CellPress
Partner Journal

Molecular Plant
Resource Article

Single-nucleus RNA and ATAC sequencing reveals the impact of chromatin accessibility on gene expression in *Arabidopsis* roots at the single-cell level

Andrew Farmer[1], Sandra Thibivilliers[2], Kook Hui Ryu[3], John Schiefelbein[3] and Marc Libault[2,*]

[1]National Center for Genome Resources, Santa Fe, NM 87505, USA

[2]Department of Agronomy and Horticulture, Center for Plant Science Innovation, University of Nebraska-Lincoln, Beadle Center, Lincoln, NE 68503, USA

[3]Department of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, MI 48109, USA

***Correspondence: Marc Libault (marc.libault@unl.edu)**

https://doi.org/10.1016/j.molp.2021.01.001

ABSTRACT

Similar to other complex organisms, plants consist of diverse and specialized cell types. The gain of unique biological functions of these different cell types is the consequence of the establishment of cell-type-specific transcriptional programs. As a necessary step in gaining a deeper understanding of the regulatory mechanisms controlling plant gene expression, we report the use of single-nucleus RNA sequencing (sNucRNA-seq) and single-nucleus assay for transposase accessible chromatin sequencing (sNucATAC-seq) technologies on *Arabidopsis* roots. The comparison of our single-nucleus transcriptomes to the published protoplast transcriptomes validated the use of nuclei as biological entities to establish plant cell-type-specific transcriptomes. Furthermore, our sNucRNA-seq results uncovered the transcriptomes of additional cell subtypes not identified by single-cell RNA-seq. Similar to our transcriptomic approach, the sNucATAC-seq approach led to the distribution of the *Arabidopsis* nuclei into distinct clusters, suggesting the differential accessibility of chromatin between groups of cells according to their identity. To reveal the impact of chromatin accessibility on gene expression, we integrated sNucRNA-seq and sNucATAC-seq data and demonstrated that cell-type-specific marker genes display cell-type-specific patterns of chromatin accessibility. Our data suggest that the differential chromatin accessibility is a critical mechanism to regulate gene activity at the cell-type level.

Key words: single-cell RNA-seq, single-nucleus ATAC-seq, *Arabidopsis*, root

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

INTRODUCTION

The gain of unique biological functions by the various cell types composing a plant depends on their differential use of the same genomic information to produce cell-type-specific transcriptional profiles. The differential use of the genomic information between cells and between cell types is thought to rely, in part, on differential chromatin accessibility. Changes in the landscape of chromatin fiber influence the accessibility of the genomic DNA for regulatory proteins such as transcription factors (TFs). This statement is supported by the human ENCODE project, which recently revealed that the establishment of a chromatin landscape at the single-cell level was highly informative to reveal putative TF-binding sites (Gulko and Siepel, 2019). Ultimately, each plant cell will activate or repress specific sets of genes to fulfill the biological functions inherent to their cell type and their response to environmental signals. In animal science, single-cell RNA sequencing (scRNA-seq) and single-cell ATAC-seq (Pott and Lieb, 2015) technologies have been successfully applied across various cell types and tissues to better understand the impact of the dynamic accessibility of chromatin on gene expression (Buenrostro et al., 2015, 2018; Norrie et al., 2019; Zhou et al., 2019)

Published by the Molecular Plant Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and CEMPS, CAS.

Recently, scRNA-seq approaches have been applied to *Arabidopsis* root protoplasts, allowing the accurate characterization of the transcriptional profiles of thousands of cells and their differential regulation in mutants or in response to a stress (Denyer et al., 2019; Jean-Baptiste et al., 2019; Ryu et al., 2019; Shulse et al., 2019; Zhang et al., 2019). These studies revealed the power of single-cell technologies to establish the transcriptomic maps of various *Arabidopsis* root cells and cell types and the dynamic regulation of gene expression during cell development. However, using plant protoplasts as biological entities to analyze gene expression is problematic. For example, some cell types are resistant to cell-wall digestion (e.g., stele cells; Denyer et al., 2019; Jean-Baptiste et al., 2019; Ryu et al., 2019; Shulse et al., 2019; Zhang et al., 2019), the protoplasting procedure itself has a significant impact on gene expression (Denyer et al., 2019), and there is a bias toward sequencing of smaller-sized cells/protoplasts (Denyer et al., 2019). In addition, the effective isolation of plant protoplasts requires the development of particular cell-wall-degrading enzymatic cocktails tailored for the differential biochemical compositions of the cell wall existing between plant species, the developmental stage of the cell (i.e., differential biochemical composition between the primary and the secondary cell wall), and the relative position of the root cell (i.e., external versus internal locations) (Sarkar et al., 2009; Somssich et al., 2016).

As an alternative to protoplasts, bulks of plant nuclei have been used to gain transcriptomic information from plant cells. For instance, transcriptomes were established from nuclei populations using isolation of nuclei tagged in specific cell types (INTACT) technology (Deal and Henikoff, 2010) on rice roots (Reynoso et al., 2018), *Arabidopsis* embryos (Palovaara and Weijers, 2019), and seed endosperm (Del Toro-De Leon and Kohler, 2019). Also, taking advantage of the nuclear ploidy of the cells composing the pericarp tissue of developing tomato fruits, Pirrello et al. (2018) analyzed the expression of tomato genes using a population of nuclei characterized by various levels of endoreduplication (Pirrello et al., 2018). Nevertheless, these methods also suffer from various limitations. For instance, the use of the INTACT technology presupposes the identification of cell-type-specific marker genes to express a reporter gene and requires the generation of transgenic material.

To overcome some of the challenges associated with the generation and manipulation of plant protoplasts, we tested the use of isolated plant nuclei to establish the transcriptomes of thousands of plant single cells. Previous single-nucleus transcriptomic experiments conducted on various animal systems suggest that nuclei can be used to establish biologically meaningful transcriptomic information compared with isolated cells (Bakken et al., 2018; Hu et al., 2018). Our data reveal high similarities between the nuclear and the protoplast transcriptomes and the discovery of additional root cell-type transcriptomes that support the use of isolated nuclei as valuable biological entities to access single-cell gene expression. To gain a deeper understanding of the regulatory mechanisms controlling gene expression in and between *Arabidopsis* root cells and cell types, we also developed and applied single-nucleus ATAC-seq (sNucATAC-seq). sNucATAC-seq revealed cell-type-specific chromatin accessibility and dynamic changes occurring in chromatin accessibility during cell differentiation and development. Upon establishing the unique and conserved chromatin landscapes for each major *Arabidopsis* root cell type, we integrated single-nucleus RNA-seq (sNucRNA-seq) and sNucATAC-seq datasets to highlight the influence of chromatin accessibility on *Arabidopsis* gene activity.

RESULTS AND DISCUSSION

Arabidopsis root nuclei generate biologically meaningful transcriptomic datasets

Nuclei from *Arabidopsis* seedling roots were purified and used with the 10X Genomics Chromium platform to create sNucRNA-seq libraries (Supplemental Figure 1). To accurately compare sNucRNA-seq and scRNA-seq transcriptomes, the *Arabidopsis* seedlings were grown, and primary roots were isolated as described by Ryu et al. (2019). Across five independent biological replicates, we sequenced the transcriptomes of 10 548 nuclei (see Supplemental Table 1 for sequencing details). Because some nuclear transcripts might not be spliced, we applied a "pre-mRNA" strategy to include intron-mapping reads. We obtained a median of 1124 expressed genes per nucleus, allowing the identification of 24 510 expressed genes of the 27 420 predicted *Arabidopsis* protein-coding genes (89.4%) (https://plants.ensembl.org/Arabidopsis_thaliana/Info/Index). In comparison, the transcriptome of 7437 selected *Arabidopsis* protoplasts by Ryu et al. (2019) allowed the detection of a median of 4739 expressed genes per cell and a total of 25 177 expressed genes (91.8%). We suggest that the higher number of expressed genes identified per protoplast versus nucleus is the consequence of the larger and more complex pool of polyadenylated transcripts in one cell versus one nucleus, due to the accumulation of transcripts and their relative stability (i.e., the half-life of the cellular mRNA is estimated at 9 h in human cells; Schwanhausser et al., 2011). This conjecture also suggests that the nuclear transcriptome represents a snapshot of the dynamic transcriptional activity of the genes, while the cellular transcriptome may represent an integration of gene activity over time.

To evaluate the biological significance of the nuclear transcriptomes obtained from the sNucRNA-seq data, we performed correlation analyses between bulk RNA-seq from intact whole roots and from protoplast suspensions and pseudo-bulk RNA-seq from root protoplasts and nuclei that were processed through scRNA-seq and sNucRNA-seq technologies (present study). The transcriptome of intact whole roots is highly correlated with the transcriptome of a suspension of protoplasts (i.e., Spearman's rank correlation coefficient [SRCC] = 0.902), with pseudo-bulked scRNA-seq datasets (i.e., 0.859 < SRCC < 0.892), and with the sNucRNA-seq dataset (i.e., SRCC = 0.879) (Supplemental Table 2). Upon exclusion of 346 known protoplast-responsive genes (Birnbaum et al., 2003), the SRCCs increased more when comparing intact whole roots and pseudo-bulked scRNA-seq datasets (i.e., from 0.862 to 0.895) than when comparing intact whole roots to our sNucRNA-seq transcriptome (i.e., SRCC = 0.880) (Supplemental Table 2). These results suggest that the sNucRNA-seq transcriptome of the *Arabidopsis* root is as highly correlated to a whole root transcriptome as protoplast-based transcriptomes. Overall, a single-nucleus-based transcriptome generated from pooled

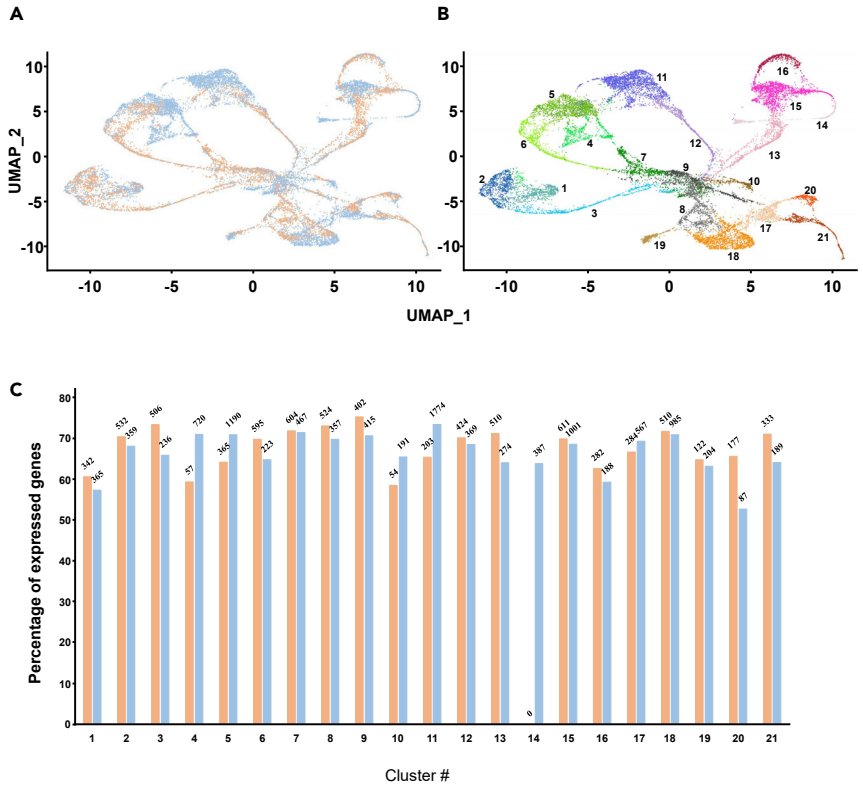

Figure 1. UMAP cluster analysis of the *Arabidopsis* single-cell and single-nucleus transcriptomes.

(A) Comparative analysis of the clustering of *Arabidopsis* cells (orange, Ryu et al., 2019) and nuclei (blue, present study) according to their transcriptomic profiles.

(B) Clustering of the *Arabidopsis* cells and nuclei in 21 different groups according to their transcriptomic profiles. Cluster 14 was exclusively identified from isolated nuclei (see Supplemental Figure 2).

(C) Percentage of *Arabidopsis* genes found expressed upon analysis of scRNA-seq (orange) and sNucRNA-seq (blue) technologies for each 21 clusters. The numbers at the top of the bars highlight the number of protoplasts and nuclei associated with each cluster.

Arabidopsis root nuclei reflects well the expected root transcriptome generated by conventional methods.

Taking advantage of the capability of the Seurat package to integrate independent datasets (Butler et al., 2018; Stuart et al., 2019), we co-clustered 10 548 *Arabidopsis* root nuclei with 7437 *Arabidopsis* root protoplasts (Ryu et al., 2019) according to their transcriptomic profiles (Figure 1A; Supplemental Figure 2). Applying uniform manifold approximation and projection (UMAP), a dimensionality reduction technique that enhances the organization of nuclei/cell clusters compared with t-distributed stochastic neighbor embedding (t-SNE) (Becht et al., 2019), the *Arabidopsis* root nuclei and protoplasts were clustered into 21 and 20 different groups, respectively (Figure 1B; Supplemental Figure 2). UMAP visualization reveals the overlapping distribution existing between the 20 co-characterized clusters. Among them, two clusters are almost exclusively represented by nuclei (i.e., Nos. 4 and 11 are composed of over 89% nuclei; Figure 1C), and one cluster was exclusively characterized by applying sNucRNA-seq technology (cluster 14; Figure 1; Supplemental Figure 2).

To further evaluate the biological relevance of the sNucRNA-seq approach, we compared the percentages of protein-coding genes found expressed or not expressed using sNucRNA-seq and scRNA-seq technologies across the 20 co-characterized clusters (Figure 1C). Among the 27 420 *Arabidopsis* protein-coding genes, an average of 60.4% were found expressed by applying both technologies (Supplemental Figure 3; minimum number of expressed genes in one cluster: 13 729 [cluster 20; 50.1%]; maximum number of expressed genes in one cluster: 18 457 [cluster 9; 67.3%]). In contrast, an average of 26.3% of the genes were not found expressed by both sNucRNA-seq and scRNA-seq technologies (Supplemental Figure 3; minimum number of unexpressed genes in one cluster: 5899 [cluster 9, 21.59%]; maximum number of unexpressed genes in one cluster: 9096 [cluster 1; 33.2%]). Among the remaining genes, an average of 6% and 7.3% were found expressed using only sNucRNA-seq or scRNA-seq, respectively (Supplemental Figure 3). Taken together, these results revealed that 73.7% of the expressed *Arabidopsis* protein-coding genes were detected by combining sNucRNA-seq and scRNA-seq technologies (Supplemental Figure 4). Excluding cluster 14, which is unique to sNucRNA-seq, the percentage of expressed genes per cluster identified using sNucRNA-seq technology (from 52.7% to 73.3%) is not significantly different compared with the percentage of expressed genes per cluster identified using scRNA-seq technology (from 58.4% to 75.2%) (i.e., Student *t*-test: $P > 0.421$; Figure 1C). These results show that the cellular and nuclear transcriptomes provide similar transcriptomic information and suggest that isolated plant nuclei can be utilized to establish meaningful transcriptomic information at the single-cell-type level. Through the identification of three new cell clusters, our data also suggest that the sNucRNA-seq approach captures a more diverse and representative population of *Arabidopsis* root cell types compared with scRNA-seq.

Functional assignment of *Arabidopsis* root cell clusters

Taking advantage of recently published *Arabidopsis* root single-cell transcriptomes (Denyer et al., 2019; Jean-Baptiste et al., 2019; Ryu et al., 2019; Shulse et al., 2019; Zhang et al., 2019) and the information on root gene expression and regulation (Bohme et al., 2004; Fendrych et al., 2014; Olvera-Carrillo et al., 2015; Turco et al., 2019), we created a list of 101 cell-type marker genes (Supplemental Table 3). To assign biological entities to the 21 *Arabidopsis* clusters, we looked for the accumulation of transcripts for these marker genes (Figure 2; Supplemental Figures 5 and 6). This strategy allowed us to characterize six major groups of cells: trichoblasts (clusters 1–3), atrichoblasts (clusters 4–7), meristematic cells (clusters 8–10), cortical cells (clusters 11 and 12), endodermal cells (clusters 13–16), and

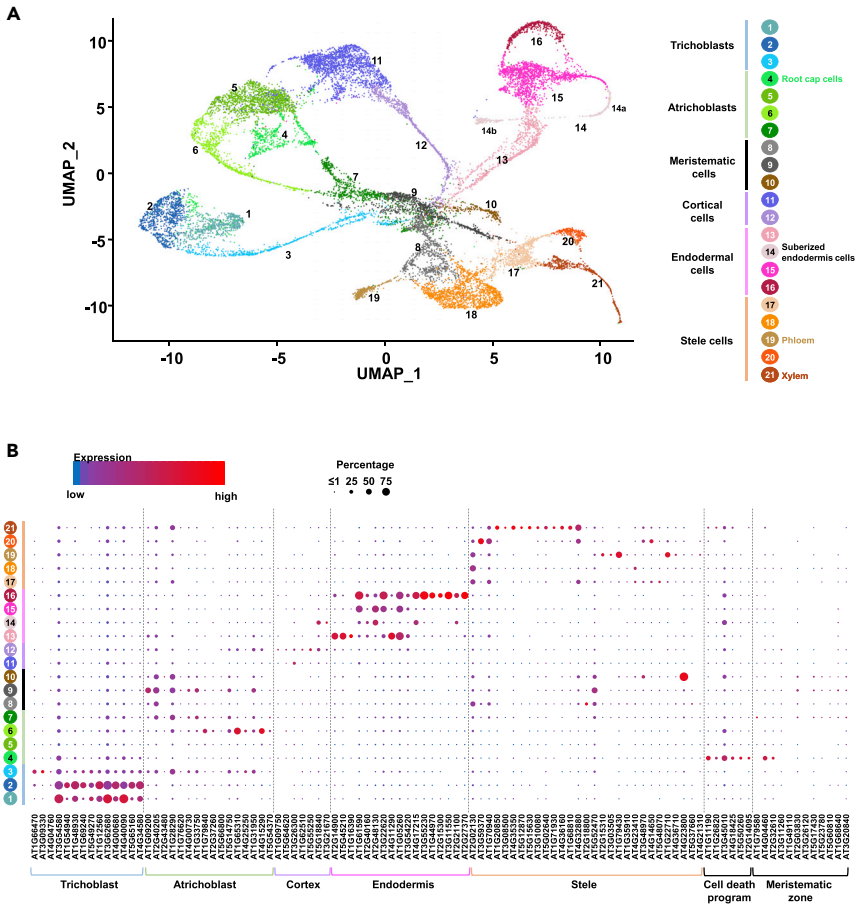

Figure 2. Functional annotation of the *Arabidopsis* root cell-types

(A) Assignment of *Arabidopsis* root cell types based on the characterization of the expression profile of cell-type and cell-death marker genes.

(B) Normalized expression levels of 101 cell-type and cell-death marker genes (*x* axis, see Supplemental Table 3) across the 21 different clusters (*y* axis). The sNucRNA-seq dataset was used to create this figure. As a comparison, the transcriptional activity of these 101 marker genes in *Arabidopsis* protoplasts (Ryu et al., 2019) is provided in Supplemental Figure 5. The percentage of nuclei expressing the gene of interest (circle size) and the mean expression (circle color) of genes are shown.

stele cells (clusters 17–21) (Figure 2A). In addition, based on the expression patterns of marker genes, we were able to assign the phloem and xylem clusters (Figure 2A).

To evaluate if sNucRNA-seq data can be analyzed alone to decode tissue heterogeneity to a level similar to that of scRNA-seq data, we independently analyzed the clustering of the *Arabidopsis* root nuclei and protoplasts based on their transcriptome. Using the same clustering parameters, the sNucRNA-seq and scRNA-seq datasets led to the identification of 19 and 17 clusters (Supplemental Figure 7). These results suggest that the nuclear transcriptome is sufficient to reveal the tissue heterogeneity of the *Arabidopsis* root.

The topology of the clusters generated by the UMAP technique reveals the functional organization of the cells and nuclei in and between the cell types (Figure 2A; Supplemental Figure 7), compared with the t-SNE technique as previously mentioned in previous reports (Jean-Baptiste et al., 2019; Zhang et al., 2019). For instance, the meristematic cells of the roots (i.e., clusters 8–10) are localized in the center of the UMAP map (Figure 2A). Originating from these meristematic cells, several elongated projections of cells (e.g., clusters 3, 6, 7, 12, and 13) end with more globular clusters (e.g., clusters 1, 2, 5, 11, 15, and 18) (Figure 2A). The elongated clusters likely reflect the progressive changes occurring in the transcriptomic programs during cell differentiation, whereas the globular clusters represent the differentiated cells composing the *Arabidopsis* roots.

Among the three clusters mostly or exclusively defined using sNucRNA-seq technology (i.e., clusters 4, 11, and 14; Figures 1C and 2A), cluster 14 could be divided into two different sub-groups, which are characterized by the activity of endodermal (cluster 14a) and cortex marker genes (cluster 14b) (Figure 2A; Supplemental Figure 8) (i.e., AT1G61590 [*PBL15*], AT2G40160 [*TBL30*], AT2G48130, and AT4G17215 as endodermal marker genes; AT5G18840 and AT3G21670 [*NPF6.4/NRT1.3*] as cortical marker genes; Supplemental Figures 6 and 8; Supplemental Table 3). Cluster 14a is also characterized by the specific expression of genes encoding peroxidases (e.g., AT1G68850, AT2G35380) and GDSL-motif esterase/acyltransferase/lipase (e.g., AT2G23540, and AT5G37690; Lai et al., 2017) (Supplemental Figure 8). Members of the *GDSL* family, such as the rice *WDL1* and the tomato *GDSL1* genes, control the process of cellular differentiation (Ding et al., 2019; Girard et al., 2012; Park et al., 2010), suggesting that cluster 14a might be composed of differentiated cells. This hypothesis is supported by the role of root peroxidases in controlling the production of reactive oxygen species to regulate cell elongation and differentiation (Tsukagoshi et al., 2010). To further support this hypothesis, the *UPBEAT1* gene (AT2G47270), a major repressor of the transcriptional activity of peroxidase genes and reactive oxygen species distribution and a negative regulator of the size of the *Arabidopsis* root apical meristem by modulating the balance between cell proliferation and differentiation (Tsukagoshi et al., 2010), is broadly expressed except for clusters 4, 10, 11, and 14 (Supplemental Figure 8). In addition to controlling cell differentiation, GDSL lipases also play a central role in cutin biosynthesis (Dominguez et al., 2015). Mining the sc/sNucRNA-seq datasets, we identified many other genes preferentially expressed in cluster 14a and involved in the biosynthesis of suberin and cutin (e.g., α/β hydrolases [AT4G24140]; *GPAT5*, which plays a central role in suberin biosynthesis [AT3G11430] [Beisson et al., 2007; Li et al., 2007b]; AT1G49430; AT2G38110; *GPTA4* [AT1G01610] and *8* [AT4G00400] [Li et al., 2007a]; and another GDSL-like lipase [AT1G74460] [Li et al., 2007a]; Supplemental

Figure 8). Previous studies revealed that suberin and cutin are strongly deposited at the location of the emergence of lateral roots (Berhin et al., 2019) and during endodermis differentiation (Andersen et al., 2015). Taken together, the transcriptional pattern of *UPBEAT1*, several genes encoding peroxidases, *GDSL* genes, and other suberin/cutin biosynthesis-related genes, associated with the activity of both cortical and endodermal marker genes, suggests that the cells composing cluster 14a are differentiated endodermis cells characterized by the suberization of their cell wall. The characterization of the transcriptome of the cells in cluster 14a from isolated nuclei and not from isolated protoplasts could be due to the low digestibility of their cell wall, a consequence of the accumulation of suberin and cutin.

Cluster 14b (Figure 2A) is characterized by the expression of a subset of the cortex-specific marker genes (e.g., AT5G18840 and AT3G21670 [*NPF6.4*/*NRT1.3*]; Supplemental Figures 6 and 8; Supplemental Table 3; Denyer et al., 2019; Jean-Baptiste et al., 2019; Ryu et al., 2019). Among the genes specifically expressed in this cluster, *SCRAMBLED/STRUBBELIG* (*SCM*, AT1G11130) plays a critical role in the patterning of the root epidermal cells (Kwak et al., 2005) (Supplemental Figure 9). The specific clustering of the *SCRAMBLED/STRUBBELIG*-expressing cells is further supported by the expression of several genes involved in lipid metabolism in cluster 14b (e.g., AT1G45201 [*TLL1*; triacylglycerol lipase-like 1]; AT5G63560 [HXXXD-type acyl-transferase]; Supplemental Figure 9). Membrane lipid remodeling has been shown to also play a critical role in root hair cell differentiation (Chen and Schmidt, 2015; Salazar-Henao et al., 2016). Taken together, the cortical cells composing cluster 14b are hypothesized to play a role in the differentiation and patterning of the *Arabidopsis* epidermal root cells. Cluster 4 (Figure 2A) is characterized by the specific expression of *CEP1* (AT5G50260) and *EXI1* (AT2G14095) (Supplemental Figure 10), two genes previously characterized as regulators of the cell death program in the root cap (Olvera-Carrillo et al., 2015). In addition, we identified *KIRA1* (AT4g28530), a gene controlling cell death during flower development (Gao et al., 2018), as specifically expressed in cluster 4. Additional cell death marker genes (i.e., *BFN1*, *RNS3*, *SCPL48*, *DMP4*, and *PASPA3*) were also mostly found expressed in cluster 4 and a subset of xylem cluster 21 (Supplemental Figure 10; Supplemental Table 3) (Olvera-Carrillo et al., 2015). Previous studies showing that cell death programs, which play critical roles in the development of the xylem and root cap (Kumpf and Nowack, 2015; Heo et al., 2017), support the specific activity of these cell death marker genes in xylem cluster 21 and the assignment of the cells composing cluster 4 as root cap cells.

Single-cell resolution ATAC-seq reveals the impact of chromatin accessibility on gene expression

Although genomic information is almost identical between somatic cells (i.e., with the exception of somatic mutations), its differential use, notably through differential chromatin accessibility between cells, is required in order to fulfill their unique biological function through cell-type-specific transcriptional gene regulation (Lai et al., 2018; Maher et al., 2018; Sijacic et al., 2018; Arendt et al., 2019). To date, bulk RNA- and ATAC-seq datasets have shown low correlations (Maher et al., 2018). This could be the consequence of the cellular heterogeneity of the samples used. This hypothesis is supported by the human ENCODE project, which recently revealed that the establishment of a chromatin landscape at the single-cell level was highly informative to reveal putative TF-binding sites (Gulko and Siepel, 2019). To better evaluate the impact of chromatin accessibility in controlling plant gene expression between cells and cell types, we applied 10X Genomics sNucATAC-seq technology on *Arabidopsis* root nuclei isolated from two independent biological replicates. Among the 6768 called nuclei, 4764 passed downstream quality control filters (Supplemental Figure 11).

Upon paired-end sequencing, a median of 10 253 independent genomic DNA fragments per nucleus were mapped against the *Arabidopsis* genome, and a total of 20 803 accessible sites were characterized (Supplemental Table 1). As a comparison, Lu et al. (2017), Tannenbaum et al. (2018), and Maher et al. (2018) identified around 20 000 and 40 000 accessible sites from bulk ATAC-seq analyses conducted on *Arabidopsis* seedling and roots, respectively (Lu et al., 2017; Maher et al., 2018; Tannenbaum et al., 2018). Among the 20 803 accessible sites, we identified 3487 and 15 730 loci characterized by cell-type-specific peaks and by "static" peaks, respectively (Supplemental Figure 12, Supplemental Table 4). As previously reported (Lu et al., 2017), we observed that the accessible regions of chromatin are mostly located in the 1000 bp windows upstream of the transcription start sites (TSSs) that contain *cis*-regulatory elements (Figure 3A) and around the transcription termination sites (TTSs) of the genes (Figure 3B). Supporting the significance of our sNucATAC-seq data, we found a high correlation between bulk ATAC-seq from intact whole *Arabidopsis* root tips (Maher et al., 2018) and our pseudo-bulked sNucATAC-seq dataset (i.e., SRCC = 0.95).

Considering that accessible chromatin sites are a prerequisite to promote gene expression, we expect that cell-type-specific ATAC-seq peaks located near the TSS contribute to the regulation of the expression of cell-type marker genes (i.e., the peaks have at least one base pair overlap with the annotated transcripts and extend upstream). Accordingly, we searched for open chromatin near the TSS of genes expressed in cell-type-specific contexts by integrating sNucATAC-seq and sc/sNucRNA-seq analysis using the Signac package v.0.2.5 (Stuart et al., 2019). This strategy led to the identification of 11 858 genes that have paired RNA-seq and ATAC-seq peaks, and to the creation of the 21 sNucATAC-seq clusters corresponding to the sc/sNucRNA-seq clusters (Figure 3C) and represented by similar distributions of their fragment length (Supplemental Figure 13). To estimate the impact of the use of the sc/sNucRNA-seq datasets to cluster the *Arabidopsis* nuclei according to their chromatin accessibility profiles, we performed a *de novo* clustering of the nuclei according to their differential peaks of open chromatin (Supplemental Figure 14). This approach led again to the identification of 21 clusters that are slightly differentially distributed compared with the clusters identified by our integrated sNucATAC-seq and sc/sNucRNA-seq analysis. This result suggests that the profiles of chromatin accessibility are sufficient to reveal the cellular complexity of the *Arabidopsis* root cells.

Focusing on our integrated sNucATAC-seq and sc/sNucRNA-seq analysis, we noticed that, overall, the sNucATAC-seq root cell clusters have a UMAP topography similar to that of the 21

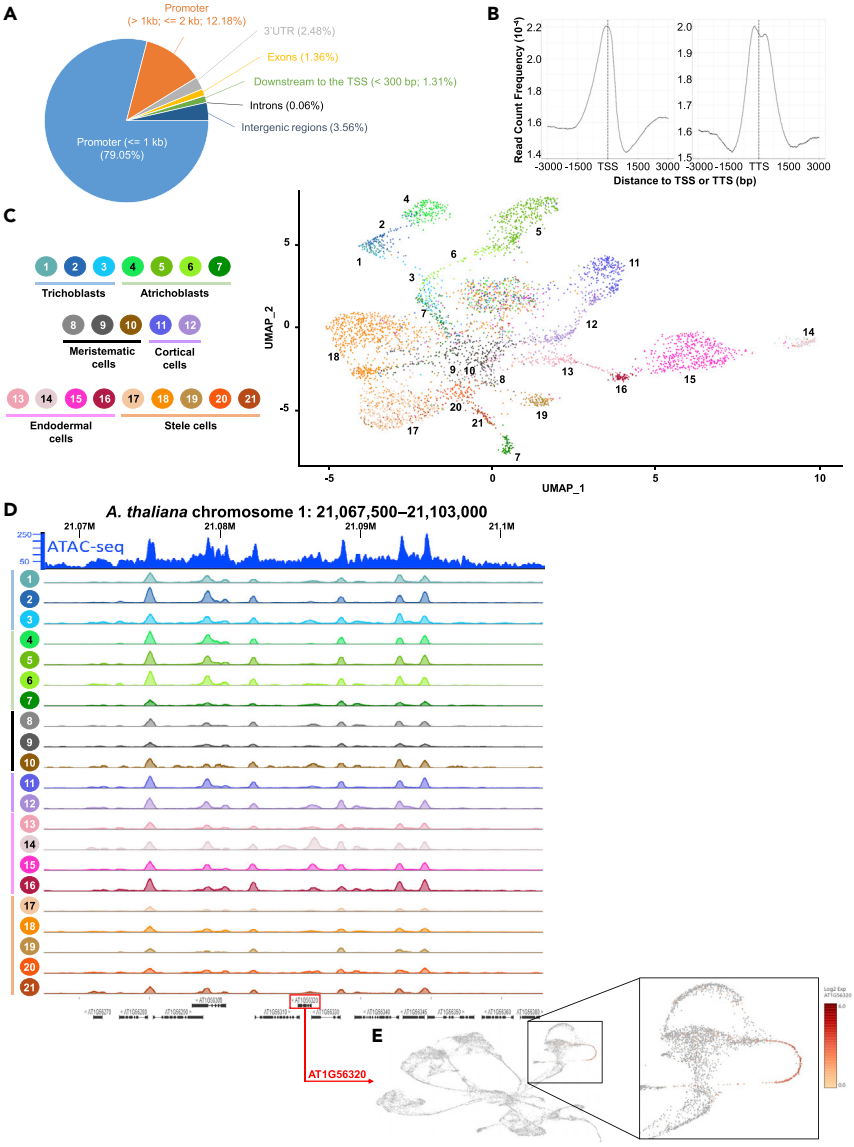

Figure 3. sNucATAC-seq to characterize the differential folding of chromatins between *Arabidopsis* root cell types.

(A) Distribution of accessible regions of the chromatin fiber after applying sNucATAC-seq technology throughout the *Arabidopsis* genome (this pie chart was created using ChIPseeker; Yu et al., 2015).

(B) Distribution of chromatin accessible regions around the TSS and TTS identified from sNucATAC-seq.

(C) Clustering of *Arabidopsis* root nuclei upon sNucATAC-seq analysis. The sNucATAC-seq clusters were annotated afterward by integrating sNucATAC-seq and sc/sNucRNA-seq analysis using the Signac package (see Methods for details).

(D) Comparative analysis of the distribution of ATAC-seq peaks on *A. thaliana* chromosome 1: 21 067 500–21 103 000. The top blue graphic was generated from a bulk ATAC-seq analysis of the *Arabidopsis* root seedling (Tannenbaum et al., 2018). Below it is shown the ATAC-seq peak profiles across the 21 sNucATAC-seq clusters.

(E) Transcriptional pattern of AT1G56320 in an integrated scRNA-seq/sNucRNA-seq UMAP dimension. The relative expression of this gene is highlighted in yellow/red.

sNuc/scRNA-seq clusters. First, sNucATAC-seq clusters 1, 2, and 3; clusters 4, 5, and 6; clusters 11 and 12; clusters 13, 14, 15, and 16; and clusters 17, 18, 19, 20, and 21, which comprise the trichoblasts, the atrichoblasts, and the cortical, endodermal, and stele cells, respectively, are grouped in superclusters that are similar to the sNucRNA-seq superclusters (Figures 2A and 3B). Second, we noticed that, similar to the organization of the sc/sNucRNA-seq clusters, the sNuc-ATAC-seq clusters associated with undifferentiated and differentiating cells (i.e., clusters 8, 9, and 10) are located in the center of the sNucATAC-seq UMAP map, while the clusters associated with differentiated cells are located at its periphery. The overall similar topographies suggest that chromatin accessibility and gene expression could both be used as molecular markers to annotate plant cell types and support some instances of a correlation for some genes between chromatin accessibility and their transcriptional activity.

To better estimate the gain in resolution of sNucATAC-seq versus bulk ATAC-seq datasets and to estimate the potential of sNucATAC-seq technology to reveal discrete changes in chromatin accessibility, we first compared the sNucATAC-seq and bulk ATAC-seq profiles generated from *Arabidopsis* root nuclei at one location (chr1: 21 067 500–21 103 000) (Tannenbaum et al., 2018). Across the 21 clusters, we can clearly identify the same major peaks revealed by bulk ATAC-seq technology (Tannenbaum et al., 2018) (Figure 3D). Furthermore, our sNucATAC-seq approach revealed additional major peaks but only in a subset of the 21 clusters (e.g., ATAC-seq peak located in the promoter region of AT1G56320 in the sNucATAC-seq clusters 14 and 15 [Figure 3D]; AT1G56320 is specifically expressed in these two clusters [Figure 3E]). This first targeted analysis suggests that a single-cell-resolution ATAC-seq analysis has the potential to reveal discrete and cell-type-specific loci of accessible chromatin. These results also support the idea that the 21 sNucATAC-seq clusters are characterized by unique combinations of open chromatin location. We conducted further analyses to reveal how cell-type chromatin accessibility plays a critical role in controlling gene expression. For instance, the heatmaps showing the scRNA-seq, sNucRNA-seq, and sNucATAC-seq profiles of 2756 *Arabidopsis* genes preferentially expressed in the trichoblasts, atrichoblasts, and meristematic, cortical, endodermal, and stele cells share similar profiles (i.e., the expression of the gene in one supercluster is at least equal to the combined expression of the gene in the five remaining superclusters; Supplemental Figure 15 and Supplemental Table 5). This result suggests that some correlations exist between chromatin accessibility and gene activity at least for these cell-type-preferential genes.

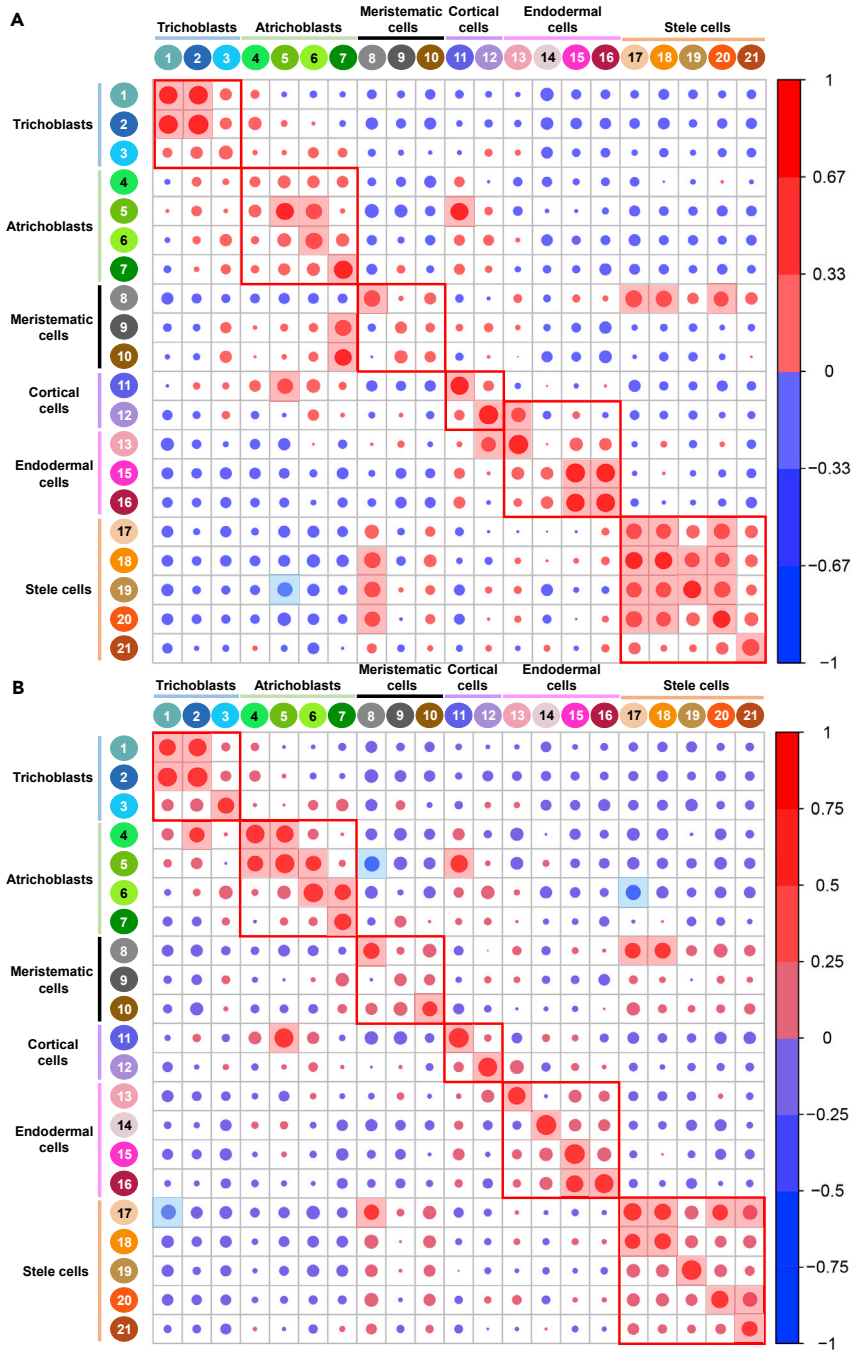

Figure 4. Correlation analyses between gene expression and chromatin accessibility (x-axis) and gene expression (y-axis) of 336 marker genes

For each correlation analysis [i.e., for the 20 sc (**A.**) and 21 sNucRNA-seq clusters (**B.**)], Kendall tau-b correlation scores were calculated based on the ranking of the cluster according to the level of accessibility of the chromatin fiber and the expression level of the gene. Positive and negative correlations are indicated by red and blue circles, respectively. The radius of the circles is proportional to the correlation score. When significant (p-value < $10E^{-05}$), positive and negative correlations are highlighted in red or blue squares, respectively. C lusters that belong to super-clusters (i.e., trichoblasts, atrichoblasts, and meristematic, cortical, endodermal, and stele cells) are highlighted in red boxes.

Consequently, to further reveal the correspondence between sc/sNucRNA-seq and sNucATAC-seq experiments, we performed a correlation analysis between the expression of *Arabidopsis* marker genes and chromatin accessibility at the location of their TSSs. Upon mining the scRNA-seq and sNucRNA-seq datasets, we selected the top 20 marker genes from each cluster based on their fold-change of expression compared with other clusters and on their lowest *p* values. Due to some redundancy between clusters, we identified 370 unique marker genes, including 32 that belong to the list of 101 marker genes used to annotate the sc/sNucRNA-seq clusters (31.7%, Supplemental Table 3). Among them, 336 are characterized by at least one sNucATAC-seq peak near their TSS (Supplemental Table 6). As a comparison, we also performed a similar analysis on 811 *Arabidopsis* root housekeeping (HK) genes (i.e., genes expressed in at least 50% of the cells and nuclei composing each cluster, across all the 20 and 21 scRNA-seq and sNucRNA-seq clusters, respectively, and with identified TSS-associated sNucATAC-seq peaks; Supplemental Table 7), and on three randomized subsets of 336 genes (Supplemental Figures 16 and 17).

Upon normalization of their expression patterns and peak distributions across the different clusters, the marker and HK genes were distributed in 20 different bins based on their transcriptional activity and chromatin accessibility profiles. Applying Kendall's tau-b rank correlation test for non-normally distributed data on the marker genes, we observed significant positive correlations between almost all the co-annotated sc/sNucRNA-seq and sNucATAC-seq datasets (Figure 4, see red squares in the descending diagonals; $p < 10E{-}05$). This result supports the idea that the differential chromatin accessibility correlates with the expression patterns of at least several cluster marker genes. The similarity between the Kendall tau-b correlation maps when using the scRNA-seq and sNucRNA-seq datasets also supports the biological relevance of a single-nucleus transcriptome compared with the single-cell transcriptome. Based on these results, we assume that, similar to their transcriptional activity, the chromatin accessibility at the location of the TSS of selected genes can be used as a molecular marker of cell-type identity. As a comparison, the sc/sNucRNA-seq and sNucATAC-seq correlation analysis of the 811 HK genes and randomized genes revealed only a few significant and more moderate correlations (Supplemental Figures 16 and 17). However, significant correlations were mostly restricted to co-annotated

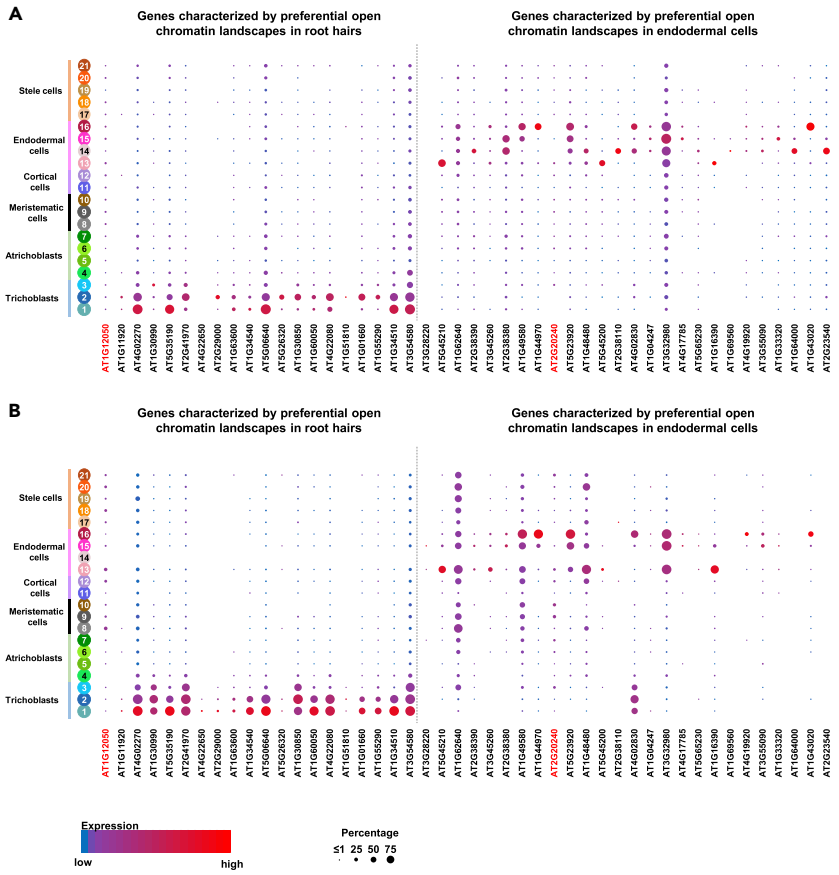

Figure 5. Normalized expression levels of 46 *Arabidopsis* genes (*x* axis) characterized by the higher accessibility of the chromatin at the location of their TSS in the root hair (left) and endodermal (right) clusters

Gene expression levels were quantified by using sNucRNA-seq **(A)** and scRNA-seq **(B)**. The percentage of cells/nuclei expressing the gene of interest (circle size) and the mean expression (circle color) of genes are shown. Genes that are not preferentially expressed in root hairs (left) or in endodermal cells (right) are highlighted in red. Preferential expression in this experiment is defined as the expression of the gene in the root hair or the endoderm superclusters being at least equal to the expression of the gene in the five remaining clusters.

sc/sNucRNA-seq and sNucATAC-seq datasets (see red squares in the descending diagonals; $p < 10E{-}05$), suggesting that chromatin accessibility near the TSS has a significant influence on gene expression.

When considering the marker genes, significant positive correlations between sc/sNucRNA-seq and sNucATAC-seq datasets were also repeatedly observed between the clusters that belong to the same superclusters (e.g., the "stele," "atrichoblast," and "trichoblast" superclusters; Figure 4, see red boxes; $p < 10E{-}05$). This result suggests that the cells composing the same tissue share, to some extent, similar transcriptomic and epigenomic signature profiles that are likely required to fulfill their tissue-specific biological functions. The significant correlations between gene expression and chromatin accessibility suggest that the position of the nucleosomes on the double strand of the genomic DNA, near the TSS of the genes, plays a critical role in controlling the activity of at least a subset of the marker genes.

Chromatin accessibility at the single-cell resolution revealed by ATAC-seq can be used as a molecular marker indicating root hair and endodermal cell developmental states

Plant single cell types are annotated based on the expression profile of marker genes (Denyer et al., 2019; Jean-Baptiste et al., 2019; Ryu et al., 2019; Shulse et al., 2019; Zhang et al., 2019). Correlation analyses between gene expression and chromatin accessibility (Figure 4) suggest that the latter could also be used as another molecular marker of the cell types composing the *Arabidopsis* root. This hypothesis is supported by similar analyses conducted on various animal organs and selected plant cells (Preissl et al., 2018; Sijacic et al., 2018; Sinnamon et al., 2019). To further test this hypothesis, we took advantage of our single-cell resolution ATAC-seq datasets and focused our analysis on the three and four clusters representative of the mature *Arabidopsis* root hair and endodermal cells (i.e., clusters 1, 2, and 3, and clusters 13, 14, 15, and 16, respectively; Figure 2A).

To validate the use of the differential chromatin accessibility as a molecular marker of plant cell identity, we first looked for the sNucATAC-seq peaks that are preferentially identified in the three and four root hair and endodermal cell clusters (fold-change of the average value across the root hair/endodermal clusters versus the next highest non-root-hair/non-endodermal cluster ≥ 2). Among the 11 858 sNucATAC-seq peaks paired with the TSSs of annotated *Arabidopsis* genes, 20 and 26 are specifically identified in the root hair and endodermal clusters, respectively (Supplemental Table 8). Mining the scRNA-seq and sNucRNA-seq datasets, we identified 19 (95%) and 25 genes (96.2%) preferentially expressed in *Arabidopsis* root hair and endodermal cells, respectively (i.e., the expression of the gene in the root hair or the endoderm supercluster is at least equal to the expression of the gene in the five remaining superclusters; Figure 5). Initiating our analysis by identifying the peaks of accessible chromatin specific to clusters 1, 2, and 3, and clusters 13, 14, 15, and 16, led to the identification of associated root hair and endodermal marker genes. In addition to supporting the role of chromatin accessibility in controlling gene expression in plant cells, this work highlights that chromatin accessibility can be used as a molecular marker to annotate specific cell types.

In this study, we first validated the use of isolated nuclei to access biologically relevant transcriptomic information. Then, we provided a comprehensive integration of transcriptomic and

chromatin accessibility datasets at the single-cell level. The study suggests that, upon their isolation, plant nuclei can be used to capture the transcriptome of various cell types. Such an approach will open new avenues to explore the dynamic transcriptomic changes occurring during cell differentiation and elongation, and between cell types. When applied across various plant species, single-nucleus-based transcriptomic and epigenomic analyses will also enable more targeted comparative analysis between plant species to reveal the evolution of gene expression at the single-cell-type level.

METHODS

Plant growth

Arabidopsis (*Arabidopsis thaliana*, Col-0 ecotype) seeds were surface sterilized using ethanol 70% and then a 30% (v/v) bleach, 0.1% (v/v) Triton X-100 solution for 5 min each. The seeds were then placed on Murashige and Skoog growth medium agar plates under 16 h light and 8 h dark conditions. Seven days after germination, the primary roots of around 150 seedlings were isolated from the seedlings.

Root nucleus isolation and single-nucleus RNA-seq and ATAC-seq library construction

To maximize the relevance of the transcriptomic information collected from plant isolated nuclei, we developed a methodology to isolate nuclei and process them for reverse transcription. The release of the plant nuclei is accomplished in no more than 15 min and exclusively occurs at 4°C to minimize RNA degradation, prevent nuclear membranes, and prevent transcriptional changes associated with the isolation process. As a comparison, 60–120 min of incubation at room temperature to 25°C is required to digest the cell wall to release *Arabidopsis* root protoplasts (Denyer et al., 2019; Jean-Baptiste et al., 2019; Ryu et al., 2019; Shulse et al., 2019; Zhang et al., 2019). Upon centrifugation, fluorescent activated nuclei sorting (FANS), and estimation of the nucleate density, the nuclei are processed into the 10X Genomics Chromium system for the construction of the sNucRNA-seq libraries in less than 90 min following the isolation of the plants from the Petri dishes.

To generate sNucRNA-seq libraries, *Arabidopsis* root nuclei were isolated by chopping the roots in 1 ml of nuclear isolation buffer (CELLYTPN1-1KT, Sigma-Aldrich) and incubated for 15 min on a rocking shaker at 4°C. The lysate was filtered through a 40 μm cell strainer (Corning) and stained with propidium iodide (20 μg/ml, Alfa Aesar). Nuclei were sorted using a FACSAria II flow cytometer (BD, Flow Cytometry Service Center), pelleted (1000 *g* for 10 min at 4°C), and resuspended in PBS buffer without calcium and magnesium and supplemented with Protector RNase inhibitor (0.2 U/μl, Roche) and ultrapure BSA (0.1%, Invitrogen). The nucleus concentration was estimated using a Countess II FL automated cell counter (Thermo Fisher) upon staining with trypan blue. The nuclei were used as an input to generate RNA-seq libraries following the 10X Genomics Chromium Single Cell 3′ v.3 protocol without modification.

To create sNucATAC-seq libraries, the nuclei were isolated by chopping and incubated for only 5 min in the nuclear isolation buffer supplemented with Triton X-100 (v/v) 0.5% final to permeabilize the nuclei. Upon incubation, the nuclei were filtered as described above. The nuclei were then centrifuged at 500 *g* for 10 min, washed with PBS supplemented with BSA 0.1%, and then pelleted again. Unlike the population of nuclei used to establish sNucRNA-seq libraries, we did not stain with propidium iodide or sort the nuclei dedicated to sNucATAC-seq technology. Preliminary work conducted on *Arabidopsis* nuclei clearly highlighted that both staining and sorting decrease the number of unique fragments per barcode, affecting the resolution of the clusters of cells (data not shown). Upon isolation and filtration, the nuclei pellet was resuspended in the ATAC-seq nucleus buffer (10X Genomics, No. 2000153). Nucleus concentration was estimated using the Countess II FL automated cell counter (Thermo Fisher). We used 6768 nuclei isolated from two independent replicates to generate sNucATAC-seq libraries following the 10X Genomics Chromium Single Cell ATAC protocol without modification.

The sNucRNA-seq and sNucATAC-seq libraries were sequenced on an Illumina HiSeq platform.

Bioinformatics analyses

sc/sNucRNA-seq

scRNA-seq data for the three wild-type replicates of previously published (Ryu et al., 2019) data (SRA experiments SRX5074330-SRX5074332) and five sNucRNA-seq replicates from the current study (Supplemental Table 1) were processed individually using the 10X Genomics Cell Ranger (v.3.1.0) count pipeline, against a reference constructed from the TAIR10 genome and Araport11 annotations. The recommended protocol for reference construction in support of pre-mRNA capture for data derived from nuclear preparations was applied.

Following generation of the unique molecular identifier (UMI) count matrices, doublet filtering was performed on each dataset using 50 iterations of the BoostClassifier prediction method of DoubletDetection (p_thresh = 1e-7, voter_thresh = 0.8) (Gayoso et al., 2018). In addition, the published scRNA-seq data were filtered to include only those cells passing quality control as performed in Ryu et al. (2019) by filtering on cell barcodes present for the wild-type replicates in the GEO dataset GSE123013. All datasets were filtered to include only called cells with more than 200 genes represented by at least one UMI. Normalization of individual datasets and their integration was performed using Seurat v.3.1.5, with genes required to be present in a minimum of five cells, and the top 2000 variable genes were used for feature selection (Supplemental Figure 18A). Integration anchors were chosen for the combined set of three scRNA-seq and five sNucRNA-seq replicates using the first 20 dimensions of the canonical correlation analysis method. Following integration, UMAP dimensionality reduction was performed on the first 20 principal components (Supplemental Figure 18C), and clustering was performed using Seurat's FindClusters method with a resolution of 0.5 (Supplemental Figure 18D). Expression values were obtained separately for the subsets of cells and nuclei belonging to each cluster using Seurat's AverageExpression method.

sNucATAC-seq

sNucATAC-seq data from two replicates were processed using the Cell Ranger ATAC (v.1.2) count pipeline, with a reference constructed using TAIR10/Araport11 and JASPAR2018_CORE_plants_non-redundant_pfms. Counts against peaks called by Cell Ranger ATAC were read into Seurat objects and subset to include only cells with between 1000 and 20 000 fragment counts in peak regions, and >15% of the cell's fragments in peak regions and a calculated nucleosome_signal of >10 (Supplemental Figure 18B). Integration of the two replicates was performed using the Harmony integration strategy (Korsunsky et al., 2019) on the latent semantic indexing reduction performed by Signac (v.0.2.5) (Stuart et al., 2019), with UMAP dimensionality reduction performed on the first 20 dimensions of the resulting integrated dataset (Supplemental Figure 18E). An unsupervised clustering was performed on the sNucATAC-seq data using Seurat's FindClusters method using the Smart Local Moving algorithm (Supplemental Figure 18F). TSS enrichment was analyzed per chromosome using the TSSEnrichment/TSSPlot functions of Signac.

Correlation analysis between bulk and pseudo-bulked RNA-seq and ATAC-seq datasets

To perform correlation analyses between bulk and pseudo-bulked RNA-seq datasets, we calculated SRCC between bulk, protoplasted bulk, and pseudo-bulk sNucRNA-seq (present study) and pseudo-bulked scRNA-seq datasets (SRA Accession Nos. SRX5128615 [Denyer et al., 2019]; SRX8089019, SRX8089020, SRX4913498, and SRX4913499 [Jean-Baptiste et al., 2019]; SRX5074330, SRX5074331, and SRX5074332 [Ryu et al., 2019]; and SRX5025979-SRX5025988 [Shulse et al., 2019]).

To perform correlation tests between bulk and pseudo-bulk ATAC-seq datasets, bulk ATAC-seq reads from *Arabidopsis* root tip samples (SRA Accession No. SRX3006634) were aligned to the genome using HiSat2, and duplicates were identified using PicardTools MarkDuplicates. Read counts for each of the peak regions from our analysis were determined with the samtools bedcov method and the read counts for the bulk sample were compared with those derived from bedcov on pseudo-bulked sNucATAC-seq data combined (samtools merge) from both replicates. Correlation tests were performed using the same methods as for bulk and pseudo-bulk RNA-seq correlations.

UMAP visualization

To relate the structure of the UMAP plot for the combined sc/sNucRNA-seq datasets to that of the sNucATAC-seq data, a putative gene activity matrix was constructed from the aggregated fragments of the two replicates, using gene body coordinates extended 2 kb upstream of the TSS. Log normalization was performed on the activity matrix, and then the Seurat FindTransferAnchors method was used with the sNucATAC-seq data as a query against the combined RNA-seq data as a reference. Finally, cluster assignments were made on the sNucATAC-seq using the TransferData method and the integration anchors determined earlier (Supplemental Figure 18G). These cluster assignments were used for subsequent extraction of averaged fragment counts across all cells assigned to the cluster, using Seurat AverageExpression on the peaks.

For visualization purposes, data from all RNA-seq data were combined using the Cell Ranger aggregate method to produce a combined cloupe file, with UMAP coordinates and cluster assignments imported from the Seurat analysis. Similarly, the two sNucATAC-seq replicates were combined using Cell Ranger ATAC aggregated, and UMAP coordinates and cluster assignments from the Harmony/Signac analysis were imported into the cloupe rendering.

Correlation analyses between sNucRNA-seq and sNucATAC-seq data

These analyses were performed by matching called peaks from the ATAC-seq data to genes by looking for any overlap of peak regions with genes using bedtools (v.2.29.2) (Quinlan and Hall, 2010) and then applying a custom script to choose among overlapped genes based on the maximum extension of the peak region into the 5′ upstream region. Correlation analysis was performed between matrices constructed from normalized average expression values for genes and average fragment count values for peaks, with the corresponding rows in the two matrices representing matched gene and peak entities. The normalization procedure was done on values in each row independently, by considering each cluster in terms of its percentage representation of the expression/fragment count for the gene/peak across all clusters, in order to represent specificity independent of the magnitude of the expression/fragment count. These normalized values were considered as mapping onto 20 quantiles, roughly corresponding to the number of clusters in the analysis. Thus, a gene with perfect specificity would be mapped into the highest quantile for its cluster context of expression and into the lowest quantile for all other clusters, while a gene expressed approximately equally across all clusters would be mapped into the next-to-least quantile for each cluster. The R cor function with method = "kendall" (implementing the Kendall tau-b method for rank-based correlation in the presence of tied ranks) was used to produce correlation matrices, which were visualized as correlograms from the corrplot package (Wei, 2017).

Creation of the sc/sNucRNA-seq and sNucATAC-seq heatmaps

The heatmaps were created in R (v.3.4.4) using the "heatmap.2" function. Upon normalization, the color scale and color breakpoint used to create the heatmaps were selected as described below: the minimum (blue) and maximum (red) color intensities were set to 0 and 71.48, the highest normalized value across the datasets. The white breakpoint color was selected to reflect the median value of the datasets.

ACCESSION NUMBERS

Expression data are available at the Gene Expression Omnibus (GEO: GSE155304).

SUPPLEMENTAL INFORMATION

Supplemental Information is available at *Molecular Plant Online*.

FUNDING

This work was supported by a grant to M.L. from the U.S. National Science Foundation (IOS #1339194 and 1854326), by grants to J.S. from the U.S. National Science Foundation (IOS #1923589) and the Department of Energy (DE-SC0020358), by the Center for Plant Science Innovation, and by the Department of Agronomy and Horticulture at the University of Nebraska-Lincoln.

AUTHOR CONTRIBUTIONS

S.T. and K.H.R. performed experiments. A.D.F., S.C., K.H.R., J.S., and M.L. carried out data analysis. M.L. oversaw the study and drafted the manuscript. All authors contributed to the preparation of the manuscript.

ACKNOWLEDGMENTS

The authors would like to acknowledge Dirk Anderson, manager of the Flow Cytometry Core Facility, and the Nebraska Center for Biotechnology at the University of Nebraska-Lincoln, for providing support in the sorting of the isolated nuclei. We also want to acknowledge Dr. Clark Boomer, Project Manager of the Genomics Core at the University of Kansas Medical Center, for the sequencing of the sNucATAC-seq libraries. In addition, we want to thank Dr. Jeffrey Doyle, from Cornell University for his constructive comments and feedback on the manuscript, and Dr. Rafal Woycicki for his help creating the heatmaps.

Received: August 4, 2020
Revised: December 14, 2020
Accepted: January 5, 2021
Published: January 6, 2021

REFERENCES

Andersen, T.G., Barberon, M., and Geldner, N. (2015). Suberization - the second life of an endodermal cell. Curr. Opin. Plant Biol. **28**:9–15.

Arendt, D., Bertucci, P.Y., Achim, K., and Musser, J.M. (2019). Evolution of neuronal types and families. Curr. Opin. Neurobiol. **56**:144–152.

Bakken, T.E., Hodge, R.D., Miller, J.A., Yao, Z., Nguyen, T.N., Aevermann, B., Barkan, E., Bertagnolli, D., Casper, T., Dee, N., et al. (2018). Single-nucleus and single-cell transcriptomes compared in matched cortical cell types. PLoS One **13**:e0209648.

Becht, E., McInnes, L., Healy, J., Dutertre, C.A., Kwok, I.W.H., Ng, L.G., Ginhoux, F., and Newell, E.W. (2019). Dimensionality reduction for visualizing single-cell data using UMAP. Nat. Biotechnol. **37**:38.

Beisson, F., Li, Y., Bonaventure, G., Pollard, M., and Ohlrogge, J.B. (2007). The acyltransferase GPAT5 is required for the synthesis of suberin in seed coat and root of Arabidopsis. Plant Cell **19**:351–368.

Berhin, A., de Bellis, D., Franke, R.B., Buono, R.A., Nowack, M.K., and Nawrath, C. (2019). The root cap cuticle: a cell wall structure for seedling establishment and lateral root formation. Cell **176**:1367–1378.e68.

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

Bohme, K., Li, Y., Charlot, F., Grierson, C., Marrocco, K., Okada, K., Laloue, M., and Nogue, F. (2004). The Arabidopsis COW1 gene encodes a phosphatidylinositol transfer protein essential for root hair tip growth. Plant J. **40**:686–698.

Buenrostro, J.D., Corces, M.R., Lareau, C.A., Wu, B., Schep, A.N., Aryee, M.J., Majeti, R., Chang, H.Y., and Greenleaf, W.J. (2018). Integrated single-cell analysis maps the continuous regulatory landscape of human hematopoietic differentiation. Cell **173**:1535–1548.e16.

Buenrostro, J.D., Wu, B., Litzenburger, U.M., Ruff, D., Gonzales, M.L., Snyder, M.P., Chang, H.Y., and Greenleaf, W.J. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. Nature **523**:486–490.

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.

Chen, C.Y., and Schmidt, W. (2015). The paralogous R3 MYB proteins CAPRICE, TRIPTYCHON and ENHANCER OF TRY AND CPC1 play pleiotropic and partly non-redundant roles in the phosphate starvation response of Arabidopsis roots. J. Exp. Bot. **66**:4821–4834.

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

Del Toro-De Leon, G., and Kohler, C. (2019). Endosperm-specific transcriptome analysis by applying the INTACT system. Plant Reprod. **32**:55–61.

Denyer, T., Ma, X., Klesen, S., Scacchi, E., Nieselt, K., and Timmermans, M.C.P. (2019). Spatiotemporal developmental trajectories in the Arabidopsis root revealed using high-throughput single-cell RNA sequencing. Dev. Cell **48**:840–852.e45.

Ding, L.N., Li, M., Wang, W.J., Cao, J., Wang, Z., Zhu, K.M., Yang, Y.H., Li, Y.L., and Tan, X.L. (2019). Advances in plant GDSL lipases: from sequences to functional mechanisms. Acta Physiol. Plant **41**:151.

Dominguez, E., Heredia-Guerrero, J.A., and Heredia, A. (2015). Plant cutin genesis: unanswered questions. Trends Plant Sci. **20**:551–558.

Fendrych, M., Van Hautegem, T., Van Durme, M., Olvera-Carrillo, Y., Huysmans, M., Karimi, M., Lippens, S., Guerin, C.J., Krebs, M., Schumacher, K., et al. (2014). Programmed cell death controlled by ANAC033/SOMBRERO determines root cap organ size in Arabidopsis. Curr. Biol. **24**:931–940.

Gao, Z., Daneva, A., Salanenka, Y., Van Durme, M., Huysmans, M., Lin, Z., De Winter, F., Vanneste, S., Karimi, M., Van de Velde, J., et al. (2018). KIRA1 and ORESARA1 terminate flower receptivity by promoting cell death in the stigma of Arabidopsis. Nat. Plants **4**:365–375.

Gayoso, A.S.J., Carr, A.J., Sharma, R., and Pe'er, D. (2018). DoubletDetection (version 2.4) (Geneva, Switzerland: Zenodo CERN).

Girard, A.L., Mounet, F., Lemaire-Chamley, M., Gaillard, C., Elmorjani, K., Vivancos, J., Runavot, J.L., Quemener, B., Petit, J., Germain, V., et al. (2012). Tomato GDSL1 is required for cutin deposition in the fruit cuticle. Plant Cell **24**:3119–3134.

Gulko, B., and Siepel, A. (2019). An evolutionary framework for measuring epigenomic information and estimating cell-type-specific fitness consequences. Nat. Genet. **51**:335–342.

Heo, J.O., Blob, B., and Helariutta, Y. (2017). Differentiation of conductive cells: a matter of life and death. Curr. Opin. Plant Biol. **35**:23–29.

Hu, P., Liu, J., Zhao, J., Wilkins, B.J., Lupino, K., Wu, H., and Pei, L. (2018). Single-nucleus transcriptomic survey of cell diversity and functional maturation in postnatal mammalian hearts. Genes Dev. **32**:1344–1357.

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.

Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.R., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. Nat. Methods **16**:1289–1296.

Kumpf, R.P., and Nowack, M.K. (2015). The root cap: a short story of life and death. J. Exp. Bot. **66**:5651–5662.

Kwak, S.H., Shen, R., and Schiefelbein, J. (2005). Positional signaling mediated by a receptor-like kinase in Arabidopsis. Science **307**:1111–1113.

Lai, B., Gao, W., Cui, K., Xie, W., Tang, Q., Jin, W., Hu, G., Ni, B., and Zhao, K. (2018). Principles of nucleosome organization revealed by single-cell micrococcal nuclease sequencing. Nature **562**:281–285.

Lai, C.P., Huang, L.M., Chen, L.O., Chan, M.T., and Shaw, J.F. (2017). Genome-wide analysis of GDSL-type esterases/lipases in Arabidopsis. Plant Mol. Biol. **95**:181–197.

Li, Y., Beisson, F., Koo, A.J., Molina, I., Pollard, M., and Ohlrogge, J. (2007a). Identification of acyltransferases required for cutin biosynthesis and production of cutin with suberin-like monomers. Proc. Natl. Acad. Sci. U S A **104**:18339–18344.

Li, Y., Beisson, F., Ohlrogge, J., and Pollard, M. (2007b). Monoacylglycerols are components of root waxes and can be produced in the aerial cuticle by ectopic expression of a suberin-associated acyltransferase. Plant Physiol. **144**:1267–1277.

Lu, Z., Hofmeister, B.T., Vollmers, C., DuBois, R.M., and Schmitz, R.J. (2017). Combining ATAC-seq with nuclei sorting for discovery of cis-regulatory regions in plant genomes. Nucleic Acids Res. **45**:e41.

Maher, K.A., Bajic, M., Kajala, K., Reynoso, M., Pauluzzi, G., West, D.A., Zumstein, K., Woodhouse, M., Bubb, K., Dorrity, M.W., et al. (2018). Profiling of accessible chromatin regions across multiple plant species and cell types reveals common gene regulatory principles and new control modules. Plant Cell **30**:15–36.

Norrie, J.L., Lupo, M.S., Xu, B., Al Diri, I., Valentine, M., Putnam, D., Griffiths, L., Zhang, J., Johnson, D., Easton, J., et al. (2019). Nucleome dynamics during retinal development. Neuron **104**:512–528 e511.

Olvera-Carrillo, Y., Van Bel, M., Van Hautegem, T., Fendrych, M., Huysmans, M., Simaskova, M., van Durme, M., Buscaill, P., Rivas, S., Coll, N.S., et al. (2015). A conserved Core of programmed cell death indicator genes discriminates developmentally and environmentally induced programmed cell death in plants. Plant Physiol. **169**:2684–2699.

Palovaara, J., and Weijers, D. (2019). Adapting INTACT to analyse cell-type-specific transcriptomes and nucleocytoplasmic mRNA dynamics in the Arabidopsis embryo. Plant Reprod. **32**:113–121.

Park, J.J., Jin, P., Yoon, J., Yang, J.I., Jeong, H.J., Ranathunge, K., Schreiber, L., Franke, R., Lee, I.J., and An, G. (2010). Mutation in Wilted Dwarf and Lethal 1 (WDL1) causes abnormal cuticle formation and rapid water loss in rice. Plant Mol. Biol. **74**:91–103.

Pirrello, J., Deluche, C., Frangne, N., Gevaudant, F., Maza, E., Djari, A., Bourge, M., Renaudin, J.P., Brown, S., Bowler, C., et al. (2018). Transcriptome profiling of sorted endoreduplicated nuclei from tomato fruits: how the global shift in expression ascribed to DNA ploidy influences RNA-Seq data normalization and interpretation. Plant J. **93**:387–398.

Pott, S., and Lieb, J.D. (2015). Single-cell ATAC-seq: strength in numbers. Genome Biol. **16**:172.

Preissl, S., Fang, R., Huang, H., Zhao, Y., Raviram, R., Gorkin, D.U., Zhang, Y., Sos, B.C., Afzal, V., Dickel, D.E., et al. (2018). Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. Nat. Neurosci. **21**:432–439.

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

Reynoso, M.A., Pauluzzi, G.C., Kajala, K., Cabanlit, S., Velasco, J., Bazin, J., Deal, R., Sinha, N.R., Brady, S.M., and Bailey-Serres, J.

(2018). Nuclear transcriptomes at high resolution using retooled INTACT. Plant Physiol. **176**:270–281.

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.

Salazar-Henao, J.E., Velez-Bermudez, I.C., and Schmidt, W. (2016). The regulation and plasticity of root hair patterning and morphogenesis. Development **143**:1848–1858.

Sarkar, P., Bosneaga, E., and Auer, M. (2009). Plant cell walls throughout evolution: towards a molecular understanding of their design principles. J. Exp. Bot. **60**:3615–3635.

Schwanhausser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature **473**:337–342.

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.

Sijacic, P., Bajic, M., McKinney, E.C., Meagher, R.B., and Deal, R.B. (2018). Changes in chromatin accessibility between Arabidopsis stem cells and mesophyll cells illuminate cell type-specific transcription factor networks. Plant J. **94**:215–231.

Sinnamon, J.R., Torkenczy, K.A., Linhoff, M.W., Vitak, S.A., Mulqueen, R.M., Pliner, H.A., Trapnell, C., Steemers, F.J., Mandel, G., and Adey, A.C. (2019). The accessible chromatin landscape of the murine hippocampus at single-cell resolution. Genome Res. **29**:857–869.

Somssich, M., Khan, G.A., and Persson, S. (2016). Cell wall heterogeneity in root development of Arabidopsis. Front. Plant Sci. **7**:1242.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. Cell **177**:1888–1902 e1821.

Tannenbaum, M., Sarusi-Portuguez, A., Krispil, R., Schwartz, M., Loza, O., Benichou, J.I.C., Mosquna, A., and Hakim, O. (2018). Regulatory chromatin landscape in Arabidopsis thaliana roots uncovered by coupling INTACT and ATAC-seq. Plant Methods **14**:113.

Tsukagoshi, H., Busch, W., and Benfey, P.N. (2010). Transcriptional regulation of ROS controls transition from proliferation to differentiation in the root. Cell **143**:606–616.

Turco, G.M., Rodriguez-Medina, J., Siebert, S., Han, D., Valderrama-Gomez, M.A., Vahldick, H., Shulse, C.N., Cole, B.J., Juliano, C.E., Dickel, D.E., et al. (2019). Molecular mechanisms driving switch behavior in xylem cell differentiation. Cell Rep. **28**:342–351 e344.

Wei, T.S.,V. (2017). R Package "corrplot": Visualization of a Correlation Matrix (version 0.84). https://github.com/taiyun/corrplot.

Yu, G., Wang, L.G., and He, Q.Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics **31**:2382–2383.

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.

Zhou, W., Ji, Z., Fang, W., and Ji, H. (2019). Global prediction of chromatin accessibility using small-cell-number and single-cell RNA-seq. Nucleic Acids Res. **47**:e121.