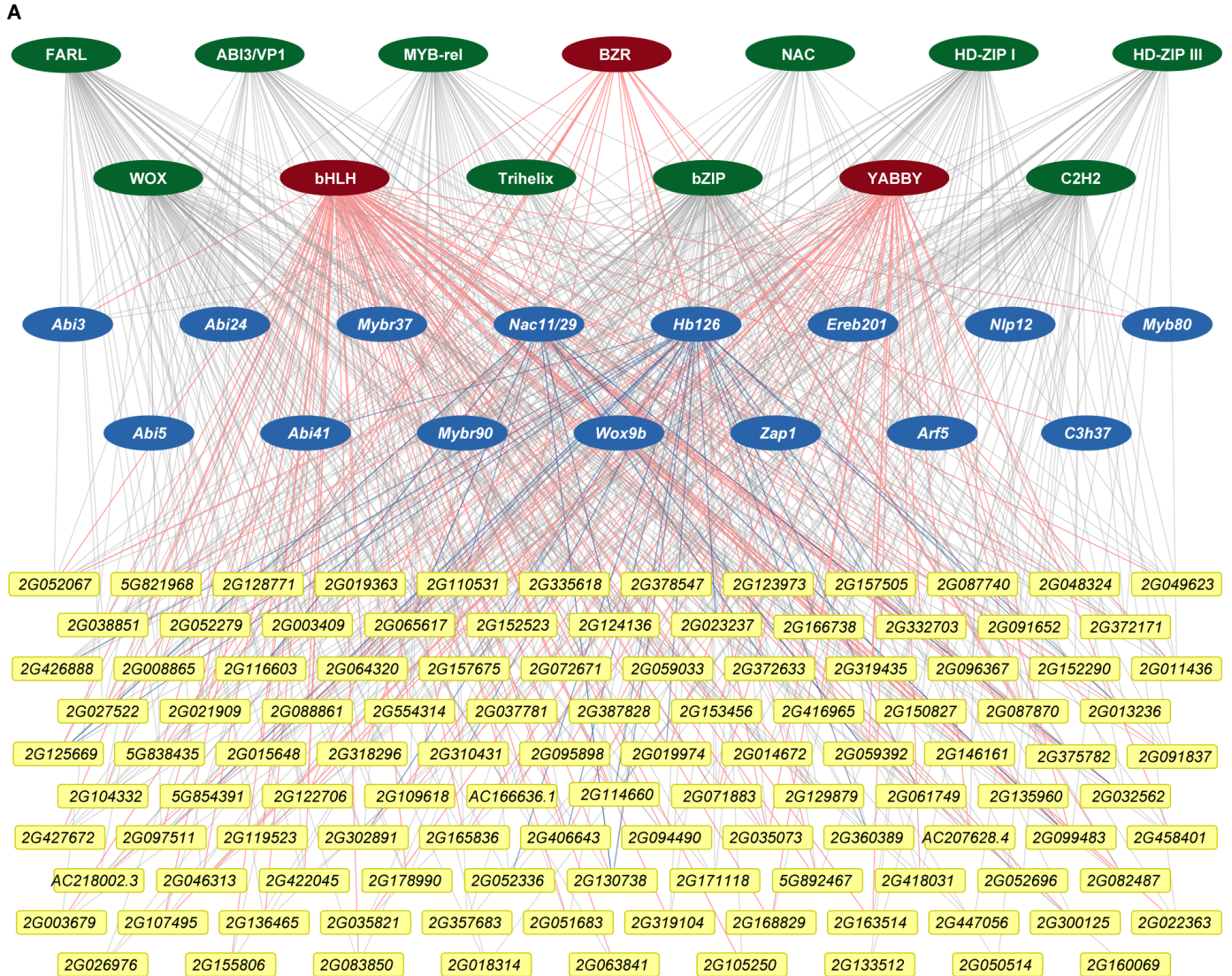
Figure S4. Combinatorial effects of diverse TFs distinguish cell identities. (*A*) Visualization of a GRN for all CZ-specific genes (yellow) with combinatorial interactions from TF families positively- (green) and negatively-correlated (red) to PC1, as well as individual CZ-specific TFs (blue). CZ-specific TFs (blue) can themselves be targeted by more broadly expressed PC1-correlated TF families revealing a hierarchy in the GRN. (*B*) Enrichment analysis on all expressed KN1 targets reveals KN1 preferentially targets genes connected to organogenesis. Color intensity reflects different Mapman category ranks. TF families shown in red are negatively-correlated to PC1 and meristem identity.
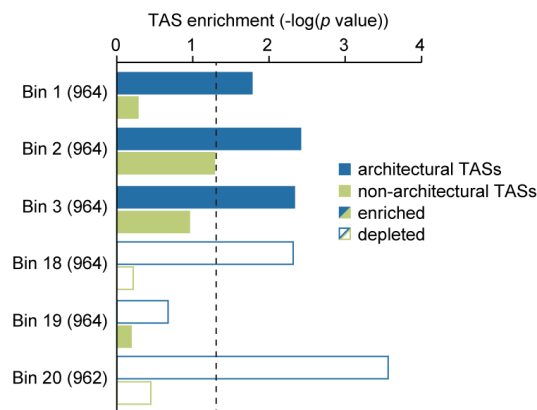
Figure S5. Dynamically expressed genes shape morphological variation. Significance plot illustrating that natural variation underlying plant architectural traits is significantly enriched near genes with highly dynamic patterns of expression across the apex. Apex-expressed genes were grouped into 20 bins based on SE score. Architectural TASs are significantly enriched near genes with low SE scores (Bins 1-3), whereas most constitutive expressed genes with high SE scores (Bins 18-20) are depleted for TASs. Dashed line, significance threshold (*p* = 0.05) based on Chi-square test with Yates' continuity correction.

Table S1 to 16 (separate file)

Appendix III. Nonrandom domain organization of the Arabidopsis genome at the nuclear periphery

Nonrandom domain organization of the *Arabidopsis* genome at the nuclear periphery

Xiuli Bi,[1,5] Ying-Juan Cheng,[2,3,5] Bo Hu,[1] Xiaoli Ma,[1] Rui Wu,[4] Jia-Wei Wang,[2] and Chang Liu[1]

[1]*Center for Plant Molecular Biology (ZMBP), University of Tübingen, Tübingen 72076, Germany;* [2]*National Key Laboratory of Plant Molecular Genetics (NKLPMG), CAS Center for Excellence in Molecular Plant Sciences, Institute of Plant Physiology and Ecology (SIPPE), Shanghai Institutes for Biological Sciences (SIBS), Shanghai 200032, People's Republic of China;* [3]*University of Chinese Academy of Sciences, Shanghai 200032, People's Republic of China;* [4]*Department of Molecular Biology, Max Planck Institute for Developmental Biology, Tübingen 72076, Germany*

The nuclear space is not a homogeneous biochemical environment. Many studies have demonstrated that the transcriptional activity of a gene is linked to its positioning within the nuclear space. Following the discovery of lamin-associated domains (LADs), which are transcriptionally repressed chromatin regions, the nonrandom positioning of chromatin at the nuclear periphery and its biological relevance have been studied extensively in animals. However, it remains unknown whether comparable chromatin organizations exist in plants. Here, using a strategy using restriction enzyme–mediated chromatin immunoprecipitation, we present genome-wide identification of nonrandom domain organization of chromatin at the peripheral zone of *Arabidopsis thaliana* nuclei. We show that in various tissues, 10%–20% of the regions on the chromosome arms are anchored at the nuclear periphery, and these regions largely overlap between different tissues. Unlike LADs in animals, the identified domains in plants are not gene-poor or A/T-rich. These domains are enriched with silenced protein-coding genes, transposable element genes, and heterochromatic marks, which collectively define a repressed environment. In addition, these domains strongly correlate with our genome-wide chromatin interaction data set (Hi-C) by largely explaining the patterns of chromatin compartments, revealed on Hi-C maps. Moreover, our results reveal a spatial compartment of different DNA methylation pathways that regulate silencing of transposable elements, where the CHH methylation of transposable elements located at the nuclear periphery and in the interior are preferentially mediated by CMT2 and DRM methyltransferases, respectively. Taken together, the results demonstrate functional partitioning of the *Arabidopsis* genome in the nuclear space.

[Supplemental material is available for this article.]

The spatial organization of the genome within the nucleus is critical for many cellular processes (Van Bortle and Corces 2012). It is broadly accepted that the packing of chromatin inside the nucleus is not random, but is structured into several hierarchical levels (Gibcus and Dekker 2013). Cytological studies have shown that within the nucleus, each chromosome occupies a distinct domain known as the chromosome territory (CT). In *Arabidopsis thaliana*, CTs in interphase nuclei were unequivocally demonstrated with chromosome painting, which further revealed a predominantly random arrangement of CTs with respect to each other (Pecinka et al. 2004). Recent *Arabidopsis* Hi-C experiments also revealed many structural features of plant chromatin packing at both the chromosomal and the local levels (Feng et al. 2014b; Grob et al. 2014; Wang et al. 2015). At a gene level, several studies in *Arabidopsis* demonstrated an association between chromatin loops and gene transcriptional activity, which involves a diverse spectrum of regulatory factors (Crevillén et al. 2013; Liu et al. 2013; Ariel et al. 2014; Cao et al. 2014). On the other hand, the nonrandom location of chromatin segments with different biological properties within the nuclear space has long been documented. In *Arabidopsis*, densely packed and aggregated heterochromatin (chromocenters) is often tethered at the nuclear envelope, whereas telomeres often cluster and reside in the nuclear interior around the nucleolus (Armstrong et al. 2001; Fransz et al. 2002). Another recent study demonstrated structural and regulatory roles of chromatin associated with the nuclear matrix in plants (Pascuzzi et al. 2014). Together with experiments showing global rearrangement of chromatin triggered by various environmental and developmental factors, such as light (Barneche et al. 2014; Bourbousse et al. 2015), microbial infection (Pavet et al. 2006), and cell differentiation (Tessadori et al. 2007), all these studies highlight a close interaction between chromatin structure and function in plants.

Chromatin positioning at the nuclear periphery in animals has been extensively studied. The nuclear lamina is a layer of meshwork beneath the nuclear envelope, consisting of lamin and lamin-associated membrane proteins (Dechat et al. 2008). The nuclear lamin was found to participate in organizing chromatin structures by serving as an anchoring site for heterochromatin (for review, see Mattout et al. 2015). Genome-wide identification of chromatin regions associated with the nuclear lamina in animals led to the discovery of lamin-associated domains (LADs), which are large-sized, depleted of active histone marks, and low in gene density (Pickersgill et al. 2006; Guelen et al. 2008). On the other hand, the nuclear pore complex (NPC), a giant protein

1162 Genome Research
www.genome.org
27:1162–1173 Published by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/17; www.genome.org

complex located at the nuclear envelope, has also been shown to play a role in tethering chromatin. Based on studies in yeast and several animal species, genes positioned close to the NPC tend to be highly transcribed (Strambio-De-Castillia et al. 2010).

Except for cytological studies showing a preferential association of chromocenters with the nuclear envelope, little is known about chromatin positioning at the nuclear periphery in plants. This is largely because plant genomes do not encode proteins with sequences similar to those of nuclear lamins in animals, although in several plant species, a meshwork similar to the nuclear lamina beneath the nuclear envelope has been observed (Ciska and Moreno Díaz de la Espina 2014). Nevertheless, over the past few years, a group of plant-specific nuclear matrix constituent proteins (NMCPs), such as CROWDED NUCLEI (CRWN) in *Arabidopsis*, have emerged as "plant lamina" components (Ciska and Moreno Díaz de la Espina 2014; Zhou et al. 2015). It appears that plant lamina components are distinct from those of animals, as another recently identified candidate, KAKU4, is also plant-specific (Goto et al. 2014). Moreover, NPC components have been systematically identified and investigated (Tamura et al. 2010; Tamura and Hara-Nishimura 2013; Parry 2015). These recent advances in knowledge provide opportunities for in-depth studies on various biological processes that occur at the plant nuclear pe-

riphery. In the present study, we identified and characterized *Arabidopsis* chromatin regions preferentially associated with the nuclear periphery on a genome-wide scale.

Results

RE-mediated ChIP reveals nonrandom chromatin distribution at the nuclear periphery

As a part of the NPC basket, the *Arabidopsis* nucleoporin NUP1 (also known as NUP136) has been shown to specifically localize at the nuclear periphery (Lu et al. 2010; Tamura et al. 2010). In our first attempt, we sought to use the NUP1 protein, tagged with green fluorescent protein (GFP), to identify chromatin that directly interacts with NPC, which might be related to the "gene gating" events that have been demonstrated in yeast and animals (Blobel 1985; Strambio-De-Castillia et al. 2010). In agreement with previously reported results, the NUP1:GFP fusion protein was localized specifically at the nuclear envelope (Fig. 1A). With a regular chromatin immunoprecipitation (ChIP) method, however, we could not identify any chromatin regions enriched by NUP1:GFP, even with our ChIP-seq libraries being sequenced more deeply than typically needed for *Arabidopsis* (Supplemental Table S1).
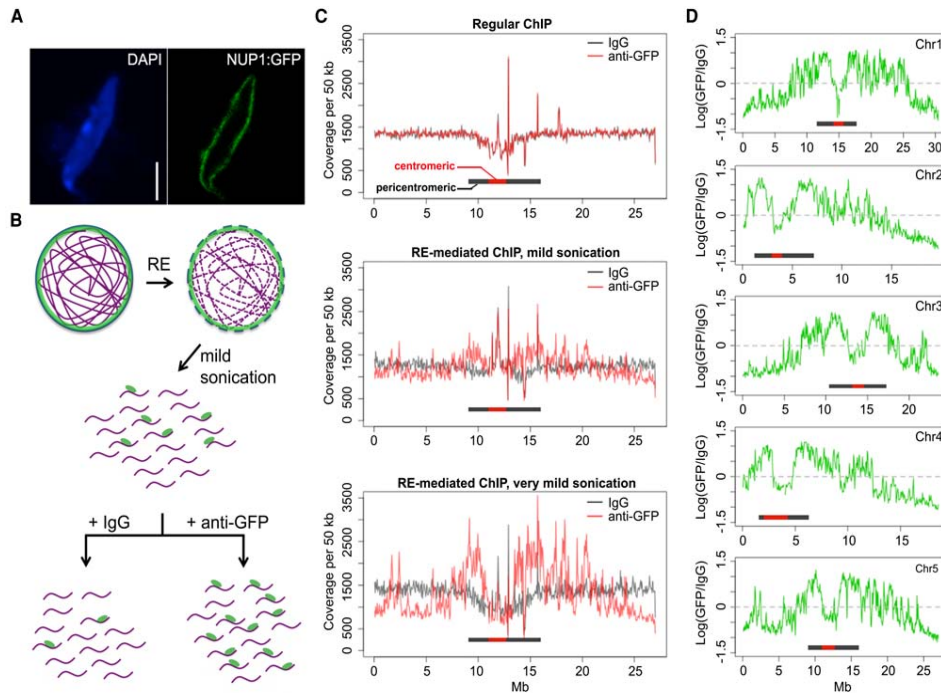
Figure 1. Identification of chromatin located at the nuclear periphery by RE-ChIP. (*A*) Localization of the NUP1:GFP protein in an *Arabidopsis* nucleus: (scale bar) 2 μm. (*B*) Procedures for RE-mediated ChIP with NUP1:GFP (green). Chromatin (purple lines) fragmentation and isolation are conducted with a combination of RE (restriction enzyme) digestion and mild sonication. (*C*) Normalized sequence coverage (50-kb window size) on Chromosome 5 from various ChIP experiments. The horizontal bars depict pericentromeric regions, within which centromeric regions are highlighted in red. (*D*) NUP1:GFP RE-mediated ChIP-seq signal (50-kb window size), represented as the $\log_2$ value of the ratio between normalized anti-GFP and IgG coverage, over all five chromosomes. Horizontal bars indicate the centromeric/pericentromeric regions, as in *C*.

In contrast, a parallel ChIP experiment conducted on the same material, but with an antibody against RNA polymerase II, showed expected enrichment on a housekeeping gene (Supplemental Fig. S1), ruling out possible technical failures in our ChIP experiments. This negative result implied that NUP1:GFP did not directly interact with chromatin, or that such interactions, if they occurred, were not efficiently preserved by the crosslinking treatment in our ChIP experiment. To enhance the sensitivity of enriching chromatin loosely interacting with NUP1:GFP, we developed a restriction enzyme (RE)-mediated ChIP protocol, in which only mild sonication was applied to break the nuclei following digestion of chromatin with RE (Fig. 1B; Methods). In principle, compared to a regular ChIP method, in which chromatin is sheared into small fragments by much stronger sonic waves, RE-mediated ChIP causes less disruption to higher-order structures; in the case of NUP1:GFP ChIP, this method allows enrichment of the chromatin positioned around the NPC, or the nuclear periphery.

We performed two RE-mediated ChIP-seq trial experiments with different sonication intensities and examined the sequence coverage with a 50-kb window setting to gain an overview of the distribution of sequencing reads. Interestingly, the RE-mediated ChIP with NUP1:GFP (hereafter referred to as NUP1 RE-ChIP-seq) showed that chromatin in pericentromeric regions was generally enriched, whereas chromatin on the distal chromosome arms tended to be depleted (Fig. 1C,D). Moreover, we found many interstitial regions on the chromosome arms showing stronger contact with NUP1:GFP (e.g., an interval corresponding to 2.0–3.0 Mb on Chromosome 5) (Fig. 1D). In contrast, there were regions close to pericentromeric chromatin but that exhibited depleted contact with NUP1:GFP (e.g., an interval corresponding to 9.8–10.2 Mb on Chromosome 3) (Fig. 1D). These patterns, which were clearly correlated with sonication strength, could not be seen using our regular ChIP-seq assay (Fig. 1C, top). To validate the RE-ChIP method, we performed RE-ChIP-seq with anti-H3K9me2 (Supplemental Table S1) and compared the results with those derived from a regular ChIP-seq (Stroud et al. 2014). Consistent with the fact that the *Arabidopsis* pericentromeric heterochromatin is heavily marked by H3K9me2, our RE-ChIP clearly captured this epigenomic feature at a global level (Supplemental Fig. 2A). In addition, chromatin regions enriched by RE-ChIP-seq largely overlapped with those enriched by the regular ChIP-seq, by which >80% of chromatin regions enriched by ChIP-seq were found enriched in each RE-ChIP-seq replicate (Supplemental Fig. 2B,C), indicating the feasibility of the RE-ChIP method in capturing chromatin features in plants. Furthermore, selected regions showing higher NUP1 RE-ChIP signals could be confirmed with fluorescence in situ hybridization (FISH) (Supplemental Fig. 3). Taken together, our results suggest that certain chromatin regions on the *Arabidopsis* chromosome arms are preferentially found near the nuclear periphery.

It has been well demonstrated that chromocenters, which consist of the centromere and pericentromeric regions, are located preferentially at the nuclear periphery (Fransz et al. 2002; Fang and Spector 2005). It was interesting that chromatin from centromeres was not enriched in our NUP1 RE-ChIP-seq experiments (Fig. 1D). A possible scenario accounting for this observation is that NPCs (or at least NUP1-containing NPCs) are not evenly distributed at the nuclear envelope, such that the NPC density is lower in regions where chromocenters are anchored. For instance, kinetochore proteins interact with *Arabidopsis* centromeres in almost all stages of the cell cycle (Lermontova et al. 2013), and Gamma-tubulin Complex Protein 3-interacting proteins (GIPs) play essential

roles in centromere assembly (Batzenschlager et al. 2015); these interactions might sequester centromeres away from NPCs. We examined nuclei in transgenic plants coexpressing NUP1:GFP and mCherry:CENH3, in which the latter was exclusively loaded to centromeres (Lermontova et al. 2006). We found that these two proteins displayed complementary patterns at the nuclear periphery, which explained our observation that centromeres were not enriched by NUP1:GFP (Supplemental Fig. 4A). Apart from this, consistent with the fact that pericentromeric chromatin is mostly found at the nuclear periphery, the chromatin regions belonging to the remaining pericentromeric regions showed the highest NUP1:GFP RE-ChIP signals (Supplemental Fig. 4B–F).

Chromatin positioned at the nuclear periphery correlates with Hi-C map

The conformation of the *Arabidopsis* genome in the nuclear space has been recently revealed by several Hi-C experiments (Feng et al. 2014b; Grob et al. 2014; Wang et al. 2015). The Hi-C method combines chromatin conformation capture (3C) and high-throughput sequencing to generate a comprehensive view of how chromatin is folded (Lieberman-Aiden et al. 2009). Due to the nature of this method, the Hi-C data only contains information on the positioning of chromatin with respect to itself (chromatin folding). As our NUP1:GFP RE-ChIP-seq data focused on chromatin localization with respect to the nuclear boundary, we considered whether this nonoverlapping information could help us gain a better understanding of chromatin organization in the nuclei. Interestingly, we found that NUP1:GFP RE-ChIP-seq signals were strongly correlated with structural domains (SDs) derived from the *Arabidopsis* Hi-C map (Grob et al. 2014), which could be visualized when the chromosome arms were partitioned into two states using principal component analysis (PCA) (Fig. 2A). It is worth noting that such two-state classification, initially referred to as "AB" compartments, was found to be strongly correlated to the demarcation of active/repressed chromatin along the chromosome arms (Lieberman-Aiden et al. 2009; Grob et al. 2014), and a connection between animal LADs and the repressed compartment was recently shown on a global scale (Vieux-Rochas et al. 2015). We found that chromatin regions that had stronger contact with the nuclear periphery (with stronger NUP1:GFP RE-ChIP-seq signals) were mostly classified as the repressed compartment (Fig. 2A). Therefore, our results indicate that the "AB" compartments of *Arabidopsis* chromatin are associated with a radial axis of nuclei, further indicating that repressed chromatin is enriched at the nuclear periphery.

The *Arabidopsis* telomeres, except for those on the short arms of Chromosomes 2 and 4, are located around nucleoli (Armstrong et al. 2001; Fransz et al. 2002; Pontvianne et al. 2016). In general, chromatin on the distal chromosome arms exhibits a positive correlation with telomeres on a Hi-C map due to physical linkages, and this correlation gradually drops when the genomic distance increases. By checking the correlation between distal chromosome arms and telomeres, we observed local valleys, most of which colocalized with the local peaks of NUP1:GFP RE-ChIP-seq data (Fig. 2B). Because these local valleys of correlation indicate decreased chromatin interactions in the 3D space, this pattern agrees with the fact that *Arabidopsis* telomeres are preferentially found in the nuclear interior. We also found that some local valleys colocalized with interactive heterochromatic islands (IHIs)/knot-engaged elements (KEEs), which are regions showing strong intra- and interchromosomal interactions on Hi-C maps (Fig. 2B; Feng et al. 2014b; Grob et al. 2014). However, a comparison between KEEs
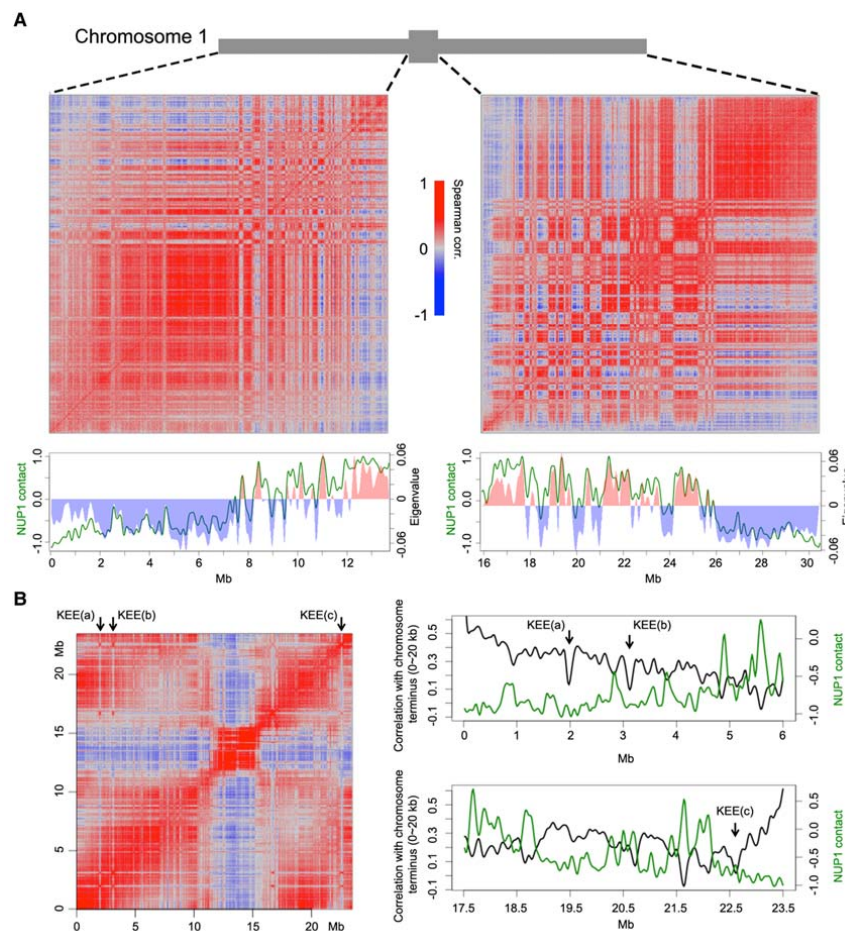
Figure 2. Correlation between chromatin anchored at the nuclear periphery and the Hi-C map. (*A*) Correlation between NUP1:GFP RE-ChIP-seq signal and Hi-C map. The Hi-C maps (normalized at 20-kb resolution) of the *left* and *right* Chromosome 1 arms are shown as Spearman correlation matrices, from which PCA was conducted; the eigenvalues of the first component are plotted *below* (red and blue bars) together with the NUP1:GFP signal (green lines, 20-kb window size), represented as the $\log_2$ value of the ratio between normalized anti-GFP and IgG coverage. (*B*) Anti-correlation between the telomeres and NUP1:GFP RE-ChIP-seq signal. The *left* panel shows a Spearman correlation matrix of Chromosome 3 derived from a Hi-C map at 20-kb resolution. Arrows depict KEE regions. The *right* panels highlight the 6-Mb distal chromosome regions, in which their correlation with the chromosome terminus (the first 20 kb of Chromosome 3) in the Hi-C map are shown as black curves. Green curves show the NUP1:GFP signal, as in *A*. Due to physical linkage, chromosome termini are expected to have strong colocalization with telomeres in the nucleus. In a Hi-C experiment, chromosome termini can be used to infer the spatial interactions between telomeres and other genomic regions.

and NUP1:GFP ChIP-seq did not reveal a connection between them (Supplemental Fig. 5), suggesting that the clustering of IHIs/KEEs does not preferentially occur at the nuclear periphery.

Chromatin positioning at the nuclear periphery has similar patterns in different tissues

Our finding of the nonrandom positioning of chromatin at the nuclear periphery prompted us to further investigate the extent to which these patterns vary among different plant tissues. In total, we examined NUP1:GFP RE-ChIP-seq data generated from four different tissues (Supplemental Table S1). The signal patterns, as well as enriched genes, between biological replicates were highly reproducible in each tissue (Fig. 3A; Supplemental Figs. 6–8). Interestingly, at a chromosomal level, NUP1:GFP RE-ChIP-seq data obtained from different tissues resembled each other (Fig. 3A; Supplemental Fig. 6). A common feature across these tissues was that chromatin close to the centromere was more frequently

175
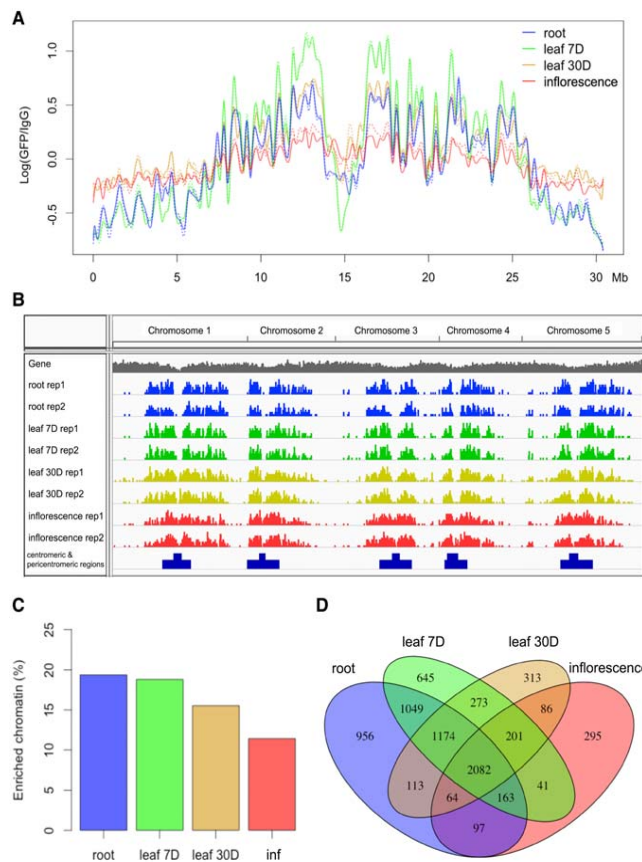
Figure 3. Genome-wide identification of NUP1-enriched regions in various tissues. (*A*) Signals of NUP1:GFP RE-ChIP-seq (20-kb window size), represented as the log$_2$ value of the ratio between normalized anti-GFP and IgG sequence coverage over Chromosome 1. For each tissue, the solid and dotted lines depict two replicates. (*B*) Distribution of NUP1-enriched domains across the genome viewed with the Integrative Genomics Viewer browser (Robinson et al. 2011). (*C*) Percentage of NUP1-enriched genomic regions: (inf) inflorescence. (*D*) Venn diagram of genes enriched in four tissues.
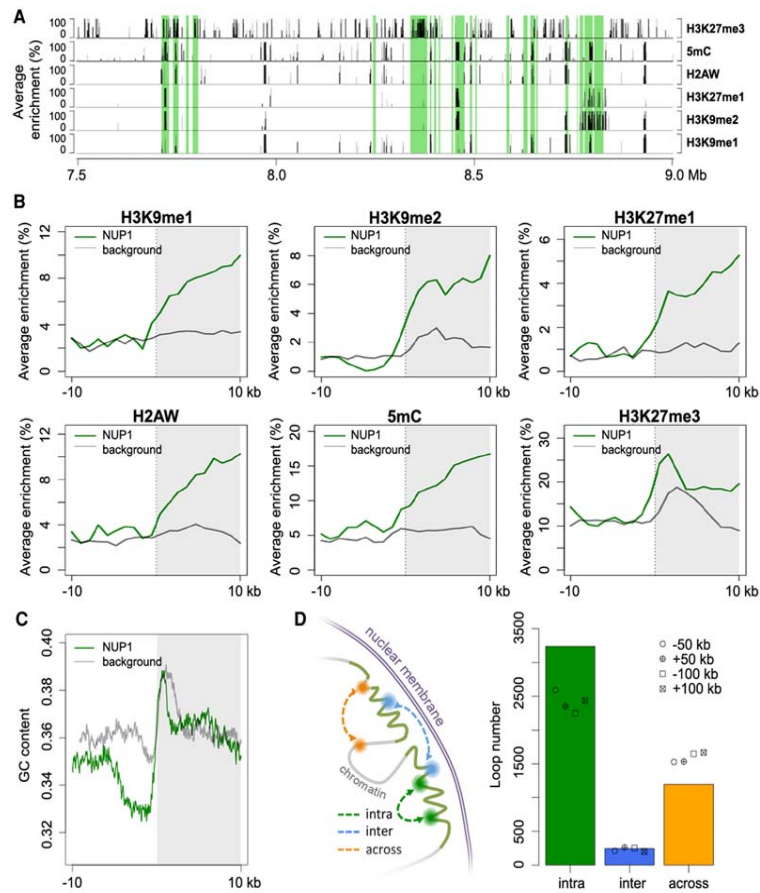
3C), with median sizes of 7–12 kb (Supplemental Fig. 9; Supplemental Table S2). A clustering analysis showed that roots and leaves from 7-d-old seedlings formed a subgroup, although from a tissue-identity point of view, leaf tissues with different ages would be expected to be clustered together (Supplemental Fig. 7). Nevertheless, due to the similar RE-ChIP-seq signal profiles on a chromosomal scale (Fig. 3A; Supplemental Fig. 6), there were substantial overlaps of enriched chromatin regions between any two given tissues (Fig. 3B,D). These results suggest that although both the linear genome structure and the tissue identity contribute to the way chromatin is tethered at the nuclear periphery, the former is the primary determinant.

Heterochromatic domains are enriched at the nuclear periphery in *Arabidopsis*

We next explored the genomic and epigenomic features associated with chromatin positioned at the nuclear periphery. As these chromatin regions were preferentially located around centromeres, we expected that features linked to the centromeric and pericentromeric regions would be enriched. To reduce such positional effects, we only included chromatin located at least 1 Mb from pericentromeric heterochromatin for all analyses described below (unless otherwise stated). Of note, our analyses in this study were not sensitive to a cutoff that we arbitrarily set. We obtained the same conclusions when we changed the cutoffs to 2 or 3 Mb, in which more genomic regions flanking the pericentromeric regions were excluded.

Our association analyses of the epigenetic and genomic features around NUP1-enriched domain boundaries showed similar epigenetic landscapes compared to those of animal LADs, but there were significant differences in terms of other genomic features. For example, the NUP1-enriched domains were enriched with classic heterochromatic marks, such as H3K9 methylation, which has been linked to LADs (Towbin et al. 2012), and H3K27me3, which has been shown to enhance the association of chromatin to the inner nuclear membrane (Fig. 4A,B; Harr et al. 2015). Accordingly, the level of euchromatic marks was lower inside NUP1-enriched domains (Supplemental Fig. 10). On the other hand, in contrast to LADs, NUP1-enriched domains were neither substantially depleted with protein-coding genes (Fig. 5A; Supplemental Figs. 11, 12A) nor enriched with A/T-rich sequences (Fig. 4C). It is not clear whether the NUP1-enriched domain boundaries are bound with insulator proteins, as they have not yet been identified in plants. Nevertheless, we found that chromatin loops connecting regions inside and outside the NUP1-enriched domains

found at the nuclear periphery than was chromatin on the distal chromosome arms, which was reflected on a density plot showing the distribution of enriched chromatin segments (Fig. 3B). We noticed that in inflorescences, the difference in RE-ChIP-seq signal amplitudes between the pericentromeric regions and distal chromosome arms became much smaller, implying a lower selectivity in positioning specific chromatin regions at the nuclear boundary in reproductive tissues (Fig. 3A; Supplemental Fig. 6). Notably, for each plant tissue used in this study, the RE-ChIP output signal was the average of a mixture of different cell types. For the inflorescence tissue, compared to roots and leaves, the lower RE-ChIP signals around the pericentromeric regions might also be attributed to a dilution effect among different cells with divergent chromatin positioning patterns. Across different tissues, the regions enriched at the nuclear periphery covered 10%–20% of the genome (Fig.

Figure 4. Epigenetic, genomic, and structural features of chromatin tethered at the nuclear periphery. (*A*) A representative genomic region from Chromosome 1 showing the distributions of NUP1-enriched chromatin identified from 7-d-old leaf tissues (shaded in green) and various epigenetic marks. Average enrichment means the percentage of regions (calculated from 100-bp windows) enriched for the respective epigenetic mark. (*B,C*) Epigenetic marks (*B*) and GC content (*C*) around NUP1-enriched domain borders, shown as a vertical line separating the white and gray blocks. For each plot, the area on the *right* indicates NUP1-enriched domains (although not all are larger than 10 kb). Average enrichment in *B* is defined as in *A*. The GC content in *C* is in a window size of 100 bp, with a step size of 20 bp. Because enrichment of gene bodies is found inward from NUP1-enriched domain boundaries (see Supplemental Fig. 12), for the background, we randomly picked 3000 genes with the same expression distribution profile as that of NUP1-enriched genes. For these control genes, we extracted the 20-kb regions flanking either their transcription start sites or their transcription termination sites, which were selected randomly. (*D*) Different types of chromatin loops associated with NUP1-enriched domains (including those in pericentromeric regions). Chromatin loops are from Liu et al. (2016). For both "intra" and "across," the number of observed chromatin loops are significantly different ($P < 2.2 \times 10^{-16}$) relative to the permutation-based null distribution of the background, which was simulated by shifting the coordinates of NUP1-enriched domains ±50 kb or ±100 kb.

were underrepresented; whereas chromatin interactions restricted within one NUP1-enriched domain were overrepresented (Fig. 4D). To a certain extent, this pattern is analogous to that of topologically associating domains (TADs), which are predominant structural units of higher-order chromatin architecture in many metazoan genomes (Dixon et al. 2012). Typically, chromatin interactions within a TAD are observed more often than expected, whereas those across TAD boundaries are underrepresented. As chromatin inside NUP-enriched domains showed suppressed inter-

actions with outside regions in our study, from a spatial point of view, these domains represented structures that were isolated from their surroundings.

The *Arabidopsis* nuclear periphery is enriched for repressed chromatin

Based on gene annotations, we found that in all tissues, transposable element (TE) genes and pseudogenes were enriched
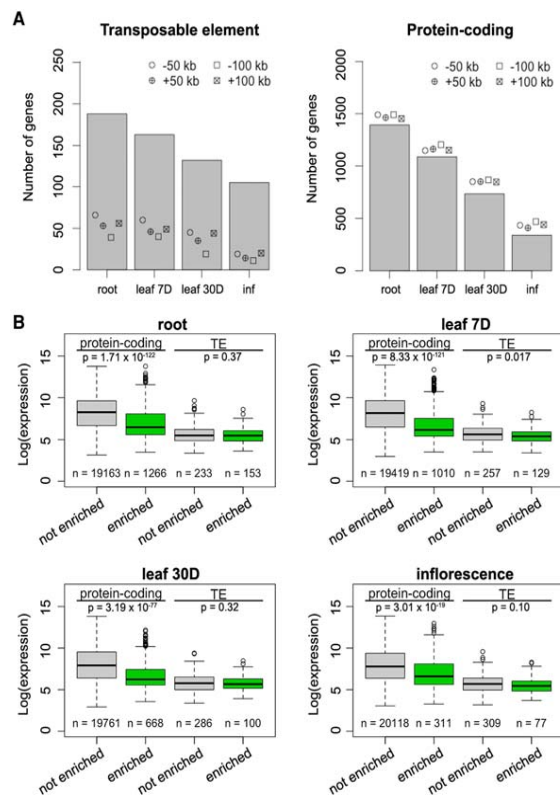
Figure 5. Enrichment of silenced genes at the nuclear periphery. (*A*) Number of TE genes (*left*) and protein-coding genes (*right*) enriched in different tissues. For each column, the observed number of genes is significantly different ($P < 0.001$) relative to the permutation-based null distribution of the background (generated as described in Fig. 4C): (inf) inflorescence. (*B*) Comparison of gene expression levels, which are from a normalized tilling array data set (Laubinger et al. 2008). The *P*-values indicate Mann-Whitney *U* test results.

compared to the control sets, which were simulated by shifting the coordinates of the enriched regions a certain distance upstream or downstream (Fig. 5A; Supplemental Fig. 11; Supplemental Table S3). We considered these control sets to be more appropriate than random permutations, as they maintained the distribution pattern of the enriched regions across the genome. In terms of transcriptional activity, the enriched protein-coding genes clearly had lower expression levels (Fig. 5B), which aligned with our finding that chromatin positioned at the nuclear periphery was generally repressed (Figs. 2, 4). Interestingly, for genes with a transcription direction toward the interior of the NUP1-enriched domains, we observed a higher occurrence of transcription start sites (TSSs) of genes with low transcription levels around domain boundaries (Supplemental Fig. 12B), suggesting a role of gene bodies in demarcating these chromatin domains. On the other hand, TE genes enriched at the nuclear periphery did not show lower expression levels than those that were not enriched (Fig. 5B). Instead, these two types of TE genes differed in terms of their lengths and locations with respect to protein-coding genes; TE genes located

at the nuclear periphery were significantly longer and were located further from protein-coding genes (Supplemental Fig. 13). Additionally, TE genes enriched at the nuclear periphery showed a higher average level of heterochromatin marks, such as H3K9me2, H3K27me1, and DNA methylation (Supplemental Fig. 14). In terms of classes, two class-II TE genes (MuDR and CACTA-like) were more likely to be found at the nuclear periphery (Supplemental Fig. 15). Taken together, the *Arabidopsis* nuclear periphery defines a domain of transcriptional repression enriched for TE genes and transcriptionally inactive protein-coding genes.

Positioning of TEs at the nuclear periphery correlates with different silencing pathways

Having shown that the nuclear peripheral zone was repressed, we next investigated whether it was connected to silencing of TEs. DNA methylation in the CG, CHG, and CHH (H representing any nucleotide except G) sequence context plays a crucial role in regulating expression and transposition of TEs. We noticed that TEs enriched at the nuclear periphery had a higher DNA methylation level in all sequence contexts (Fig. 6). We next examined several DNA methylation mutants by asking how the corresponding types of methylation would change in these two types of TEs. Regardless of TE location in the nuclear space, mutations impairing CG or CHG methylation showed similar effects (Fig. 6). Interestingly, when comparing the CHH methylation patterns, we found that TEs located at the nuclear periphery lost more DNA methylation in the *cmt2* mutant; in contrast, TEs not located at the nuclear periphery lost more DNA methylation in the *drm1/2* double mutant (Fig. 6). These patterns were also observed when we focused on TEs located in the pericentromeric regions (Supplemental Fig. 16). CHH methylation over TE bodies is mediated by two partially overlapping pathways: RNA-directed DNA methylation (RdDM) and RdDM-independent (Zemach et al. 2013; Stroud et al. 2014). However, it is not clear how these two pathways branch to target different TEs (for review, see Sigman and Slotkin 2016). DOMAINS REARRANGED METHYLASE 1 (DRM1) and DRM2 are responsible for CHH methylation in the RdDM pathway, whereas CHROMOMETHYLASE 2 (CMT2) is required for the RdDM-independent pathway (Cao et al. 2003; Stroud et al. 2013; Zemach et al. 2013). Our results reveal a spatial association between TE locations and the demand on different CHH methylation pathways, in which CHH methylation of TEs located at the nuclear periphery tends to be more dependent on CMT2, whereas the other type of TE relies more on RdDM.
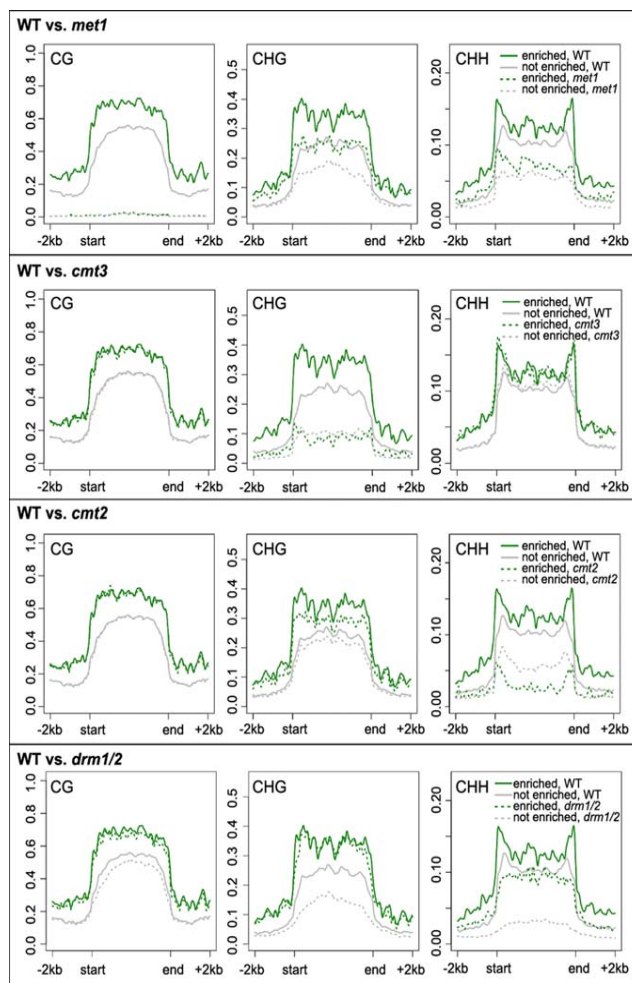
Figure 6. Comparison of DNA methylation over TEs. Patterns of TE DNA methylation (CpG, CHG, CHH) in wild-type (WT) and mutants. The grouping of TEs is according to the enrichment results of NUP1:GFP RE-ChIP-seq from 30-d-old leaf tissues. The methylation ratio is calculated in 100-bp windows. The signal over each TE is linearly transformed so that the boundaries of all TEs are aligned.

Discussion

It has long been recognized that the *Arabidopsis* chromocenters are preferentially found at the nuclear periphery, but it was unclear whether such nonrandom localization was restricted to chromocenters. The present study demonstrates that the peripheral zone of the *Arabidopsis* nucleus is also enriched with interstitial regions on the chromosome arms, which are mainly heterochromatic. This is reflected by the fact that these regions have higher percentages of TE genes and silenced protein-coding genes (Figs. 4, 5). In this regard, the nuclear periphery in plants is a functional compartment for docking repressed chromatin; therefore, the biologi-

cal properties of the nuclear periphery in eukaryotes are highly conserved. An intriguing question is which factors are involved in specifically tethering chromatin to the plant nuclear periphery. In animals, lamins and lamin-associated proteins have been identified as key factors involved in these events (for review, see Harr et al. 2016). Unfortunately, to identify the counterparts in plants, a strategy based on a protein-sequence similarity search might be of little use compared to forward genetics approaches, because plants lack orthologs of lamins and most lamin-associated proteins (Ciska and Moreno Díaz de la Espina 2014). Nevertheless, the CRWN and KAKU4 proteins in *Arabidopsis* have been suggested as plant lamin candidates, and *CRWN* mutations result in altered chromocenter structure (Wang et al. 2013). In the *crwn4* mutant, although the chromocenters decondensed, the regions corresponding to the 5S RNA repeats remained anchored at the nuclear periphery (Wang et al. 2013). This suggests the need to investigate higher-order *crwn* mutants to clarify their potential roles in tethering chromatin.

In this study, many highly expressed genes were also found to be enriched at the nuclear periphery (Supplemental Table S3), which could be at least attributed to following reasons. First, both the gene expression and RE-ChIP-seq experiments were conducted on tissues with a certain degree of cell-type heterogeneity; therefore, even if active transcription and positioning at the nuclear periphery are mutually exclusive, both events might be eventually captured at a gene locus in a mixed cell population. Second, the nuclear periphery does not absolutely inhibit transcription. This has been demonstrated in human cells by tracking the expression of a pool of genes after artificially anchoring them to the nuclear envelope; only a subset of the targeted genes showed reduced expression (Finlan et al. 2008). Specifically in *Arabidopsis*, a recently reported case study on the *CHLOROPHYLL A/B BINDING (CAB) PROTEIN* locus showed that it is repositioned from the nuclear interior to the nuclear periphery together with robust transcriptional activation in response to light stimuli (Feng et al. 2014a). Third, in plants, there may be "gene gating" events that position actively transcribed genes at the nuclear periphery through interactions with nucleoporins (Blobel 1985). Several potential interactions between transcription regulators (such as the TREX2 complex and SUMO proteases) and NPC have recently been discussed (Parry 2015). Although we did not detect any direct binding of NUP1 to chromatin with a regular ChIP-seq method, it remains

179

unknown whether other NPC components directly interact with chromatin.

Interestingly, the RE-ChIP signals from the root tissue, which consisted of nonmesophyll cells, were highly similar to those from leaves, with the majority cell type being mesophyll cells (Fig. 3A,B). Compared to other tissues, RE-ChIP signals from inflorescences showed a much lower extent of enrichment for chromatin at the nuclear periphery, which might be attributed to a dilution effect due to cell-type heterogeneity. Overall, the chromatin regions positioned at the nuclear periphery in plants tend to be conserved among different tissues, suggesting that in general, the linear genome per se contributes substantially to how it is deployed with respect to the nuclear peripheral zone on the chromosomal scale. By showing a tight association between this pattern and the "AB" compartment derived from Hi-C maps (Fig. 2A), we provide an additional way to visualize and understand plant Hi-C maps in the context of the nuclear space. Although Hi-C maps from *Arabidopsis* roots or inflorescences are not presently available, we suspect that the chromatin packing in these tissues would follow a scheme in common with that of leaves, and all these Hi-C maps would be strongly correlated with each other on a global level.

Having "AB" compartments in the nuclear space implies the existence of a radial gradient, offering a spatial specificity with which different regulatory pathways can regulate chromatin activity. Globally, this is reflected by the observation that chromocenters and telomeres preferentially reside at the nuclear periphery and nuclear interior (around the nucleolus), respectively (Fransz et al. 2002). By showing that TEs are selectively tethered at the nuclear periphery, our results reveal additional features of this spatial compartment. The differential loss of CHH DNA methylation on TEs in the RdDM and RdDM-independent mutants implies a spatial preference of these two TE-silencing pathways, in which RdDM is under more demand in the nuclear interior (Fig. 6). This correlates with observations that many small RNA pathway components are concentrated around nucleoli in *Arabidopsis* (Li et al. 2008; Pontes et al. 2013). From a spatial point of view, our results provide insights into how these two silencing pathways might collaborate to regulate TEs (Zemach et al. 2013), as well as how certain components of one pathway interact with each other (e.g., a recently reported positive feedback loop between Pol IV-dependent small RNA biogenesis and DRM2-dependent CHH methylation (Li et al. 2015)). It would also be of great interest to further investigate the possible dynamic locations of TEs in mutants of TE-silencing pathways.

Conventionally, chromatin regions that are preferentially tethered to the nuclear periphery could be identified by the ChIP method, such as those showcased in recent studies on animal nuclear lamin A and lamin B (Kubben et al. 2012; Lund et al. 2013; Sadaie et al. 2013; Shah et al. 2013). Additionally, an alternative approach is to anchor a modification enzyme on the nuclear envelope and trace its footprint on the genome (Kind et al. 2013). This method, which utilizes a DNA adenine methyltransferase (Dam) that methylates DNA on the N^6-adenine residue, has been applied to *Arabidopsis* to identify targets of LIKE HETEROCHROMATIN PROTEIN 1 (LHP1) as a complementary approach to the conventional ChIP method (Zhang et al. 2007). A potential limitation of these approaches is that the protein of interest must be in close contact with chromatin. In plants, however, these proteins' identities remain unknown. By performing RE-mediated chromatin fragmentation in combination with mild sonication, our RE-ChIP protocol alleviates the requirement that proteins of interest must directly interact with chromatin, because in principle the

RE-ChIP would better protect higher-order structures from destruction by strong sonic waves and would permit recovery of chromatin in the proximity of a protein of interest even when the interaction is not direct. In our opinion, this is a feasible method for identifying chromatin, if it is positioned close to other subnuclear structures, such as the nuclear matrix, nucleolus, and various nuclear bodies. The RE-ChIP method uses a restriction enzyme to digest chromatin; therefore, the chromatin fragmentation pattern is not random and is dependent on both the restriction-cutting site density and the digestion efficiency (Wang et al. 2015). Compared to regular ChIP methods, RE-ChIP cannot achieve resolution at the nucleosomal level and is not suitable for genome-wide identification of narrow peaks, such as the typical binding sites of transcription factors.

Methods

Plant material

Arabidopsis thaliana transgenic plants *NUP1:GFP* in the *nup136-1* (Salk_104728) background were grown at 23°C in long days (16 h light/8 h dark) on half-strength Murashige and Skoog (MS) medium supplemented with 1% sucrose and 0.3% Phytagel. The aerial and root tissues of 7-d-old seedlings were harvested at Zeitgeber time (ZT) 6 h. Other tissues, including 30-d-old leaf and inflorescence with flower bud up to stage 9 (Smyth et al. 1990), were collected from plants grown in growth rooms under long days at 23°C.

Plasmid construction

NUP1:GFP was constructed with an overlapping PCR strategy. The genomic fragment spanning 600 bp upstream of *NUP1* to the *NUP1* stop codon was amplified with primers 5′-GTTCGTTAG ACTGGTTTAGGT-3′ and 5′-TTTCTTCCTGGTGGATTTCTT-3′; the genomic fragment spanning the *NUP1* stop codon to 150 bp downstream from *NUP1* was amplified with primers 5′-TTTGGA GAAGAAGGCTTCTCT-3′ and 5′-TAAGAAAAACACATTGTTCAA G-3′; and GFP cDNA was amplified with primers 5′-AAGAAATCC ACCAGGAAGAAAGCGGCCGCTGTGAGCAAGGG-3′ and 5′-CTT GAACAATGTGTTTTTCTTAAGATCCACCAGTATCCTCAC-3′. These PCR products were mixed and assembled by overlapping PCR and amplified with primers 5′-GTTCGTTAGACTGGTTTAGGT-3′ and 5′-TTTGGAGAAGAAGGCTTCTCT-3′. The final PCR product, in which *GFP* was fused with *NUP1*, was cloned into a Gateway-compatible pGREEN-IIS binary destination vector (Karlsson et al. 2015). Similarly, to make the mCherry:CENH3 fusion protein, mCherry was amplified with primers 5′-GTAAAAATCAATGG CCATCATCAAGGAGTT-3′ and 5′-ACGCGATGCTTGGTTCTCGC ACCGCCACCCTTGTACAGCTCGTCCATGC-3′, cenH3 (AT1G 01370) was amplified with primers 5′-GCGAGAACCAAGCATCG CGT-3′ and 5′-TCACCATGGTCTGCCTTTTC-3′, these PCR products were assembled by overlapping PCR and amplified with primers 5′-GTAAAAATCAATGGCCATCATCAAGGAGTT-3′ and 5′-TC ACCATGGTCTGCCTTTTC-3′. The PCR product was cloned into a Gateway-compatible pGREEN-IIS binary destination vector containing a 35S promoter (Karlsson et al. 2015).

RE-ChIP-seq library preparation

Tissues were collected and fixed under vacuum for 30 min with 1% formaldehyde in MC buffer (10 mM potassium phosphate, pH 7.0; 50 mM NaCl; 0.1 M sucrose) at room temperature. After fixation, tissues were incubated at room temperature for 5 min under vacuum in MC buffer with 0.15 M glycine. Nuclei from 0.5 g fixed material were used for each round of ChIP. Nuclei were isolated as

described (Wang et al. 2015). Nuclei were permeabilized through incubation with 150 µL 0.5% SDS for 5 min at 62°C, and SDS was quenched with addition of 75 µL of 10% Triton X-100. Following the nuclei permeabilization treatment, chromatin was digested overnight with 150 units DpnII at 37°C, which was deactivated the next morning for 20 min at 62°C. Next, nuclei were collected after spinning at 1000*g* for 3 min, and suspended with 1 mL sonication buffer (10 mM potassium phosphate, pH 7.0; 0.1 mM NaCl; 0.5% Sarkosyl; 10 mM EDTA) and sheared by sonication with a Covaris S220 instrument (set at 20dc, 1i, 200cpb, 15 sec). The sonicated sample was centrifuged at 14,000 rpm for 5 min, and the supernatant was mixed with 100 µL 10% Triton X-100. Next, the sheared chromatin was mixed with an equal volume of IP buffer (50 mM Hepes, pH 7.5; 150 mM NaCl; 5 mM $MgCl_2$; 10 µM $ZnSO_4$; 1% Triton X-100; 0.05% SDS) and then equally divided and incubated with anti-GFP antibody (Abcam, ab290) or normal rabbit IgG (Santa Cruz, sc-2027), respectively. After overnight incubation at 4°C, 10 µL Protein A/G magnetic beads (Pierce) were added and incubated for 2 h at 4°C. The beads were washed at 4°C as follows: 3× with IP buffer, 1× with IP buffer having 500 mM NaCl, and 1× with LiCl buffer (0.25 M LiCl; 1% NP-40; 1% deoxycholate; 1 mM EDTA; 10 mM Tris pH 8.0), for 5 min each. Chromatin retained on beads was incubated in 200 µL elution buffer (50 mM Tris, pH 8.0; 200 mM NaCl; 1% SDS; 10 mM EDTA) for 6 h at 65°C, followed by Proteinase K treatment for 1 h at 37°C. DNA was extracted with a standard phenol-chloroform method. To increase sequence diversity at the ends of DNA, the immunoprecipitated DNA was incubated with dsDNA Fragmentase (NEB) for 25 min at 37°C, which randomly cut DNA into ~100- to 200-bp fragments. The digested DNA was purified with AMPure XP beads (Beckman Coulter), and all subsequent end repairing, A-tailing, adaptor ligation, library amplification steps were done through following a standard protocol (Illumina). The final libraries were sequenced on an Illumina HiSeq 3000 instrument with 2×150-bp reads.

Sequencing reads analysis

Paired-end reads were aligned against the *Arabidopsis thaliana* reference genome (TAIR10) using Bowtie 2 v2.2.4 (Langmead and Salzberg 2012) with a "very sensitive" mapping mode. For each replicate, the mapped reads were analyzed by SICER v1.1 (Zang et al. 2009) to call enriched regions (parameters: W = 1000; G = 3000; FDR < 0.01). For each type of tissue, regions shared between the two replicates were extracted, which were classified as domains enriched at the nuclear periphery (or NUP1-enriched domains). The *Arabidopsis* gene annotation was retrieved from Ensembl Genomes (ftp://ftp.ensemblgenomes.org/) (release-24) (Kersey et al. 2016). A gene was claimed enriched if >80% of its transcribed region overlapped with NUP1-enriched domains.

FISH

The FISH experiment was performed as previously published (Prieto et al. 2007; Wegel et al. 2009) with modifications, in which instead of biotin-16-dUTP, dinitrophenol-11-dUTP (DNP-11-dUTP) (PerkinElmer) was used to label BAC probes in the nick translation reaction. Slide pretreatment, hybridization, and post-hybridization wash were carried out as described (Prieto et al. 2007). Detection of the digoxigenin-11-dUTP was done with 1:1000 mouse anti-digoxin antibody (Sigma, D-8156) and followed by 1:150 goat anti-mouse antibody coupled to Alexa Fluor 488 (Invitrogen, A11017). Detection of the DNP-11-dUTP was done with 1:500 rabbit anti-dinitrophenyl antibody (Invitrogen, A6430) and followed by 1:150 goat anti-rabbit antibody coupled

to Alexa Fluor 546 (Invitrogen, A-11071). After the final wash step, slides were mounted with SlowFade Diamond Antifade Mountant with DAPI (Thermo Fisher Scientific).

Fluorescence microscopy

Confocal images were acquired with the Leica SP8 AOBS system. The detection of various fluorophores (DAPI, Alexa Fluor 488, and Alexa Fluor 546) and fluorescent proteins (GFP and mCherry) was according to the default settings. Image processing was done with the Fiji software and final assembly in Photoshop. For the distance measurement of FISH signals, first, a *Z*-stack image was obtained by maximum projection of signals from five optical sections. Then, the nuclear edge was defined by adjusting the threshold of DAPI channel, and the distance between a FISH signal and the nuclear edge was determined as described (Feng et al. 2014a). Only nuclei containing hybridization signals of both probes were included in analyses.

Published genomic data

Data for gene expression in various tissues were from Laubinger et al. (2008), bisulfite sequencing from Stroud et al. (2013), Hi-C matrix (20-kb window setting) from Wang et al. (2015), and chromatin loops and other processed epigenetic data sets from Liu et al. (2016). Definition of centromeric regions (Chr 1, ~14.08–15.61 Mb; Chr 2, ~2.93–3.95 Mb; Chr 3, ~13.16–14.55 Mb; Chr 4, ~2.00–4.26 Mb; and Chr 5, ~10.93–12.66 Mb) was according to coordinates of BAC clones on the TAIR10 genome that overlapped with the boundaries of estimated centromeric regions (The Arabidopsis Genome Initiative 2000). The definition of pericentromeric heterochromatin (Chr 1, ~11.5–17.7 Mb; Chr 2, ~1.1–7.2 Mb; Chr 3, ~10.3–17.3 Mb; Chr 4, ~1.5–6.3 Mb; and Chr 5, ~9.0–16.0 Mb) was according to Stroud et al. (2013).

Data access

All sequence data from this study have been submitted to the NCBI Sequence Read Archive (SRA; http://www.ncbi.nlm.nih.gov/sra) under accession number SRP079108.

References

Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408:** 796–815.

Ariel F, Jegu T, Latrasse D, Romero-Barrios N, Christ A, Benhamed M, Crespi M. 2014. Noncoding transcription by alternative RNA polymerases dynamically regulates an auxin-driven chromatin loop. *Mol Cell* **55:** 383–396.

Armstrong SJ, Franklin FC, Jones GH. 2001. Nucleolus-associated telomere clustering and pairing precede meiotic chromosome synapsis in *Arabidopsis thaliana*. *J Cell Sci* **114:** 4207–4217.

Barneche F, Malapeira J, Mas P. 2014. The impact of chromatin dynamics on plant light responses and circadian clock function. *J Exp Bot* **65:** 2895–2913.

Batzenschlager M, Lermontova I, Schubert V, Fuchs J, Berr A, Koini MA, Houlné G, Herzog E, Rutten T, Alioua A, et al. 2015. *Arabidopsis* MZT1 homologs GIP1 and GIP2 are essential for centromere architecture. *Proc Natl Acad Sci* **112:** 8656–8660.

Blobel G. 1985. Gene gating: a hypothesis. *Proc Natl Acad Sci* **82:** 8527–8529.

Bourbousse C, Mestiri I, Zabulon G, Bourge M, Formiggini F, Koini MA, Brown SC, Fransz P, Bowler C, Barneche F. 2015. Light signaling controls nuclear architecture reorganization during seedling establishment. *Proc Natl Acad Sci* **112:** E2836–E2844.

Cao X, Aufsatz W, Zilberman D, Mette MF, Huang MS, Matzke M, Jacobsen SE. 2003. Role of the *DRM* and *CMT3* methyltransferases in RNA-directed DNA methylation. *Curr Biol* **13:** 2212–2217.

Cao S, Kumimoto RW, Gnesutta N, Calogero AM, Mantovani R, Holt BF III. 2014. A distal *CCAAT/NUCLEAR FACTOR Y* complex promotes chromatin looping at the *FLOWERING LOCUS T* promoter and regulates the timing of flowering in *Arabidopsis*. *Plant Cell* **26:** 1009–1017.

Ciska M, Moreno Díaz de la Espina S. 2014. The intriguing plant nuclear lamina. *Front Plant Sci* **5:** 166.

Crevillén P, Sonmez C, Wu Z, Dean C. 2013. A gene loop containing the floral repressor *FLC* is disrupted in the early phase of vernalization. *EMBO J* **32:** 140–148.

Dechat T, Pfleghaar K, Sengupta K, Shimi T, Shumaker DK, Solimando L, Goldman RD. 2008. Nuclear lamins: major factors in the structural organization and function of the nucleus and chromatin. *Genes Dev* **22:** 832–853.

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485:** 376–380.

Fang Y, Spector DL. 2005. Centromere positioning and dynamics in living *Arabidopsis* plants. *Mol Biol Cell* **16:** 5710–5718.

Feng CM, Qiu Y, Van Buskirk EK, Yang EJ, Chen M. 2014a. Light-regulated gene repositioning in *Arabidopsis*. *Nat Commun* **5:** 3027.

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

Finlan LE, Sproul D, Thomson I, Boyle S, Kerr E, Perry P, Ylstra B, Chubb JR, Bickmore WA. 2008. Recruitment to the nuclear periphery can alter expression of genes in human cells. *PLoS Genet* **4:** e1000039.

Fransz P, De Jong JH, Lysak M, Castiglione MR, Schubert I. 2002. Interphase chromosomes in *Arabidopsis* are organized as well defined chromocenters from which euchromatin loops emanate. *Proc Natl Acad Sci* **99:** 14584–14589.

Gibcus JH, Dekker J. 2013. The hierarchy of the 3D genome. *Mol Cell* **49:** 773–782.

Goto C, Tamura K, Fukao Y, Shimada T, Hara-Nishimura I. 2014. The novel nuclear envelope protein KAKU4 modulates nuclear morphology in *Arabidopsis*. *Plant Cell* **26:** 2143–2155.

Grob S, Schmid MW, 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.

Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, Eussen BH, de Klein A, Wessels L, de Laat W, et al. 2008. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453:** 948–951.

Harr JC, Luperchio TR, Wong X, Cohen E, Wheelan SJ, Reddy KL. 2015. Directed targeting of chromatin to the nuclear lamina is mediated by chromatin state and A-type lamins. *J Cell Biol* **208:** 33–52.

Harr JC, Gonzalez-Sandoval A, Gasser SM. 2016. Histones and histone modifications in perinuclear chromatin anchoring: from yeast to man. *EMBO Rep* **17:** 139–155.

Karlsson P, Christie MD, Seymour DK, Wang H, Wang X, Hagmann J, Kulcheski F, Manavella PA. 2015. KH domain protein RCF3 is a tissue-biased regulator of the plant miRNA biogenesis cofactor HYL1. *Proc Natl Acad Sci* **112:** 14096–14101.

Kersey PJ, Allen JE, Armean I, Boddu S, Bolt BJ, Carvalho-Silva D, Christensen M, Davis P, Falin LJ, Grabmueller C, et al. 2016. Ensembl Genomes 2016: more genomes, more complexity. *Nucleic Acids Res* **44:** D574–D580.

Kind J, Pagie L, Ortabozkoyun H, Boyle S, de Vries SS, Janssen H, Amendola M, Nolen LD, Bickmore WA, van Steensel B. 2013. Single-cell dynamics of genome-nuclear lamina interactions. *Cell* **153:** 178–192.

Kubben N, Adriaens M, Meuleman W, Voncken JW, van Steensel B, Misteli T. 2012. Mapping of lamin A- and progerin-interacting genome regions. *Chromosoma* **121:** 447–464.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9:** 357–359.

Laubinger S, Zeller G, Henz SR, Sachsenberg T, Widmer CK, Naouar N, Vuylsteke M, Schölkopf B, Rätsch G, Weigel D. 2008. At-TAX: a whole genome tiling array resource for developmental expression analysis and transcript identification in *Arabidopsis thaliana*. *Genome Biol* **9:** R112.

Lermontova I, Schubert V, Fuchs J, Klatte S, Macas J, Schubert I. 2006. Loading of *Arabidopsis* centromeric histone CENH3 occurs mainly during G2 and requires the presence of the histone fold domain. *Plant Cell* **18:** 2443–2451.

Lermontova I, Kuhlmann M, Friedel S, Rutten T, Heckmann S, Sandmann M, Demidov D, Schubert V, Schubert I. 2013. *Arabidopsis* KINETOCHORE NULL2 is an upstream component for centromeric histone H3 variant cenH3 deposition at centromeres. *Plant Cell* **25:** 3389–3404.

Li CF, Henderson IR, Song L, Fedoroff N, Lagrange T, Jacobsen SE. 2008. Dynamic regulation of ARGONAUTE4 within multiple nuclear bodies in *Arabidopsis thaliana*. *PLoS Genet* **4:** e27.

Li S, Vandivier LE, Tu B, Gao L, Won SY, Li S, Zheng B, Gregory BD, Chen X. 2015. Detection of Pol IV/RDR2-dependent transcripts at the genomic scale in *Arabidopsis* reveals features and regulation of siRNA biogenesis. *Genome Res* **25:** 235–245.

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326:** 289–293.

Liu C, Teo ZW, Bi Y, Song S, Xi W, Yang X, Yin Z, Yu H. 2013. A conserved genetic pathway determines inflorescence architecture in *Arabidopsis* and rice. *Dev Cell* **24:** 612–622.

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

Lu Q, Tang X, Tian G, Wang F, Liu K, Nguyen V, Kohalmi SE, Keller WA, Tsang EW, Harada JJ, et al. 2010. *Arabidopsis* homolog of the yeast TREX-2 mRNA export complex: components and anchoring nucleoporin. *Plant J* **61:** 259–270.

Lund E, Oldenburg AR, Delbarre E, Freberg CT, Duband-Goulet I, Eskeland R, Buendia B, Collas P. 2013. Lamin A/C-promoter interactions specify chromatin state-dependent transcription outcomes. *Genome Res* **23:** 1580–1589.

Mattout A, Cabianca DS, Gasser SM. 2015. Chromatin states and nuclear organization in development—a view from the nuclear lamina. *Genome Biol* **16:** 174.

Parry G. 2015. The plant nuclear envelope and regulation of gene expression. *J Exp Bot* **66:** 1673–1685.

Pascuzzi PE, Flores-Vergara MA, Lee TJ, Sosinski B, Vaughn MW, Hanley-Bowdoin L, Thompson WF, Allen GC. 2014. In vivo mapping of *Arabidopsis* scaffold/matrix attachment regions reveals link to nucleosome-disfavoring poly(dA:dT) tracts. *Plant Cell* **26:** 102–120.

Pavet V, Quintero C, Cecchini NM, Rosa AL, Alvarez ME. 2006. *Arabidopsis* displays centromeric DNA hypomethylation and cytological alterations of heterochromatin upon attack by *pseudomonas syringae*. *Mol Plant Microbe Interact* **19:** 577–587.

Pecinka A, Schubert V, Meister A, Kreth G, Klatte M, Lysak MA, Fuchs J, Schubert I. 2004. Chromosome territory arrangement and homologous pairing in nuclei of *Arabidopsis thaliana* are predominantly random except for NOR-bearing chromosomes. *Chromosoma* **113:** 258–269.

Pickersgill H, Kalverda B, de Wit E, Talhout W, Fornerod M, van Steensel B. 2006. Characterization of the *Drosophila melanogaster* genome at the nuclear lamina. *Nat Genet* **38:** 1005–1014.

Pontes O, Vitins A, Ream TS, Hong E, Pikaard CS, Costa-Nunes P. 2013. Intersection of small RNA pathways in *Arabidopsis thaliana* sub-nuclear domains. *PLoS One* **8:** e65652.

Pontvianne F, Carpentier MC, Durut N, Pavlištová V, Jaške K, Schořová S, Parrinello H, Rohmer M, Pikaard CS, Fojtová M, et al. 2016. Identification of nucleolus-associated chromatin domains reveals a role for the nucleolus in 3D organization of the *A. thaliana* genome. *Cell Rep* **16:** 1574–1587.

Prieto P, Moore G, Shaw P. 2007. Fluorescence in situ hybridization on vibratome sections of plant tissues. *Nat Protoc* **2:** 1831–1838.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol* **29:** 24–26.

Sadaie M, Salama R, Carroll T, Tomimatsu K, Chandra T, Young AR, Narita M, Pérez-Mancera PA, Bennett DC, Chong H, et al. 2013. Redistribution of the Lamin B1 genomic binding profile affects rearrangement of heterochromatic domains and SAHF formation during senescence. *Genes Dev* **27:** 1800–1808.

Shah PP, Donahue G, Otte GL, Capell BC, Nelson DM, Cao K, Aggarwala V, Cruickshanks HA, Rai TS, McBryan T, et al. 2013. Lamin B1 depletion in senescent cells triggers large-scale changes in gene expression and the chromatin landscape. *Genes Dev* **27:** 1787–1799.

Sigman MJ, Slotkin RK. 2016. The first rule of plant transposable element silencing: location, location, location. *Plant Cell* **28:** 304–313.

182

Smyth DR, Bowman JL, Meyerowitz EM. 1990. Early flower development in *Arabidopsis*. *Plant Cell* **2:** 755–767.

Strambio-De-Castillia C, Niepel M, Rout MP. 2010. The nuclear pore complex: bridging nuclear transport and gene regulation. *Nat Rev Mol Cell Biol* **11:** 490–501.

Stroud H, Greenberg MV, Feng S, Bernatavichute YV, Jacobsen SE. 2013. Comprehensive analysis of silencing mutants reveals complex regulation of the *Arabidopsis* methylome. *Cell* **152:** 352–364.

Stroud H, Do T, Du J, Zhong X, Feng S, Johnson L, Patel DJ, Jacobsen SE. 2014. Non-CG methylation patterns shape the epigenetic landscape in *Arabidopsis*. *Nat Struct Mol Biol* **21:** 64–72.

Tamura K, Hara-Nishimura I. 2013. The molecular architecture of the plant nuclear pore complex. *J Exp Bot* **64:** 823–832.

Tamura K, Fukao Y, Iwamoto M, Haraguchi T, Hara-Nishimura I. 2010. Identification and characterization of nuclear pore complex components in *Arabidopsis thaliana*. *Plant Cell* **22:** 4084–4097.

Tessadori F, Chupeau MC, Chupeau Y, Knip M, Germann S, van Driel R, Fransz P, Gaudin V. 2007. Large-scale dissociation and sequential reassembly of pericentric heterochromatin in dedifferentiated *Arabidopsis* cells. *J Cell Sci* **120:** 1200–1208.

Towbin BD, González-Aguilera C, Sack R, Gaidatzis D, Kalck V, Meister P, Askjaer P, Gasser SM. 2012. Step-wise methylation of histone H3K9 positions heterochromatin at the nuclear periphery. *Cell* **150:** 934–947.

Van Bortle K, Corces VG. 2012. Nuclear organization and genome function. *Annu Rev Cell Dev Biol* **28:** 163–187.

Vieux-Rochas M, Fabre PJ, Leleu M, Duboule D, Noordermeer D. 2015. Clustering of mammalian *Hox* genes with other H3K27me3 targets within an active nuclear domain. *Proc Natl Acad Sci* **112:** 4672–4677.

Wang H, Dittmer TA, Richards EJ. 2013. *Arabidopsis* CROWDED NUCLEI (CRWN) proteins are required for nuclear size control and heterochromatin organization. *BMC Plant Biol* **13:** 200.

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

Wegel E, Koumproglou R, Shaw P, Osbourn A. 2009. Cell type–specific chromatin decondensation of a metabolic gene cluster in oats. *Plant Cell* **21:** 3926–3936.

Zang C, Schones DE, Zeng C, Cui K, Zhao K, Peng W. 2009. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics* **25:** 1952–1958.

Zemach A, Kim MY, Hsieh PH, Coleman-Derr D, Eshed-Williams L, Thao K, Harmer SL, Zilberman D. 2013. The *Arabidopsis* nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* **153:** 193–205.

Zhang X, Germann S, Blus BJ, Khorasanizadeh S, Gaudin V, Jacobsen SE. 2007. The *Arabidopsis* LHP1 protein colocalizes with histone H3 Lys27 trimethylation. *Nat Struct Mol Biol* **14:** 869–871.

Zhou X, Graumann K, Meier I. 2015. The plant nuclear envelope as a multifunctional platform LINCed by SUN and KASH. *J Exp Bot* **66:** 1649–1659.

183