Title

Molecular Mechanisms Driving Switch Behavior in Xylem Cell Differentiation

Journal

Cell Reports, 28(2)

ISSN

2639-1856

Authors

Turco, Gina M

Rodriguez-Medina, Joel

Siebert, Stefan

et al.

Publication Date

2019-07-01

DOI

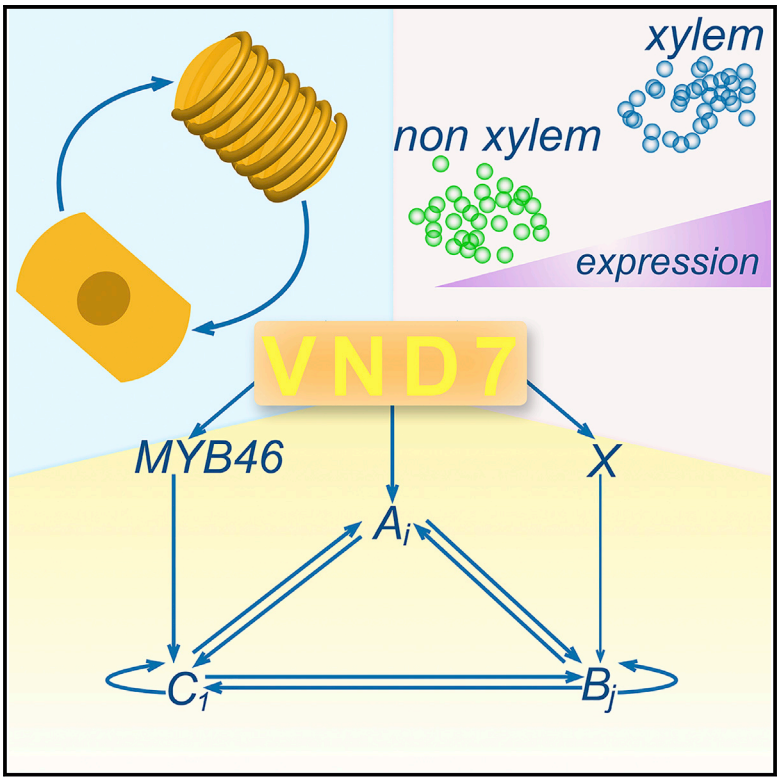
10.1016/j.celrep.2019.06.041

Peer reviewed

Cell Reports

Molecular Mechanisms Driving Switch Behavior in Xylem Cell Differentiation

Graphical Abstract



Authors

Gina M. Turco, Joel Rodriguez-Medina, Stefan Siebert, ..., Diane E. Dickel, Michael A. Savageau, Siobhan M. Brady

Correspondence

sbrady@ucdavis.edu

In Brief

Turco et al. demonstrate that *VND7* initiates sharp switching of root cells to xylem cell identity. The authors identify four candidate *VND7* downstream target genes capable of generating this switch. This work provides a framework for models that give rise to totipotency relative to terminal differentiation.

Highlights

- *VND7* modulates switching to xylem fate, as observed in whole roots and single cells
- Single-cell sequencing reveals xylem cell subtypes
- Identification of genes and the network capable of generating a switch to xylem fate

Turco et al., 2019, Cell Reports *28*, 342–351
July 9, 2019
https://doi.org/10.1016/j.celrep.2019.06.041

CellPress

OPEN ACCESS
CellPress

Cell Reports
Report

Molecular Mechanisms Driving Switch Behavior in Xylem Cell Differentiation

Gina M. Turco,[1] Joel Rodriguez-Medina,[1] Stefan Siebert,[4] Diane Han,[1] Miguel Á. Valderrama-Gómez,[2] Hannah Vahldick,[1] Christine N. Shulse,[5] Benjamin J. Cole,[5] Celina E. Juliano,[4] Diane E. Dickel,[6] Michael A. Savageau,[2,3] and Siobhan M. Brady[1,7,*]

[1]Department of Plant Biology and Genome Center, University of California, Davis, Davis, CA 95616, USA
[2]Department of Microbiology & Molecular Genetics, University of California, Davis, Davis, CA 95616, USA
[3]Department of Biomedical Engineering, University of California, Davis, Davis, CA 95616, USA
[4]Department of Molecular and Cellular Biology, University of California, Davis, Davis, CA 95616, USA
[5]Department of Energy Joint Genome Institute, Walnut Creek, CA 94598, USA
[6]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[7]Lead Contact
*Correspondence: sbrady@ucdavis.edu
https://doi.org/10.1016/j.celrep.2019.06.041

SUMMARY

Plant xylem cells conduct water and mineral nutrients. Although most plant cells are totipotent, xylem cells are unusual and undergo terminal differentiation. Many genes regulating this process are well characterized, including the *Vascular-related NAC Domain 7* (*VND7*), *MYB46*, and *MYB83* transcription factors, which are proposed to act in interconnected feedforward loops (FFLs). Less is known regarding the molecular mechanisms underlying the terminal transition to xylem cell differentiation. Here, we generate whole-root and single-cell data, which demonstrate that *VND7* initiates sharp switching of root cells to xylem cell identity. Based on these data, we identified 4 candidate *VND7* downstream target genes capable of generating this switch. Although *MYB46* responds to *VND7* induction, it is not among these targets. This system provides an important model to study the emergent properties that may give rise to totipotency relative to terminal differentiation and reveals xylem cell subtypes.

INTRODUCTION

Xylem cells are critical for the transport of water, sugars, and mineral nutrients from the soil to the aboveground tissues of the plant. The development of these xylem vessels and tracheary elements has facilitated the evolution of vascular plants from bryophytes, which need to grow on or near the surface of water, to trees, allowing the transport of water over 100 meters (Raven, 1993). This essential role has led to the secondary cell wall (SCW) surrounding xylem cells comprising most terrestrial biomass. The xylem SCW is composed primarily of polymers: cellulose, hemicellulose, and lignin, which provide the structural support needed for long-distance transport. In addition, these polymers determine both cell shape and cell specialization. Xylem vessel type can be distinguished by SCW deposition patterns. Protoxylem cells have spiral and annular thickenings, while metaxylem cells have pitted thickenings.

In the *Arabidopsis* root, the xylem developmental trajectory can be separated into 3 distinct stages. First, xylem vessels are specified by positional signals from stem cell precursor cells (De Rybel et al., 2014). They then undergo proliferation and elongation and finally enter the differentiation zone, where they undergo SCW deposition and programmed cell death to form a long, hollow vessel (Déjardin et al., 2010; Fukuda, 1996). Although most plant cells are totipotent, meaning that they can differentiate into any other cell type within the organism, xylem cells are unusual in that they undergo a unidirectional developmental trajectory, ending with terminal differentiation and programmed cell death. This is required for function—that is, its ability to conduct water and nutrients. This raises an important question: what molecular mechanisms facilitate the commitment to terminal cell differentiation in xylem cells, and how are these mechanisms distinct from the differentiation of plant cells that retain totipotency?

In contrast to plants, totipotency is not a feature of most animal cells. Instead, as animal cells continue along their developmental trajectory, their ability to switch fates becomes progressively reduced. Most examples of cell-fate determination that have been well characterized at the molecular level in organisms from microbes to mammals involve bistable hysteretic switches. Examples include the lysis-lysogeny decision in the λ virus, tracheal cell specification from a field of progenitor cells in *Drosophila melanogaster*, gut differentiation from the endoderm in *Caenorhabditis elegans*, differentiation of the Veg_2 lineage in *Strongylocentrotus purpuratus* embryos, and switching of neural progenitor cells to oligodendroglia in the brains of *Rattus norvegicus* (Ptashne, 2004; Fukushige et al., 1998; Lai et al., 2004; Maduro and Rothman, 2002; Marshall and McGhee, 2001; Metzger and Krasnow, 1999; Zelzer and Shilo, 2000; Zhu et al., 1997). A bistable hysteretic switch facilitates this developmental commitment by creating a sharp threshold value that a developmental signal must first exceed to switch fates and a buffer zone that would discourage inappropriate commitment or reversal of commitment in the presence of noise.

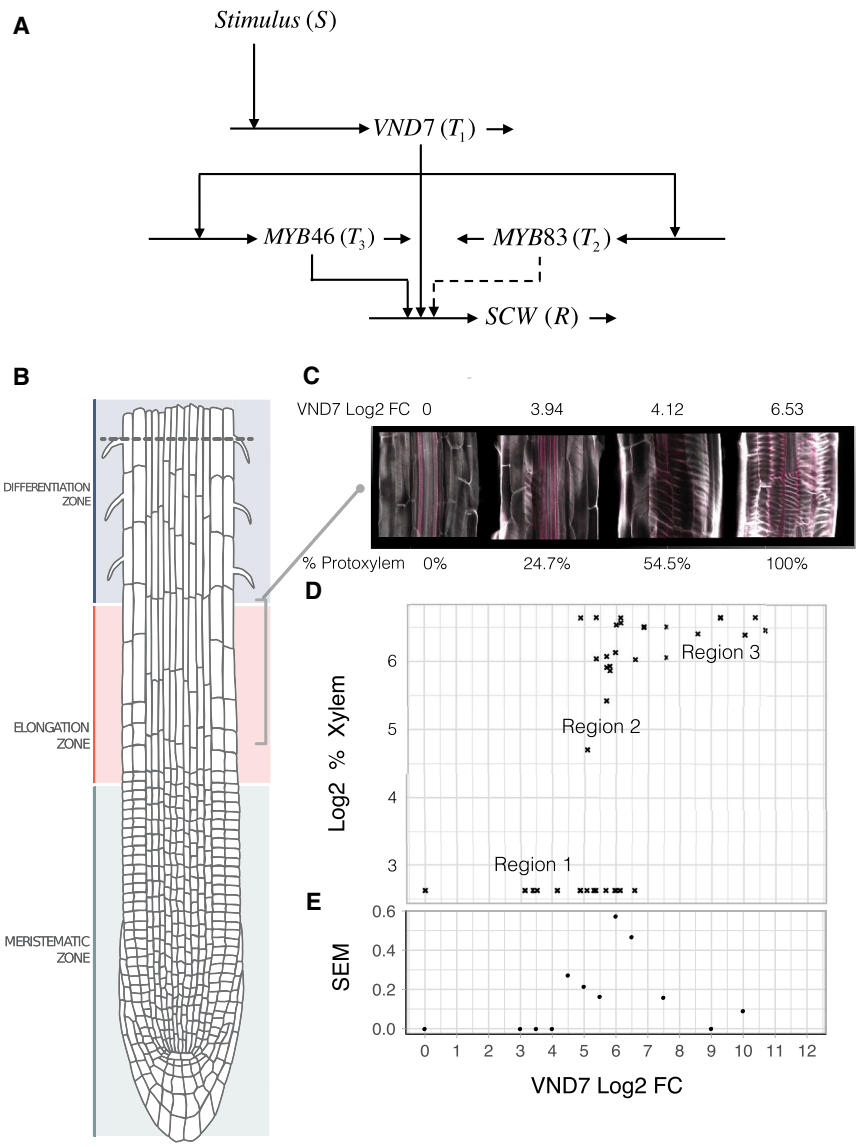


Figure 1. *VND7* Expression Stimulates a Sharp Switch in Xylem Cell Differentiation

(A) Here VND7 is turned on in response to a stimulus (S), where solid lines indicate known binding and regulation of a transcription factor (T_i) to the promoter region of indicated targets (R) (marked by the directionality of the arrow) and dotted lines represent inferred interactions.

(B) Xylem cell differentiation was measured in the root elongation zone. Image of *Arabidopsis* root (Frédéric, 2017).

(C) Differentiation was measured using calcofluor white and fuchsin to image cellulose (white) and lignin (pink) as hallmarks of xylem cell differentiation.

(D) Each X mark represents an individual root's xylem phenotype and its estimated VND7 fold change, inferred from RNA extracted in bulk from the plate of origin. Plates are grown at varying estradiol concentrations, each containing at least 3 biological replicates (separate plates). The percentage of xylem cells is represented in log format for a log-log comparison to *VND7* log fold change in response β-estradiol induction.

(E) To plot variance as a function of *VND7* fold change, the SEM was calculated for each *VND7* log fold change (n data points > 10).

Whether plants differ in a major way from other organisms with regard to developmental switches is unclear, because comparable molecular studies have not been carried out. One suggestion comes from Cruz-Ramírez et al. (2012), who demonstrated that the potential for an asymmetric cell division state was linked via mathematical modeling to bistability of nuclear localized SHORT-ROOT and SCARECROW transcription factor levels generated through feedback within feedforward loops (FFLs). Bistable hysteretic switches require positive feedback; monostable graded switches do not. But in the presence of sufficient error and with sufficient sensitivity, monostable graded mechanisms might provide an alternative to account for developmental switching in plants.

Genes involved in xylem cell differentiation have been identified through transdifferentiation experiments. Transdifferentiation of non-xylem cells—that is, any cell without xylem identity, independent of its developmental state—into xylem cells can be established through the addition of phytohormones (auxin, cytokinin, and brassinosteroids) (Fukuda, 1992; Fukuda and Komamine, 1980; Roberts, 1976; Sachs, 1969). Alternatively, xylem transdifferentiation can be induced by the Shaggy/Glycogen synthase kinase III inhibitor bikinin (Kondo et al., 2015). A hormone-mediated transdifferentiation study led to the identification of the transcription factors *Vascular-related NAC Domain proteins 6* and *7* (*VND6* and *VND7*) as being sufficient to drive xylem cell differentiation (Kondo et al., 2015; Kubo et al., 2005; Yamaguchi et al., 2010); thus, these genes are considered master regulators of xylem differentiation. Transdifferentiation studies were also used to identify direct downstream targets of VND7; the transcription factors *MYB83* and *MYB46* are suggested to act redundantly in the second tier of a transcriptional cascade controlling xylem SCW biosynthesis (McCarthy et al., 2009; Yamaguchi et al., 2011) through a series of FFLs (Figure 1A) (Mangan and Alon, 2003; Mangan et al., 2003; McCarthy et al., 2009; Taylor-Teeples et al., 2015).

Here, we generate whole-root and single-cell transcription data, formulate a quantitative method (separation score) to screen for potential VND7 downstream target genes capable of generating the developmental switch to xylem cell identity, and apply this screening method to identify 4 candidate genes. Although MYB46 responds to VND7 induction, it does not

display the switching behavior required to be among the candidate genes. We describe the necessary conditions for the realization of alternative switching mechanisms. These data provide insight into a design principle that underlies the formation of a cell type critical for plant growth and development and the transition of plants from water to land.

RESULTS

Xylem Formation in the Whole-Root Exhibits a Distinct Switching Pattern in Response to *VND7* Expression

To investigate how VND7 influences xylem differentiation, we used β-estradiol-*VND7*-inducible transgenic *Arabidopsis* plants (Coego et al., 2014). This two-component system drives expression of *VND7* in all cell types under the G10-90 constitutive promoter (Zuo et al., 2000). We varied the expression of *VND7* in a dose-dependent manner (using estradiol as the inducer) and then quantified the number of xylem cells in the elongation zone of the root, where protoxylem differentiation typically occurs (Figures 1B and 1C) (Benfey and Scheres, 2000; Esau, 1977). Roots were stained with calcofluor white to stain for cellulose and with basic fuchsin to mark lignin deposition, both a hallmark of xylem differentiation associated with SCW biosynthesis (Figure 1C) (Ursache et al., 2018). The ratio of cells with spiral patterning to the total number of cells in each root image was used as a measure of the percentage of transdifferentiation to xylem and ranged from 0% to 100% (Figures 1C and 1D; Table S1). When *VND7* expression was increased approximately 5-fold, a dramatic jump in xylem cell differentiation was observed. Cell types closer to the root vascular tissue were more likely to transdifferentiate into xylem cell types upon β-estradiol treatment, suggesting these cells may have more plasticity in their response to *VND7* induction (Figure S1).

The observed relationship between *VND7* and xylem differentiation (Figure 1D) exhibited 3 distinct regions: a region of minimal (wild-type, uninduced) levels of xylem cell differentiation when the value of *VND7* expression is below a lower threshold (5-fold change, region 1); a region of fully induce xylem differentiation, when the value of *VND7* expression is above an upper threshold (7-fold change, region 3); and a region in which xylem differentiation is at its maximum or minimum, with few intermediate points, when *VND7* expression is between the upper and the lower thresholds of *VND7* expression (region 2).

We used a sliding window to determine the variance in xylem formation via SEM for each binned *VND7* fold change expression value (Figure 1E). A sharp increase in variance was observed within the middle bin between values of 5- and 7-fold change (Figure 1E). Altogether, these data provide evidence that VND7 is a sufficient input stimulus for the switch in xylem differentiation.

Single-Cell Sequencing Reveals Switching to Xylem Cell Identity in Response to *VND7* Expression

Transdifferentiation is the process by which differentiated cells transform into different cell types. In the root elongation zone, we observed evidence of a distinctive pattern of switching among the population of cells. To test how *VND7* changes the transcriptional and developmental trajectory of individual cells with different identities, we carried out single-cell RNA sequencing (RNA-seq) profiling using Drop-seq at the minimum and uninduced (0 μM β-estradiol) and at the maximum and induced (20 μM β-estradiol) concentrations (Macosko et al., 2015), where 100% conversion was observed (Figure 1D). From Drop-seq, 250 cells were collected for each of the two conditions (Figure 2A; Table S2), and 374 transcriptomes were obtained. Cell identity was determined in an unsupervised manner and then compared between the minimum and the maximum *VND7* expression levels to determine how VND7 influences the transformation of cell identity. To determine identity, cells were clustered using a shared nearest-neighbor-based algorithm in Seurat (Satija et al., 2015). The identities of the 11 resulting clusters (Figure 2A) were determined by their average gene expression profiles using a predefined index of cell identity (ICI) algorithm for *Arabidopsis* roots (Figures 2A and 2B; Table S2) (Birnbaum et al., 2003; Birnbaum and Kussell, 2011; Brady et al., 2007; Dinneny et al., 2008; Efroni et al., 2015; Gifford et al., 2008; Lee et al., 2006; Nawy et al., 2005). As expected, a significant increase in the number of maturing xylem cells (protoxylem identity) was observed in the *VND7*-induced cell population (adjusted $p < 2.2 \times 10^{-16}$). A less dramatic although significant increase was observed in the number of phloem cells (adjusted $p = 0.0002$) (Figures 2A and 2B). *VND7* is also expressed in a subset of procambium cells (Yamaguchi et al., 2008). Procambium cells can differentiate into either phloem or xylem cells depending on positional signals, including brassinosteroid, auxin, and cytokinin (De Rybel et al., 2014; Kondo et al., 2015, 2016). These results suggest that VND7 is sufficient to drive xylem cell differentiation and potentially vascular initial cell specification as an intermediate step during VND7-mediated root cell transdifferentiation. In comparison, the numbers of cortex, trichoblast, and endodermis cells were significantly reduced in the 20 μM β-estradiol-treated cells relative to the uninduced cells (Figures 2A and 2B). Based on this change in cell proportions, we can infer that cortex, trichoblast, and endodermis cells are more likely to transform into differentiating xylem cells. This would need to be confirmed by measurement of *VND7* expression or protein levels in these cell types after induction.

The single-cell transcriptomes of these 374 cells provide a higher resolution (relative to whole-root expression profiling) to quantify the switching threshold in relation to *VND7* expression. Using this increased power to quantify the cell identity switch, we calculated the ICI score for each cell's transcriptome (Figures 2C and S2). We then plotted *VND7* gene expression relative to the p value associated with an individual xylem cell ICI score (Figure 2C). The number of cells with a significant protoxylem ICI score is very low for normalized *VND7* levels below a relative $\log_2$ value of 2, either very low or very high for *VND7* levels between 2 and 3.5, and very high for *VND7* levels above 3.5, consistent with the pattern and thresholds of switching in the whole root (Figure 2C). A similar switch was observed in the native context using 12,000 Col-0 single-cell sequences (Shulse et al., 2019) (Figure S3A) and in native xylem temporal data of the *Arabidopsis* root (Cartwright et al., 2009) (Figure S3B).

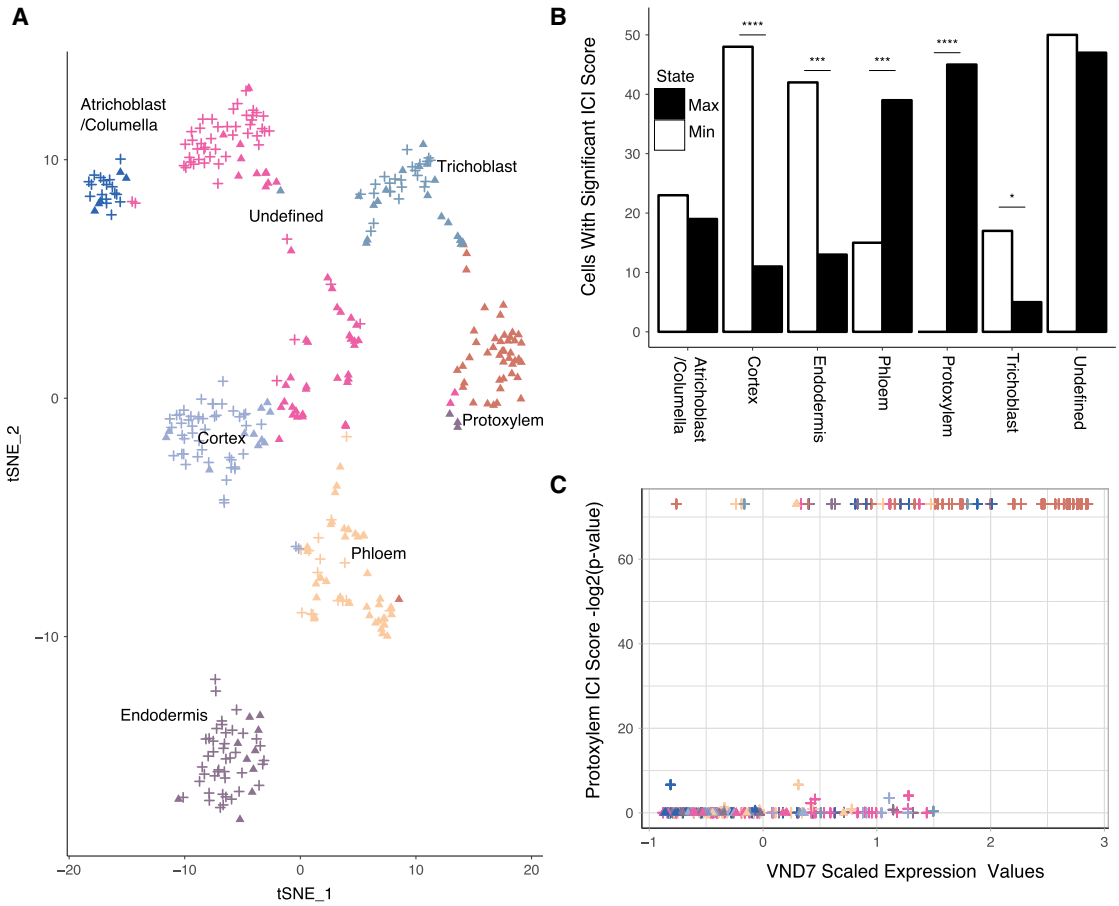


Figure 2. Evidence for a Transcriptional Switch in Single-Cell Sequencing Profiles

(A) t-distributed Stochastic Neighbor Embedding (t-SNE) clustering of 374 single cells from *Arabidopsis* roots either overexpressing *VND7* (triangle) or with wild-type levels of *VND7* expression (+). Clusters are labeled according to their cell identity score (ICI).

(B) Quantification of cell type with cell identity determined from (A) in the induced population (20 μM β-estradiol, denoted as VND7 overexpression [OE]), compared with the total number of cells with that identity in the uninduced cells (0 μM β-estradiol, denoted as control). ****adjusted $p \leq 0.0001$, ***$p \leq 0.001$, *$p \leq 0.05$, as determined by a Fisher's exact test with Benjamini-Hochberg correction.

(C) Significant protoxylem ICI scores determined for individual cells for induced (+) and uninduced (triangle) populations, plotted against *VND7* normalized expression data.

A Separation Score to Identify Candidate Genes Downstream of VND7 that Participate in the Developmental Switch

If VND7 stimulates a switch in xylem differentiation, then we expect to observe similar switching of some downstream targets as an emergent property of the network. In an effort to link the switching pattern in the whole root, as well as in the single-cell transcriptome data, to the underlying molecular mechanisms, we introduced a quantitative method (separation score, with values between 0 and 1) to identify candidate target genes with a similar pattern of switching in response to *VND7* (STAR Methods). A separation score near 1/3 is indicative of a uniform distribution, whereas the distinctive pattern of switching observed would produce larger scores. Application of this method to Figure 1D shows that at low expression of *VND7* (1–4 fold change [FC]), the separation score is 0; at moderate *VND7* expression (4–7 FC), the separation score is 0.82; and at high *VND7* expression (7–10 FC), the separation score is 0.4. The separation scores in the single-cell data show a similar pattern. The score for low (0–2) and high (4–6) *VND7* expression levels was 0, while for middle *VND7* levels (4–6), the score was 0.98.

Depending on the network topology, we may also observe a subset of VND7 targets that do not participate in the switch, because these possibilities are not mutually exclusive. To determine the downstream targets exhibiting a switch-like behavior, we took advantage of the single-cell variance in *VND7* expression. The distribution of *VND7* expression levels across cells is not bimodal (Figure S2D), and its separation score in induced cells with a xylem identity is 0.34, which is the expected for a uniformly random response. Thus, there is no evidence of a sharp increase in *VND7*. We used these data to examine the relationship between *VND7* and potential target genes among 203 correlated genes ($r > 0.5$) to identify candidate genes with the

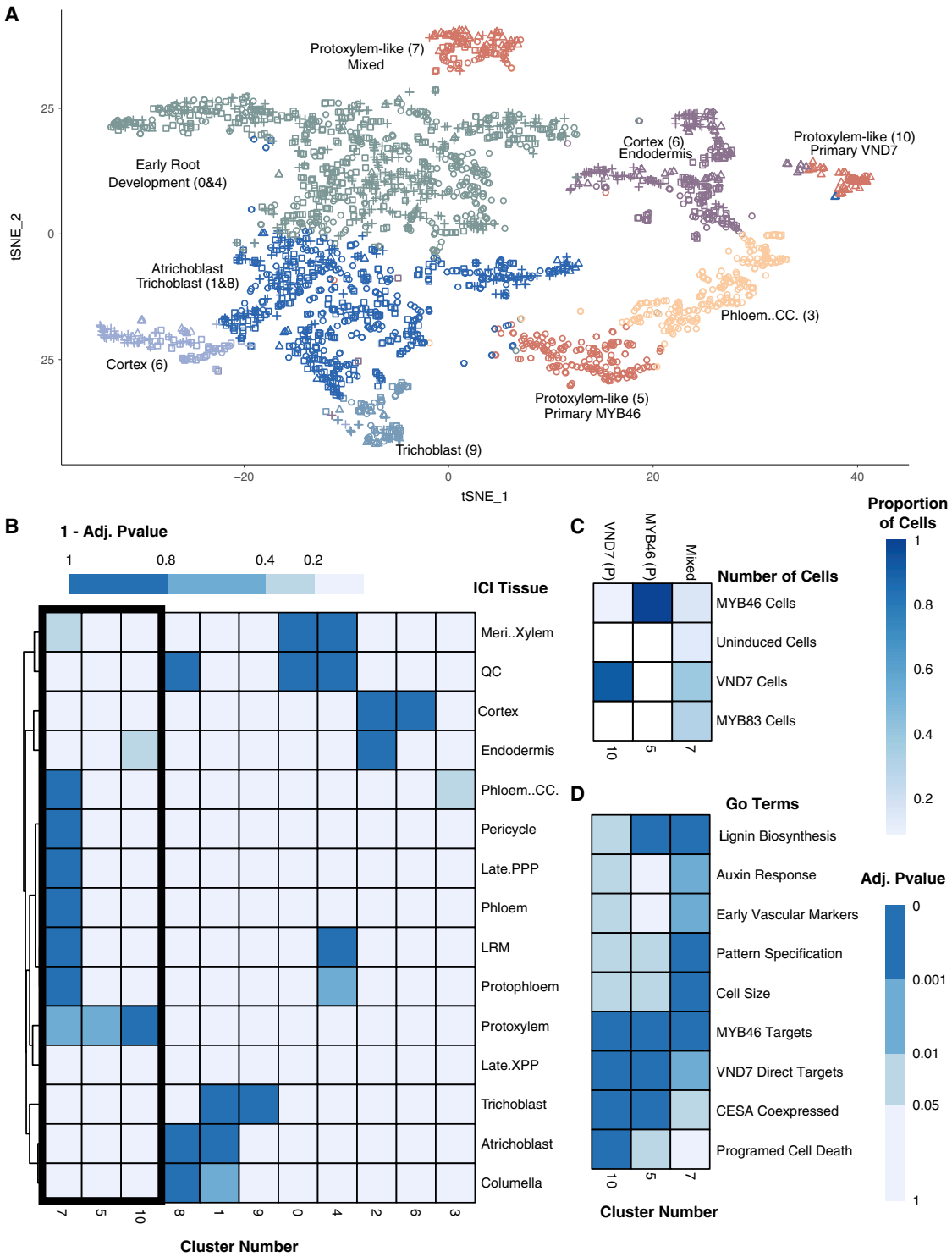


Figure 3. Single-Cell Transcriptomes Overexpressing *VND7*, *MYB46*, and *MYB83* Separate Distinct Vascular Developmental Processes

(A) t-SNE clustering of single cells from all *Arabidopsis* root overexpression experiments for *VND7* (triangle), MYB46 (circle), MYB83 (square), or uninduced cells (+). Clusters are labeled according to their cell identity score (ICI) and corresponding cluster number.

(B) Heatmap showing enrichment for ICI scores within each cluster, with the 3 protoxylem clusters highlighted.

(legend continued on next page)

distinctive pattern of switching observed at the population and single-cell identity.

Given these 3 expression bins (low, middle, and high), 116 of the 203 positively correlated genes had significantly different VND7 induction at middle and high levels of induction relative to that in uninduced cells. If a gene was responding to VND7, but not participating in the switch mechanism, target gene expression should increase in a graded manner concomitant with an increase in VND7 expression and should have a low separation score. If participating in the switch, the target gene should change its expression suddenly and dramatically between the low and the middle levels of VND7 expression, but there should be little difference in the expression of the target gene between the middle and the high VND7 expression values (STAR Methods). 161 genes changed their expression in a graded fashion, as determined by a statistically significant difference between the middle and the high expression bins (Table S3). In comparison, 15 genes showed a dramatic increase in expression between the low and the middle VND7 expression bins but no statistically significant change between the middle and the high bins, and they are candidate genes that participate directly in the switch (Figure S3C; Table S3). No defined gene ontology (GO) terms were significantly over-represented in the switch-responsive category. GO terms associated with SCW synthesis were identified for the non-switch categories (Table S3). VND7, MYB46, and VND7/MYB46 joint direct targets were not enriched in the switch gene set (Kim et al., 2013; Ko et al., 2009; Yamaguchi et al., 2010, 2011). Given that MYB46 expression is not switch-like, and the lack of enrichment of the direct targets in the switch gene set, the VND7-MYB46 FFL is not directly involved in generating the switch observed. Furthermore, given that MYB83 expression is not induced in response to VND7 induction (greater than r = 0.5), the VND7-MYB83 FFL is also likely not involved in xylem cell transdifferentiation.

Single-Cell Sequencing Identifies Cohorts of Cells that Are Targets of VND7-, MYB46-, and MYB83-Dependent Vascular Processes

MYB83 is sufficient to drive lignin biosynthesis associated with the xylem SCW and shares many targets with *MYB46* (McCarthy et al., 2009). These data suggest that MYB83 acts in a partially redundant manner to MYB46 and in a parallel FFL. However, unlike *MYB46*, *MYB83* expression is not induced upon *VND7* induction (Table S2). To further explore the role of these transcription factors in VND7-dependent xylem transdifferentiation, we carried out Drop-seq experiments on *MYB46* and *MYB83* β-estradiol-inducible plants. The transcriptomes of 776 *MYB46*-induced cells and 989 *MYB83*-induced cells were collected (Table S2), and the extent of transcriptome variation was determined by combining the *MYB46*-, *MYB83*-, and *VND7*-induced cell transcriptomes and their corresponding controls in Seurat (Figure 3A). Within the 11 resulting clusters (Figures S2A–S2C), 3 clusters with a high protoxylem ICI score were identified (Figures 3A, 3B, and S4). To assess whether these clusters could be used to differentiate VND7-MYB46 FFL and VND7-MYB83 FFL regulation, we quantified the number of cells originating from each overexpression experiment within the 3 protoxylem clusters (Figure 3C). Clusters were labeled accordingly: mixed (cluster 7, containing induced and uninduced cells from *VND7*, *MYB83*, and *MYB46* overexpression experiments), *VND7* primary (cluster 10), or *MYB46* primary (cluster 5) (Figure 3C). Single-cell sequencing has been used to classify novel cell subtypes (Macosko et al., 2015), and we propose that these 3 clusters represent subtypes of xylem cells. To confirm that these xylem subtypes occur in natural cell populations, we included 67 native xylem cell transcriptomes from *Arabidopsis* root cells (Shulse et al., 2019). Each of our 3 xylem cell subtypes show expression similarity to native xylem cell populations (Figures S4C and S4D; Table S2). To determine the molecular function of each cluster/cell subtype, a range of xylem-related gene sets was assessed for over-representation (Figure 3D; Table S4).

The *VND7*-primary xylem cell subtype contains an over-representation of genes associated with cells undergoing late stages of xylem development, including programmed cell death marker genes, protoxylem cell identity genes (Figures 3B and 3C) and genes defined by the GO terms "cell wall polysaccharides" and "xylem" (Table S4). The *MYB46*-primary xylem cell subtype showed similar patterns of enrichment for programmed cell death marker genes but also more extreme enrichment in genes associated with lignin biosynthesis than the VND7-primary cluster (Figure 3D). The *MYB46*-primary xylem cell subtype differs from the mixed and the *VND7*-primary xylem cell subtypes in the absence of enrichment of developmental genes, including auxin-responsive genes and genes associated with cell expansion (Figure 3D). In contrast, genes of the mixed xylem cell subtype showed significant enrichment in developmental categories such as auxin response (including *PIN3* and *PIN7*), cell expansion, and pattern specification (Figure 3D; Table S4). In addition, a significant increase in expression was observed for the early vascular development genes *TMO5* and *WOL* in the mixed cluster (Caño-Delgado et al., 2000, 2010; De Rybel et al., 2013; Mähönen et al., 2000). Altogether, these distinct molecular signatures reflect unique aspects of transcription factor-mediated regulation of vascular cell development and SCW biosynthesis.

Network Topology for Xylem Cell Differentiation

The morphological and molecular data presented here show that xylem cell differentiation occurs via a switch-like mechanism stimulated by VND7, but not MYB46 or MYB83. The single-cell transcriptome data provide an opportunity to identify the network topology needed for xylem cell differentiation using the separation score. Of these 15 candidate switch genes, 4 are VND7 direct targets (class A), 1 is a MYB46 direct target (class C), and 10 are neither VND7 nor MYB46 direct targets (class B) (Figure 4A; Table S3). Thus, expression of a set of

(C) 3 protoxylem clusters and proportion of cells within each cluster that correspond to each overexpression experiment. Cluster numbers are the same as those in (D). White boxes indicate no cells are present for the corresponding experiment. Clusters were also labeled based on their experimental distribution as mixed, VND7 P (*VND7* primary) or MYB46 P (*MYB46* primary).
(D) Enrichment of xylem-related genes within the 3 protoxylem clusters.

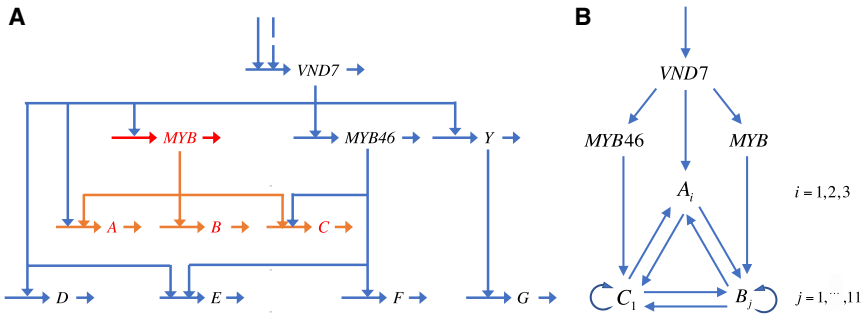


Figure 4. Classes of Downstream Target Genes Identified from Single-Cell Transcription Data

(A–C) Classes of switch genes: (A) 4 VND7 direct targets, (B) 10 VND7 indirect targets, and (C) 1 MYB46 direct target.

(D–G) Classes of non-switch genes: (D) 13 VND7 direct targets, (E) 5 VND7 and MYB46 direct targets, (F) 24 MYB46 direct targets, and (G) 37 VND7 indirect targets. Potential intermediate genes: X (targets of VND7, signals for downstream switch genes) and Y (targets of VND7, signals for downstream non-switch genes).

hypothetical genes X must respond to VND7 induction and provide the intermediary link to the 15 candidate switch genes.

Class C contains one gene with a MYB46 binding site, but no VND7 binding site, in its regulatory region (Figure 4B). Given that *MYB46* transcriptionally responds in a non-switch fashion, the response of the single gene in this class must be switch-like or be receiving input from another candidate switch gene in class A or B; alternatively, MYB46 may be post-translationally regulated (Figure 4B). Class B contains 10 genes, which generates too many possibilities to examine with the resolution of data available. Thus, we have identified 5 candidate genes for participation in the developmental switch to xylem cells that can be the focus for more detailed molecular studies.

DISCUSSION

In this study, 65 of 195 xylem cells were observed upon induction within the whole-root population (Figure 2B), which is less than the 100% observed in the elongation zone using confocal microscopy. The discrepancy is due to several factors, including differences in the root spatiotemporal zone characterized, as well as the developmental nature of xylem cells. Hair cells (trichoblasts), columella cells, and undefined cells were not imaged in Figures 1C and 1D. However, when these cell populations were removed from the single-cell sequencing experiment demonstrated in Figure 2B, vascular cell types comprise 77.8% of induced cells in comparison to 14.6% in the controls. This is well within the 70%–90% ($\log_2$ 6–6.5) range seen in region 3 of Figure 1D. In addition, because xylem cells undergo programmed cell death, we are unable to sequence the transcriptome of fully differentiated cells. Thus, we anticipate a decrease in the number of xylem cells identified in Figure 2B relative to Figure 1D.

Even when the underlying genes generate a switch in expression, there is no way for a population of cells acting independently to show the results seen for the whole root. In a population of hundreds of cells, some fully ON (nearly all differentiated) and others fully OFF (a very small number of xylem cells), an average somewhere between ON and OFF would occur at the level of the whole population. This suggests that within the whole-root induction experiment, there must be some mechanism that coordinates the results for all cells in the population to behave identically in the root elongation zone—fully uninduced or induced (Figure 1D). We propose that the first cells to produce levels of transcription sufficient for cell differentiation are pacemaker cells that can facilitate the response of neighboring cells. This would provide a mechanism to synchronize the population of cells and potentially give the experimentally observed switch response. Given that such behavior was not observed in the single-cell profiling (all ON or all OFF), we predict that disruption of the cell wall via protoplasting provides a hint that such a synchronization mechanism will likely involve cell-cell signaling.

Many TFs involved in xylem differentiation form a series of FFLs, including both MYB46 and MYB83 (McCarthy et al., 2009; Taylor-Teeples et al., 2015). However, only *MYB46* expression was induced in response to *VND7* induction, and common VND7 and MYB46 direct targets were not enriched within the set of genes exhibiting switch behavior. These data support our conclusion that the switch observed is not caused by the VND7-MYB46 FFL. This reflects the observation that MYB46 is sufficient to regulate ectopic SCW deposition (including cellulose, xylan, and lignin), but not xylem cell transdifferentiation (Zhong et al., 2007). Although MYB83 is also sufficient to only regulate ectopic SCW deposition like MYB46, it is only upon loss of both MYB83 and MYB46 function that a SCW phenotype is observed (McCarthy et al., 2009).

In addition to those genes identified to act in a switch, several genes in 4 classes (classes D–G) were identified as showing a non-switch response upon *VND7* induction (Figure 4A). Here, expression of a set of hypothetical genes Y must respond to VND7 induction and provide the intermediary link to 37 non-switch genes. Although our data demonstrate that VND7 stimulates genes participating in the xylem transdifferentiation switch, the enrichment of these non-switch genes suggests that VND7 may have independent functions not associated with xylem cell transdifferentiation.

The classification of cell identity from single-cell RNA-seq profiles in this study relied on methods previously employed to characterize regeneration of a root stem cell niche (Efroni et al., 2015, 2016). The clusters obtained by combining cells from *VND7*, *MYB83*, and *MYB46* overexpression lines revealed molecular differences that could not be completely differentiated given the general vascular ICI classifications and suggest potential xylem cell subtypes. Further characterization of the marker genes that define these clusters would provide insight into the subtle differences by which VND7 and MYB46 regulate xylem cell differentiation. Enrichment of programmed cell death marker genes and genes co-expressed with *CESA4* and mild enrichment of genes associated with lignin biosynthesis within the *VND7*-primary cluster suggest that the cells within this cluster represent

the end of the xylem cell transdifferentiation developmental trajectory, which includes programmed cell death, but not necessarily lignification (Figure 3D). The *MYB46*-primary xylem cell subtype and associated enrichments show that MYB46 has the capacity to regulate some aspects of xylem cell differentiation or SCW synthesis distinctly from VND7—specifically, genes associated with hormone signaling and pattern differentiation (Figure 3D). The mixed xylem cell subtype suggests that MYB46, MYB83, and VND7 coordinately regulate earlier stages of vascular development, because this cluster is enriched in multiple cell identities, in addition to the patterning genes and the auxin-responsive genes (Figure 3D). However, the enrichment of lignin biosynthesis genes is surprising and indicates that lignin biosynthesis may occur in vascular cells earlier than previously understood.

Molecular characterization of the total number of possible gene circuits by which switch behavior could be achieved was beyond the scope of this study. Thus, the regulatory logic that determines the switch behavior of the 15 candidate genes remains to be determined. The logic for the class C gene (MYB46 direct target) is fairly straightforward in that it responds to VND7 induction either by itself, generating switch behavior, or by receiving input from another switch gene. The logic of the 4 class A genes also provides a tractable focus for future work. It will be interesting to determine whether the mechanisms for commitment to xylem cell development are similar to or different from those of animal cells, for which commitment to a terminally differentiated cell fate is common and totipotency is rare.

Finally, while our results do not distinguish between a monostable graded switch and a bistable hysteretic switch, either of these mechanisms could be consistent with our data and the generation of a sufficiently sharp threshold as observed in the middle bin. It is possible to predict the biophysical conditions associated with both the monostable graded and the bistable hysteretic mechanisms. The key biophysical parameter that quantifies the sharp threshold is the Hill number n, which is the maximum number of transcription factor (TF) binding events involved with their target regulatory regions.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2019.06.041.

ACKNOWLEDGMENTS

G.M.T. was funded by an NSF predoctoral fellowship and AAUW Dissertation Completion Fellowship. G.M.T., J.R.-M., and S.M.B. were partially funded by a Howard Hughes Medical Institute (55108506) Faculty Scholar Fellowship. J.R.-M. was partially supported by a UC-MEXUS and CONACYT fellowship. M.A.V.-G. was supported by a grant from the U.S. National Science Foundation (MCB 1716833). M.A.S. was supported in part by a grant from the National Institutes of Health (RO1-GM30054). Single-cell RNA-seq work was supported by the Laboratory Directed Research and Development Program of Lawrence Berkeley National Laboratory and performed under U.S. Department of Energy contract DE-AC02-05CH11231.

AUTHOR CONTRIBUTIONS

G.M.T. and H.V. constructed and confirmed T3 homozygous estradiol-inducible plant lines. G.M.T., J.R.-M., M.A.V.-G., M.A.S., and S.M.B. designed experiments to test VND7 as a bistable switch. G.M.T. and D.H. contributed to Figure 1. M.A.S. wrote kinetic equations to model the system responses and analyzed global robustness of phenotypes. G.M.T., D.H., S.S., and C.N.S. conducted Drop-seq experiments. G.M.T., S.S., C.N.S., C.J., D.E.D., B.J.C., and S.M.B. designed Drop-seq experiments and analysis. All authors contributed to writing of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 26, 2018
Revised: May 1, 2019
Accepted: June 11, 2019
Published: July 9, 2019

REFERENCES

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

Bechtold, N., and Pelletier, G. (1998). In planta Agrobacterium-mediated transformation of adult *Arabidopsis thaliana* plants by vacuum infiltration. Methods Mol. Biol. *82*, 259–266.

Benfey, P.N., and Scheres, B. (2000). Root development. Curr. Biol. *10*, R813–R815.

Birnbaum, K.D., and Kussell, E. (2011). Measuring cell identity in noisy biological systems. Nucleic Acids Res. *39*, 9093–9107.

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

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

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.

Caño-Delgado, A.I., Metzlaff, K., and Bevan, M.W. (2000). The eli1 mutation reveals a link between cell expansion and secondary cell wall formation in *Arabidopsis thaliana*. Development *127*, 3395–3405.

Caño-Delgado, A., Lee, J.Y., and Demura, T. (2010). Regulatory mechanisms for specification and patterning of plant vascular tissues. Annu. Rev. Cell Dev. Biol. *26*, 605–637.

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

Coego, A., Brizuela, E., Castillejo, P., Ruíz, S., Koncz, C., del Pozo, J.C., Piñeiro, M., Jarillo, J.A., Paz-Ares, J., and León, J.; TRANSPLANTA Consortium (2014). The TRANSPLANTA collection of *Arabidopsis* lines: a resource for functional analysis of transcription factors based on their conditional overexpression. Plant J. *77*, 944–953.

Cruz-Ramírez, A., Díaz-Triviño, S., Blilou, I., Grieneisen, V.A., Sozzani, R., Zamioudis, C., Miskolczi, P., Nieuwland, J., Benjamins, R., Dhonukshe, P., et al. (2012). A bistable circuit involving SCARECROW-RETINOBLASTOMA integrates cues to inform asymmetric stem cell division. Cell *150*, 1002–1015.

De Rybel, B., Möller, B., Yoshida, S., Grabowicz, I., Barbier de Reuille, P., Boeren, S., Smith, R.S., Borst, J.W., and Weijers, D. (2013). A bHLH complex controls embryonic vascular tissue establishment and indeterminate growth in *Arabidopsis*. Dev. Cell *24*, 426–437.

De Rybel, B., Adibi, M., Breda, A.S., Wendrich, J.R., Smit, M.E., Novák, O., Yamaguchi, N., Yoshida, S., Van Isterdael, G., Palovaara, J., et al. (2014). Plant development. Integration of growth and patterning during vascular tissue formation in *Arabidopsis*. Science *345*, 1255215.

Déjardin, A., Laurans, F., Arnaud, D., Breton, C., Pilate, G., and Leplé, J.C. (2010). Wood formation in Angiosperms. C. R. Biol. *333*, 325–334.

Dinneny, J.R., Long, T.A., Wang, J.Y., Jung, J.W., Mace, D., Pointer, S., Barron, C., Brady, S.M., Schiefelbein, J., and Benfey, P.N. (2008). Cell identity mediates the response of *Arabidopsis* roots to abiotic stress. Science *320*, 942–945.

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.

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.

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.

Esau, K. (1977). Anatomy of seed plants, Second Edition (Wiley).

Frédéric, B. (2017). Arabidopsis—Root cell types. figshare. https://doi.org/10.6084/m9.figshare.4688752.v1.

Fukuda, H. (1992). Tracheary element formation as a model system of cell differentiation. Int. Rev. Cytol. *136*, 289–332.

Fukuda, H. (1996). Xylogenesis: Initiation, Progression, and Cell Death. Annu. Rev. Plant Physiol. Plant Mol. Biol. *47*, 299–325.

Fukuda, H., and Komamine, A. (1980). Establishment of an Experimental System for the Study of Tracheary Element Differentiation from Single Cells Isolated from the Mesophyll of *Zinnia elegans*. Plant Physiol. *65*, 57–60.

Fukushige, T., Hawkins, M.G., and McGhee, J.D. (1998). The GATA-factor elt-2 is essential for formation of the *Caenorhabditis elegans* intestine. Dev. Biol. *198*, 286–302.

Gifford, M.L., Dean, A., Gutierrez, R.A., Coruzzi, G.M., and Birnbaum, K.D. (2008). Cell-specific nitrogen responses mediate developmental plasticity. Proc. Natl. Acad. Sci. USA *105*, 803–808.

Kim, W.C., Ko, J.H., Kim, J.Y., Kim, J., Bae, H.J., and Han, K.H. (2013). MYB46 directly regulates the gene expression of secondary wall-associated cellulose synthases in *Arabidopsis*. Plant J. *73*, 26–36.

Ko, J.H., Kim, W.C., and Han, K.H. (2009). Ectopic expression of MYB46 identifies transcriptional regulatory genes involved in secondary wall biosynthesis in *Arabidopsis*. Plant J. *60*, 649–665.

Kondo, Y., Fujita, T., Sugiyama, M., and Fukuda, H. (2015). A novel system for xylem cell differentiation in *Arabidopsis thaliana*. Mol. Plant *8*, 612–621.

Kondo, Y., Nurani, A.M., Saito, C., Ichihashi, Y., Saito, M., Yamazaki, K., Mitsuda, N., Ohme-Takagi, M., and Fukuda, H. (2016). Vascular Cell Induction Culture System Using *Arabidopsis* Leaves (VISUAL) Reveals the Sequential Differentiation of Sieve Element-Like Cells. Plant Cell *28*, 1250–1262.

Kubo, M., Udagawa, M., Nishikubo, N., Horiguchi, G., Yamaguchi, M., Ito, J., Mimura, T., Fukuda, H., and Demura, T. (2005). Transcription switches for protoxylem and metaxylem vessel formation. Genes Dev. *19*, 1855–1860.

Kuhn, M. (2008). Building predictive models in R using the caret package. Journal of Statistical Software, Published online April 22, 2008. https://doi.org/10.18637/jss.v028.i05.

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

Lai, K., Robertson, M.J., and Schaffer, D.V. (2004). The sonic hedgehog signaling system as a bistable genetic switch. Biophys. J. *86*, 2748–2757.

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

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.

Maduro, M.F., and Rothman, J.H. (2002). Making worm guts: the gene regulatory network of the *Caenorhabditis elegans* endoderm. Dev. Biol. *246*, 68–85.

Mähönen, A.P., Bonke, M., Kauppinen, L., Riikonen, M., Benfey, P.N., and Helariutta, Y. (2000). A novel two-component hybrid molecule regulates vascular morphogenesis of the *Arabidopsis* root. Genes Dev. *14*, 2938–2943.

Mangan, S., and Alon, U. (2003). Structure and function of the feed-forward loop network motif. Proc. Natl. Acad. Sci. USA *100*, 11980–11985.

Mangan, S., Zaslaver, A., and Alon, U. (2003). The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks. J. Mol. Biol. *334*, 197–204.

Marshall, S.D., and McGhee, J.D. (2001). Coordination of ges-1 expression between the *Caenorhabditis* pharynx and intestine. Dev. Biol. *239*, 350–363.

McCarthy, R.L., Zhong, R., and Ye, Z.H. (2009). MYB83 is a direct target of SND1 and acts redundantly with MYB46 in the regulation of secondary cell wall biosynthesis in *Arabidopsis*. Plant Cell Physiol. *50*, 1950–1964.

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

Metzger, R.J., and Krasnow, M.A. (1999). Genetic control of branching morphogenesis. Science *284*, 1635–1639.

Mi, H., Muruganujan, A., and Thomas, P.D. (2013). PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic Acids Res. *D1*, D377–D386.

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

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.

Persson, S., Wei, H., Milne, J., Page, G.P., and Somerville, C.R. (2005). Identification of genes required for cellulose synthesis by regression analysis of public microarray data sets. Proc. Natl. Acad. Sci. USA *102*, 8633–8638.

Ptashne, M. (2004). A Genetic Switch: Phage Lambda Revisited (Cold Spring Harbor Laboratory Press).

Raes, J., Rohde, A., Christensen, J.H., Van de Peer, Y., and Boerjan, W. (2003). Genome-wide characterization of the lignification toolbox in *Arabidopsis*. Plant Physiol. *133*, 1051–1071.

Raven, J.A. (1993). The evolution of vascular plants in relation to quantitative functioning of dead water-conducting cells and stomata. Biol. Rev. *68*, 337–363.

Roberts, L.W. (1976). Cytodifferentiation in Plants: Xylogenesis as a Model System (Cambridge University Press).

Sachs, T. (1969). Polarity and the induction of organized vascular tissues. Ann. Bot. *33*, 263–275.

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.

Schmittgen, T.D., and Livak, K.J. (2008). Analyzing real-time PCR data by the comparative C(T) method. Nat. Protoc. *3*, 1101–1108.

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

Taylor-Teeples, M., Lin, L., de Lucas, M., Turco, G., Toal, T.W., Gaudinier, A., Young, N.F., Trabucco, G.M., Veling, M.T., Lamothe, R., et al. (2015). An *Arabidopsis* gene regulatory network for secondary cell wall synthesis. Nature *517*, 571–575.

Ursache, R., Andersen, T.G., Marhavy, P., and Geldner, N. (2018). A protocol for combining fluorescent proteins with histological stains for diverse cell wall components. Plant J. *93*, 399–412.

Yamaguchi, M., Kubo, M., Fukuda, H., and Demura, T. (2008). Vascular-related NAC-DOMAIN7 is involved in the differentiation of all types of xylem vessels in *Arabidopsis* roots and shoots. Plant J. *55*, 652–664.

Yamaguchi, M., Goué, N., Igarashi, H., Ohtani, M., Nakano, Y., Mortimer, J.C., Nishikubo, N., Kubo, M., Katayama, Y., Kakegawa, K., et al. (2010). VASCULAR-RELATED NAC-DOMAIN6 and VASCULAR-RELATED NAC-DOMAIN7 effectively induce transdifferentiation into xylem vessel elements under control of an induction system. Plant Physiol. *153*, 906–914.

Yamaguchi, M., Mitsuda, N., Ohtani, M., Ohme-Takagi, M., Kato, K., and Demura, T. (2011). VASCULAR-RELATED NAC-DOMAIN7 directly regulates the expression of a broad range of genes for xylem vessel formation. Plant J. *66*, 579–590.

Zelzer, E., and Shilo, B.Z. (2000). Cell fate choices in *Drosophila* tracheal morphogenesis. BioEssays *22*, 219–226.

Zhong, R., Richardson, E.A., and Ye, Z.H. (2007). The MYB46 transcription factor is a direct target of SND1 and regulates secondary wall biosynthesis in *Arabidopsis*. Plant Cell *19*, 2776–2792.

Zhu, J., Hill, R.J., Heid, P.J., Fukuyama, M., Sugimoto, A., Priess, J.R., and Rothman, J.H. (1997). end-1 encodes an apparent GATA factor that specifies the endoderm precursor in *Caenorhabditis elegans* embryos. Genes Dev. *11*, 2883–2896.

Zuo, J., Niu, Q.W., and Chua, N.H. (2000). Technical advance: An estrogen receptor-based transactivator XVE mediates highly inducible gene expression in transgenic plants. Plant J. *24*, 265–273.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
β-Estradiol	Sigma-Aldrich	E-8875
Critical Commercial Assays		
Nextera XT DNA Library Prep (24 samples)	Illumina	FC-121-1030
ChemGene Barcoded Beads	ChemGenes Corp.	N/A
Macosko sequence B – N13V Beads	LGC Biosearch	N/A
Drop-seq	Macosko et al., 2015	http://mccarrolllab.com/download/905/
Deposited Data		
Raw and analyzed data	This paper	GEO: GSE114615
Arabidopsis Drop-Seq raw and analyzed data	High-throughput single-cell transcriptome profiling of plant cell types	GEO: GSE122687
Experimental Models: Organisms/Strains		
Arabidopsis: pMDC7: AT1G71930 (VND7)	TRANSPLANTA	NASC code: 2101676 TPT_1.71930.1C)
Arabidopsis: pMDC7: AT5G12870 (MYB46)	This paper	N/A
Arabidopsis: pMDC7: AT3G08500 (MYB83)	This paper	N/A
Oligonucleotides		
VND7_F: CGAAGGAACCTCAGTGTTCATCA	This paper	N/A
VND7_R: GGGAAGCATCCAAGAGAA	This paper	N/A
MYB46_F: GAATGTGAAGAAGGTGATTGGTACA	This paper	N/A
MYB46_R: CGAAGGAACCTCAGTGTTCATCA	This paper	N/A
MYB83_F: AGCAACAACCTAAACTTACAGGCC	This paper	N/A
MYB83_R: CCCAACACATTCCCTACCTT	This paper	N/A
Recombinant DNA		
Plasmid: pMDC7	Zuo et al., 2000	N/A
Software and Algorithms		
Seurat	Butler et al., 2018	https://satijalab.org/seurat/install.html
Drop-seq_tools-1.13	Macosko et al., 2015	http://mccarrolllab.com/download/1276/
Cell Identity Score	Efroni et al., 2015	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4354993/bin/13059_2015_580_MOESM5_ESM.zip
Other		
Arabidopsis Drop-seq analyses and modified code for calculating cell identity	This paper	https://github.com/gturco/xylem_scRNA/

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to sbrady@ucdavis.edu.

Construction of Transgenic Lines

The *VND7* β–estradiol-inducible line was acquired from the TRANSPLANTA collection (NASC code: 2101676 TPT_1.71930.1C) (Coego et al., 2014). *MYB83* and *MYB46* TFs were cloned in the Gateway vector derived from the β–estradiol-inducible vector PER8 (Zuo et al., 2000). *Arabidopsis* plants (Col-0) were floral dipped in *Agrobacterium tumefaciens* strain GV3101 (Bechtold and Pelletier, 1998). Transgenic plants were selected for on hygromycin and T3 homozygous plants were used for experiments. Increased expression of transgenics was confirmed with estradiol induction followed by qPCR (Key Resources Table).

Growth Conditions

Seeds were sterilized in a 50% bleach 10% tween solution and then stored at 4°C for 3 days. Sterilized seeds were germinated on nylon mesh (100 uM) filters on square MS Petri dish plates and vertically grown at 22°C, 74% humidity in a 12 hr light cycle chamber at a 100-120 μmol m^{-2} s^{-1} light intensity. Seeds were plated in rows with approximately 100 plants per row and two rows per plate. Plates where randomized using a random block design where each tray contained an equal number of control and treatment plates. After 7 days of growth, plants were transferred by mesh filters to either MS plates buffered to pH 5.8 with KOH or MS containing estradiol (0.01, 0.05, 0.075, 0.1, 0.13, 0.15, 0.175, 0.2, 1, 2, 10, 20uM) plates and grown for an additional 24 hours (induction). All induction experiments were carried out at 7 am and collected at 7 am. Whole root samples were imaged, frozen for RNA extraction and qRT-PCR or protoplasted for Drop-seq. In all induction experiments, we made the assumption that each cell within the root (the population of cells) encounters approximately the same mean concentration of β-estradiol and thus, *VND7* expression.

METHOD DETAILS

Histology and Confocal Microscopy

Seven-day-old *VND7* overexpression roots were stained following a modified version of the ClearSee protocol (Kurihara et al., 2015; Ursache et al., 2018). Following 1 hour fixation with 4% paraformaldehyde in 1 x PBS, roots were transferred to the ClearSee solution overnight. Basic Fuchsin dye was used to mark lignified cells for identification of fully differentiated xylem. Roots were stained overnight in Basic Fuchsin followed by one ClearSee wash and an additional hour in the ClearSee solution. Roots were stained for an additional for 45 mins with Calcofluor White to stain the cellulose of the primary cell wall as well as the emergence of cellulose in SCWs, allowing us to count the total number of cells in each image. Roots were visualized on the Zeiss LSM780 microscope: Basic Fuchsin: 550-561 nm excitation and 570-650 nm detection and Calcofluor White: 405 nm excitation and 425-475 nm detection. Z stack images were collected from the start of the elongation zone.

Confocal Laser Scanning Microscopy Analysis

The total number of cells was determined by the presence of an outline of a Calcofluor White stained cell. The Basic Fuchsin dye was used to mark lignified cells for identification of fully differentiated xylem. The percentage of cells experiencing transdifferentiation into protoxylem cells was determined by drawing a box around each cell in ImageJ and color coding based on presence or absence of a helical pattern (marked by Calcofluor or Fuchsin) as observed in Figure 1C. For each root, the area for each xylem marked cell was summed and divided by the total area of all cells. The percentage of cells experiencing transdifferentiation into protoxylem was determined for each estradiol concentration with 10-20 roots per concentration. Xylem differentiation was quantified for z stacks throughout the entire root (Table S1).

RNA extraction and qRT-PCR

Whole roots were separated from shoots and collected in bulk from each plate for RNA extraction. RNA for qRT-PCR was extracted using Trizol reagent (Life Technologies, USA) according to the manufacturer's instructions. RNA was DNase treated prior to cDNA synthesis using M-MLV reverse transcriptase (Invitrogen). cDNA was used as a template for real-time quantitative PCR analysis. Each plate was considered an independent biological replicate with at least three biological replicates (same estradiol concentration & grown same day) per dilution series. Log_2 Fold Change (FC) of VND7 varied 0.041 - 0.57 standard deviations (SD) among biological replicates. The largest variance was observed under lower induction of VND7 (2-4 FC) with the SD ranging from 0.08-0.57 SD, in mid-expression lines (5-7 FC) 0.041-0.064 SD. The $2^{\Delta\Delta CT}$ method was used to measure relative transcript abundance (Schmittgen and Livak, 2008). Three technical replicates were carried out and averaged for each plate. VND7 expression was normalized to Actin to determine the delta CT and the delta delta CT was determined from uninduced roots grown on the same rack. Primers used for RT-PCR are listed in the Key Resources Table.

Protoplasting and Drop-seq Assays

For Drop-seq runs, plants were grown as described above; the entire root was then separated and harvested for protoplasting (Benfey and Scheres, 2000; Birnbaum et al., 2003). A single Drop-seq sample consisted of ten plates with ~200 roots per plate protoplasted in bulk as previously described. Individual cells were resuspended in Solution A containing: 600 mM Mannitol, 2 mM MgCl2*6H20, 0.10% BSA, 2 mM CaCl2*2H20, 2 mM MES Hydrate, 10 mM KCl and pH to 5.5 with 1 M Tris. Each sample of cells was filtered through a 40 μM Cell strainer (BD Falcon) into a 50 mL Falcon tube. Cells were then counted and diluted to 100-180 cells per μl immediately before Drop-seq (Table S2). All Drop-seq steps followed the standard protocol outlined by the McCarroll Lab (http://mccarrolllab.org/wp-content/uploads/2016/03/Drop-seqAlignmentCookbookv1.2Jan2016.pdf). The droplet diameter ranged from 0.78 - 1 nl, barcoded beads (Barcoded Beads SeqB; ChemGenes Corp., Wilmington, MA, USA or Macosko sequence B – N13V, LGC Biosearch, Petaluma CA) and cells were loaded at concentrations specified in Table S2, with a collection target of 1 mL according to standard Drop-seq protocol. Compared to the ChemGene beads, the LGC Biosearch beads had a 13 bp UMI as the only modification to the linker.

Drop-seq was run with a collection target of 1 mL according to standard Drop-seq protocol. cDNA was amplified using 13-17 cycles and double-purified in 0.6x volume of Agencourt AMPure XP beads. Tagmentation was carried out with 1200 pg of DNA as input. Samples were run on the Bioanalyzer prior to NextSeq sequencing.

QUANTIFICATION AND STATISTICAL ANALYSIS

Alignment and Normalization of Drop-seq Data

All Drop-seq data was pre-processed and aligned to the TAIR10 *Arabidopsis* genome with STAR (Dobin et al., 2013) via the Drop-seq Tools v1.12 software (http://mccarrolllab.org/wp-content/uploads/2016/03/Drop-seqAlignmentCookbookv1.2Jan2016.pdf). Gene expression matrices were further processed and normalized using the Seurat R package, with a cutoff of at least 500 genes/cell and for genes seen in at least 3 cells (Satija et al., 2015). For each experiment the number of cells before and after processing are reported in Table S2. When scaling the data in Seurat for downstream analysis, cells with a high percentage of mitochondrial, chloroplast and protoplasting-induced genes were regressed out prior to clustering of cells (Birnbaum et al., 2003). Experiments in Figure 3A were merged using the MergeSeurat function, providing us more power to cluster cells but also the potential risk of clustering by experiment. Overall, we observed no large batch effects (Figure 3A).

Determining Cell Identity

Clustering Analysis

Reducing data dimensionality and clustering cell types was performed as described by Andrews and Hemberg (2018). Code is provided in github. To identify cell types via clustering we first reduced the data to the top 1,000 highly variable genes across cell types. We then further reduced the data dimensionality through principal component analysis (PCA). PCs 1-12 were used for further clustering as determined by the elbow of the PC standard deviations plot. Clusters were then determined from these PCs using the shared nearest neighbor modularity optimization based algorithm, Louvain clustering in FindClusters. To determine each cluster's cell identity, the average expression value of each cluster was determined using Seurat's AverageExpression function, requiring at least 25% of cells within the cluster to express each gene. These average expression profiles were then compared to ~100 predefined markers for cell identity in the *Arabidopsis* root and an ICI score calculated (Birnbaum and Kussell, 2011; Efroni et al., 2015). Tissues with significant p values (< 0.01) for cell identity were used to define the corresponding clusters and cells within clusters. VND7, MYB83 and MYB46 were not among the 100 predefined markers for protoxylem identity in the ICI algorithm and therefore did not bias assignment of cluster identity. To preserve the original data structure tSNE was used for plotting with cells color coded by cluster identity.

The number of cells in each DropSeq experiment before and after induction that could be assigned an individual cell identity given a significant ICI score are quantified in Table S2.

The accuracy of cluster assignment was determined with the R Classification And REgression Training package; caret (Kuhn, 2008). First, the single cell data was resampled using cross validation, with 10 resampling iterations resulting in 10 training and 10 testing splits. The data was then trained using the support vector machine with linear kernel algorithm (svmLinear), as used by Seurat, to predict cluster identity given the expression of a subset of genes. Specifically, we focused on those with the strongest contribution to the first twelve principal components. After ten iterations of resampling the model prediction, accuracy and kappa were reported. When applied to the three protoxylem clusters in Figures S4A and S4B we observed 90% accuracy with clustering of only protoxylem cells and 99% accuracy in calling the protoxylem cluster when using all the data in Figure 3A.

Cluster accuracy was also determined by applying two clustering algorithms - Smart Local Moving (SLM) and Louvain clustering. When applied to the three protoxylem clusters of Figures S4A and S4B the two clustering algorithms agreed 100% of the time indicating that these three clusters indeed have distinct transcriptome profiles which is unbiased by the clustering algorithm used. The cells within these three clusters show 92% similarity to those seen in Figure 3A. Cluster 2 shows 100% similarity to cluster 10 of Figure 3A (the VND7 primary xylem cell sub-type), and cluster 1 showing the largest deviation from clusters of Figure 3 with 90% similarity to cluster 7 (Mixed). Note that the cluster numbering here ONLY refers to that of Figures S4A and S4B.

Clustering of Protoxylem Cluster Sub-Types with Native Cells

Cells were identified as protoxylem cells based on the ICI scores of clustered cells from Figure 3A (Birnbaum and Kussell, 2011; Efroni et al., 2015). To determine if these protoxylem subtypes from Figure 3A (clusters 5, 7 and 10) occur in native protoxylem populations we used 67 native protoxylem cells (cells with a protoxylem ICI score < 0.01) independently identified in the *Arabidopsis* root (Shulse et al., 2019). The induced and native protoxylem cell type populations were merged into a single dataset and scaled as described above with cells with a high percentage of mitochondrial, chloroplast and protoplasting-induced genes were regressed out. Variable genes were re-identified, and PCs 1-12 were used for Louvain clustering. All clustering was performed with forced recalculation of clusters, resulting clusters are reported in Figures S4C and S4D and Table S2.

Enrichment of Vascular Categories

Marker genes for each cluster were identified using Seurat with a log fold change threshold of plus two. Marker genes were compared globally using GO enrichment analysis in python (https://github.com/tanghaibao/goatools) or with predefined gene lists. In both

cases a Fisher's exact test was carried out comparing the study (marker genes of a cluster) to the population (all genes expressed in the single cell experiment), and p values were adjusted with Benjamini-Hochberg. Gene lists were obtained from previously published work: lignin biosynthesis genes (Raes et al., 2003), CESA co-expressed genes (Persson et al., 2005), PCD (Olvera-Carrillo et al., 2015) and early vascular markers (Caño-Delgado et al., 2010). Genes related to auxin, cell size and patterning were determined by GO annotation.

Identification of switch-like behavior – the Separation Score

The separation score is defined as the 2/3 y axis tercile value - 1/3 y axis tercile value divided by the difference between the maximum and minimum y axis values. Our null hypothesis is that the mean separation score should be 1/3 if all data is evenly distributed. To determine VND7 downstream targets exhibiting switch versus non-switch behavior we took advantage of the variance in *VND7* expression observed within single cell types. *VND7* expression was ''binned'' into three groups: low (0-2), middle (2-4) and high (4-6) VND7 normalized expression values. For the whole root data in Figure 1D, or the single cell data in Figure 2C, we measured the spread of xylem cell differentiation or of ICI score determination, respectively, by grouping these y-axes data into thirds. We next calculated a ''separation score'' for each of these measures.

In order to identify genes that respond to VND7 expression as a switch or non-switch, we identified genes from the single cell data which had a positive Pearson correlation with *VND7* ($r > 0.5$). Next, we divided cells into ''xylem'' cell populations (as determined using significant ICI scores) and non-xylem cells. Using custom scripts in R, we interrogated the correlated genes with VND7 expression by two means: a) compared expression of these genes in protoxylem induced cells against all uninduced cells on a bin-by-bin basis using the base R function *t.test()*. Next, we selected genes that were significantly different ($p \leq 0.05$) in the medium and high bins but not in the low bin ($p > 0.05$). For further classification, the absolute difference between the mean expression of the middle and high bins was taken and the first (25%) quartile was selected (n = 29 genes) (i.e., genes with minimal difference between the middle and high bins in the protoxylem induced cells, but significantly different between induced and uninduced). b) comparison between the middle and high bins using the 203 set of VND7 correlated genes, only in protoxylem induced cells; genes with no significant difference (t test, $p > 0.05$) were selected (n = 43 genes); The overlap between the sets in a) and b) was taken for further analysis as the set of potential switch genes (n = 15 genes). Conversely, the overlap between genes in the rest of the quartiles from **a)** and genes with significant difference ($p < 0.05$) in **b)** were selected as potential non-switch genes. Enrichment of the clusters 5, 7 and 10 as well as the targets of VND7, MYB46 or VND7/MYB46 was done using the *fisher.test()* function in R with the non-switch and switch sets as defined above. GO enrichment was done using the gene list analysis tool in TAIR powered by PANTHER (https://www.arabidopsis.org/, Mi et al., 2013). AME (McLeay and Bailey, 2010) was used for motif enrichment of the switch gene set.

DATA AND CODE AVAILABILITY

The accession number for the raw and processed Drop-seq data of VND7, MYB83 and MYB46 overexpression lines reported in this paper is GEO: GSE114615. Code for determining cell identity from Drop-seq data and determining cell clusters were made available on github (https://github.com/gturco/xylem_scRNA).