Title
Transcriptional Regulation of Xylem Development and Secondary Cell Wall Composition

Author
Gouran, Mona

Publication Date
2023

Peer reviewed|Thesis/dissertation

Transcriptional Regulation of Xylem Development and Secondary Cell Wall Composition

By

MONA GOURAN
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Plant Biology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Dr. Siobhan Brady

Dr. Andrew Groover

Dr. John Harada

Committee in Charge

2023

ABSTRACT

The plant vascular system is comprised of specialized tissues known as xylem and phloem. The interconnected network of xylem vessel elements, with their elongated structures and reinforced secondary cell walls, form a hydraulic system that enables efficient and uninterrupted water transport throughout the organism, contributing to plant growth and fitness.

In the first chapter, genetic and genomic methodologies are employed to illustrate the conservation and repurposing of transcriptional regulation within the xylem network in *Solanum lycopersicum*. Key transcriptional regulators of xylem cell differentiation are identified in the tomato root, uncovering a partial conservation of the xylem developmental master regulators (*VND6* and *VND7*) between *Arabidopsis* and tomato. Furthermore, through functional validation of putative orthologs of known xylem patterning and differentiation genes, examples of conservation (*HD-ZIPIII* TFs) as well as a novel regulator (*SlKNAT1*) are revealed in the tomato xylem regulatory network.

The focus of the second chapter is directed towards the function of *VND6* and *VND7* in the *Arabidopsis* inflorescence stem. An integrated approach combining glycome profiling, an *in vitro* immunoanalytical platform, and *in situ* immunolocalization was employed to identify, for the first time, the differential abundance of specific cell wall biopolymers at cellular resolution in *vnd6* and *vnd7* mutants. Further gene expression profiling reveals perturbed expression of multiple cell-wall associated genes in the mutant backgrounds that could partially contribute to the observed cell wall phenotypes in the mutants.

The third chapter delves deeper into functional conservation within the *VND* transcription factor family in the tomato root. It is demonstrated that various tomato *VND* TFs have

expression in the vascular tissue in the root and are sufficient to induce ectopic xylem differentiation. The results here emphasize both the evolutionary conservation and distinct variations within the *VND* TF family concerning xylem development between two evolutionarily distant plant species.

Lastly, utilizing an inducible *VND7* system, an investigation is conducted on the hysteretic or memory feature associated with a bistable switch system linked to xylem cell fate determination. The findings obtained from this study indicate the necessity of an alternative inducible system, one that possesses both the "on" and "off" features. Such a system is crucial for the effective exploration of *VND7*-dependent hysteresis in the xylem cell differentiation.

The research presented here has broad implications for understanding the molecular mechanisms underlying the plant vascular system, which is crucial for efficient water transport and plant growth. This understanding can have applications in areas such as agriculture and forestry, where efficient water use and plant growth are essential. Additionally, the insights gained from this research can inform the development of new strategies for improving plant growth and crop yields in the face of climate change and other environmental challenges. Overall, this research has significant potential for advancing our knowledge of plant biology and contributing to the development of sustainable agriculture practices.

This study expands our knowledge of the molecular mechanisms governing xylem cell differentiation. It highlights both similarities and differences in the gene regulatory network of xylem across different plant species. Additionally, this research sheds light on the

functions of *VND6* and *VND7* in regulating secondary cell wall composition, challenging the earlier notions regarding functional redundancy within the *VND* gene family.

ACKNOWLEDGMENTS

First and foremost, I like to express my sincere gratitude to you, Dr. Siobhan Brady. Your enthusiasm for science was inspiring to me when I took your class years ago and that inspiration continues to this day. From the moment I approached you for advice on pursuing my academic goals, your guidance, advice, support, and kindness have been invaluable to me. Joining your group was a pivotal point in my career, and I am forever thankful for the opportunity. Your encouragement and belief in my abilities kept me motivated, even during the most challenging times along this path.

I would also like to extend my appreciation to Dr. Andrew Groover and Dr. John Harada for their valuable contributions as members of my dissertation committee. I also like to thank the department of Plant Biology at UC Davis for their support.

I am incredibly grateful to the members of the Brady lab, both past and present, who have been instrumental in making this journey both successful and enjoyable. This experience would not have been the same without the amazing group of people (Conchi, Anne-Maarit, Alex CP, Lidor, Dorota, Alex M, Chao, Tamera, Kevin, Jiregna and many more..) that I met and had the privilege of calling lab mates. The collaborative and supportive environment we shared has been a source of inspiration and motivation to me.

Above all, thank you to my family and especially to my parents, who made tremendous sacrifices by leaving their beloved homeland behind and embracing the challenges of a new life to provide us with a better future. Their selflessness and determination have paved the way for my success, and I am eternally thankful for their unconditional love and the sacrifices they have made on my behalf. Also, a special thanks to my brother, Hossein,

who served as an inspiring example by showing me that pursuing a PhD can be fulfilling, enjoyable and rewarding.

TABLE OF CONTENTS

Page

INTRODUCTION

Plant Cell Wall

Land plants have acquired specific structures allowing them to cope with their sessile nature. One such structure is the plant cell wall, a tough extracellular matrix that is crucial for plant survival. Plant cell walls are essential for plant development and perform a variety of functions throughout the plant life cycle including mechanical strength, determining organ and cell type morphology, regulating turgor pressure and water relations, adhesion, intercellular communication, plant-microbe interactions, and serving as the first line of defense when plants encounter pathogens[1]. Moreover, cell walls are widely exploited for human use relating to food, fiber, textile, lumber, and biofuel industries. This wide reach of interest in such a small and elegant nanostructure places plant cell walls amongst the center of importance in plant biology research[2].

Plant Primary and Secondary Cell Walls

The plant cell wall is categorized into two groups: primary and secondary cell walls (SCW). The primary wall is a thin, dynamic, and extensible structure surrounding dividing cells. Primary walls allow for cell expansion through controlled plasticity without losing mechanical integrity. Primary wall size and shape dictates the majority of the externally visible characteristics of plant cells[3]. Primary walls are mainly composed of carbohydrate polymers including cellulose, hemicellulose, and pectin. Once growth stops the primary wall structure and composition is often retained, but in some defined cell types such as xylem, fiber, and anther cells, a thickened and rigid structure known as SCW is deposited inside the primary wall. The major components of SCW are cellulose, hemicellulose

(xylan and glucomannan), and lignin. In addition to providing physical strength for the plant, SCW facilitates critical biological processes such as water transport in xylem vessel elements, anther dehiscence, silique shattering, and response to pathogens[4]. Even with these general differences in cell wall components, there is additional variation in their respective molecular structures and compositions[5]. The proportion of the polymers contributing to cell walls can vary depending on plant species, tissue or even within a single cell type. Cell wall composition can be modified during development but also upon environmental conditions[6]. Impairment in production of any of the polymers or modifications in the structure or proportion of the wall epitopes could be consequential for plants. For instance, inhibition of cellulose production in *Arabidopsi*s mutant lines leads to constitutive expression of immune response genes[7]. In a recent in-depth study of a set of *Arabidopsis* cell wall mutants it was confirmed that the abundance of specific pectin epitopes in the cell wall (e.g., fucosylated-xyloglucan) is correlated with opposite patterns of pathogen resistance in comparison to those in wild-type[8]. Given the evidence for importance of cell walls, it is not hard to imagine that throughout a plant's life cycle, the needs of individual cells are in part achieved by fine-tuning of the molecular factors that regulate cell wall polysaccharide synthesis, modification, and deposition.

Plant Glycan-Directed Antibodies as Tools for Analysis of Cell Wall Composition

An estimated 10% of the plant genome is dedicated to plant cell wall metabolism and biosynthesis[9]. Although the plant cell wall community has made great advances in defining the genetic and biochemical pathways that give rise to these structures, one of the biggest gaps in our knowledge of cell wall biology are the plethora of molecular players that regulate and catalyze the synthesis and remodeling of various cell wall constituents[9].

As the main constituent of plant renewable biomass, SCWs, have been central in our efforts to understand the molecular basis of its formation. Steps toward identifying transcription factors involved in SCW development has revealed a complex network of genes controlling this process[10–12]. In parallel, progress has been made in establishing methods to explore variation in cell wall composition variations. Current protocols in cell wall composition can only use whole organs which do not allow the investigation of the heterogeneity of polysaccharide composition between cell types and tissues using the existing high-throughput methodologies. SCWs are only deposited in a smaller sub-set of cells within an organ making their isolation more laborious and challenging for high resolution cell wall compositions analysis. There are no studies that specifically explore SCW composition in a high-resolution or comprehensive manner. Current approaches usually involve cell wall decomposition with chemical removal of cell wall polymers followed by methods for detection of monomers such as mass spectroscopy, and nuclear magnetic resonance (NMR) spectrometry [13,14]. Such methods usually have major limitations as they require expensive machinery and special training to operate.

More recently, immunological approaches with plant cell wall-directed antibodies have emerged as a powerful and systematic method to study plant cell wall carbohydrates. The development of a comprehensive collection of plant cell wall glycan-directed monoclonal antibodies (mAbs) has enabled direct high-throughput analyses of the plant cell wall ultrastructure in an approach known as glycome profiling. In this ELISA-based approach (Enzyme-Linked Immunoabsorbent Assay), the cell wall carbohydrates are isolated by alcohol insoluble residue (AIR) extraction and is further fractionated by sequential extraction steps which employ increasingly harsh chemical treatments. These extractive

chemical reagents disrupt the interactions among cell wall components producing fractions that are rich in pectin, hemicellulose, and lignin. This is followed by ELISA screening of the extracts using mAbs to produce a measure of abundance of each epitope relative to the produced fluorescence units. This platform has broad applications and has contributed to our understanding of cell wall heterogeneity, its molecular players as well as its plasticity to water stress, and pathogens[15,16]. However, there are limitations to this approach; cellulosic polymers cannot be detected due to the large size of these molecules and their inability to bind properly to the solid surface of plates used for the high throughput ELISA assay used in this platform. In addition, glycans might get altered and therefore not detected while being extracted by chemicals – such as loss of acetylation or methylation – therefore absence of signal cannot be automatically interpreted as the absence of the constituent itself. Further, this method is reliant upon chemical extraction but not all tissue types have the same extractability and complete extraction is not often possible. Also, glycome profiling is dependent upon the affinity of mAbs available, so absolute quantification of glycans is not possible[17]. Despite these limitations, this methodology has proven to be a valuable tool in functional characterization of genes involved in cell wall biosynthesis[18].

Because of the above-mentioned limitations in glycome profiling, it is important to complement and verify the data driven by this platform with alternative *in situ* experiments such as immunolabelling also known as immunohistochemistry. This technique involves fixing, embedding, sectioning, and finally labeling the thin sections from the tissue of interest. Vast methodological differences exist in each step depending on the scientific question investigated[19]. This powerful technique takes advantage of the highly specific

cell wall-directed mAbs to access the cellular location, relative abundance, and distribution of wall-associated glycans *in vivo*. This approach provides another angle to investigate cell wall composition because, unlike glycome profiling, it does not involve the use of harsh chemicals (relative to chemicals used in cell wall fractionation for glycome profiling) so minimal if any modifications of glycans is expected. On the other hand, some epitopes in the wall structure can be masked by other components and inaccessible to bind to mAbs while in glycome profiling, the cell wall is fractionated so glycans are freely available for detection. Also, the strength of detection is dependent upon the plane of sectioning (for specifically buried epitopes in certain planes) and the quality of sections obtained as not all tissues are amenable to producing high-quality sections[20]. Despite these limitations, a great amount of information can still be obtained about cell wall composition by this approach, however, it is often overlooked in research projects due to its laborious nature and low throughput. Thus, it is often useful to integrate data from glycome profiling to narrow down the selection of antibodies to be tested with immunohistochemistry.

Cell and Tissue Type-specific Cell Wall Composition in *Arabidopsis* Inflorescence Stem

The elongation of the developing inflorescence stem in *Arabidopsis* is accompanied by rapid axial growth and at maturity is composed of cell-types with visibly different cell wall structures. However, cell wall deposition and analysis of the molecular players at a cell type-resolution has been hampered due to challenges in sampling distinct cell types along the developing stem tissue. To gain insight into the spatiotemporal compositional divergence of cell walls in *Arabidopsis* stem, *H.C. Hall et al, 2013*, have performed one

of the more comprehensive studies where seven cell types were assayed for abundance and distribution of 18 major cell-wall glycan classes (55 mAbs) at three developmental stages along the developing inflorescence stem. In this work, they developed a pipeline for a semi-high throughput comparative immunolabelling strategy[21]. This study highlights major shifts in epitope abundance and accessibility across developmentally distinct tissues and cell types in the stem. As an example, they identified a cluster of antibodies that are restricted to protoxylem in the apical regions but over the course of stem maturation, the signal expands to the interfascicular fiber region and finally is detected in most cell types. This data also demonstrates clusters of antibodies with a pectic backbone that have similar binding patterns in different tissues and developmental zones. There are also examples of novel cell type-specific epitopes, as shown for xylem and bundle sheath in specific developmental zones. These findings along with many other available data sets do reflect the existence of a highly dynamic and tissue-specific cell wall structure in *Arabidopsis* stem which is orchestrated by an underlying molecular mechanism yet to be delineated at tissue resolution.

The Development of the Vascular Tissue in the Arabidopsis Inflorescence Stem

Arabidopsis has served as a versatile model organism to study plant cell walls. The *Arabidopsis* inflorescence stem is a load-bearing organ and produces SCWs throughout its development making it a well-suited accessible organ to study cell wall formation in a tissue or cell type-specific manner[22]. In comparison with animals, plants possess the ability for post-embryonic growth and development throughout life. Above ground parts of the plant are developed from the shoot apical meristem (SAM). The SAM is a collection

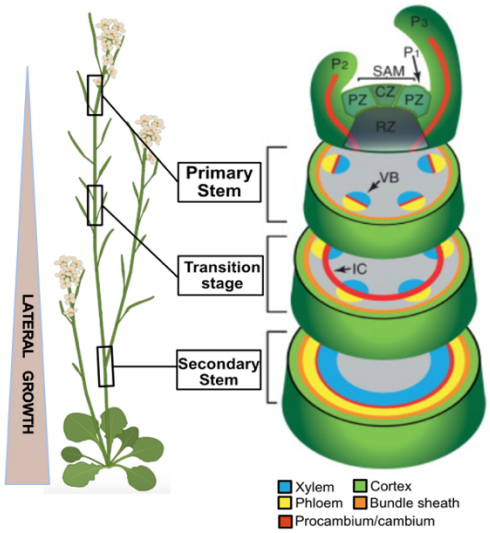


Figure.1 Schematic representation of inflorescence stem showing structure of primary vascular bundles. Base of inflorescence stem indicating growth of the fascicular and interfascicular cambia and the initiation of secondary growth. Abbreviations: CZ, central zone; IC, interfascicular cambium; Px, leaf primordial; PZ, peripheral zone; RZ, rib zone; SAM, shoot apical meristem; VB, vascular bundle. Part of image adapted from *Pablo-Sanchez et. al. Cell. 2011*.

of cells that retain the ability to continuously renew itself by cell division, and to generate new above-ground tissues and organs. When vegetative growth ceases, the primary stem is formed from the Rib Zone in SAM. The inflorescence stem is composed of various tissues including the epidermis, cortex, starch sheath, vascular bundles, and pith[23] (Figure.1). A prominent aspect of stem development is the formation of vascular tissue composed of xylem and phloem. The apical meristem in both the root and shoot contains procambium, the primary vascular meristem. *Arabidopsis* inflorescence stems develop a vascular pattern similar to most dicots; upon vascular initiation, procambial cell division and differentiation gives rise to the xylem, the water conducting tissues, and the phloem, through which photosynthetic compounds and signaling molecules are transported. Discrete collateral vascular bundles form with primary xylem facing inside, and the primary phloem facing the epidermis are formed along the periphery of the stem. Interfascicular fiber (IF) composed of 3-4 layers of fiber cells are positioned between vascular bundles[23](Figure.1). Fiber cells have a thick SCW that are mostly responsible for the mechanical strength of the mature stem. The vascular procambium is located between the phloem and xylem cells and restricted to vascular bundles in primary growth[24]. In *Arabidopsis* and other species that undergo secondary growth for radial expansion, a transition stage is described that leads to formation of a ring-like domain of

cambium through *de novo* recruitment of interfascicular parenchyma cells to dedifferentiate into interfascicular cambium[25](Figure.1). This represents a specialized vascular meristem that gives rise exclusively to secondary xylem and phloem. In the secondary growth stage this continuous cambium band facilitates the transformation of the primary into secondary stem[23,25].

Vascular Patterning and Organization in the Inflorescence Stem

Many factors including plant hormones, peptide signaling factors, and transcriptional regulators are implicated in procambium/cambium proliferation, vascular patterning, and xylem differentiation in the stem tissue.

I. Hormone Signaling: Auxin is central in vascular tissue specification and it's response is mediated by *MONOPTEROS (MP)*, an auxin-responsive transcription factor which is thought to be an activator of vascular proliferation. During early development, *MP* is expressed in procambial cells and its expression is upregulated by auxin. In *MP* knockout mutant embryos there is a lack of procambium or irregular vascular development[26,27]. Such a defect seems to be related to the lack of polar auxin transport during early embryogenesis. *PIN-FORMED1 (PIN1)* proteins, main auxin efflux carriers, are polarly localized in in procambial cells. *PIN1* expression is dramatically reduced in a loss-of-function *MP* mutants, suggesting that *MP* might regulate its transcription[28]. Further, integrating computational modeling with analyzing synthetic auxin-response element DR5::GUS has shown that that vascular bundle spacing and distribution in the stem tissue is controlled in part by auxin maxima derived from polar transport[29]. Brassinosteroids (BRs) are also involved in vascular differentiation[30,31]. BR receptors BRI1-like receptor kinase (BRL1) and BRL3 are expressed in the vascular tissue and promote xylem

differentiation while repressing phloem formation. BR synthesis or signaling mutants exhibit alterations of the vascular pattern, implicating a role for BRs in promoting the formation of vascular bundles. Further, the interplay of auxin and BRs can predict the number of vascular bundles in the stem[29].

II. Peptide Signaling : In the vascular meristem, a few layers of stem cells proliferate and their daughter cells differentiate into xylem and phloem cell on opposite sides forming highly organized vascular bundles with clear boundaries between xylem and phloem identities[32]. *CLAVATA3/ESR LIKE 41(CLE41)* and *PHLOEM INTERCALATED WITH XYLEM (PXY)* are a ligand-receptor pair that regulates this organization in the vascular tissue. *PXY* gene encodes for a leucine-rich repeat receptor-like kinase that is expressed in the cambium and was identified through forward mutagenesis screen for mutants affected in primary stem anatomy[31]. In these mutants the clear boundary between xylem and phloem is diminished and the collateral bundle organization is disrupted as phloem tissue is mixed with xylem. This is attributed to defects in the orientation of procambial cell divisions. The *TRACHERY ELEMENT DIFFERENTIATION INHIBITORY FACTOR (TDIF)* encoded by *CLE41/42/44* genes is a ligand synthesized in the phloem and travels to the cambium, where it binds to its receptor PXY on the plasma membrane and activates PXY/TDIF signaling module[33](Figure.2). This dual functioning pathway regulates cambium proliferation by regulating the expression of the *WUSCHEL HOMEOBOX RELATED14* (*WOX4/14)*

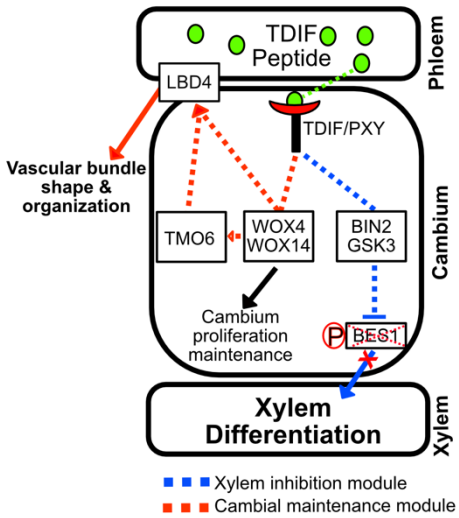


Figure. 2 Peptide signaling pathways controlling vascular development

transcription factors and suppresses xylem differentiation mediated by the GSK3 pathway (Figure. 2). Overexpression of *CLE41/42/44* peptide genes promotes procambial cell proliferation and inhibits xylem cell differentiation[34,35]. In a recent study, a regulatory network downstream of TDIF-PXY was identified using a yeast one-hybrid system[35]. A feedforward loop containing transcription factors *WOX14/WOX4* and *TARGET OF MONOPTEROS6 (TMO6)*, each of which regulates the expression of a third transcription factor, *LATERAL ORGAN BOUNDARIES DOMAIN4 (LBD4). PXY* in turn regulates the expression of *WOX14, TMO6*, and *LBD4* within a feedforward loop to control vascular proliferation. *LBD4* is found in the phloem-cambium boundary and *PXY-LBD4* feedforward loop defines the phloem-cambium boundary and thus the shape and organization of the vascular bundle[36]. Another genetic pathway is identified to act downstream of TDIF/PXY which suppresses xylem differentiation in the bundles and is mediated by *GLYCOGEN SYNTHASE KINASE 3 PROTEINS (GSK3)*. This model proposes activation of *GSK3* by TDIF/PXY module. GSK3 activation leads to phosphorylation and degradation of *BRI1-EMS SUPPRESSOR 1 (BES1)* which is a well-known target of GSK3 and a positive regulator of xylem differentiation. This maintains procambial cells by suppressing xylem differentiation[37]. Taken together, these findings have established distinct function of the TDIF-PXY module in cambial cell proliferation, maintenance, xylem differentiation, and overall vascular patterning.

III. Class III HD-ZIP and KANADI Genes Regulate Radial Patterning in Vascular Bundles

Radial patterning of the *Arabidopsis* inflorescence stem is also known to be regulated by *CLASS III HOMEODOMAIN-LEUCINE ZIPPER* (*HD-ZIPIII*) and *KANADI* transcription

factors[38]. *KANADI* and *HD-ZIPIII* factors *REVOLUTA* (*REV*), *PHABULOSA* (*PHB*), and *PHAVOLUTA* (*PHV*) exhibit complementary expression patterns in the vasculature. *KANADI* expression is restricted to the developing phloem and *HD-ZIPIII* expression is limited to the developing xylem. It is hypothesized that *HD-ZIPIII* activity is antagonistic with that of *KANADI* activity and these complementary expression patterns are shown to have functional significance. Interactions between the two gene classes pattern the arrangement of xylem and phloem tissues in the stem[37]. Loss-of-function *KANADI* mutants phenocopy stem vascular patterning defects indistinguishable from the *REV (*a member of the *class III HD-ZIP* family) gain-of-function alleles; both lines display indistinguishable alterations in the radial patterning in the stem with radialized and amphivasal vascular bundles, with xylem tissue surrounding phloem, in contrast to collateral bundles seen in the wild type[38]. In addition, *REV* loss-of-function mutant allele(s) exhibits alterations in the position of interfascicular fibers in the stem. The altered vascular patterning in these genotypes indicate a direct role for *HD-ZIP III* and *KANADI* gene families in stem vascular development[38].

Transcriptional Regulators of SCW Development in the *Arabidopsis* Stem

One of the best studied events of xylem and fiber cell differentiation is the production and deposition of secondary wall thickening controlled by a cascade of transcription factors[12]. The transcriptional network regulating SCW biosynthesis employs a feed-forward loop structure, in which the top-level NAC master switches activate the second level MYB master genes and they together induce the expression of downstream transcription factors and secondary wall biosynthesis genes. Arguably, the most studied group of

genes in this network are the first tier *Secondary Wall NAC (SWN)* domain transcription factors, including *NAC SECONDARY WALL THICKENING PROMOTING FACTOR1 (NST1)* , *NST2, SECONDARY WALL ASSOCIATED NAC DOMAIN PROTEIN1(SND1)* , also called *NST3, VASCULAR-RELATED NAC-DOMAIN6(VND6)* , *and VND7*. Expression analysis of *SWN*s has revealed that they are expressed in various *Arabidopsis* organs and cell types[12,39,40]. These five genes function at the top of the SCW genetic network and directly activate the expression of not only downstream transcription factors but also several genes involved in SCW biosynthesis and programmed cell death[11,40,41](Figure. 3). As the second-tier master switches, *MYB46/MYB83* are also able to activate the entire SCW biosynthetic program. These transcription factors are directly activated by SWNs. Dominant repression lines of *MYB83* and *MYB46* cause a lack of secondary walls in vessels and fibers and therefore suggests that they are functional homologs. Despite the tissue specificity of their upstream master switches, *MYB46* and *MYB83* seem to function in both vessels and fibers in a redundant manner[42]. As mentioned above, first tier SCW master regulators differ from one another depending on the tissue or cell type. In primary root, *VND6 and VND7* are xylem-specific and are

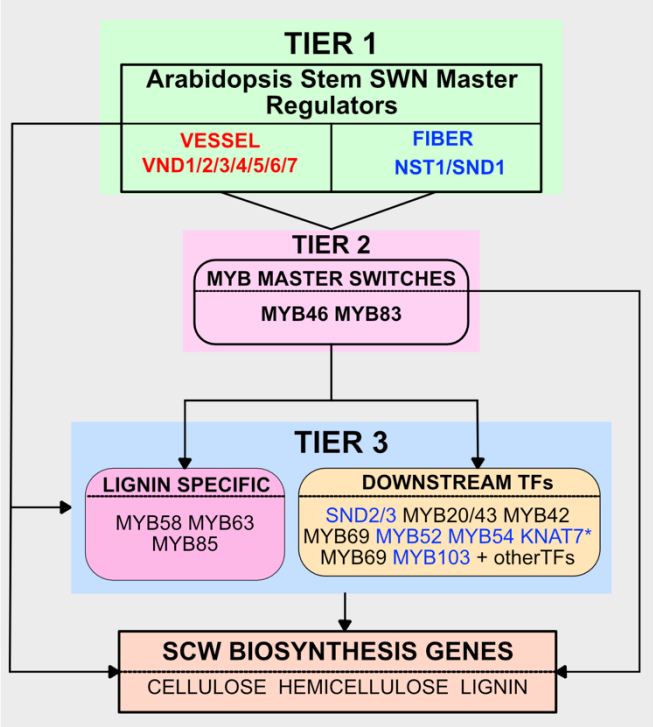


Figure. 3 Diagram of the Transcriptional Network Regulating Secondary Wall Biosynthesis in Arabidopsis Stem. Genes in blue are direct targets of SWNs.

considered master switches activating xylem development and secondary wall biosynthesis in vessel elements. Studies in *Arabidopsis* and tomato (chapter I) have shown that overexpression of *VND6* and/or *VND7* induces ectopic differentiation of metaxylem and protoxylem-like cells in roots.[39,43,44]. These data have been used as evidence to demonstrate that *VND6* and *VND7* are the main regulators of root xylem differentiation. On the other hand, knock-out mutations in either of these genes presented no detectable defects in root xylem cell morphology or growth, suggesting functional redundancy within this gene family[11]. Previous studies in the *Arabidopsis* root support this theory by demonstrating that the transcripts of other related VND transcription factors (*VND1* to *VND5*) are also xylem vessel-enriched and that their overexpression lines activate xylem transdifferentiation and SCW deposition[39]. Thus, *VND6* and *VND7* are sufficient but not necessary in xylem vessel SCW deposition. It should be noted that *VND6* and *VND7* loss-of-function phenotypes in roots were only demonstrated using the SRDX dominant repression system[11] which has succeeded in cases when single gene knock-outs do not produce a distinct phenotype. This system exhibits a phenotype similar to the loss-of-function of the target gene and all its functionally redundant paralogs[45].

In the *Arabidopsis* inflorescence stem, there are two secondary cell wall-forming cell types: vessel elements and fibers (xylary or interfascicular). Vessel elements are the main water conducting cells, while interfascicular and xylary fibers play a critical role in mechanical strength and load-bearing. *SND1* and it's close homolog *NST1* act redundantly as regulators of fiber SCW development in *Arabidopsis*[40,46]. *SND1* transcript is exclusively expressed in the stem, and expression analysis with beta-glucuronidase (GUS)-reporter lines demonstrate that the expression of this gene is restricted to

interfascicular fibers and xylary fibers. In a similar scenario to VND6/7 in vessel elements, a single *SND1* loss-of-function mutant allele does not produce any detectable phenotype, suggesting that other NAC genes might compensate for the loss of *SND1* expression. Dominant repression lines have been used to show specific reduction in the secondary wall thickening of fiber cells[39]. Further, overexpression of this gene results in a large number of epidermal and mesophyll cells developing a thick wall leading to stunted growth and developmental abnormalities in the flower such as shortened sepals, petals, stamens, and carpels[39]. Interestingly, despite these phenotypes, excess SND1 seems to inhibit normal SCW thickening in fiber cells in the stem [39]. *SND1* over-expression does not induce the expression of genes associated with programmed cell death such as *XCP1 (XYLEM CYSTEINE PEPTIDASE1), XCP2, and BFN1(BIFUNCTIONAL NUCLEASE1)*. Thus, *SND1* activates the developmental program for secondary wall thickening without induction of programmed cell death. This is in line with the fact that over-expression of *SND1* does not cause chlorosis in the leaves or other organs [39]. Another interesting fact about *SND1* overexpression is the different patterns of SCW deposited in different cell types. For instance, upon *SND1* overexpression, the SCWs in leaf epidermal cells are deposited as thick bands, whereas those in leaf mesophyll cells form a reticulated wall suggesting these patterns are independent of *SND1* and possibly dictated by microtubule organization in a cell type-dependent manner[40]. Simultaneous repression of *SND1* and its closet homolog *NST1* results in a loss of SCW formation in the stem, which is accompanied by a severe reduction in the expression of genes involved in their biosynthesis therefore lacking all three major secondary wall components, including cellulose, xylan, and lignin. Thus, *SND1*

and *NST1* function redundantly in the regulation of secondary wall synthesis in fibers[46].

Similar to the primary root, *VND6* and *VND7* are expressed in inflorescence stem vessel elements in the stem with *VND6* only expressed in non-elongating internodes while *VND7* is expressed in early and late vessel elements as well as xylary fibers[12]. Thus, the expression domain of these two genes seems to be developmentally regulated in an organ- and cell type-specific manner. Utilizing an inducible system, direct targets of these two genes have been previously reported; in the inflorescence stem *MYB46, SND3, MYB103,* and *KNAT7* are examples of genes activated directly by *VND7*. Interestingly, the same set of genes is also activated by master regulators of fiber development *SND*1 and *NST1*, revealing NAC master switches have a set of common targets regardless of their cell type-specific expression. Other VND family transcription factors, *VND1* to *VND5*, were also shown to be specifically expressed in vessels but not in interfascicular fibers in the stem[39]. When overexpressed, these transcription factors are also able to induce the transcription of other SCW-associated transcription factors as well as biosynthesis and cell death related genes[38]. As a result, ectopic deposition of all the three major secondary wall components, including lignin, xylan and cellulose is observed in stem. In contrast to this, dominant repression of *VND3* exhibits reduced SCW thickening and collapsed vessel elements in the vascular bundles. Just as in the case of *VND7*, *VND1* to *VND5* are also able to activate the same downstream transcription factors as *SND1*, suggesting that they are functional homologs of *SND1*. Moreover, complementation assays have revealed that *VND1* to *VND5* as well as *VND7* can rescue the secondary wall defects in fibers of the *nst1 nst3* double mutants[39,47]. These findings demonstrate that VND transcription factors

can also act as functional homologs of fiber SCW master regulators. As previously mentioned, single or double knockouts for either *VND6* or *VND7* do not show any morphological or growth phenotypes in the root or shoot. However, these results should not rule out that perturbations in these transcription factors could still lead to aberrant phenotypes, especially, subtle cell wall composition-related alterations, which is not detectable at the spatial resolution previously investigated.

Secondary Cell Wall Biosynthetic Enzymes

The third level transcription factors in the SCW network directly regulate the transcription of genes governing lignin, cellulose, and hemicellulose biosynthesis (Figure. 3). The major load bearing component in SCWs is cellulose; a linear polysaccharide comprised of β-1, 4-linked D-glucopyranosyl residues. Three cellulose synthases (*CESA4, CESA7* and *CESA8*) are necessary for cellulose production in SCWs specifically[48]. Plants carrying mutations in these genes lack SCW in the inflorescence stem. It is important to note, that the lignin and hemicellulose content in these plans were unaffected[48,49]. In addition to third level transcription factors, *MYB46, VND6* and *VND7* are also shown to directly activate secondary wall *CESA*s. This is demonstrated *in planta* by increased cellulose content in plants upon overexpression of these transcription factors[50]. It is likely that additional regulators regulate the transcription of secondary wall *CESA* genes. Indeed, yeast one-hybrid screening using the promoter sequences of *CESA4*, *CESA7* and *CESA8* as baits identified multiple transcription factors that bind to their promoter sequences and the exact function of many of these transcription factors remain unknown[50]. It has proven challenging to specifically uncouple the transcriptional regulation of cellulose and hemicellulose biopolymers within the SCW network[4].

Currently, only the lignin pathway seems to be specifically targeted by transcription factors such as *MYB58, MYB63,* and *MYB85* during xylem and fiber differentiation in root and inflorescence stem[51].(Figure. 3). In *Arabidopsis*, the dominant repression of *MYB58* and results in reduced lignin deposition and caused distorted vessels and reduced thickness of fiber cells, whereas *MYB58/63* overexpression produces ectopic lignin deposition (but not cellulose or xylan) in in both interfascicular fibers and xylary fibers in the inflorescence stem. Vessels in the MYB63 repressors were also deformed. Consistent with this phenotype, *MYB58/63* further activate the expression of downstream lignin biosynthesis-related genes but not those involved in xylan or cellulose biosynthesis[52]. In *Arabidopsis*, the overexpression of *MYB85* promotes lignin biosynthesis, leading to the ectopic deposition of lignin in the epidermal and cortex cells in the stem[12]. Moreover, *MYB85* controls the expression of *4CL1* (lignin specific) but not *IRX9* or *CesA8* (implicated in hemicellulose and cellulose synthesis respectively), suggesting that its role is in regulating lignin biosynthesis[12]. *MYB58* and *MYB63* transcription are regulated by the *SND1* and its homologs *NST1, NST2, VND6, and VND7* and their downstream target *MYB46*[50].These results suggest that *MYB85*, *MYB58*, and *MYB63* downstream of Tier 1 and 2 transcription factors in the SCW network and regulate lignin biosynthesis in both vessel and fiber cells[53].

One of the major hemicelluloses in SCW is xylan; a linear cell wall polymer that consists of a backbone made up of b-(1,4)-linked xylose residues that is generally substituted with arabinose, acetyl, glucuronic acid (GlcA), and 4-O-methylglucuronic acid residues to varying degrees depending on the plant species[53]. In *Arabidopsis* members of Glycosyltransferase family proteins *IRREGULAR XYLEM 9 (IRX9), IRX10 and IRX14* are

the key proteins involved in SCW xylan backbone biosynthesis[54–56]. Loss of function mutations in these genes lead to a shorter xylan chain length, reduced xylan content and SCW weakness sufficient to result in collapsed xylem vessels in the stem. all three genes have close homologs (*IRX9L*, *IRX14L*, *IRX10L*) that are partially functionally redundant with *IRX9, IRX14 and IRX10*, respectively. These homologs seem to differ mostly in their organ expression and are able to make a reduced quantity of xylan in various tissues. *IRX9, IRX10* and *IRX14L* are directly regulated by SWNs including *VND6, VND7* and *SND1*[57,58]. Overexpression of *IRX9* leads to increased xylan synthase activity in *Arabidopsis* plants[59], and other reports have shown a synergistic effect of *IRX9* and *IRX14* overexpression on xylan synthase activity[60]. *Arabidopsis* SCW xylan is predominantly composed of (methyl)glucuronoxylan where *GLUCURONIC ACID SUBSITUTION OF XYLAN1 (GUX1)* and *GUX2* are responsible for glucuronic acid (GlcA) sidechain addition and subsequently a proportion of these molecules undergo methylation[61]. Another two *Arabidopsis* glycosyltransferases, *FRAGILE FIBER8 (FRA8) or IRX7* (and *IRX7-L*), *IRX8* have also been implicated in glucuronoxylan (GX) biosynthesis and if mutated lead to decreased xylan content, but still xylan with extended backbones is present. Thus, it is suggested that this set of genes might be involved in synthesizing a primer or the reducing end oligosaccharide structure found in many plants[62].

Conclusion

In this section, the intricate and diverse nature of the plant secondary cell wall in Arabidopsis was covered. Molecular mechanism shaping *Arabidopsis* vascular tissue were highlighted by delving into various aspects of vascular tissue formation and

secondary cell wall development. Our knowledge of these processes in *Arabidopsis* has served as a framework for understanding the underlying molecular mechanisms. However, much more remains to be uncovered in this field. One significant knowledge gap pertains to the conservation of the gene network involved in plant vascular formation and secondary cell wall development across different plant species. Understanding the extent to which these regulatory mechanisms are conserved would shed light on the evolutionary aspects of vascular tissue development. Furthermore, the composition and regulation of plant cell walls display substantial variations among different cell types. The lack of high-resolution studies focused on specific cell types poses a limitation in the field. Conducting such studies would provide valuable insights into the composition and regulation of cell walls in a cell-type specific manner, thus enhancing our understanding of their diverse functions and adaptations.

1. Lucas, W. J. *et al.* The Plant Vascular System: Evolution, Development and FunctionsF. *J. Integr. Plant Biol.* **55**, 294–388 (2013).
2. Lee, K. J. D., Marcus, S. E. & Knox, J. P. Cell Wall Biology: Perspectives from Cell Wall Imaging. *Mol. Plant* **4**, 212–219 (2011).
3. Keegstra, K. Plant Cell Walls. *Plant Physiol.* **154**, 483–486 (2010).
4. Hussey, S., Mizrachi, E., Creux, N. & Myburg, A. Navigating the transcriptional roadmap regulating plant secondary cell wall deposition. *Front. Plant Sci.* **4**, (2013).
5. Burton, R. A., Gidley, M. J. & Fincher, G. B. Heterogeneity in the chemistry, structure and function of plant cell walls. *Nat. Chem. Biol.* **6**, 724–732 (2010).
6. Kumar, M., Campbell, L. & Turner, S. Secondary cell walls: biosynthesis and manipulation. *J. Exp. Bot.* **67**, 515–531 (2016).
7. Hernández-Blanco, C. *et al.* Impairment of cellulose synthases required for Arabidopsis secondary cell wall formation enhances disease resistance. *Plant Cell* **19**, 890–903 (2007).
8. Molina, A. *et al.* Arabidopsis cell wall composition determines disease resistance specificity and fitness. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2010243118 (2021).
9. Carpita, N. C. Update on Mechanisms of Plant Cell Wall Biosynthesis: How Plants Make Cellulose and Other (1→4)-β-d-Glycans1. *Plant Physiol.* **155**, 171–184 (2011).
10. Taylor-Teeples, M. *et al.* An Arabidopsis gene regulatory network for secondary cell wall synthesis. *Nature* **517**, 571–575 (2015).
11. Kubo, M. *et al.* Transcription switches for protoxylem and metaxylem vessel formation. *Genes Dev.* **19**, 1855–1860 (2005).

12. Zhong, R., Lee, C., Zhou, J., McCarthy, R. L. & Ye, Z.-H. A battery of transcription factors involved in the regulation of secondary cell wall biosynthesis in Arabidopsis. *Plant Cell* **20**, 2763–2782 (2008).

13. Peña, M. J., Darvill, A. G., Eberhard, S., York, W. S. & O'Neill, M. A. Moss and liverwort xyloglucans contain galacturonic acid and are structurally distinct from the xyloglucans synthesized by hornworts and vascular plants*. *Glycobiology* **18**, 891–904 (2008).

14. Haghighat, M., Teng, Q., Zhong, R. & Ye, Z.-H. Evolutionary Conservation of Xylan Biosynthetic Genes in Selaginella moellendorffii and Physcomitrella patens. *Plant Cell Physiol.* **57**, 1707–1719 (2016).

15. Zhao, Q. *et al.* Loss of function of cinnamyl alcohol dehydrogenase 1 leads to unconventional lignin and a temperature-sensitive growth defect in Medicago truncatula. *Proc. Natl. Acad. Sci.* **110**, 13660–13665 (2013).

16. Pattathil, S. *et al.* Cell Wall Ultrastructure of Stem Wood, Roots, and Needles of a Conifer Varies in Response to Moisture Availability. *Front. Plant Sci.* **7**, (2016).

17. Pattathil, S. *et al.* A comprehensive toolkit of plant cell wall glycan-directed monoclonal antibodies. *Plant Physiol.* **153**, 514–525 (2010).

18. Popper, Z. A. *et al.* Evolution and diversity of plant cell walls: from algae to flowering plants. *Annu. Rev. Plant Biol.* **62**, 567–590 (2011).

19. Verhertbruggen, Y., Walker, J. L., Guillon, F. & Scheller, H. V. A Comparative Study of Sample Preparation for Staining and Immunodetection of Plant Cell Walls by Light Microscopy. *Front. Plant Sci.* **8**, (2017).

20. Avci, U., Pattathil, S. & Hahn, M. G. Immunological Approaches to Plant Cell Wall and Biomass Characterization: Immunolocalization of Glycan Epitopes. in *Biomass Conversion: Methods and Protocols* (ed. Himmel, M. E.) 73–82 (Humana Press, 2012). doi:10.1007/978-1-61779-956-3_7.

21. Hall, H. C., Cheung, J. & Ellis, B. E. Immunoprofiling reveals unique cell-specific patterns of wall epitopes in the expanding Arabidopsis stem. *Plant J.* **74**, 134–147 (2013).

22. Strabala, T. J. & MacMillan, C. P. The Arabidopsis wood model—the case for the inflorescence stem. *Plant Sci.* **210**, 193–205 (2013).

23. Sanchez, P., Nehlin, L. & Greb, T. From thin to thick: major transitions during stem development. *Trends Plant Sci.* **17**, 113–121 (2012).

24. Fibers. A Model for Studying Cell Differentiation, Cell Elongation, and Cell Wall Biosynthesis | Plant Physiology | Oxford Academic. https://academic.oup.com/plphys/article/126/2/477/6102714.

25. Sehr, E. M. *et al.* Analysis of secondary growth in the Arabidopsis shoot reveals a positive role of jasmonate signalling in cambium formation. *Plant J.* **63**, 811–822 (2010).

26. Berleth, T. & Jurgens, G. The role of the monopteros gene in organising the basal body region of the Arabidopsis embryo. *Development* **118**, 575–587 (1993).

27. Hardtke, C. S. & Berleth, T. The Arabidopsis gene MONOPTEROS encodes a transcription factor mediating embryo axis formation and vascular development. *EMBO J.* **17**, 1405–1411 (1998).

28. Miyashima, S., Sebastian, J., Lee, J.-Y. & Helariutta, Y. Stem cell function during plant vascular development. *EMBO J.* **32**, 178–193 (2013).

29. Ibañes, M., Fàbregas, N., Chory, J. & Caño-Delgado, A. I. Brassinosteroid signaling and auxin transport are required to establish the periodic pattern of Arabidopsis shoot vascular bundles. *Proc. Natl. Acad. Sci.* **106**, 13630–13635 (2009).

30. Kang, Y. H., Breda, A. & Hardtke, C. S. Brassinosteroid signaling directs formative cell divisions and protophloem differentiation in Arabidopsis root meristems. *Development* **144**, 272–280 (2017).

31. Yamamoto, R., Demura, T. & Fukuda, H. Brassinosteroids Induce Entry into the Final Stage of Tracheary Element Differentiation in Cultured Zinnia Cells. *Plant Cell Physiol.* **38**, 980–983 (1997).

32. Fisher, K. & Turner, S. PXY, a Receptor-like Kinase Essential for Maintaining Polarity during Plant Vascular-Tissue Development. *Curr. Biol.* **17**, 1061–1066 (2007).

33. Yuan, B. & Wang, H. Peptide Signaling Pathways Regulate Plant Vascular Development. *Front. Plant Sci.* **12**, (2021).

34. Etchells, J. P. & Turner, S. R. The PXY-CLE41 receptor ligand pair defines a multifunctional pathway that controls the rate and orientation of vascular cell division. *Development* **137**, 767–774 (2010).

35. Hirakawa, Y. *et al.* Non-cell-autonomous control of vascular stem cell fate by a CLE peptide/receptor system. *Proc. Natl. Acad. Sci.* **105**, 15208–15213 (2008).

36. Smit, M. E. *et al.* A PXY-Mediated Transcriptional Network Integrates Signaling Mechanisms to Control Vascular Development in Arabidopsis[OPEN]. *Plant Cell* **32**, 319–335 (2020).

37. Kondo, Y. *et al.* Plant GSK3 proteins regulate xylem cell differentiation downstream of TDIF–TDR signalling. *Nat. Commun.* **5**, 3504 (2014).

38. Emery, J. F. *et al.* Radial Patterning of Arabidopsis Shoots by Class III HD-ZIP and KANADI Genes. *Curr. Biol.* **13**, 1768–1774 (2003).

39. Zhou, J., Zhong, R. & Ye, Z.-H. Arabidopsis NAC Domain Proteins, VND1 to VND5, Are Transcriptional Regulators of Secondary Wall Biosynthesis in Vessels. *PLOS ONE* **9**, e105726 (2014).

40. Zhong, R., Demura, T. & Ye, Z.-H. SND1, a NAC Domain Transcription Factor, Is a Key Regulator of Secondary Wall Synthesis in Fibers of Arabidopsis. *Plant Cell* **18**, 3158–3170 (2006).

41. Mitsuda, N. *et al.* NAC transcription factors, NST1 and NST3, are key regulators of the formation of secondary walls in woody tissues of Arabidopsis. *Plant Cell* **19**, 270–280 (2007).

42. McCarthy, R. L., Zhong, R. & Ye, Z.-H. 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 (2009).

43. Endo, H. *et al.* Multiple Classes of Transcription Factors Regulate the Expression of VASCULAR-RELATED NAC-DOMAIN7, a Master Switch of Xylem Vessel Differentiation. *Plant Cell Physiol.* **56**, 242–254 (2015).

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

45. Mitsuda, N. & Ohme-Takagi, M. Functional Analysis of Transcription Factors in Arabidopsis. *Plant Cell Physiol.* **50**, 1232–1248 (2009).

46. Zhong, R., Richardson, E. A. & Ye, Z.-H. Two NAC domain transcription factors, SND1 and NST1, function redundantly in regulation of secondary wall synthesis in fibers of Arabidopsis. *Planta* **225**, 1603–1611 (2007).

47. Yamaguchi, M. *et al.* VASCULAR-RELATED NAC-DOMAIN 7 directly regulates the expression of a broad range of genes for xylem vessel formation. *Plant J.* **66**, 579–590 (2011).

48. Taylor, N. G. Cellulose biosynthesis and deposition in higher plants. *New Phytol.* **178**, 239–252 (2008).

49. Turner, S. R. & Somerville, C. R. Collapsed xylem phenotype of Arabidopsis identifies mutants deficient in cellulose deposition in the secondary cell wall. *Plant Cell* **9**, 689–701 (1997).

50. Kim, W.-C. *et al.* MYB46 directly regulates the gene expression of secondary wall-associated cellulose synthases in Arabidopsis. *Plant J.* **73**, 26–36 (2013).

51. Zhou, J., Lee, C., Zhong, R. & Ye, Z.-H. MYB58 and MYB63 Are Transcriptional Activators of the Lignin Biosynthetic Pathway during Secondary Cell Wall Formation in Arabidopsis. *Plant Cell* **21**, 248–266 (2009).

52. Zhong, R. & Ye, Z.-H. Transcriptional regulation of lignin biosynthesis. *Plant Signal. Behav.* **4**, 1028–1034 (2009).

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

54. The Arabidopsis IRX10 and IRX10-LIKE glycosyltransferases are critical for glucuronoxylan biosynthesis during secondary cell wall formation - Wu - 2009 - The Plant

Journal - Wiley Online Library. https://onlinelibrary.wiley.com/doi/full/10.1111/j.1365-313X.2008.03724.x.

55. Arabidopsis irregular xylem8 and irregular xylem9: Implications for the Complexity of Glucuronoxylan Biosynthesis | The Plant Cell | Oxford Academic. https://academic.oup.com/plcell/article/19/2/549/6092251?login=false.

56. Brown, D. M. *et al.* Comparison of five xylan synthesis mutants reveals new insight into the mechanisms of xylan synthesis. *Plant J.* **52**, 1154–1168 (2007).

57. Zhong, R., Lee, C. & Ye, Z.-H. Global Analysis of Direct Targets of Secondary Wall NAC Master Switches in Arabidopsis. *Mol. Plant* **3**, 1087–1103 (2010).

58. Ohashi-Ito, K., Oda, Y. & Fukuda, H. Arabidopsis VASCULAR-RELATED NAC-DOMAIN6 Directly Regulates the Genes That Govern Programmed Cell Death and Secondary Wall Formation during Xylem Differentiation[C][W]. *Plant Cell* **22**, 3461–3473 (2010).

59. Lee, C., Teng, Q., Huang, W., Zhong, R. & Ye, Z.-H. The Arabidopsis Family GT43 Glycosyltransferases Form Two Functionally Nonredundant Groups Essential for the Elongation of Glucuronoxylan Backbone. *Plant Physiol.* **153**, 526–541 (2010).

60. Lee, C., Zhong, R. & Ye, Z.-H. Arabidopsis Family GT43 Members are Xylan Xylosyltransferases Required for the Elongation of the Xylan Backbone. *Plant Cell Physiol.* **53**, 135–143 (2012).

61. Mortimer, J. C. *et al.* Absence of branches from xylan in Arabidopsis gux mutants reveals potential for simplification of lignocellulosic biomass. *Proc. Natl. Acad. Sci.* **107**, 17409–17414 (2010).

62. York, W. S. & O'Neill, M. A. Biochemical control of xylan biosynthesis — which end is up? *Curr. Opin. Plant Biol.* **11**, 258–265 (2008).

Chapter I

Identification of Transcriptional Regulators of Xylem Cell Development in *Solanum lycopersicum*

Kaisa Kajala*, **Mona Gouran***, Lidor Shaar-Moshe*, G. Alex Mason*, Joel Rodriguez-Medina*, Dorota Kawa,Germain Pauluzzi, Mauricio Reynoso, Alex Canto-Pastor, Concepcion Manzano, Vincent Lau, Mariana A. S. Artur, Donnelly A. West, Sharon B. Gray, Andrew I. Yao, Marko Bajic, Elide Formentin, Niba Nirmal, Alan Rodriguez, Asher Pasha, Alexander T. Borowsky, Roger B. Deal, Daniel Kliebenstein, Torgeir R. Hvidsten, Nicholas J. Provart, Neelima Sinha, Daniel E. Runcie, Julia Bailey-Serres, & Siobhan M. Brady

* These authors contributed equally

The first chapter of this thesis was published in the journal *Cel Volume 184, Issue 12, 10 June 2021, Pages 3333-3348.* In this study, I identified a group of transcription factors (TF) involved in tomato root xylem cell development.

Arabidopsis research is the framework for our understanding of cell type development in other organisms, however, wiring of these regulatory networks is not identical across species. To test this hypothesis, I used xylem development as a case study as our lab has made significant contributions to understanding regulatory networks underlying xylem and secondary cell wall development[1].

I characterized cell type resolution transcriptome of Solanum lycopersicum using Translating Ribosome Affinity Purification (TRAP-seq) technology and generated a list of xylem-enriched TFs in tomato roots. To uncover the conserved and unique genetic mechanisms driving xylem development in tomato, I used phylogenomic analysis of candidate TFs to determine their most likely orthologs and further incorporated the data

available on natural variation in xylem cell number in two species of tomato to identify a set of tomato xylem enriched TFs that coincide with significant *cis*-expression quantitative trait loci (eQTLs) in roots of introgression lines of *Solanum lycopersicum* and *Solanum pennellii*. Many key TFs regulating xylem development have been identified in *Arabidopsis* through overexpression studies, therefore, I then generated overexpression lines of the tomato candidate genes in the hairy root system for functional validation. From the combination of phylogenomic, translatome, and overexpression data, I identified several conserved and unique xylem transcriptional regulators in tomato. My data suggests that *SlVND6* is a likely functional ortholog of *AtVND6* in tomato. I also show in this paper that two of HDZIP III orthologs namely *SlPHB/PHV-LIKE1* and *SlCNAL1* regulate xylem patterning both in *Arabidopsis* and tomato in a semi-conserved manner. My further investigation of tomato xylem enriched TFs revealed an expanded expression domain for *SlKNAT1* compared to *AtKNAT1*[2]. Xylem developmental abnormalities in overexpression lines of *SlKNAT1* compared to a lack of phenotype in *Arabidopsis KNAT1* overexpression lines suggests that *SlKNAT1* function is repurposed to control xylem cell development in tomato root.

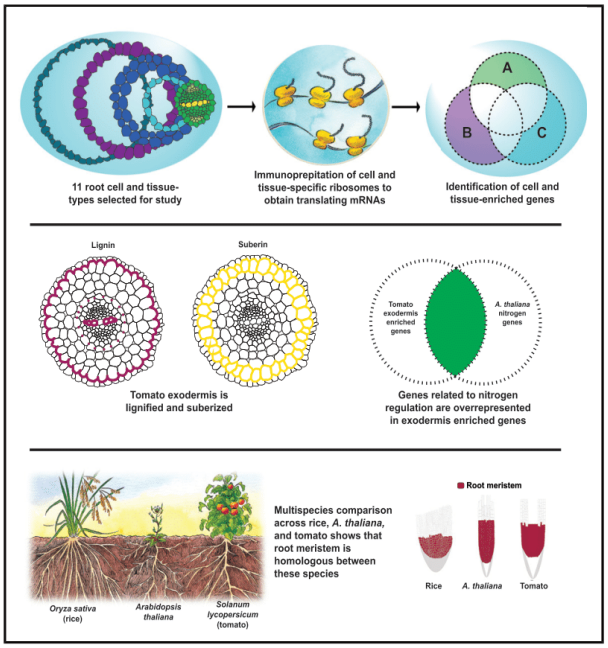
Taken together, my data illustrate conservation and repurposing of transcriptional regulation in xylem development in tomato relative to the *Arabidopsis* network.

Cell

Resource

Innovation, conservation, and repurposing of gene function in root cell type development

Graphical abstract



Highlights

- Tomato cell type-resolution translatome atlas reveals cell type function
- Conservation and repurposing in gene regulation between *Arabidopsis* and tomato
- The tomato exodermis is lignified, suberized, and enriched for nitrogen regulation
- The root meristem is molecularly homologous across plant species

Authors

Kaisa Kajala, Mona Gouran, Lidor Shaar-Moshe, ..., Daniel E. Runcie, Julia Bailey-Serres, Siobhan M. Brady

Correspondence

sbrady@ucdavis.edu

In brief

The integration of pan-species cell type data reveals molecular signatures across growth conditions and sheds light on novelty, conservation, and repurposing of gene function relevant to crop engineering.

Kajala et al., 2021, Cell *184*, 3333–3348
June 10, 2021
https://doi.org/10.1016/j.cell.2021.04.024

CellPress

Cell

CellPress
OPEN ACCESS

Resource

Innovation, conservation, and repurposing of gene function in root cell type development

Kaisa Kajala,[1,2,15] Mona Gouran,[1,3,15] Lidor Shaar-Moshe,[1,15] G. Alex Mason,[1,15] Joel Rodriguez-Medina,[1,4,15] Dorota Kawa,[1,16] Germain Pauluzzi,[5,16] Mauricio Reynoso,[5,6,16] Alex Canto-Pastor,[1,16] Concepcion Manzano,[1] Vincent Lau,[7] Mariana A.S. Artur,[2] Donnelly A. West,[4,8] Sharon B. Gray,[1,17] Alexander T. Borowsky,[5] Bryshal P. Moore,[9] Andrew I. Yao,[10] Kevin W. Morimoto,[1] Marko Bajic,[11] Elide Formentin,[5,12] Niba A. Nirmal,[1] Alan Rodriguez,[1] Asher Pasha,[7] Roger B. Deal,[11] Daniel J. Kliebenstein,[13] Torgeir R. Hvidsten,[14] Nicholas J. Provart,[7] Neelima R. Sinha,[8] Daniel E. Runcie,[13] Julia Bailey-Serres,[2,5] and Siobhan M. Brady[1,18,*]

[1]Department of Plant Biology and Genome Center, University of California, Davis, Davis, CA 95616, USA
[2]Plant Ecophysiology, Institute of Environmental Biology, Utrecht University, 3584 Utrecht, the Netherlands
[3]Plant Biology Graduate Group, University of California, Davis, Davis, CA 95616, USA
[4]Integrative Genetics and Genomics Graduate Group, University of California, Davis, Davis, CA 95616, USA
[5]Center for Plant Cell Biology, Department of Botany and Plant Sciences, University of California, Riverside, Riverside, CA 92521, USA
[6]IBBM, FCE-UNLP CONICET, La Plata 1900, Argentina
[7]Department of Cell and Systems Biology/Centre for the Analysis of Genome Evolution and Function, University of Toronto, 25 Willcocks St., Toronto, ON M5S 3B2, Canada
[8]Department of Plant Biology, University of California, Davis, Davis, CA 95616, USA
[9]Fort Valley State University, Fort Valley, GA 31030, USA
[10]Department of Biomedical Engineering and Genome Center, University of California, Davis, Davis, CA 95616, USA
[11]Department of Biology, Emory University, Atlanta, GA 30322, USA
[12]Department of Biology, University of Padova, Padova, Italy
[13]Department of Plant Sciences, University of California, Davis, Davis, CA 95616, USA
[14]Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, 1432 Ås, Norway
[15]These authors contributed equally
[16]These authors contributed equally
[17]Deceased
[18]Lead contact
*Correspondence: sbrady@ucdavis.edu
https://doi.org/10.1016/j.cell.2021.04.024

SUMMARY

Plant species have evolved myriads of solutions, including complex cell type development and regulation, to adapt to dynamic environments. To understand this cellular diversity, we profiled tomato root cell type translatomes. Using xylem differentiation in tomato, examples of functional innovation, repurposing, and conservation of transcription factors are described, relative to the model plant *Arabidopsis*. Repurposing and innovation of genes are further observed within an exodermis regulatory network and illustrate its function. Comparative translatome analyses of rice, tomato, and *Arabidopsis* cell populations suggest increased expression conservation of root meristems compared with other homologous populations. In addition, the functions of constitutively expressed genes are more conserved than those of cell type/tissue-enriched genes. These observations suggest that higher order properties of cell type and pan-cell type regulation are evolutionarily conserved between plants and animals.

INTRODUCTION

Irrespective of species, all vascular plant roots contain a stem cell niche at the root tip and cell types along the radial axis that are arranged in concentric cylinders. These cell types are constrained within files along the root longitudinal axis. After production from initial (stem) cells, the epidermis, cortex, endodermis, and vascular cells progress through 3 defined developmental zones: the root meristem (including the stem cell niche and proliferating cells), elongation zone, and maturation zone. Epidermal cells uptake water and nutrients from the rhizosphere. Ground tissue contains the cortex and endodermis, the latter of which produces an intercellular barrier to regulate the apoplastic movement of water and nutrients to and from the vascular tissue. The xylem transports water and mineral nutrients, while the phloem transports photosynthetic sugars and other molecules. Based on morphology and expression data, many plant cell types and developmental zones are considered homologous (i.e., derived

from a common ancestor) (Cridge et al., 2016; Kenrick and Strullu-Derrien, 2014). However, the degree to which root cell type developmental programs are molecularly or functionally conserved across plant species is unknown.

Such questions of developmental conservation have long been considered in animals. The developmental hourglass model hypothesizes that body plans, as described by anatomical and morphological features, are established at the most conserved embryonic or phylotypic period (Duboule, 1994; Raff, 2012; Smith et al., 1985). More recently, orthologous gene expression profiles across animals were used to identify the phylotypic period (Cruickshank and Wade, 2008; Gilad and Mizrahi-Man, 2015). In plants, transcriptomic analyses suggest an analogous hourglass model in plant embryos, with the phylotypic period occurring during the embryonic stage when the body plan is being established (Drost et al., 2015; Quint et al., 2012). Given tissue-type (epidermal, ground tissue, vascular) and temporal (developmental zone) conservation (Cridge et al., 2016; Kenrick and Strullu-Derrien, 2014), similar questions arise as to the molecular similarity between these spatiotemporal aspects of root development.

There is also diversity in root cell types as well as in cell signaling and metabolic programs across species. This diversity can remain uncharacterized if a given cell type, signaling, or metabolic program is not present in a reference species. For example, the exodermis is an outer cortex layer, which can produce an apoplastic barrier, that is present in a reported 93% of angiosperms, but absent in the model species *Arabidopsis thaliana* and thus molecularly uncharacterized (Perumalla et al., 1990).

Transcriptome as well as ribosome-associated mRNA profiles (translatomes) have provided insight into the regulatory mechanisms underlying root cell type development and its interaction with the environment in *Arabidopsis* (Brady et al., 2007; Denyer et al., 2019; Dinneny et al., 2008; Iyer-Pascuzzi et al., 2011; Jean-Baptiste et al., 2019; Li et al., 2016; Mustroph et al., 2009; Ryu et al., 2019; Shulse et al., 2019; Turco et al., 2019). Typically, transcriptomes of root cell types are obtained by cell protoplasting coupled with fluorescence-activated cell sorting (Birnbaum et al., 2003; Brady et al., 2007; Li et al., 2016) or immunopurification of tagged nuclei within specific cell populations (Deal and Henikoff, 2010). Cell protoplasting has also been used to characterize transcriptomes of individual root cells (Denyer et al., 2019; Jean-Baptiste et al., 2019; Ryu et al., 2019; Shulse et al., 2019; Turco et al., 2019). In comparison, translatomes comprise transcripts associated with tagged ribosomes within specific cell populations (translating ribosome affinity purification [TRAP]) (Mustroph et al., 2009, 2014) and thus can be considered a proxy for translation.

Here, we use TRAP to profile a variety of cell populations in tomato and rice of distinct developmental stages and growth conditions. We then test hypotheses generated from these data with *Rhizobium-rhizogenes*-transformed (hairy) roots that resemble the cellular architecture of primary roots and provide a rapid mode of functional validation (i.e., weeks compared to months with *Agrobacterium-tumefaciens*-mediated transformation) (Ron et al., 2014). These data illustrate conservation and repurposing of transcriptional regulation in xylem of tomato and *Arabidopsis*. Exodermis-enriched transcripts and associated networks reveal exodermis function, and multi-species analyses reveal the degree of molecular conservation of four homologous cell types and tissues across tomato, *Arabidopsis*, and rice.

RESULTS

Tomato cell type and tissue-resolution translatomes

The tomato (*Solanum lycopersicum* cv. M82) root contains the same cell types as *Arabidopsis*, with the exception of three cortex layers (the exodermis and two inner cortex layers) versus one cortical layer in *Arabidopsis* (Ron et al., 2013). We previously identified 11 promoters in tomato that drive expression in distinct or overlapping cell type domains of *Rhizobium-rhizogenes*-transformed roots (Ron et al., 2014). The primary domains marked include the epidermis and lateral root cap (*pAtWER*); the two non-exodermis (inner) cortex layers throughout all developmental zones (*pAtPEP*); the inner cortex layers in the root meristematic zone (*pSlCO2*); the endodermis and a single tier of the quiescent center (QC) (*pSlSCR*); the stele (*pSlSHR*); the phloem and vascular initials (*pAtS32*); xylem and epidermis in the maturation zone (*pAtS18*); the QC, initials, and pericycle in the root meristem (*pSlWOX5*); the meristematic zone (*pSlRPL11C*); and two constitutive promoters (*p35S* and *pSlACT2*) (Ron et al., 2014) (Figures 1A and S1; for detailed expression profiles, see Table S1). An additional promoter, *SlPEP*, was newly identified that marks the exodermis and the inner cortex layers in all developmental zones (Figures 1A and S1). Comparisons of transcripts in cells marked by *SlPEP* versus *AtPEP* facilitate characterization of exodermis function.

Cell type translatomes are easily accessed by ribosome immunopurification. These 12 promoters were fused to a FLAG-GFP-tagged ribosomal protein (RPL18) to enable TRAP of mRNA coupled with sequencing (TRAP-seq) (Figure 1B; *35S* and *SlACT2* datasets profile the same cell populations; thus, 11 cell populations were profiled) (Mustroph et al., 2009; Reynoso et al., 2019; Ron et al., 2014). We confirmed GFP patterns to be largely similar between the TRAP lines (transformed with *A. tumefaciens*) and those previously observed in hairy roots (Ron et al., 2014) and further described TRAP-GFP patterns across the developmental zones (Figure S1; Table S1). Principal-component analysis (PCA) revealed a clear grouping of the samples based on cell populations, confirming the reproducibility of the marker-line-derived translatomes (Figure S1C). Expression patterns of known cell type markers (Brady et al., 2007; Li et al., 2016) in tomato marker line translatomes largely recapitulated expected expression patterns, thus providing a first validation of our approach to quantify ribosome-associated transcripts at cell population resolution (Table S1). Normalized transcript abundance is visualized on a gene-by-gene basis in the ePlant browser (Figures 1C and 1D) with *SlACT2* data not included due to redundancy with 35S.

Inference of CTEGs

A number of these promoters drive expression in specific cell types, while others are expressed in overlapping domains. Using the spatial and temporal domains driven by the 12 promoters (Table S1), we formulated arguments to infer cell type-enriched

CellPress
OPEN ACCESS

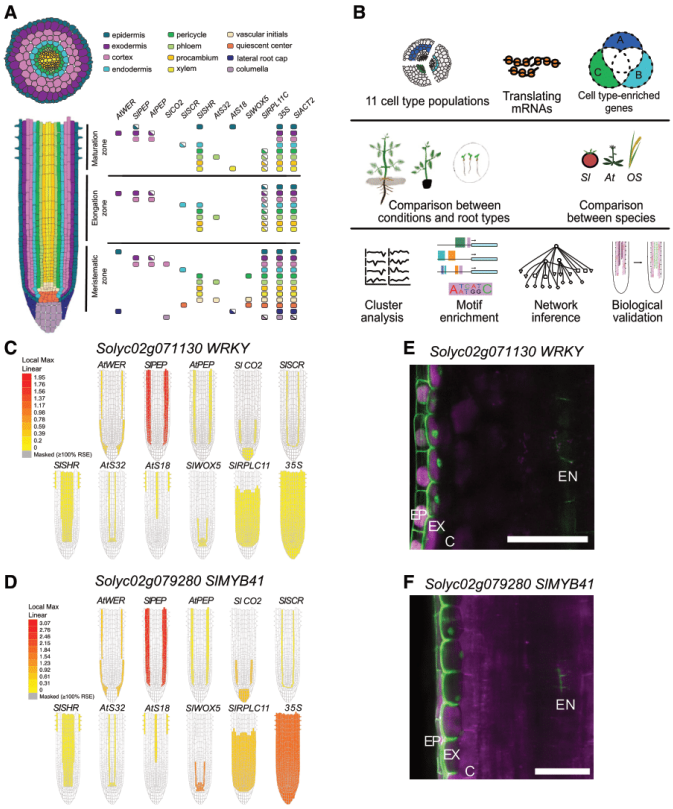


Figure 1. Tomato translatome root atlas at cell type resolution

(A) Tomato root cell types profiled and the promoters that mark them.

(B) Overview of experimental approach.

(C and D) Relative TRAP-RNA abundance of *Solyc02g071130* and *Solyc02g079280* (*SlMYB41*) in 11 marker line translatomes in the Tomato ePlant "Experiment eFP." The colored regions represent the domain marked by the given promoter, according to the normalized heatmap scale.

(E and F) *Solyc02g071130* (E) and *Solyc02g079280* (*SlMYB41*) (F) were inferred to be exodermis enriched, as confirmed by their promoter:nuclear localized GFP pattern (green). Green signal in the cell wall autofluorescence. Magenta signal TagRFP (membrane-tagged red fluorescent protein [RFP]). EP, epidermis; EX, exodermis; C, cortex; EN, endodermis. Scale bar, 100 μm.

See also Figure S1 and Tables S1 and S2.

genes (CTEGs) that display significant translatome enrichment in one cell type relative to the others (STAR Methods). We utilized these CTEG lists to explore the molecular signatures enriched in tomato root cell types, including that of the exodermis. We validated our computational approach using transcriptional fusions of selected CTEGs in tomato hairy roots. This approach was validated with *MYB* (*Solyc02g079280*; *SlMYB41*) and *WRKY* (*Solyc02g071130*) transcription factor (TF) promoters driving nuclear-localized GFP (Figures 1E and 1F) solely in the exodermis. Thus, our CTEGs provide an opportunity to infer cell type function.

Condition-specific and "core" CTEGs

The most complete analyses of single cell or cell type-resolution gene expression in plant roots comes from studies of *Arabidopsis* seedlings grown in sterile conditions, due to experimental tractability (Brady et al., 2007; Denyer et al., 2019; Jean-Baptiste et al., 2019; Li et al., 2016; Mustroph et al., 2009; Ryu et al., 2019; Shulse et al., 2019). However, plants in their natural environment grow in soil with a composition that is heterogeneous. Furthermore, the seedling root system is primarily composed of a single primary root, while the mature plant root system architecture is complex (Figure S1D). The system elaborates with roots of different developmental origins including the primary root, lateral roots, and shoot-borne roots. This architecture is highly plastic and dependent on dynamic interactions between cell type regulatory networks and the environment. The degree to which cell type expression patterns are conserved or divergent in their natural soil environment or in roots of different developmental origins is unknown.

Although there are limitations in interpretation of cell type-resolution data from seedlings grown in sterile conditions, these conditions have enabled characterization of the environmental responsiveness of genes within *Arabidopsis* root cell types. Indeed, a subset of *Arabidopsis* CTEGs maintain their expression patterns in response to stress conditions (Dinneny et al., 2008; Iyer-Pascuzzi et al., 2011). Yet, maintenance of these patterns in pots or field conditions is unknown. To identify such candidate tomato genes, we explored cell population-specific expression dynamics between sterile conditions and the field. We sequenced the meristematic zone (*pSlRPL11C*), meristematic inner cortex (*SlCO2*), and endodermis/QC (*SlSCR*) translatomes from 2-month-old plants grown in the field under standard cultivation practices. CTEGs were defined as previously described, but from comparisons only involving these marked cell populations. We compared CTEGs from sterile-grown seedlings and field-grown plants to determine the extent of overlap (Table S2). Despite differences in plant age, we found overlapping genes between the meristematic zone (*SlRPL11C*, 50 genes), endodermis/QC (*SlSCR*, 47 genes), and the meristematic inner cortex (*SlCO2*, two genes). We call these overlapping genes "core" CTEGs (Figures 2A–2C). The endodermis/QC "core" CTEGs were enriched

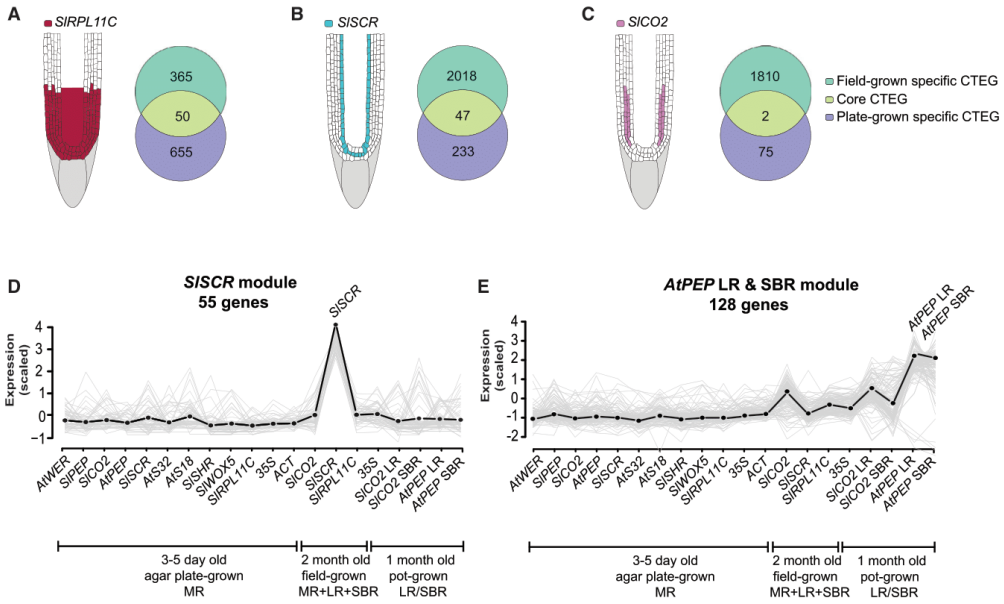


Figure 2. Cell type-enriched genes (CTEGs) across multiple developmental zones, conditions, and root types

(A–C) Core and condition-specific CTEGs in (A) meristematic zone, (B) endodermis/quiescent center (QC), and (C) meristematic cortex.

(D and E) WGCNA co-expression modules with scaled expression values (y axis) across translatome profiles derived from different (i) promoters (Figure 1A), (ii) conditions (3- to 5-day-old plants grown on sterile agar plates in a growth chamber; 2-month-old plants grown in the field; 1-month-old plants grown in a growth chamber), and (iii) individual root types (MR, main root; LR, lateral roots; SBR, shoot-borne roots). Black dotted line indicates eigengene expression profile. The maximum peak of expression within the module is indicated by black font on top of the eigengene expression line. Gray line indicates expression values of all genes within the module. Most genes in these modules were positively correlated to the eigengene.

See also Figure S2 and Tables S2 and S3.

for ontology terms associated with nucleic acid binding (p value = 0.036) and the CYS2-HIS2 ZINC FINGER (C2H2-ZF) family (adjusted [adj] p = 0.05), while meristematic zone "core" CTEGs were associated with zinc ion binding (p value = 0.016) and calcium ion binding (p value = 0.045) (Table S2). Core endodermis/QC-enriched genes included *SCARECROW* (*SlSCR*, *Solyc10g074680*), a homolog of a core endodermis-enriched gene in *Arabidopsis* (*At3g54220*) (Iyer-Pascuzzi et al., 2011); two zinc-finger TFs (*Solyc01g090840*, a single ortholog of *ZINC-FINGER ARABIDOPSIS THALIANA GENE*s) (*AtZAT4* and *AtZAT9*); and *Solyc06-g054600*, the *Solanum* zinc-finger (C-x8-C-x5-C-x3-H) family protein (Table S2; Data S1A–S1C).

We complemented the analyses of "core" root CTEGs from different conditions with that of root types of different developmental origin. To this end, we obtained additional translatome data from the meristematic inner cortex (*pSlCO2*) and the inner cortex layers throughout all developmental zones (*pAtPEP*) of lateral and shoot-borne roots from 1-month-old plants grown in pots (Figure S1D). By compiling these translatome data with the translatome of sterile-grown seedlings (primary root) and field-grown plants (whole root system) and using weighted gene correlation network analysis (WGCNA) (Langfelder and Horvath, 2008), we identified modules of co-expressed genes enriched within a root type or environment. The predominant effect of the environment on gene expression is captured by large modules of co-expressed genes whose relative expression is higher for all or most cell populations in a specific condition, i.e., field-enriched expression of 4,006 genes (Figure S2A; Table S2) and a module of 2,896 co-expressed genes in more typical cultivation conditions (pot- and field-grown plants; Figure S2B; Table S2). Co-expression modules with expression limited to a particular cell type or root type comprise a smaller number of genes (Figures 2D and 2E). Consistent with the role of the endodermis in interactions with the environment (Robbins et al., 2014), we found a group of genes co-expressed only in the endodermis/QC, but specifically in field conditions. These include genes whose function is linked with response to the environment, such as *Solyc10g080310*, a dehydration-responsive element binding TF and a CASP-like protein (*Solyc07g056040*) (Figure 2D; Table S2) (Agarwal et al., 2017; Lee et al., 2019). One module represents genes co-expressed in cortex cells specifically in lateral and shoot-borne roots of mature plants, but not in the main root of plate-grown plants (seedlings). Genes co-expressed in inner cortex layers (*AtPEP* and *SlCO2*) are more lowly expressed in the meristematic inner

cortex (*SlCO2*) of lateral roots in pot-grown plants and the whole root system in field-grown plants (Figure 2E; Table S2). The root system of the field-grown plants consists primarily of lateral roots. These same genes are then more highly expressed within this module in mature inner cortex cells (*AtPEP*) of shoot-borne and lateral roots (Figure 2E; Table S2). The gradual increase in the expression of genes in this module could reflect the temporal trajectory of cortical cells from the meristem to the maturation zone in lateral and shoot-borne roots. Functions associated with these genes include calcium signaling and hydrolase activity (Figure 2E; Table S2). Together, by profiling translatomes of a subset of cell or tissue types of several growth stages and under several conditions, we identified three classes of root cell type-enriched signatures. While (i) "core" cell type signatures maintain expression over time and in dynamic environments, most of the cell type signatures are (ii) specific to a given root type or (iii) depend on external conditions. In the future, analyzing representative marker genes from these three signatures in roots of the same developmental origin in sterile conditions, pots, and the field will determine their validity as archetypes for tomato root cell type spatiotemporal patterns.

Conservation and divergence of xylem regulation between *Arabidopsis* and tomato

Next, we explored conservation and differences in cell type regulatory networks between tomato and *Arabidopsis* with xylem regulation as a case study. When differentiated, xylem cells are elongated, primarily hollow cells encased by a secondary cell wall. The xylem secondary cell wall is a critical component of wood, and sugars within this wall are harnessed for biofuels. Xylem cell development is a critical feature of land plant evolution. Bryophytes lack xylem and obtain water by growing on or near its surface. By contrast, plants with xylem are able to transport water over great distances and thus exploit different ecological niches than bryophytes (Raven, 1993).

Much of what we know about xylem patterning and differentiation has been from studies of the *Arabidopsis* root. On either side of the vascular cylinder, a single protoxylem vessel differentiates in the root meristem, with up to three intervening metaxylem vessels that differentiate in the root maturation zone (Figure 3A). Protoxylem vessels have an annular or helical secondary cell wall morphology, while metaxylem cells have a reticulate or pitted morphology. The five *Arabidopsis* Class III HOMEODOMAIN-LEUCINE ZIPPER (HD-ZIP III) TFs, i.e., *ATHB8*, *CORONA* (*CNA*), *PHABULOSA* (*PHB*), *PHAVOLUTA* (*PHV*), and *REVOLUTA* (*REV*), specify protoxylem and metaxylem vessel patterning in a combinatorial, dose-dependent manner by a microRNA (miRNA)-mediated transcript gradient (Carlsbecker et al., 2010; Miyashima et al., 2011). Disruption of this gradient by the production of high levels of miRNA-resistant *PHB* transcript throughout the root vasculature results in protoxylem cells mis-specified as metaxylem (Miyashima et al., 2011). Transcriptional regulation also determines the final steps of xylem cell differentiation, including the coordinated transcription of secondary cell wall biosynthetic enzymes. The VASCULAR-RELATED NAC-DOMAIN6 (VND6) and VND7 TFs act at the top of this hierarchy and are sufficient to specify xylem differentiation in *Arabidopsis* within and outside of the vascular cylinder (Kubo et al., 2005). While these studies in *Arabidopsis* have provided a framework for understanding xylem patterning in trees and maize (Dong et al., 2020; Ohtani et al., 2011; Robischon et al., 2011), little is known about the players in tomato root xylem. Given the importance of xylem to plant evolution, we formulated and tested a hypothesis that critical regulators of xylem differentiation are conserved between *Arabidopsis* and tomato.

The tomato genome encodes six HD-ZIPIII TFs (Data S1F). Two of these, *Solyc02g069830* (*SlPHB/PHV-LIKE1*) and *Solyc03g120910* (*SlCORONA-LIKE1*, *SlCNAL1*), coincide with *cis*-expression quantitative trait loci (eQTLs) in roots of an introgressed population between tomato and *Solanum pennellii* and are located within intervals significantly associated with natural variation in xylem cell number in the same population (Ron et al., 2013; Toal et al., 2018) (Table S3). In tomato, xylem cells are patterned similarly as in *Arabidopsis*, with the exception of two protoxylem vessel cells (as opposed to one in *Arabidopsis*) differentiating on either side of the central axis of the vascular cylinder (Figure 3B). We observed that *SlPHB/PHV-LIKE1* and *SlCNAL1* have high transcript levels in the tomato root vasculature and decreasing levels toward outer root tissues in the translatome data, similar to that found in *Arabidopsis* (Carlsbecker et al., 2010; Miyashima et al., 2011). If these genes regulate xylem differentiation similarly in *Arabidopsis* and tomato, we reasoned that constitutive levels of these transcripts would result in ectopic metaxylem specification. Constitutive expression of miRNA-resistant versions of *SlCNAL1* and *SlPHB/PHV-LIKE1* was indeed sufficient to regulate xylem vessel identity. Constitutive expression of *SlCNAL1* results in a change in xylem patterning from diarch to triarch (Figures 3C, 3D, and S3A; Table S3). Thus, these two HD-ZIPIII TFs regulate xylem patterning in a likely conserved manner between *Arabidopsis* and tomato.

We reasoned that a functional ortholog of *VND6* or *VND7* in tomato would (i) show transcript abundance in the xylem; and (ii) when overexpressed, be sufficient to drive ectopic xylem differentiation like in *Arabidopsis* (Endo et al., 2015; Kubo et al., 2005; Yamaguchi et al., 2008). Using phylogenetic analyses, we identified 2 genes (*Solyc06g034340*, *Solyc03g083880*) as potential orthologs of *AtVND6* and 2 genes (*Solyc06g065410*, *Solyc11g018660*) as potential orthologs of *AtVND7* in tomato (Figure S3B). Only 1 (*Solyc06 g034340*; Figure S3B) of these 4 genes was expressed in xylem and vascular translatomes; thus, we pursued experiments to determine whether it is a functional ortholog of *AtVND6*. As a control, we quantified secondary cell wall deposition in β-estradiol-*AtVND6*-inducible transgenic *Arabidopsis* plants (Coego et al., 2014). In parallel, we drove expression of *Solyc06g034340* under the near-constitutive *35S* promoter in tomato hairy roots. Similar hallmarks of ectopic xylem vessel differentiation were observed with overexpression of *AtVND6* and *Solyc06g034340* (Figures 3E–3G). We further confirmed, with a transcriptional reporter-GFP fusion, that the other putative *AtVND6* ortholog, *Solyc03g083880*, is not expressed in tomato root xylem cells (Figure S3C). From this combination of phylogenomic, translatome, and overexpression data, we conclude that *Solyc06g034340* is the most likely functional ortholog of *AtVND6* and assign it the name *SlVND6*. These experimental validations demonstrate likely

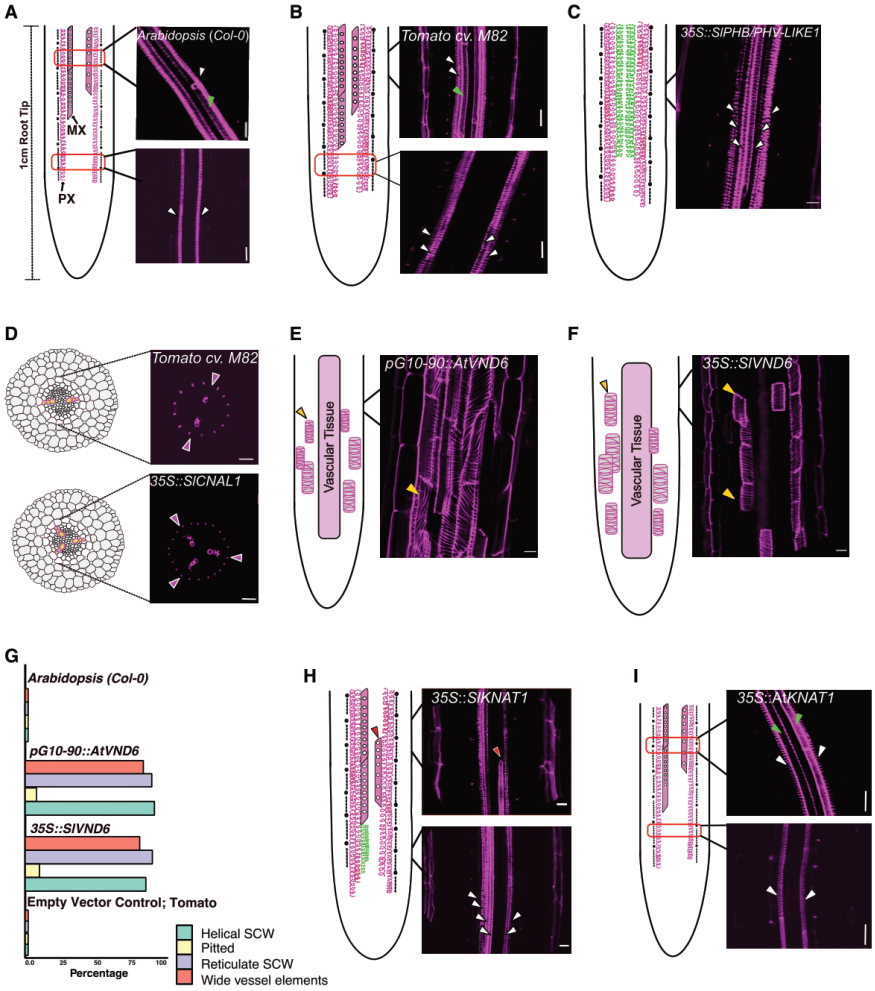


Figure 3. Identification of xylem vessel transcriptional regulators in tomato

Schematic and confocal images of basic fuchsin-stained roots of wild-type (WT) and overexpression lines.

(A) Xylem cell development in a WT *Arabidopsis* (Columbia-0 [Col-0]) root.

(B) Xylem cell development in a WT tomato root.

(C) *35S::SlPHB/PHV-LIKE*1 promotes protoxylem differentiation in the metaxylem position (green in schematic).

(D) Schematic and confocal images of a tomato root cross section in WT and *35S::SlCNAL1*; purple arrowheads mark the xylem axis with a diarch (top) and a triarch symmetry (bottom).

(E and F) *pG10-90::AtVND6* (E) and *35S::SlVND6* (F); yellow arrowheads indicate ectopic xylem cells.

(G) Frequency of ectopic xylem secondary cell wall (SCW) features in *VND6* overexpression lines in *Arabidopsis* primary roots and tomato hairy roots.

(H) *35S::SlKNAT1* with ectopic protoxylem strands (bottom image, green in schematic) and metaxylem break in continuity indicated with the red arrowhead.

(I) *35S::AtKNAT*1 with WT-like phenotype.

PX, protoxylem; MX, metaxylem. White arrowheads indicate protoxylem. Green arrowheads indicate metaxylem. Red boxes indicate zoomed-in region excluding the epidermis. Scale bars, 20 μm.

See also Figure S3 and Table S3.

conservation of VND6 function between *Arabidopsis* and tomato.

Repurposing of KNAT1 function for primary root xylem development in tomato

In a complementary approach, we set out to identify putative novel regulators within the tomato xylem regulatory network. *SlKNAT1*, a putative *AtKNAT1* ortholog (*Solyc04g077210*) (Data S1G), is a xylem CTEG (Table S1). However, *AtKNAT1* (*At4g08150*) is not expressed in *Arabidopsis* primary root xylem (Truernit et al., 2006). Instead, *AtKNAT1* regulates spatial boundaries within the *Arabidopsis* shoot meristem (Douglas et al., 2002) and inflorescence secondary cell wall biosynthesis (Woerlen et al., 2017). To test whether *SlKNAT1* regulates tomato root xylem development, we overexpressed *SlKNAT1* in tomato hairy roots and found it is sufficient to specify additional protoxylem cell files in the place of metaxylem and to cause breaks in metaxylem continuity (Figures 3H and S3A; Table S3). To determine whether *AtKNAT1* and *SlKNAT1* function is conserved in root development, we tested the effect of overexpression of *AtKNAT1* in *Arabidopsis* (Lincoln et al., 1994). No change in xylem patterning was observed (Figure 3I). The additional *KNAT1* expression domain in tomato and overexpression phenotype in comparison with *Arabidopsis* suggests that *SlKNAT1* function is repurposed (adapted for a different function) to control primary root xylem development.

Conservation and divergence of *cis*-regulation across CTEGs

In the case of cell type-enriched transcriptomes, transcript abundance is largely determined by the activity of TFs that bind to *cis*-regulatory motifs contained within gene upstream regulatory regions. To assess differences and similarities in factors that regulate transcription of CTEGs, we surveyed the promoters of CTEGs for enriched *cis*-regulatory motifs.

WRKY and basic-helix-loop-helix (bHLH) TFs are known to regulate *Arabidopsis* epidermal cell fate (Bernhardt et al., 2003; Rishmawi et al., 2014), and we correspondingly found their TF binding sites enriched in the promoters of epidermis-enriched genes (Figure 4A). MYB TFs play an important role in *Arabidopsis* xylem differentiation (Kim et al., 2014). We found that MYB domain binding sites were significantly over-represented in tomato xylem-enriched genes (false discovery rate [FDR] adj p value $\leq$ 0.01), demonstrating likely conservation in regulation of xylem development by MYB TFs between *Arabidopsis* and tomato. We also found highly significant MYB and bHLH TF binding site enrichment in the exodermis-enriched genes (p adj $\leq$ 0.01), which suggests that these factors are important in regulating exodermis development (Figure 4A). As *Arabidopsis* lacks an exodermis, this represents diversification of MYB and bHLH regulatory roles in tomato. Collectively, these cell type-enriched motifs suggest both conservation and divergence of TF-mediated regulation of cell type development.

Inference of cell type-unique regulatory networks reveals exodermis function

Motif enrichment within regulatory regions of CTEGs provides an excellent opportunity to infer cell type regulatory networks. We identified TF motifs from target gene promoters and complemented these datasets with nearby transposase hypersensitive sites (THSs; Figure S4; Table S4; STAR Methods). We combined these data with the previously discovered promoter motifs and filtered for motifs that were unique to a given set of CTEGs (Figures 4B and 4C; Table S4; STAR Methods). We next searched for the most likely tomato ortholog of the motif's cognate *Arabidopsis* TF (i.e., expressologs; STAR Methods) and included motifs only if its cognate expressolog was expressed at $\geq$1 transcript per million (TPM) in the translatome of a given cell type (STAR Methods). These networks were particularly informative in generating hypotheses regarding exodermal regulation.

Within the unique exodermis regulatory network, regulatory connections were inferred between exodermis-enriched regulatory sites and the expressolog of *AtMYB4* (Figure 4B), which is associated with lignin metabolism in *Arabidopsis* (Panda et al., 2020). Additionally, *cis*-regulatory motifs targeted by AtMYB41 are significantly over-represented in the exodermis-enriched gene set (Figure 4A); AtMYB41 is sufficient to ectopically induce suberin in *Arabidopsis* (Kosma et al., 2014). The tomato ortholog of *AtMYB41* (*Solyc02g079280*; Du et al., 2015) is also exodermis enriched (Figure 1F). In addition to the repurposing of this regulatory module to the exodermis, exodermis CTEGs provide insight into its function. Gene function enrichment analysis using MapMan Ontology terms (Table S1) of exodermis-enriched genes supports production of lignin and suberin in the exodermis. Exodermis-enriched genes have an over-representation for terms associated with lipid metabolism (including an enzyme responsible for the linkage of fatty acyl precursors to glycerol [*GPAT4*, *Solyc01g094700.1*] that makes up polyester compounds such as suberin; adj p = 0.11), phenylpropanoid biosynthesis (adj p = 0.01), and phenylpropanoid biosynthesis associated with lignin (adj p = 0.14). We previously observed secondary cell wall substances associated with the tomato exodermis that are either lignin, suberin, or callose (Brundrett et al., 1988; Ron et al., 2013). The ontology enrichments lead to our hypothesis that the exodermis is both lignified and suberized in tomato. We found lignin deposition in the exodermis in the first centimeter of the root (Figure 4D). The first suberized cells were detected at 4 cm distal from the root tip in the exodermis, but not in the endodermis (Figure 4E). These findings support co-option of lignin and suberin regulatory modules to the tomato exodermis. The finding of exclusive exodermal suberin suggests that, unlike *Arabidopsis*, tomato might rely primarily on non-endodermal (i.e., exodermal) suberin to control molecular diffusion in the root.

Lignin and suberin associated with endodermis differentiation regulate nutrient uptake in *Arabidopsis* (Barberon et al., 2016; Baxter et al., 2009). The exodermis lignin and suberin regulatory modules and presence (Figures 4D and 4E) suggest that the exodermis has an analogous function. Exodermis-enriched Gene/MapMan Ontology terms include nitrate reductase activity (p value = 0.03), transporter activity (p value = 0), as well as sugar and nutrient signaling (adj p = 0.14) (Table S1). Additionally, genes associated with nitrogen metabolism were detected in the unique exodermis network (Figure 4B; Table S4). The *Arabidopsis* root pericycle and lateral root cap as the most transcriptionally responsive to nitrogen (Gifford et al., 2008). However, regulatory motifs bound by nitrogen-associated AtNIN-LIKE PROTEIN 7

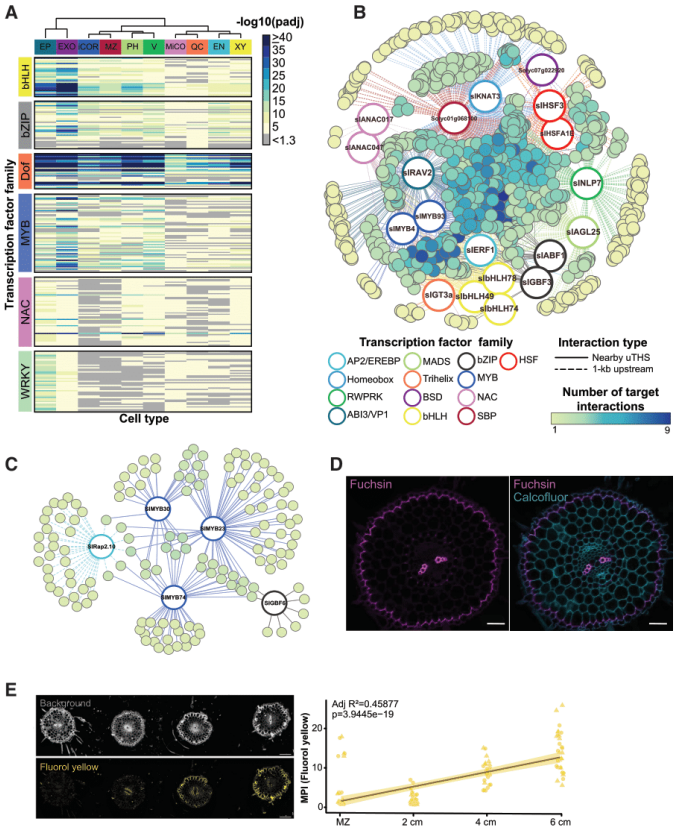


Figure 4. Inferred exodermis and inner cortex regulatory network and function

(A) Each row represents a transcription factor (TF) binding site (motif) significantly enriched within a 1-kb promoter region of at least 1 CTEG. Rows are grouped according to motifs associated with a given TF family. Each column represents the adjusted p value for that motif in a given CTEG. Significant adjusted (adj) p values ($-\log_{10}$) are indicated according to the heatmap scale. EP, epidermis; EN, endodermis; EXO, exodermis; iCOR, inner cortex; MiCO, meristematic inner cortex; MZ, meristematic zone; PH, phloem; V, vasculature; XY, xylem.

(B) Unique inferred exodermis regulatory network. Solid edges indicate motif-THS interaction; dashed edges indicate motif-1-kb upstream regulatory region interaction; large circles indicate TF expressolog for cognate TF motif; colored edges indicate TF family. Small circles indicate exodermis-enriched target genes that contain the motif in either the union THS (uTHS) or 1-kb upstream regulatory region; color scale indicates the number of target interactions.

(C) Unique inferred inner cortex network. Solid edges indicate motif-THS interaction; dashed edges indicate motif-1-kb upstream regulatory region interaction; large circles indicate TF expressolog for cognate TF motif; and colored edges indicate TF family. Small circles indicate inner cortex-enriched target genes that contain the motif in either the uTHS or 1-kb upstream regulatory region; color scale indicates the number of target interactions.

(D) Representative cross section taken from the middle of a 1-cm segment of the root tip. Cellulose is stained by calcofluor (blue), and lignin is stained by fuchsin (purple). Scale bar, 50 μm.

(E) Exodermal suberin deposition across the tomato primary root. (Top) Representative cross sections of root visualized with background autofluorescence, and suberin (stained by fluorol yellow). (Bottom) Fluorol yellow quantification of suberized exodermal cells in cross sections (3 cells/root section; 6 roots/position; n = 18). Scale bar, 100 μm. MPI, mean pixel intensity. Experiment was repeated twice, as indicated triangles and circles. Adjusted R-square (adj R^2) and p value (p) were calculated using a linear regression model and indicate a significant relationship between position and MPI signal of the plotted data.

See also Figure S6 and Table S4.

(NLP7) and AtRELATED TO ABI3/VP1 2 (RAV2) were over-represented in exodermis-enriched gene regulatory regions (Konishi and Yanagisawa, 2013; Li et al., 2020; Schommer et al., 2008). NLP7 is predicted to bind to promoters of genes associated with lignin biosynthesis and polymerization, while RAV2 is predicted to bind to the promoters of a nitrate reductase and amino acid transporter (Table S4) in the unique exodermis network. In addition, we found more expressolog overlap than expected by chance between transcriptional regulatory network genes of *Arabidopsis* nitrogen metabolism (Gaudinier et al., 2018) and the exodermis network (odds ratio = 2.8, $p < 0.01$) (Table S3). These data support repurposing of nitrogen regulation to the exodermis.

Similar to the exodermis, little is known regarding the function of inner cortex cells. The unique inner cortex regulatory network (Figure 4C) is defined by interconnection of several MYB TFs, as well as tomato homologs of RAP2.10 and GBF6. Their collective target genes are associated with primary metabolism and energy acquisition, supporting previous observations of cortex function in *Arabidopsis* (Table S4) (Brady et al., 2007).

The meristem translatome is more similar across multiple species

The functional and network analysis of cortical cell layers in tomato and its comparison with *Arabidopsis* provide insights on both the evolutionary conservation and divergence of cell type processes and regulation. To gain a more comprehensive understanding of gene expression conservation at cell population resolution, we utilized our translatome data for a systematic multi-species analysis (Figure 5A; STAR Methods). Comparative transcriptome studies of homologous tissues in vertebrates

CellPress
OPEN ACCESS

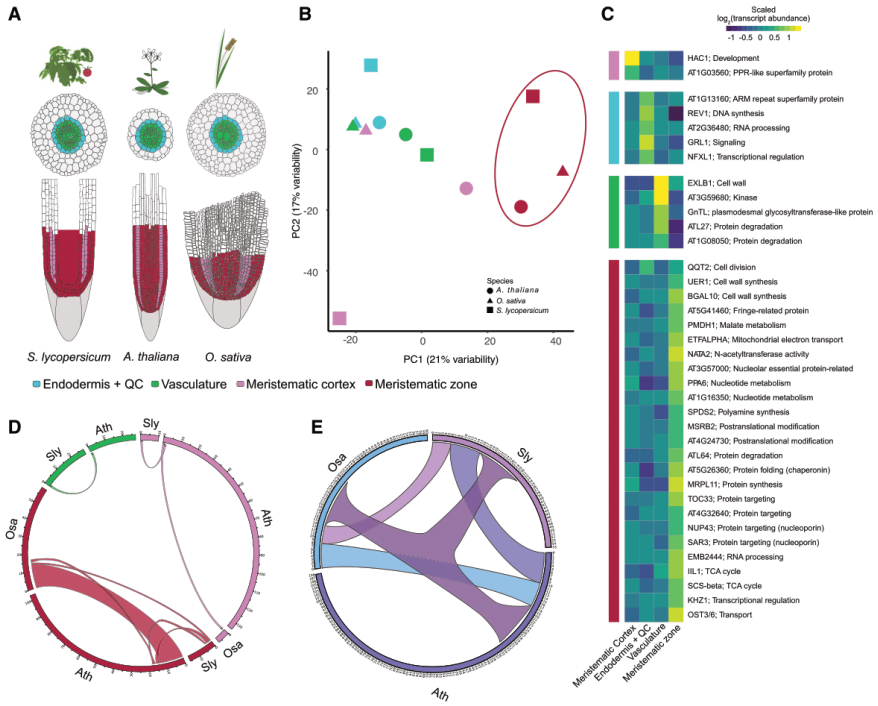


Figure 5. Homologous cell populations show limited conservation of gene expression

(A) Species and cell populations selected for comparative translatome analysis. Colors in legend are used throughout Figure 5.

(B) Grouping of cell population expression profiles between *Arabidopsis* (circle), rice (triangle), and tomato (square). Colors are cell populations as described in (A). Plot of principal component (PC) analysis of cell population expression of 2,642 1:1:1 orthologs.

(C) Thirty-seven conserved cell type/tissue-enriched expressologs. The mean expression of each consensus expressolog in tomato, *Arabidopsis*, and rice is presented for each cell population. Transcript abundance is scaled across the 4 cell populations. EN+QC, endodermis and quiescent center; MCO, meristematic cortex; MZ, meristematic zone; V, vasculature.

(D) Overlap of MapMan Ontology terms between homologous cell populations. The width of the ribbon is proportional to the number of common ontology terms. Ath, *Arabidopsis thaliana*; Osa, *Oryza sativa*; Sly, *Solanum lycopersicum*. The numbers in the circle represent the number of terms within each group.

(E) Overlaps of MapMan Ontology terms for constitutively expressed genes (CEGs). Color palette is chosen to maximally differentiate pairwise comparisons between species, and three-way overlap is shown in dark purple.

See also Figure S5 and Table S5.

demonstrate that gene expression data tend to cluster by homologous tissue rather than by species (Gilad and Mizrahi-Man, 2015). These studies suggest functional equivalency of these tissues and support the hypothesis that conserved gene regulatory networks drive homologous cell population identity in vertebrates where identity is largely determined by cell lineage. The similarity in root cell type patterning implies a similar phenomenon in plants. Here, we sought to define the degree of expression similarity and functional equivalency of homologous cell populations among three evolutionary distinct plant species. We generated and collected translatome profiles of the meristematic cortex, endodermis (which includes the QC), vasculature, and meristematic zone of tomato, *Arabidopsis* (Mustroph et al., 2009), and rice (Table S5; STAR Methods) as marked by similar promoter expression domains (Figures 5A, S1E, and S1F). Rice translatome data were confirmed to represent previously characterized cell type expression patterns (Table S5). To explore translatome similarities, we focused on 2,642 1:1:1 orthologs (Table S5; STAR Methods). PCA showed that the translatome profiles of the meristematic zone from all three species grouped together and were distinct from the other tissues (Figure 5B). This pattern was largely recapitulated by two additional independently derived orthology maps (Figures S5B and S5C; STAR Methods). Similarities between the endodermis and vasculature of *Arabidopsis* and tomato were supported by some of these additional orthology maps (Figures 5B, S5B, and S5C).

To find similarities between homologous tissues, we focused on genes with conserved cell type/tissue-enriched expression

among the three species. To this end, we constructed a fourth orthology map based on root TRAP-expressologs (Figure S5D; STAR Methods). Using a subset of 1,555 "consensus expressologs" (STAR Methods), we detected 37 "consensus expressologs" whose cell type or tissue-enriched expression is conserved across the 3 species (Figure 5C; Table S5; STAR Methods). In concordance with the PCA of the 1:1:1 orthologs presented in Figure 5B, 68% of these genes showed enriched expression in the meristematic zone. Among these genes is *QQT2*, a gene that is essential for correct cell divisions during embryogenesis (Lahmy et al., 2007) and required for the assembly of RNA polymerases II, IV, and V (Li et al., 2018). Additional conserved meristematic zone-enriched genes encode two nucleoporins, *SAR3* and *NUP43*, subunits of the nuclear pore complex, that regulate nucleocytoplasmic transport of protein and RNA and play important roles in hormone signaling and developmental processes (Parry, 2013). Genes associated with tricarboxylic acid (TCA) metabolism and cell wall biogenesis were also enriched in the meristematic zone. The *GLR1.1* glutamate receptor was enriched in the endodermis/QC. These conserved cell type/tissue-enriched genes provide an avenue for gene discovery with respect to cell type/tissue function.

Meristem functional equivalency between species

Detection of similarities based on orthology (i.e., the 1:1:1 orthologs or "consensus expressologs") limits the number of genes that could be assessed to those conserved among all three species. To circumvent this limitation, we applied a complementary approach to assess functional similarity of cell populations by identifying CTEGs within each species (Table S5) and assessing their function via Gene/MapMan Ontology enrichment (Table S5; STAR Methods). The meristematic zone of all three species was enriched for leucine-rich repeat receptor kinases (LRR-RKs; adj $p < 0.14$; Table S5), shown to regulate diverse signal transduction pathways including root development. At the level of individual genes, *RGF1 INSENSITIVE 5* (*RGI5*; Table S5) was enriched in tomato and rice meristem translatomes. In *Arabidopsis*, this gene is a receptor for root meristem growth factor 1 which, with additional LRR receptor-like kinases, is essential for meristem development (Ou et al., 2016). Despite its meristematic characteristics, the meristematic cortex demonstrates few-to-no overlaps of enriched terms and expressologs (Figures 5D, S5E, and S5F; Table S5). One explanation for this finding is due to differences in tissue composition between the species, meaning the variable number of cortical cell files in each species. The *Arabidopsis* cortex consists of a single cell layer, while in tomato (cv. M82) it consists of three (including the exodermis), and in rice, up to ten layers have been reported (Henry et al., 2016; Figure 5A). The endodermis of all three species is a single cell layer surrounding the vasculature and contains a lignified Casparian strip. Despite these similar morphological characteristics, only a limited number of endodermis-enriched genes and processes overlapped between the species, implying that, similar to vertebrates, distinct molecular programs can give rise to similar cellular morphologies and function (Alam et al., 2020).

The similarity observed in the translatome profiles of the meristematic zone (1:1:1 orthologs; Figure 5B) and the number of conserved meristematic-zone-enriched genes ("consensus expressologs"; Figure 5C) was further reflected in the relatively high overlap of meristematic zone-enriched ontology terms (Figures 5D and S5E; STAR Methods) across species. The lower similarity between the endodermis/QC, vasculature, and meristematic cortex of the three species (Figure 5B) was also reflected in the limited overlaps of expressologs and enriched gene functions (ontologies) (Figures S5E and S5F; Table S5; STAR Methods). Therefore, using these 3 complementary approaches to assess similarity among the 4 homologous cell populations, the root meristem consistently demonstrated higher functional conservation. These data suggest that, molecularly, the meristem is truly homologous and more evolutionarily conserved relative to the other cell populations examined. Similar observations have been made in animals, where embryonic tissues or early developmental stages of homologous cell types show higher similarity across species than mature cell types/tissues (Liang et al., 2018).

Constitutively expressed genes within each species have similar function

A comparative transcriptome study in mammals demonstrated that genes with low expression variation across tissues are enriched for housekeeping genes (Chen et al., 2019), which tend to evolve more slowly than tissue-specific genes (Zhang and Li, 2004). To test whether this observation is also true for plants, that is, that genes with low expression variation have housekeeping function, we identified a set of genes with minimal expression variation within each species, referred to as constitutively expressed genes (CEGs) (Table S5; STAR Methods). In concordance with the literature, overlapping ontology terms and expressologs between the CEGs were involved in housekeeping functions (e.g., cell division, chromatin remodeling, RNA binding, and protein metabolism) (Figures S5G–S5I; Table S5). In addition, a larger number of ontology terms overlapped between the CEGs (Figure 5E) compared with the CTEGs, even when considering only the meristematic-zone-enriched genes (odds ratio = 1.9, $p < 0.03$), suggesting that the expression patterns of CTEGs are more affected by speciation than CEGs.

DISCUSSION

Our integration of multiple cell type-resolution datasets sheds light for the first time as to how cell type molecular signatures in a single species change between *in vitro* culture conditions relative to their more natural soil environment. In several cases, our observations of CTEGs, their networks, or functions have led to the proposal of repurposing. A repurposed gene or network is one that has been adapted for a different function. For example, the endodermis is present in all vascular plants, while the exodermis occurs unevenly in the species studied thus far (Perumalla et al., 1990). This, along with the presence of MYB and bHLH site enrichment within the unique exodermis network, leads to our hypothesis of repurposed gene regulation in the exodermis. Nitrogen regulation may also be repurposed in the tomato exodermis. Our observation of exodermal nitrogen gene regulation is the first report of this molecular function for this cell type. However, nitrogen inducibility of exodermis differentiation has been observed in other species, suggesting that

nitrogen signaling also plays a role in exodermal differentiation (Armand et al., 2019; Namyslov et al., 2020; Schreiber et al., 2005). Our data also support repurposing of *SlKNAT1* function to the primary root xylem. This is based on the observation that *SlKNAT1* is present as a single tomato gene within a well-supported clade including *Arabidopsis KNAT1* (*AtKNAT1*), maize *KNOTTED-1*, and several of its homologs (Data S1G). The maize homologs and *AtKNAT1* are all expressed in shoot meristem tissue and in occasional vascular tissue within the meristem and inflorescence stem (Douglas et al., 2002; Jackson et al., 1994; Truernit et al., 2006; Woerlen et al., 2017). However, none are expressed in primary root xylem. Relative to *Arabidopsis* and maize, we posit that this repurposing is an invention in tomato. Single orthologs of *SlKNAT1* exist in potato, pepper, tobacco, petunia, coffee, and mimulus (Data S1G). Thus, it is possible that this repurposed root xylem function exists in the most recent common ancestor of this group.

Gene-by-gene functional validation of putative xylem cell regulators revealed examples of conservation (HD-ZIPIII TFs) and partial conservation (*VND6*, but not *VND7*) of known xylem patterning and differentiation genes between *Arabidopsis* and tomato. In the evolutionary context, this conservation and partial conservation has been observed in several tree species, and maize, and perhaps point to the critical importance of xylem to plant growth and development (Dong et al., 2020; Ohtani et al., 2011; Robischon et al., 2011). These collective observations of gene conservation and repurposing are supported by conservation and divergence in gene family member responses to submergence in tomato, *S. pennellii*, rice, and *Medicago* (Reynoso et al., 2019). By contrast, the partial conservation of *SlVND6* function to *AtVND6*, and lack of xylem/vascular expression of the At*VND7* ortholog in tomato, suggests that other genes likely contribute to xylem differentiation. It remains to be determined which genes contribute as well as their evolutionary context; however, our xylem-enriched genes provide an avenue for hypothesis generation.

Our multi-species analyses confirm that translation of research between *Arabidopsis* and other dicots or monocots is not straightforward. Extensive translatome similarity was observed between the root meristem of these divergent species, relative to other cell populations. The root meristem is a population of cells comprising the stem cell niche and proliferating or transit amplifying cells and thus represents a discrete location (the stem cell niche) and temporal period (proliferating cells). The meristem is morphologically recognizable across plant species, and our translatome data suggest that this cell population is more developmentally constrained than the others that we characterized. Our results suggest some similarities to the phylotypic period as observed in animals and plants (morphologically and molecularly similar), as well as a major difference in that it encompasses both developmental space and time, and not just a discrete stage of an organism. It is also intriguing to consider that this developmentally constrained stage is associated with root indeterminacy, a conserved property of root growth.

In the context of animal developmental biology, Davidson and Levin (2005) have previously discussed network "architecture" and its emergent properties that can only be appreciated at the higher order organizational level. They propose that the functions of a particular regulatory module within a network may not be understood by observing the individual genes within the module, but instead from the pattern that results from the aggregation of regulatory linkages associated with the network module. Examples of such aggregate patterns include the observation that homologous tissue transcriptomes of different vertebrate species are more similar to each other than to other tissues from the same species (Gilad and Mizrahi-Man, 2015). Additionally, early developmental stages of homologous animal tissues show higher gene expression correlation than mature tissues (Liang et al., 2018). We propose that these aggregate patterns observed in animals are recapitulated in plants for the root meristem, the earliest of 3 developmental stages. The translatome profiles of the meristem cluster together and are distinct compared with translatomes of other cell populations. Furthermore, we provide supporting evidence for this similarity in aggregate pattern with ontology terms. Thus, higher order organizational properties that determine similarities in the transcriptome or translatome of homologous tissues likely reside in the "architecture" of their associated networks in plants and animals. We also observed similarities in aggregate patterns for CEGs as reported for animals (Chan et al., 2009). Again, network "architecture" for housekeeping genes must ensure that these genes have low expression variation. In the future, identification of factors that give rise to this similarity could reveal deeply conserved mechanisms associated with the development of multicellular organisms. Finally, these data and resources serve as powerful tools for evaluating cell type processes relevant to breeding stress-resilient crops where such applications are limited.

Limitations of study

The 3 molecular signatures identified ("core" CTEGs and root-type-/condition-dependent genes) are potentially confounded with plant age. Further cell population profiling at each plant age and condition is needed to determine their consistency. Xylem developmental regulation is combinatorial and thus proof of conservation will require generation of higher order loss-of-function mutant alleles. Network conservation should be proven by a combination of chromatin immunoprecipitation and transcriptome profiling of TF mutants or inducible TF assays.

Our multi-species analyses are limited due to the confounding effect of the laboratories in which experiments were performed; differences in cell populations between species; differences in profiling methodologies; and differences in gene family expansions, orthology relationships, gene annotation, and ontologies. The rice root systems sampled include crown, lateral, and primary roots and their associated marked cell populations that have distinct root anatomy.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2021.04.024.

ACKNOWLEDGMENTS

We would like to thank H. Masson, R.M. Sanz, A. Gothberg, S. Vitales, J. Zheng, K. Zumstein, and B. Waring for experimental assistance; T. Demura, E. Kellogg, and K. Bubb for discussion; A. Hay for seed; and P. Benfey, N. Geldner, L. Strader, and J. Harada for manuscript critique. Funding was as follows: S.M.B. by HHMI 55108506; S.B.G. by NSF PGRP IOS-1306848; R.B.D., N.R.S., J.B.-S., and S.M.B. by NSF PGRP IOS-123824 and IOS-1856749; J.B.-S. and A.T.B. by NSF PGRP IOS-1810468; T.R.H. by the US-Norway Fulbright Foundation; K.K. by an SKR Postdoctoral Fellowship and MSCA RI Fellowship 790057; J.R.-M. by UC-MEXUS/CONACYT; A.T.B. by NSF DGE-1922642; D.J.K. by NSF MCB 1906486 and SDA NIFA Hatch CA-D-PLS-7033-H; V.L. by a Michael Smith Foreign Studies Supplement; V.L. and N.J.P. by an NSERC and OG-128; C.M. by MSCA Global Fellowship 655406; G.A.M. by NSF PGRP IOS-1907088; D.E.R. by USDA NIFA Hatch 1010469; L.S.-M. by BARD FI-570-2018; N.R.S. by NSF IOS-1558900; and B.P.M. by the PABGAP grant, funded by the UCOP/UC-HBCU Initiative.

AUTHOR CONTRIBUTIONS

Conceptualization, K.K., L.S.-M., G.A.M., G.P., M.R., D.A.W., S.B.G., N.R.S., D.E.R., J.B.-S., and S.M.B.; data curation, K.K., L.S.-M., G.A.M., J.R.-M., D.K., G.P., M.R., A.C.-P., C.M., K.M., and M.G.; formal analysis, K.K., L.S.-M., G.A.M., M.G., J.R.-M., D.K., A.C.-P., V.L., A.T.B., D.J.K., T.R.H., N.R.S., and D.E.R.; funding acquisition, K.K., L.S.-M., G.A.M., R.B.D., N.J.P., N.R.S., J.B.-S., and S.M.B.; investigation, K.K., L.S.-M., G.A.M., M.G., J.R.-M., D.K., G.P., M.R., V.L., M.A.S.A., D.A.W., C.M., S.B.G., A.I.Y., M.B., E.F., N.A.N., A.R., A.T.B., and S.M.B.; methodology, K.K., L.S.-M., G.A.M., J.R.-M., M.G., G.P., M.R., A.T.B., M.B., D.E.R., and S.M.B.; project administration, R.B.D., N.R.S., J.B.-S., and S.M.B.; software, L.S.-M., G.A.M., J.R.-M., A.C.-P., V.L., A.T.B., T.R.H., N.J.P., and D.E.R.; supervision, K.K., L.S.-M., G.A.M., J.R.-M., C.M., R.B.D., T.R.H., N.J.P., N.R.S., D.E.R., J.B.-S., and S.M.B.; validation, K.K., L.S.-M., G.A.M., M.G., J.R.-M., D.K., G.P., M.R., C.M., K.M. A.C.-P., M.A.S.A., D.E.R., and S.M.B.; visualization, K.K., L.S.-M., G.A.M., M.G., J.R.-M., D.K., G.P., M.R., V.L., M.A.S.A., D.W., C.M., N.J.P., and S.M.B.; writing – original draft, K.K., L.S.-M., G.A.M., M.G., J.R.-M., A.T.B., D.K., M.A.S.A., and S.M.B.; writing – review & editing, K.K., L.S.-M., G.A.M., M.G., J.R.-M., D.K., M.R., A.C.-P., M.A.S.A., D.W., C.M., A.T.B., R.B.D., D.J.K., N.R.S., D.E.R., J.B.-S., and S.M.B.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in science. One or more of the authors of this paper self-identifies as a member of the LGBTQ+ community. One or more of the authors of this paper received support from a program designed to increase minority representation in science. The author list of this paper includes contributors from the location where the research was conducted who participated in the data collection, design, analysis, and/or interpretation of the work.

Received: November 7, 2020
Revised: January 19, 2021
Accepted: April 14, 2021
Published: May 18, 2021

REFERENCES

Agarwal, P.K., Gupta, K., Lopato, S., and Agarwal, P. (2017). Dehydration responsive element binding transcription factors and their applications for the engineering of stress tolerance. J. Exp. Bot. *68*, 2135–2148.

Alam, T., Agrawal, S., Severin, J., Young, R.S., Andersson, R., Arner, E., Hasegawa, A., Lizio, M., Ramilowski, J.A., Abugessaisa, I., et al. (2020). Comparative transcriptomics of primary cells in vertebrates. Genome Res. *30*, 951–961.

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

Armand, T., Cullen, M., Boiziot, F., Li, L., and Fricke, W. (2019). Cortex cell hydraulic conductivity, endodermal apoplastic barriers and root hydraulics change in barley (Hordeum vulgare L.) in response to a low supply of N and P. Ann. Bot. *124*, 1091–1107.

Barberon, M., Vermeer, J.E.M., De Bellis, D., Wang, P., Naseer, S., Andersen, T.G., Humbel, B.M., Nawrath, C., Takano, J., Salt, D.E., and Geldner, N. (2016). Adaptation of Root Function by Nutrient-Induced Plasticity of Endodermal Differentiation. Cell *164*, 447–459.

Baxter, I., Hosmani, P.S., Rus, A., Lahner, B., Borevitz, J.O., Muthukumar, B., Mickelbart, M.V., Schreiber, L., Franke, R.B., and Salt, D.E. (2009). Root suberin forms an extracellular barrier that affects water relations and mineral nutrition in Arabidopsis. PLoS Genet. *5*, e1000492.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. B Stat. Methodol. *57*, 289–300.

Bernhardt, C., Lee, M.M., Gonzalez, A., Zhang, F., Lloyd, A., and Schiefelbein, J. (2003). The bHLH genes GLABRA3 (GL3) and ENHANCER OF GLABRA3 (EGL3) specify epidermal cell fate in the Arabidopsis root. Development *130*, 6431–6439.

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.

Bolger, A., Scossa, F., Bolger, M.E., Lanz, C., Maumus, F., Tohge, T., Quesneville, H., Alseekh, S., Sørensen, I., Lichtenstein, G., et al. (2014). The genome of the stress-tolerant wild tomato species Solanum pennellii. Nat. Genet. *46*, 1034–1038.

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.

Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. *34*, 525–527.

Brundrett, M.C., Enstone, D.E., and Peterson, C.A. (1988). A berberine-aniline blue fluorescent staining procedure for suberin, lignin, and callose in plant tissue. Protoplasma *146*, 133–142.

Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics *25*, 1972–1973.

Carlsbecker, A., Lee, J.-Y., Roberts, C.J., Dettmer, J., Lehesranta, S., Zhou, J., Lindgren, O., Moreno-Risueno, M.A., Vatén, A., Thitamadee, S., et al. (2010). Cell signalling by microRNA165/6 directs gene dose-dependent root cell fate. Nature *465*, 316–321.

Chan, E.T., Quon, G.T., Chua, G., Babak, T., Trochesset, M., Zirngibl, R.A., Aubin, J., Ratcliffe, M.J.H., Wilde, A., Brudno, M., et al. (2009). Conservation of core gene expression in vertebrate tissues. J. Biol. *8*, 33.

Chen, J., Swofford, R., Johnson, J., Cummings, B.B., Rogel, N., Lindblad-Toh, K., Haerty, W., Palma, F.D., and Regev, A. (2019). A quantitative framework for characterizing the evolutionary history of mammalian gene expression. Genome Res. *29*, 53–63.

Chung, W.-Y., Albert, R., Albert, I., Nekrutenko, A., and Makova, K.D. (2006). Rapid and asymmetric divergence of duplicate genes in the human gene coexpression network. BMC Bioinformatics *7*, 46.

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.

Cridge, A.G., Dearden, P.K., and Brownfield, L.R. (2016). Convergent occurrence of the developmental hourglass in plant and animal embryogenesis? Ann. Bot. *117*, 833–843.

Cruickshank, T., and Wade, M.J. (2008). Microevolutionary support for a developmental hourglass: gene expression patterns shape sequence variation and divergence in Drosophila. Evol. Dev. *10*, 583–590.

Davidson, E., and Levin, M. (2005). Gene regulatory networks. Proc. Natl. Acad. Sci. USA *102*, 4935.

de Mendiburu, F. (2019). agricolae: Statistical Procedures for Agricultural Research. R package version.

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.

Denyer, T., Ma, X., Klesen, S., Scacchi, E., Nieselt, K., and Timmermans, M.C.P. (2019). Spatiotemporal Developmental Trajectories in the Arabidopsis Root Revealed Using High-Throughput Single-Cell RNA Sequencing. Dev. Cell *48*, 840–852.e5.

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

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.

Dong, Z., Xu, Z., Xu, L., Galli, M., Gallavotti, A., Dooner, H.K., and Chuck, G. (2020). *Necrotic upper tips1* mimics heat and drought stress and encodes a protoxylem-specific transcription factor in maize. Proc. Natl. Acad. Sci. USA *117*, 20908–20919.

Douglas, S.J., Chuck, G., Dengler, R.E., Pelecanda, L., and Riggs, C.D. (2002). KNAT1 and ERECTA regulate inflorescence architecture in Arabidopsis. Plant Cell *14*, 547–558.

Drost, H.-G., Gabel, A., Grosse, I., and Quint, M. (2015). Evidence for active maintenance of phylotranscriptomic hourglass patterns in animal and plant embryogenesis. Mol. Biol. Evol. *32*, 1221–1231.

Du, H., Liang, Z., Zhao, S., Nan, M.-G., Tran, L.-S.P., Lu, K., Huang, Y.-B., and Li, J.-N. (2015). The Evolutionary History of R2R3-MYB Proteins Across 50 Eukaryotes: New Insights Into Subfamily Classification and Expansion. Sci. Rep. *5*, 11037.

Duboule, D. (1994). Temporal colinearity and the phylotypic progression: a basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. Dev. Suppl. *1994*, 135–142.

Durinck, S., Spellman, P.T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat. Protoc. *4*, 1184–1191.

Endo, H., Yamaguchi, M., Tamura, T., Nakano, Y., Nishikubo, N., Yoneda, A., Kato, K., Kubo, M., Kajita, S., Katayama, Y., et al. (2015). Multiple classes of transcription factors regulate the expression of VASCULAR-RELATED NAC-DOMAIN7, a master switch of xylem vessel differentiation. Plant Cell Physiol. *56*, 242–254.

Franco-Zorrilla, J.M., and Solano, R. (2017). Identification of plant transcription factor target sequences. Biochim. Biophys. Acta. Gene Regul. Mech. *1860*, 21–30.

Gaudinier, A., Rodriguez-Medina, J., Zhang, L., Olson, A., Liseron-Monfils, C., Bågman, A.-M., Foret, J., Abbitt, S., Tang, M., Li, B., et al. (2018). Transcriptional regulation of nitrogen-associated metabolism and growth. Nature *563*, 259–264.

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.

Gilad, Y., and Mizrahi-Man, O. (2015). A reanalysis of mouse ENCODE comparative gene expression data. F1000Res. *4*, 121.

Gu, X., Zhang, Z., and Huang, W. (2005). Rapid evolution of expression and regulatory divergences after yeast gene duplication. Proc. Natl. Acad. Sci. USA *102*, 707–712.

Hay, A., Barkoulas, M., and Tsiantis, M. (2006). ASYMMETRIC LEAVES1 and auxin activities converge to repress BREVIPEDICELLUS expression and promote leaf development in Arabidopsis. Development *133*, 3955–3961.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Henry, S., Divol, F., Bettembourg, M., Bureau, C., Guiderdoni, E., Périn, C., and Diévart, A. (2016). Immunoprofiling of Rice Root Cortex Reveals Two Cortical Subdomains. Front. Plant Sci. *6*, 1139.

Iyer-Pascuzzi, A.S., Jackson, T., Cui, H., Petricka, J.J., Busch, W., Tsukagoshi, H., and Benfey, P.N. (2011). Cell identity regulators link development and stress responses in the Arabidopsis root. Dev. Cell *21*, 770–782.

Jackson, D., Veit, B., and Hake, S. (1994). Expression of maize KNOTTED1 related homeobox genes in the shoot apical meristem predicts patterns of morphogenesis in the vegetative shoot. Development *120*, 405–413.

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.

Kadota, K., Ye, J., Nakai, Y., Terada, T., and Shimizu, K. (2006). ROKU: a novel method for identification of tissue-specific genes. BMC Bioinformatics *7*, 294.

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.

Kenrick, P., and Strullu-Derrien, C. (2014). The origin and early evolution of roots. Plant Physiol. *166*, 570–580.

Kim, W.-C., Kim, J.-Y., Ko, J.-H., Kang, H., and Han, K.-H. (2014). Identification of direct targets of transcription factor MYB46 provides insights into the transcriptional regulation of secondary wall biosynthesis. Plant Mol. Biol. *85*, 589–599.

Klie, S., and Nikoloski, Z. (2012). The Choice between MapMan and Gene Ontology for Automated Gene Function Prediction in Plant Science. Front. Genet. *3*, 115.

Konishi, M., and Yanagisawa, S. (2013). Arabidopsis NIN-like transcription factors have a central role in nitrate signalling. Nat. Commun. *4*, 1617.

Kosma, D.K., Murmu, J., Razeq, F.M., Santos, P., Bourgault, R., Molina, I., and Rowland, O. (2014). AtMYB41 activates ectopic suberin synthesis and assembly in multiple plant species and cell types. Plant J. *80*, 216–229.

Krueger, F. (2012). Trim Galore: a wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files, with some extra functionality for MspI-digested RRBS-type (Reduced Representation Bisufite-Seq) libraries. http://www.Bioinformatics.Babraham.Ac.Uk/projects/trim_galore/.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: an information aesthetic for comparative genomics. Genome Res. *19*, 1639–1645.

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.

Lahmy, S., Guilleminot, J., Schmit, A.-C., Pelletier, G., Chaboute, M.-E., and Devic, M. (2007). QQT proteins colocalize with microtubules and are essential for early embryo development in Arabidopsis. Plant J. *50*, 615–626.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559.

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

Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. *15*, R29.

Lee, M.-H., Jeon, H.S., Kim, S.H., Chung, J.H., Roppolo, D., Lee, H.-J., Cho, H.J., Tobimatsu, Y., Ralph, J., and Park, O.K. (2019). Lignin-based barrier restricts pathogens to the infection site and confers resistance in plants. EMBO J. *38*, e101948.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Li, L., Stoeckert, C.J., Jr., and Roos, D.S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. *13*, 2178–2189.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map (SAM) format and SAMtools. Bioinformatics *25*, 2078–2079.

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

Li, Y., Yuan, Y., Fang, X., Lu, X., Lian, B., Zhao, G., and Qi, Y. (2018). A Role for MINIYO and QUATRE-QUART2 in the Assembly of RNA Polymerases II, IV, and V in Arabidopsis. Plant Cell *30*, 466–480.

Li, X., Sanagi, M., Lu, Y., Nomura, Y., Stolze, S.C., Yasuda, S., Saijo, Y., Schulze, W.X., Feil, R., Stitt, M., et al. (2020). Protein Phosphorylation Dynamics Under Carbon/Nitrogen-Nutrient Stress and Identification of a Cell Death-Related Receptor-Like Kinase in Arabidopsis. Front. Plant Sci. *11*, 377.

Liang, C., Musser, J.M., Cloutier, A., Prum, R.O., and Wagner, G.P. (2018). Pervasive Correlated Evolution in Gene Expression Shapes Cell and Tissue Type Transcriptomes. Genome Biol. Evol. *10*, 538–552.

Lincoln, C., Long, J., Yamaguchi, J., Serikawa, K., and Hake, S. (1994). A knotted1-like homeobox gene in Arabidopsis is expressed in the vegetative meristem and dramatically alters leaf morphology when overexpressed in transgenic plants. Plant Cell *6*, 1859–1876.

Lux, A., Morita, S., Abe, J., and Ito, K. (2005). An improved method for clearing and staining free-hand sections and whole-mount samples. Ann. Bot. *96*, 989–996.

Madden, T. (2013). The BLAST Sequence Analysis Tool (National Center for Biotechnology Information).

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.

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

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

Mi, G., Di, Y., Emerson, S., Cumbie, J.S., and Chang, J.H. (2012). Length bias correction in gene ontology enrichment analysis using logistic regression. PLoS ONE *7*, e46128.

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

Mustroph, A., Zanetti, M.E., Jang, C.J.H., Holtan, H.E., Repetti, P.P., Galbraith, D.W., Girke, T., and Bailey-Serres, J. (2009). Profiling translatomes of discrete cell populations resolves altered cellular priorities during hypoxia in Arabidopsis. Proc. Natl. Acad. Sci. USA *106*, 18843–18848.

Mustroph, A., Barding, G.A., Jr., Kaiser, K.A., Larive, C.K., and Bailey-Serres, J. (2014). Characterization of distinct root and shoot responses to low-oxygen stress in Arabidopsis with a focus on primary C- and N-metabolism. Plant Cell Environ. *37*, 2366–2380.

Nakagawa, T., Suzuki, T., Murata, S., Nakamura, S., Hino, T., Maeo, K., Tabata, R., Kawai, T., Tanaka, K., Niwa, Y., et al. (2007). Improved Gateway binary vectors: high-performance vectors for creation of fusion constructs in transgenic analysis of plants. Biosci. Biotechnol. Biochem. *71*, 2095–2100.

Namyslov, J., Bauriedlová, Z., Janoušková, J., Soukup, A., and Tylová, E. (2020). Exodermis and Endodermis Respond to Nutrient Deficiency in Nutrient-Specific and Localized Manner. Plants *9*, 201.

Neph, S., Kuehn, M.S., Reynolds, A.P., Haugen, E., Thurman, R.E., Johnson, A.K., Rynes, E., Maurano, M.T., Vierstra, J., Thomas, S., et al. (2012). BEDOPS: high-performance genomic feature operations. Bioinformatics *28*, 1919–1920.

O'Malley, R.C., Huang, S.C., Song, L., Lewsey, M.G., Bartlett, A., Nery, J.R., Galli, M., Gallavotti, A., and Ecker, J.R. (2016). Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. Cell *166*, 1598.

Ogawa, D., Abe, K., Miyao, A., Kojima, M., Sakakibara, H., Mizutani, M., Morita, H., Toda, Y., Hobo, T., Sato, Y., et al. (2011). RSS1 regulates the cell cycle and maintains meristematic activity under stress conditions in rice. Nat. Commun. *2*, 278.

Ohtani, M., Nishikubo, N., Xu, B., Yamaguchi, M., Mitsuda, N., Goué, N., Shi, F., Ohme-Takagi, M., and Demura, T. (2011). A NAC domain protein family contributing to the regulation of wood formation in poplar. Plant J. *67*, 499–512.

Ou, Y., Lu, X., Zi, Q., Xun, Q., Zhang, J., Wu, Y., Shi, H., Wei, Z., Zhao, B., Zhang, X., et al. (2016). RGF1 INSENSITIVE 1 to 5, a group of LRR receptor-like

kinases, are essential for the perception of root meristem growth factor 1 in Arabidopsis thaliana. Cell Res. *26*, 686–698.

Panda, C., Li, X., Wager, A., Chen, H.-Y., and Li, X. (2020). An importin-beta-like protein mediates lignin-modification-induced dwarfism in Arabidopsis. Plant J. *102*, 1281–1293.

Parry, G. (2013). Assessing the function of the plant nuclear pore complex and the search for specificity. J. Exp. Bot. *64*, 833–845.

Patel, R.V., Nahal, H.K., Breit, R., and Provart, N.J. (2012). BAR expressolog identification: expression profile similarity ranking of homologous genes in plant species. Plant J. *71*, 1038–1050.

Pathan, M., Keerthikumar, S., Ang, C.-S., Gangoda, L., Quek, C.Y.J., Williamson, N.A., Mouradov, D., Sieber, O.M., Simpson, R.J., Salim, A., et al. (2015). FunRich: An open access standalone functional enrichment and interaction network analysis tool. Proteomics *15*, 2597–2601.

Perumalla, C.J., Peterson, C.A., and Enstone, D.E. (1990). A survey of angiosperm species to detect hypodermal Casparian bands. I. Roots with a uniseriate hypodermis and epidermis. Bot. J. Linn. Soc. *103*, 93–112.

Pickrell, J.K., Gaffney, D.J., Gilad, Y., and Pritchard, J.K. (2011). False positive peaks in ChIP-seq and other sequencing-based functional assays caused by unannotated high copy number regions. Bioinformatics *27*, 2144–2146.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS ONE *5*, e9490.

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

Quint, M., Drost, H.-G., Gabel, A., Ullrich, K.K., Bönn, M., and Grosse, I. (2012). A transcriptomic hourglass in plant embryogenesis. Nature *490*, 98–101.

Raff, R.A. (2012). The Shape of Life: Genes, Development, and the Evolution of Animal Form (University of Chicago Press).

Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. *42*, W187–W191.

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

Reynoso, M.A., Juntawong, P., Lancia, M., Blanco, F.A., Bailey-Serres, J., and Zanetti, M.E. (2015). Translating Ribosome Affinity Purification (TRAP) followed by RNA sequencing technology (TRAP-SEQ) for quantitative assessment of plant translatomes. Methods Mol. Biol. *1284*, 185–207.

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.

Reynoso, M.A., Kajala, K., Bajic, M., West, D.A., Pauluzzi, G., Yao, A.I., Hatch, K., Zumstein, K., Woodhouse, M., Rodriguez-Medina, J., et al. (2019). Evolutionary flexibility in flooding response circuitry in angiosperms. Science *365*, 1291–1295.

Rishmawi, L., Pesch, M., Juengst, C., Schauss, A.C., Schrader, A., and Hülskamp, M. (2014). Non-cell-autonomous regulation of root hair patterning genes by WRKY75 in Arabidopsis. Plant Physiol. *165*, 186–195.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

Robbins, N.E., 2nd, Trontin, C., Duan, L., and Dinneny, J.R. (2014). Beyond the barrier: communication in the root through the endodermis. Plant Physiol. *166*, 551–559.

Robischon, M., Du, J., Miura, E., and Groover, A. (2011). The Populus class III HD ZIP, popREVOLUTA, influences cambium initiation and patterning of woody stems. Plant Physiol. *155*, 1214–1225.

Rokas, A. (2011). Phylogenetic analysis of protein sequence data using the Randomized Axelerated Maximum Likelihood (RAXML) Program. Curr. Protoc. Mol. Biol. *19*, Unit19.11.

Ron, M., Dorrity, M.W., de Lucas, M., Toal, T., Hernandez, R.I., Little, S.A., Maloof, J.N., Kliebenstein, D.J., and Brady, S.M. (2013). Identification of novel loci regulating interspecific variation in root morphology and cellular development in tomato. Plant Physiol. *162*, 755–768.

Ron, M., Kajala, K., Pauluzzi, G., Wang, D., Reynoso, M.A., Zumstein, K., Garcha, J., Winte, S., Masson, H., Inagaki, S., et al. (2014). Hairy root transformation using Agrobacterium rhizogenes as a tool for exploring cell type-specific gene expression and function using tomato as a model. Plant Physiol. *166*, 455–469.

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.

Sallaud, C., Meynard, D., van Boxtel, J., Gay, C., Bès, M., Brizard, J.P., Larmande, P., Ortega, D., Raynal, M., Portefaix, M., et al. (2003). Highly efficient production and characterization of T-DNA plants for rice (Oryza sativa L.) functional genomics. Theor. Appl. Genet. *106*, 1396–1408.

Schommer, C., Palatnik, J.F., Aggarwal, P., Chételat, A., Cubas, P., Farmer, E.E., Nath, U., and Weigel, D. (2008). Control of jasmonate biosynthesis and senescence by miR319 targets. PLoS Biol. *6*, e230.

Schreiber, L., Franke, R., and Hartmann, K. (2005). Effects of NO 3 deficiency and NaCl stress on suberin deposition in rhizo- and hypodermal (RHCW) and endodermal cell walls (ECW) of castor bean (Ricinus communis L.) roots. Plant Soil *269*, 333–339.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.

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.

Smith, J.M., Burian, R., Kauffman, S., Alberch, P., Campbell, J., Goodwin, B., Lande, R., Raup, D., and Wolpert, L. (1985). Developmental Constraints and Evolution: A Perspective from the Mountain Lake Conference on Development and Evolution. Q. Rev. Biol. *60*, 265–287.

Sonnhammer, E.L.L., and Östlund, G. (2015). InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. Nucleic Acids Res. *43*, D234–D239.

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

Thimm, O., Bläsing, O., Gibon, Y., Nagel, A., Meyer, S., Krüger, P., Selbig, J., Müller, L.A., Rhee, S.Y., and Stitt, M. (2004). MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J. *37*, 914–939.

Toal, T.W., Ron, M., Gibson, D., Kajala, K., Splitt, B., Johnson, L.S., Miller, N.D., Slovak, R., Gaudinier, A., Patel, R., et al. (2018). Regulation of Root Angle and Gravitropism. G3 (Bethesda) *8*, 3841–3855.

Tomato Genome Consortium (2012). The tomato genome sequence provides insights into fleshy fruit evolution. Nature *485*, 635–641.

Townsley, B.T., Covington, M.F., Ichihashi, Y., Zumstein, K., and Sinha, N.R. (2015). BrAD-seq: Breath Adapter Directional sequencing: a streamlined, ultra-simple and fast library preparation protocol for strand specific mRNA library construction. Front. Plant Sci. *6*, 366.

Truernit, E., Siemering, K.R., Hodge, S., Grbic, V., and Haseloff, J. (2006). A map of KNAT gene expression in the Arabidopsis root. Plant Mol. Biol. *60*, 1–20.

Turco, G.M., Rodriguez-Medina, J., Siebert, S., Han, D., Valderrama-Gómez, M.Á., 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.e4.

Turner, S. (2012). Faculty Opinions recommendation of [Dobin A et al., Bioinformatics 2013 29(1):15–21]. In Faculty Opinions, 09 Nov 2012; 10.3410/f.717961569.793464455.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., and Rozen, S.G. (2012). Primer3–new capabilities and interfaces. Nucleic Acids Res. *40*, e115.

Urbanczyk-Wochniak, E., Usadel, B., Thimm, O., Nunes-Nesi, A., Carrari, F., Davy, M., Bläsing, O., Kowalczyk, M., Weicht, D., Polinceusz, A., et al. (2006). Conversion of MapMan to allow the analysis of transcript data from Solanaceous species: effects of genetic and environmental alterations in energy metabolism in the leaf. Plant Mol. Biol. *60*, 773–792.

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

Wagner, G.P., Kin, K., and Lynch, V.J. (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory Biosci. *131*, 281–285.

Weirauch, M.T., Yang, A., Albu, M., Cote, A.G., Montenegro-Montero, A., Drewe, P., Najafabadi, H.S., Lambert, S.A., Mann, I., Cook, K., et al. (2014). Determination and inference of eukaryotic transcription factor sequence specificity. Cell *158*, 1431–1443.

Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis (Springer Science & Business Media).

Woerlen, N., Allam, G., Popescu, A., Corrigan, L., Pautot, V., and Hepworth, S.R. (2017). Repression of BLADE-ON-PETIOLE genes by KNOX homeodomain protein BREVIPEDICELLUS is essential for differentiation of secondary xylem in Arabidopsis root. Planta *245*, 1079–1090.

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.

Young, M.D., Wakefield, M.J., Smyth, G.K., and Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol. *11*, R14.

Zhang, L., and Li, W.-H. (2004). Mammalian housekeeping genes evolve more slowly than tissue-specific genes. Mol. Biol. Evol. *21*, 236–239.

Zhao, D., Hamilton, J.P., Hardigan, M., Yin, D., He, T., Vaillancourt, B., Reynoso, M., Pauluzzi, G., Funkhouser, S., Cui, Y., et al. (2017). Analysis of Ribosome-Associated mRNAs in Rice Reveals the Importance of Transcript Size and GC Content in Translation. G3 (Bethesda) *7*, 203–219.

CellPress
OPEN ACCESS

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
Rhizobium rhizogenes	American Type Culture Collection	American Type Culture Collection Strain: 15834
Agrobacterium tumefaciens	Plant Transformation Facility, UC Davis	Strain GV3101
Chemicals, peptides, and recombinant proteins		
Monoclonal ANTI-FLAG® M2 antibody produced in mouse	Sigma-Aldrich	Catalog# F1804; RRID: AB_262044
Dynabeads Protein G for immunoprecipitation	Thermo Fisher Scientific	Catalog# 1003D
β–estradiol	Sigma-Aldrich	SKU# E8875
Fluorol yellow	Santa Cruz Biotech.	Catalog# sc-215052
Basic Fuchsin	Fisher Scientific	Catalog# 632-99-5
Calcofluor White	Sigma-Aldrich	SKU# 18909
Critical commercial assays		
Nextera DNA library kit	Illumina	Catalog# FC-121-1030
pENTR-D/TOPO cloning kit	Thermo-Fisher	Catalog# K240020
LR Clonase II Enzyme mix	Thermo-Fisher	Catalog# 11791020
QuikChange II XL Site-Directed Mutagenesis Kit	Agilent	Catalog# 200522
Deposited data		
Raw and analyzed tomato data	This study	NCBI: GSE149217
Arabidopsis TRAP data	Mustroph et al., 2009	NCBI: GSE14493
Rice TRAP data	This study	NCBI: GSE149217
Experimental models: Organisms/strains		
S. lycopersicum: AtWER TRAP	This study	Line EP-TR-7
S. lycopersicum: SlPEP TRAP	This study	Lines EXO-TR-10-22, EXO-TR-10-22-2
S. lycopersicum: AtPEP TRAP	This study	Lines COR-TR-2, COR-TR-2-4, COR-TR-6
S. lycopersicum: SlCO2 TRAP	This study	Lines MCO-TR-4, MCO-TR-4-1
S. lycopersicum: SlSCR TRAP	This study	Line EN-TR-3
S. lycopersicum: SlSHR TRAP	This study	Lines V-TR-13, V-TR-13-1
S. lycopersicum: AtS32 TRAP	This study	Line PH-TR-3
S. lycopersicum: AtS18 TRAP	This study	Line XY-TR-1
S. lycopersicum: SlWOX5 TRAP	This study	Line WOX-TR-6
S. lycopersicum: SlRPL11C TRAP	This study	Line MZ-TR-8
*S. lycopersicum: 35S*TRAP	This study	Line 35S-TR-5, 35S-TR-5-2
S. lycopersicum: SlACT2 TRAP	This study	Line ACT-TR-2
A. thaliana: 35S TRAP	Mustroph et al., 2009	NCBI: GSE14493
A. thaliana: AtRPL11C TRAP	Mustroph et al., 2009	NCBI: GSE14493
A. thaliana:AtCO2 TRAP	Mustroph et al., 2009	NCBI: GSE14493
A. thaliana: AtSCR TRAP	Mustroph et al., 2009	NCBI: GSE14493
A. thaliana: AtSHR TRAP	Mustroph et al., 2009	NCBI: GSE14493
O. sativa: 35S TRAP	This study	Line TRAP_C_3
O. sativa: OsRSS1 TRAP	This study	Line 57_26
O. sativa: OsCMZ TRAP	This study	Lines 66_6_2, 66_2_4

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
O. sativa: AtSCR TRAP	This study	Line 46_19_2
O. sativa: OsSHR1 TRAP	This study	Lines 24_5_21, 24_5_22, 24_5_23
S. lycopersicum: AtWER INTACT	This study	Line EP-IN-7
S. lycopersicum: SlPEP INTACT	This study	Line EXO-IN-6
S. lycopersicum: AtPEP INTACT	This study	Line COR-IN-1
S. lycopersicum: SlCO2 INTACT	This study	Lines MCO-IN-3, MCO-IN-3-12
S. lycopersicum: SlSCR INTACT	This study	Lines EN-IN-7, EN-IN-7-1
S. lycopersicum: SlSHR INTACT	This study	Line V-IN-7
S. lycopersicum: AtS32 INTACT	This study	Line PH-IN-8
S. lycopersicum: AtS18 INTACT	This study	Lines XY-IN-1, XY-IN-1-3
S. lycopersicum: SlWOX5 INTACT	This study	Line WOX-IN-6
S. lycopersicum: SlRPL11C INTACT	This study	Line MZ-IN-10
S. lycopersicum: 35S INTACT	This study	Lines 35S-IN-1, 35S-IN-1-4
S. lycopersicum: 35S:SlVND6	This study	Line 35S:SlVND6
S. lycopersicum: 35S:SlPHB/PHV-Like1	This study	Line 35S:SlPHB/PHV-Like1
S. lycopersicum: 35S:SlCNAL1	This study	Line 35S:SlCNAL1
S. lycopersicum: 35S:SlKNAT1	This study	Line 35S:SlKNAT1
A. thaliana: β–estradiol-inducible *VND6*	TRANSPLANTA; Coego et al., 2014	ABRC: stock #CS2102542
A. thaliana: 35S:atKNAT1	Hay et al., 2006	*35S:atKNAT1*
Oligonucleotides		
See Table S4	This study	
Recombinant DNA		
Plasmid: pK7WG-TRAP	Ron et al., 2014	VIB-UGent: Vector ID: 6_26
Plasmid: pK7WG-INTACT-SI	Ron et al., 2014	VIB-UGent: Vector ID: 6_25
Plasmid: pH7WG	VIB-UGent	VIB-UGent: Vector_ID:4_40
Plasmid: pMR074	Ron et al., 2014	N/A
Plasmid: pMR099	Ron et al., 2014	N/A
PGWB417	Addgene; Nakagawa et al., 2007	Addgene: Stock #74811
CDS Synthesized and cloned into pENTR by TwistBioSciences	This study	Solyc08g079120
CDS Synthesized and cloned into pENTR by TwistBioSciences	This study	Solyc03g120910
Software and algorithms		
Code used is freely available on github	This study	https://github.com/plant-plasticity/tomato-root-atlas-2020
Primer3Plus software	Untergasser et al., 2012	http://www.primer3plus.com/
FunRich tool v3.1.3	Pathan et al., 2015	www.funrich.org

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Siobhan M. Brady (sbrady@ucdavis.edu).

Materials availability

- Plasmids generated in this study are available upon request with completion of an MTA for third-party components.
- Seed lines generated in this study are available upon request with completion of appropriate governmental regulatory paperwork and a fee to cover the cost of seed bulking and phytosanitary certificate acquisition.
- This study did not generate new unique reagents.

Data and code availability

- The accession number for the raw TRAP-Seq libraries and genomic DNA-based ATAC-seq libraries reported in this paper is NCBI: GSE149217.
- Code used to generate and analyze all datasets during this study is available at https://github.com/plant-plasticity/tomato-root-atlas-2020
- Tomato translatome abundance data can be viewed on a gene-by-gene basis for the 11 cell populations (Root eFP) and for the field or pot data (Root Field Pot eFP) at http://bar.utoronto.ca/eplant_tomato/ by clicking on the "Tissue and Experiment eFP Viewers."
- Rice translatome abundance can be viewed on a gene-by-gene basis for the rice cell populations (Root eFP) at http://bar.utoronto.ca/eplant_rice/ by clicking on the "Tissue and Experiment eFP Viewers."

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Tomato material and growth conditions

Transgenic INTACT (isolation of nuclei tagged in cell types) and TRAP marker lines of *Solanum lycopersicum* cultivar M82 (LA3475) were generated by *Agrobacterium tumefaciens* transformation at the UC Davis Plant Transformation Facility. The pK7WG-TRAP and pK7WG-INTACT-SI binary vectors (Ron et al., 2014; https://gatewayvectors.vib.be) were used with a range of promoters to drive the expression of either the nuclear tagging fusion (NTF; *WPP-GFP-BLRP*) for INTACT or the polysome tag (*His6-FLAG-RPL18-GFP*) for TRAP. The promoters used were the previously published *SlACT2, 35S, SlRPL11C, AtWER, AtPEP, SlCO2, SlSCR, SlSHR, SlWOX5, AtS18* and *AtS32* (Ron et al., 2014), and *SlPEP* (*Solyc04g076190*) amplified using CACCTTCTCCAACAACGTAGAAGCTCCTCGCT and GGTGTGCTTTTTCCTTATCAACAAC. The promoters were recombined into pENTR-D/TOPO (Invitrogen) and introduced into pK7WG-TRAP and pK7WG-INTACT-SI vectors using LR Clonase II Enzyme mix (Invitrogen). In order to visually confirm cell type specificity, the expression patterns of all the promoters driving the GFP-containing INTACT and TRAP tags in tomato (Figure S1; Table S1) were imaged using an LSM 700 laser scanning microscope (Carl Zeiss) with the following settings: 488-nm excitation laser, the preset eGFP emission spectrum, 70% laser power, 1.87-Airy unit pinhole and gain optimized to the signal strength (450-1200). Additionally, the 561-nm laser and the preset RFP emission spectrum were used to capture autofluorescence.

The nuclear and translating ribosome affinity purification experiments were conducted with T1 seed stocks (and T2 as needed) from one independent line per construct (line IDs listed in Table S1). Plate-based experiments were conducted with four independent replicates of each line, and for each replicate, 1 cm of primary root tips were pooled from up to 200 seedlings. The seeds were surface sterilized with 3% hypochlorite (Clorox) for 20 minutes and rinsed three times with sterile water. Seven seeds were planted per 12 cm x 12 cm square plate containing 1x MS without vitamins (Caisson), 1% (w/v) sucrose, 0.5 g/L MES, pH = 5.8 and 1% (w/v) agar (Difco). Plates were placed vertically into racks using a completely random design in a growth chamber with a 16:8 light:dark cycle at 25°C and 50%–75% humidity with a light intensity of 55-75 μE. As tomato germination is uneven, the germination day of each seedling was scored and 1 cm of root tip was harvested from 3-5 days after germination (Figure S1D). The tissue was harvested at relative noon and placed immediately into liquid nitrogen.

Experiments with 1-month-old plants were conducted as follows. Transgenic seeds (Table S2) were surface sterilized and germinated on 1xMS media as described above, with the addition of 200 μg/ml kanamycin to screen for the presence of the transgenic construct. After 7 days, seedlings were transplanted into pots with Turface Athletic Profile Field & Fairway clay substrate (Turface Athletics) that was pre-wetted with a nutrient water solution containing 4% nitrogen, 18% phosphoric acid, and 38% soluble potash. Plants were grown in a completely randomized design for 31 days in a Conviron Growth Chamber at 22°C, 70% RH, 16/8 hour light/dark cycle and light intensity of 150-200 $\mu mol/m^2/s$. The root systems were harvested as close to relative noon as feasible (±2h) by immersing the pot into cool water, massaging the rootball free, rinsing three times sequentially with water, and then dissecting the root tissues and flash-freezing with liquid nitrogen. The harvested tissues were the lateral roots at the depth of 6-12 cm, and the shoot-borne (hypocotyl-derived) roots (Figure S1D).

Tomato plants were grown in the field as follows: transgenic seeds (Table S2) were surface sterilized and germinated on 1xMS media as described above, and the root tips were dissected for microscopy-based screening for the correct GFP pattern. The remaining seedlings were transplanted on soil and grown in a growth chamber with a 16:8 light:dark cycle at 25°C and 50%–75% humidity with a light intensity of 55-75 μE for one week. The plants were then transferred into a screen house for two weeks prior to transplanting into the field in Davis, California, USA on August 25, 2016 in a randomized block design with six replicate blocks, each block consisting of five plants of each genotype. Plants were grown in the field for 32 days with furrow irrigation once weekly with biweekly removal of flower buds to follow the local genetic modification guidelines. The root systems were harvested by digging the plant out, immersing the root ball with soil into water, massaging the rootball free, and three sequential water rinses prior to flash-freezing the entire root ball with liquid nitrogen (Figure S1D).

Rice material and growth conditions

Transgenic marker lines of rice (*Oryza sativa* cv. Nipponbare) were generated by *Agrobacterium tumefaciens* transformation as described by Sallaud et al. (2003) or at the UC Davis Plant Transformation Facility. The Rice TRAP binary vector was constructed

as described by Ron et al. (2014) using the Gateway binary vector pH7WG, for hygromycin resistance, as a backbone instead of pK7WG (https://gatewayvectors.vib.be) and incorporating rice *OsRPL18-2* as described in Zhao et al. (2017). Promoters were incorporated by LR recombination as performed for *S.lycopersicum* constructs to drive the expression of *His6-FLAG-RPL18-GFP* for TRAP. The promoters used were the previously published *35S* (Ron et al., 2014), *AtSCR* (Mustroph et al., 2009), *OsRSS1* (Ogawa et al., 2011), as well as *OsCMZ (Os01g0957100)* and *OsSHR1 (Os07g0586900)*. In order to visually confirm cell type specificity, the expression patterns of all the promoters driving the GFP-containing TRAP tags were imaged (Figure S1F) using a Leica SP5 laser scanning microscope (Leica) with a 488-nm excitation laser at 50% power, 56.7 μm pinhole, the preset eGFP emission, and Smart Gain 650-1100. Additionally, brightfield images were captured to show localization of GFP within the root.

Rice (*Oryza sativa* cv. Nipponbare) seeds from transgenic lines (Table S5) were dehulled and surface sterilized with 3% hypochlorite (Clorox) for 30 min and then rinsed with sterile distilled water. Seedlings were grown on plates (10 cm x 10 cm) containing half-strength Murashige and Skoog standard medium (MS) agar (1% w/v) and 1% w/v sucrose, for 7 days in a growth chamber (16 h day / 8 h night; at 28°C/25°C day/night; 110 $\mu Em^{-2}s^{-1}$). The whole root system was placed immediately into liquid nitrogen upon harvesting.

Arabidopsis material and growth conditions

Arabidopsis (Col-0) and the β–estradiol-inducible *VND6* line (ABRC: stock # CS2102542) seeds were sterilized in 50% bleach(V/V) for 10 minutes and then stored at 4°C for 3 days. Sterilized seeds were germinated on nylon mesh (100 μM) on MS Petri dish plates and grown at 22°C in a 12 hr light cycle chamber. After 7 days of growth, plants were transferred to MS plates containing 20 μM estradiol and grown for an additional 24 hours for induction. Whole root samples from Col-0 and the inducible line were then sampled and transferred to ClearSee buffer for clearing. 35S:AtKNAT1 line was sterilized and germinated as above without the induction steps and imaged after 7 days.

METHOD DETAILS

TRAP & RNA-seq libraries

These steps were conducted as described in Reynoso et al. (2019) (https://github.com/plant-plasticity/tomato-root-atlas-2020/tree/master/Protocols). In brief, cell type-specific ribosome-associated mRNAs were isolated from the frozen root tip material using TRAP (Reynoso et al., 2015, 2018, 2019; Ron et al., 2014; Zhao et al., 2017) and mRNA was isolated from the ribosome complexes for non-strand specific random primer-primed RNA-seq library construction (Townsley et al., 2015). Barcoded libraries were pooled together and sequenced on the Illumina HiSeq 4000 at the UC Davis DNA Technologies Core to obtain 50-bp reads.

Transcriptional reporter construction and imaging

Promoters of the exodermis-enriched *WRKY* (*Solyc02g071130*) and *MYB* (*Solyc02g079280*) and putative VND6 ortholog (*Solyc03g083880)* TFs were cloned from *Solanum lycopersicum* cultivar M82 genomic DNA. Cloning primers were designed to amplify 2,130 bp, 3,408 bp and 2,101 bp upstream of the translational start site of WRKY, MYB and VND6, respectively, using the tomato reference genome annotation ITAG3.2 (https://solgenomics.net). The promoters were amplified from genomic DNA using Phusion DNA polymerase (New England Biolabs). Amplified fragments were cloned into pENTR5′TOPO (Invitrogen) and sequences were confirmed by Sanger sequencing. LR Clonase II Enzyme mix (Invitrogen) was used to clone the promoters upstream of a *nlsGFP-GUS* reporter gene fusion in the binary vector pMR074 (MYB and WRKY) and pMR99 (VND6) (Ron et al., 2014) which also contains a ubiquitously expressing plasma membrane marker TagRFP-LTI6b. The binary vectors were used for hairy root (*Rhizobium rhizogenes*) transformation as described below. Transgenic hairy root fluorescence was visualized using Confocal Laser Scanning Microscopy with a Zeiss Observer Z1 LSM700 (Zeiss) microscope (water immersion, × 20 objective) with excitation at 488 nm and emission at 493–550 nm for GFP and excitation at 555 nm and emission at 560–800 nm for mRFP. Images were taken at approximately 1 cm from the root tip.

Overexpression construct design and cloning

The coding sequence (CDS) for target genes was obtained from the Sol Genomics database (https://solgenomics.net - ITAG3.2). CDS were amplified from tomato (*Solanum lycopersicum* cv. M82) cDNA. In brief, total RNA was isolated from 50 mg of tomato root tissue using the Zymo-Direct-Zol RNA Miniprep Plus Kit (Zymo Research- catalog#R2071) according to manufacturer's instructions and treated with RNase-Free DNase (1unit/10μl). 1μg of DNase-treated RNA was reverse-transcribed into cDNA using oligo(dT) primers and SuperScript III Reverse Transcriptase (SuperScript III First-Strand Synthesis System; Invitrogen) per kit instructions. Cloning primers were designed to PCR amplify the CDS without the stop codon. PCR products were purified from the agarose gel (QIAquick Gel Extraction kit; Catalog#28704) for subsequent recombination and cloning.

Purified cDNAs were introduced into the pENTR/D-Topo vector (Invitrogen). The resulting pENTR plasmids were then LR recombined (LR Clonase II Enzyme mix; Invitrogen) into the pGWB417 binary destination vector (Addgene plasmid #74811; http://addgene.org/74811; RRID:Addgene_74811) containing a 35S promoter driving the expression of the CDS. All constructs were confirmed by Sanger sequencing.

Site directed mutagenesis for miRNA resistant HD-ZIPIII TF constructs

A point mutation causing a silent substitution in predicted miRNA binding site of *Solyc03g120910, Solyc02g069830* was created with the QuikChange II XL following the provided protocol (Agilent; Catalog no. 200521). This mutated cDNA was then cloned into PGWB417 as described earlier. Mutagenesis was confirmed by Sanger sequencing.

Rhizobium (Agrobacterium) rhizogenes transformation

Rhizobium rhizogenes (ATCC: Strain 15834) transformation followed the protocol previously described (Ron et al., 2014). Briefly, competent *R. rhizogenes* was transformed by electroporation with the desired binary vector, plated on nutrient agar (BD 247940) plates with the appropriate antibiotics (spectinomycin, 100 mg L−1), and incubated for 2-3 days at 28-30°C. *R. rhizogenes* colonies passing selection were inoculated from plates into 10 mL nutrient broth liquid medium (BD 90002-660) with the appropriate antibiotics (spectinomycin, 100 mg L−1) and were grown overnight at 30°C with shaking at 200 rpm. This culture was used to transform 40 to 50 fully expanded tomato cotyledons grown in sterile conditions for 8-10 days (just before the first true leaves emerge). Using a scalpel, 8-10 day old M82 cotyledons were cut and immediately immersed in the bacterial suspension at an optical density of 600 nm in Murashige and Skoog (MS, 1X) liquid medium for 20 minutes and then blotted on sterile Whatman filter paper and transferred (adaxial side down) onto MS agar plates (1X with vitamins, 3% sucrose, 1% agar) without antibiotic selection and incubated for 3 days at 25°C in dark. The cotyledons were then transferred to MS plates with Vitamins (MSP09-10LT), 1% agar and 3% sucrose with a broad spectrum antibiotic cefotaxime (200 mg L^{-1}) and kanamycin (100 mg L^{-1}) for selection of successfully transformed roots and returned to 25°C. At least three to five independent roots develop from each cotyledon. Antibiotic-resistant roots that emerged were further transferred to new selection media. Fifteen independent roots, representing 15 independent transgenic events, were subcloned for each construct for further analysis (genotyping and imaging).

Quantitative RT–PCR of overexpression lines

All quantitative RT-PCR primers were designed with Primer3Plus software (http://www.primer3plus.com/). Primers were designed to amplify a 100-150 bp region near the 3′ end of each target TF coding sequence. qRT-PCR was performed by setting up a 20 μL PCR reaction containing 5ul of cDNA (100ng/reaction) and 200 nM of each primer (PCRBIO Taq DNA Polymerase/Mix; Catalog no. PB10.11-05 and EvaGreen dye; PCRBIO; Catalog no. 89138-982). qRT–PCR was performed in a Bio-RAD CFX384-Real Time System with the following thermal cycling conditions: 5 min at 95°C, followed by 40 cycles of 20 s at 95°C, 20 s at 60°C, and 20 s at 72°C. To ensure that PCR products were unique, a melting-curve analysis was performed after the amplification step. The experiment was carried out on a minimum of three independent lines and three technical replicates for each overexpression line. To determine the fold change of the overexpression line relative to the wild-type control (tomato transformed with *R. rhizogenes* with no plasmid), an absolute quantification method was conducted by generating a standard curve for each primer set. Values were normalized to the Ct value of an endogenous control gene (*Solyc07g025390*). The qPCR data for each gene is shown as a relative expression with respect to a control hairy root sample to which an expression value of 1 was assigned. Standard error of the mean (SEM) was then calculated from the normalized expression for each sample represented in the graphs. P values were determined by performing a simple t test; subtracting Ct number of the target gene for 3 replicates from that of the reference gene, which provides ΔCt values for overexpression lines and the wild-type control to be subject for a t.test (Table S3).

Histochemistry and imaging of xylem phenotypes and exodermis characterization

Hairy root tissue and seven-day-old *Arabidopsis* primary roots from Col-0, the *VND6* inducible line and a mock control were cleared for 4-5 days in ClearSee buffer (Ursache et al., 2018). The mock control recapitulated the phenotype observed in wild-type, and thus we only include wild-type in Figure 3G. Detection of xylem vessel elements was conducted by incubation of cleared roots in Basic Fuchsin (0.04% w/v in ClearSee; Fuchsin stains lignin and phenylpropanoid molecules) for 24 hours followed by a 1-2 hour wash in the ClearSee buffer before imaging as previously described (Turco et al., 2019). Confocal Laser Scanning microscopy was performed on a Zeiss LSM700 confocal with the 20X objective, Basic Fuchsin: 550-561 nm excitation and 570-650 nm detection. Root samples were mounted in ClearSee (Ursache et al., 2018) and scanned. Protoxylem vessel differentiation was first observed at 0.2-0.4 mm distance from the tip, while metaxylem vessels differentiate up to 1 cm from the root tip, in the maturation zone. Secondary cell wall quantification for the β–estradiol-inducible *VND6* line and *35S::SlVND6* was performed by characterizing 3 ectopic xylem cells (per root) for width and secondary cell wall pattern observed in the root tip (1 cm). Traits quantified were determined based on discussions with Dr. Taku Demura (NAIST). A minimum of 10 roots were imaged per line. Results were reported as percentages (Table S3). For the exodermis lignin staining we used 1cm root tips from five-day-old *Solanum lycopersicum* roots. The root tips were embedded in 3% agarose and the blocks were sectioned using a vibratome. The root sections were stained with Basic Fuchsin for lignin and Calcofluor White for the cell wall in the Clearsee buffer (Ursache et al., 2018). Confocal Laser Scanning microscopy was performed on a Zeiss Observer.Z1 confocal with the 20X objective, Basic Fuchsin: 550-561 nm excitation and 570-650 nm detection and Calcofluor: 405 nm excitation and 425-475 nm detection. Exodermal suberin was observed in seven-day-old *S. lycopersicum* cv. M82 roots after Fluorol Yellow (FY) staining as described in Lux et al. (2005). In short, roots were divided in 1 cm segments, embedded in 3% agarose, and sectioned using a vibratome. Sections were then incubated in FY088 (0.01%w/v, dissolved in lactic acid) for 1 hour at RT in darkness, rinsed three times with water, and counterstained with aniline blue

(0.5% w/v, dissolved in water) for 1 hour at RT in darkness. Confocal Laser Scanning microscopy was performed on a Zeiss Observer Z1 confocal with the 20X objective and GFP filter (488nm excitation, 500-550nm emission).

Nuclei purification by INTACT for ATAC-seq

These steps were conducted as described in Reynoso et al. (2019). In brief, nuclei from cell type populations were isolated from the frozen root tip material using INTACT (Deal and Henikoff, 2010; Maher et al., 2018; Reynoso et al., 2018), and the nuclei were counted and used for ATAC-seq library preparation (Maher et al., 2018). Libraries were size selected for under 750-nt and up to 24 barcoded libraries were pooled together. ATAC-seq libraries were sequenced on the NextSeq 500 at the UC Davis DNA Technologies Core to obtain 40-bp paired-end reads.

QUANTIFICATION AND STATISTICAL ANALYSIS

Tomato RNA-seq data processing and analysis

Sequences were pooled, and then trimmed and filtered using Trim Galore! (v0.4.5) (Krueger, 2012) with parameter -a GATCGGAA GAGCACA, resulting in removal of 7.8% of the reads on average. Trimmed reads were pseudo-aligned to the ITAG3.2 transcriptome (cDNA) (Tomato Genome Consortium, 2012) using Kallisto (v0.43.1) (Bray et al., 2016), with the parameters -b 100–single -l 200 -s 30, to obtain count estimates and transcript per million (TPM) values. On average 62% of the trimmed reads were aligned to the tomato transcriptome. As a quality control we used STAR (Turner, 2012) to map the entire genome (including organelles), with default parameters. This approach resulted in additional mapping of 19% of the trimmed reads (to a total of 81%), which include expressed transposons or organelle transcripts that are beyond the scope of this study (Table S1).

Tomato RNA-seq quality control and relative differential expression

Raw RNA-seq read counts were filtered to remove genes with zero counts across all samples. Reads were converted to count per million (CPM) using the cpm() function in edgeR. Genes with CPM > 0.5 in at least 4 biological replicates were kept, thus removing genes that were consistently lowly expressed across all samples. In order to perform data quality control, we conducted exploratory data analysis with the filtered CPM values as recommended by Dillies et al. (2013) and demonstrated by (Gilad and Mizrahi-Man, 2015). The data were $\log_2$ transformed with a prior count of 3 to reduce the contribution of low-abundance genes. Batch effects due to sequencing date were corrected with the removeBatchEffect function (Ritchie et al., 2015). Similarities and dissimilarities between samples were assessed with principal component analysis (PCA) using the function 'prcomp' in R. PCA plots were generated with the ggplot2 package (Wickham, 2009) (Figure S1C).

Data is available for review on a gene-by-gene basis using the following instructions:

- Search a tomato gene/click on example
- Go to Tissue & Experiment eFP viewers
- Preview "Root eFP" or "Root Field Pot eFP"
- Hover over colored areas to see expression calculation

Expression of SlPHB/PHV-LIKE1 and *SlCNAL1* are found at: (http://bar.utoronto.ca/eplant_tomato/?ActiveSpecies=Solanum%20lycopersicum&Genes=Solyc02g069830&ActiveGene=Solyc02g069830&ActiveView=RootView) and (http://bar.utoronto.ca/eplant_tomato/?ActiveSpecies=Solanum%20lycopersicum&Genes=Solyc03g120910&ActiveGene=Solyc03g120910&ActiveView=RootView).

Inference of tomato cell type-enriched genes and ontology terms

To identify genes with enriched expression in each cell type we combined two independent approaches described above. *Approach 1*: Differentially expressed genes (DEGs) were detected with the limma R package, using normalized CPM values as required by the package (Ritchie et al., 2015). CPM values were normalized with the voom function (Law et al., 2014) using quantile normalization with a design matrix that included identifiers for the marker line populations and the sequencing replicates (batch). The functions lmfit, contrasts.fit, and ebayes were used to fit a linear model and calculate differential gene expression between the different contrasts. Genes with a $\log_2$ fold change (FC) value ≥ 2 and FDR adjusted P value (adj.P.Val) ≤ 0.15 were considered as differentially expressed. The fdr method was used to control the false discovery rate (FDR) (Benjamini and Hochberg, 1995). DEGs, as determined by limma's contrasts, were processed with the Brady method (described in Brady et al., 2007) to identify genes with enriched expression ($\log_2$FC ≥ 2, FDR adjusted P value ≤ 0.15) in each cell type compared with all other non-overlapping cell types (see Table S1 for these contrasts). *Approach 2*: ROKU, an approach based on Shannon entropy statistics, has previously been used to identify genes enriched in a tissue specific manner (Kadota et al., 2006; Li et al., 2016). This approach calculates an entropy score of 1, 0 and -1, for depleted, no change, or enriched, respectively, for each gene across cell or tissue specific samples. A gene could be considered as enriched or depleted in no more than half of the cell types. ROKU uses a subset of constitutively expressed genes to determine empirical baseline distributions of entropy scores and to calculate a threshold to call significantly enriched genes. We used TPM

values for the ROKU method since they reliably depict the proportion of a sample's reads that were mapped to the transcriptome (Wagner et al., 2012). Since the batch effect cannot be modeled for the ROKU method, and since batch correction changes the expression data (i.e., DEGs, based on batch corrected TPM values, have low correlations with batch modeled DEGs [$r \leq 0.5$, $p < 0.01$, data not shown]), we used upper quartile normalized TPM values to calculate gene entropy. The parameters to determine enriched genes using the Shannon entropy approach were *delta* = 0.08, *lowexp* = 0.05, *bgfold* = 2, *bgmedian* = 0.5, and *pvalue* = 0.001. The R script and functions are hosted in {https://github.com/plant-plasticity/tomato-root-atlas-2020}. *Combining Datasets*: next, a union gene set, based on both the Brady and ROKU methods, was obtained for each cell type. A non-redundant list of enriched genes was curated by including only genes with a TPM value ≥ 2 that have the highest expression in the target cell type compared with all other cell types, excluding 35S and Actin (Table S1). To differentiate between the general cortex (gCOR), which includes the exodermis, and the inner cortex (iCOR), which includes only the two inner cell files of the cortex, the union set of enriched cortex genes was not filtered against the exodermis, resulting in a partially redundant list with the exodermis of gCOR-enriched genes.

Ontology enrichment analyses were done using two different ontologies: i) Gene ontology (GO) and ii) MapMan ontology to identify enriched terms within each cell type/tissue enriched gene list. GO enrichment analysis was done with the GOseq R package (Young et al., 2010), using the effective transcript length (Kallisto output) for correction of the length bias present in the data. Gene Ontology annotation (ITAG3.2) was downloaded from Sol Genomics Network (solgenomics.net). A term was considered significantly enriched if it has a p value < 0.05 and a fold enrichment > 1. Multiple testing correction is not recommended for GO enrichment due to the graph structure of the GO terms (Mi et al., 2012). Fold enrichment was calculated as (genes annotated with a term in the query dataset / total genes in the dataset) / (genes annotated with a term in the background set / total expressed genes) (Table S1). The hierarchical and non-redundant MapMan bin terms (Urbanczyk-Wochniak et al., 2006) were used as a reference database for functional enrichment analysis using the FunRich tool (v3.1.3 www.funrich.org; Pathan et al., 2015). Mapping files (ITAG2.3) were retrieved from the MapMan Store (mapman.gabipd.org). To create a structure that resembles GO, enrichment analysis was carried out independently for four hierarchy levels; the two top-level terms, which tend to be similar to the "biological processes" and the two-lower-level terms of the MapMan hierarchy, which are more similar to the "molecular functions" associated with GO terms (Klie and Nikoloski, 2012). Terms with a fold enrichment > 1 were selected for FDR adjustment of their p values using p.adjust function in R. Only terms with an FDR < 0.15 were considered significantly enriched (Table S1). This cutoff was selected based on known cell type processes or genes, including enrichment of the WRKY domain transcription factor family in the epidermis (FDR = 0.09), DOF zinc finger family in the vasculature (FDR = 0.1), and lignin biosynthesis (4CL) and MYB domain transcription factor family in the exodermis (FDR = 0.14).

Identification of tomato cell type-enriched genes in field and pot-grown plants

Four TRAP lines profiled in agar plate-grown plants were also profiled in a field experiment (driving expression in the endodermis and quiescent center *(SlSCR)*, meristematic zone (*SlRPL11C*), meristematic cortex (*SlCO2*) and whole root (*35S*)). The cell type-enriched genes were derived from comparisons involving only these marked cell type populations and were performed as described for the whole atlas dataset. Gene lists were filtered for FC ≥ 1 in the case of the field experiment and a FC ≥ 2 for the tomato atlas experiment (FDR adjusted P value ≤ 0.15) and can be found in Table S2. Genes identified as cell type-enriched in both the field and atlas experiments were considered as "core" cell type genes (Table S2). GO and MapMan enrichment analysis was carried out for the cell type-enriched genes derived from four cell type comparisons (separate for atlas and field experiment as well for the list of core genes) in the same manner as for the full dataset (Table S2). Enriched categories and annotations shared between the atlas and field experiment (meaning enriched among CTEGs in both the field and atlas, respectively) can be found in Table S2.

Co-expression network analysis

Co-expression network modules were created with the WGCNA R package version 1.68 (Langfelder and Horvath, 2008). Individual libraries from each growth condition (agar plates, field, pots) were quantile normalized together and 75% of the most variable genes were used for analysis. A soft threshold of 5 was used to create a scale-free network. An unsigned network was created using the *blockwiseModules*-function with the bicor correlation measure and the following parameters: maxPOutliers = 0.05, mergeCutHeight = 0.35 and maxBlockSize = 25000. Gene Ontology and MapMan enrichment analysis for genes from each individual module was carried out in the same manner as for the cell type-enriched genes. A list of genes assigned to each module, as well as GO and MapMan annotations enriched in each module, can be found in Table S2.

Phylogenetic tree construction

First, 42 representative proteomes were downloaded from Phytozome, Ensembl, or consortia sites depending on availability (Table S3). These include early-diverging taxa, and broadly representative taxa from angiosperms. Next, blastp (Madden, 2013) was used to identify homologous sequences within each proteome based on a sequence of interest, with options "-max_target_seqs 15 -evalue 10E-6 -qcov_hsp_perc 0.5 -outfmt 6." To refine this set of sequences, a multiple sequence alignment was generated with MAFFT v7 (Katoh and Standley, 2013) (option–auto), trimmed with trimal (Capella-Gutiérrez et al., 2009) with

setting "-gappyout," and a draft tree was generated with FastTree (Price et al., 2010). A monophyletic subtree containing the relevant sequences of interest was selected and more distantly related sequences were removed from the list of sequences. Tree construction methodology was informed by Rokas (2011). For the final trees, MAFFT v7 using L-INS-i strategy was used to generate a multiple sequence alignment. Next, trimal was used with the -gappyout option. To generate a phylogenetic tree using maximum likelihood, RAxML was used with the option -m PROTGAMMAAUTO and 100 bootstraps. Finally, bipartitions with bootstrap values less than 25% were collapsed using TreeCollapserCL4 (http://emmahodcroft.com/TreeCollapseCL.html). Resulting trees were rooted on sequences from the earliest-diverging species represented in the tree. Phylogenetic trees can be found in Data S1.

Gene orthology determination

To identify the best orthologs between *Arabidopsis* and tomato we used a phylogenetic approach as described above. In the cases where orthology was defined, it was done so based on the position of the target tomato gene relative to its closet *Arabidopsis* ortholog.

We identified the closest possible orthologs as follows: We identified *At3g54220 (AtSCR)* as a 1:1 ortholog of SlSCR, Solyc10g074680 (Data S1A). *At1g02030- ZAT 4 and Atg4512- ZAT9* (paralogs in the same clade) are orthologs to *Solyc01g090840* (Data S1B). There were three possible tomato orthologs to *At3g06410* (Zinc finger C-x8-C-x5-C-x3-H type family protein), including *Solyc06g054600* (Data S1C). We identified *At3g57600 (AtDREB2F)* as a 1:1 ortholog with *Solyc10g080310 (SlDREB2F)* (Data S1D). *Solyc07g056040* has two possible *Arabidopsis* orthologs, *At1g17200 (CASP-LIKE2A1)* and *At3g14380 (CASP-LIKE2A2)*, and no other tomato gene is closely related to these two sequences so we call *Solyc07g056040 SlCASP-LIKE2A* (Data S1E); for HD-ZIPIII transcription factors (Data S1F) AT1g30490 *(PHV)* or AT2g34710 *(PHB)* are paralogs in the same clade and each are possible orthologs for *Solyc02g069830.* Therefore this gene was named *SlPHB/PHVLIKE1. Solyc03g0120910* was named as a possible ortholog for *CORONA (At1g52150)*. Since there is another CORONA paralog, we named *Solyc03g120910, SlCORONA-LIKE1. At4g08150 (AtKNAT1)* is a 1:1 ortholog for *SlKNAT1* (*Solyc04g077210)* (Data S1G).

Ranking candidate xylem regulatory TFs – Intersection of QTL and eQTL data

Genetic intervals significantly associated with variation in xylem cell number were identified using data reported in Ron et al. (2013). Introgression lines containing these significant genetic intervals were then screened for significant *cis*-eQTL (Toal et al., 2018) of (1) TF loci enriched in tomato xylem cells or vascular tissue, or of (2) *HD-ZIPIII* family putative orthologs (Table S3).

Statistical analyses for overexpression lines

Comparisons and significance of aberrant xylem phenotype frequencies (*SlKNAT1* - extra protoxylem or xylem breaks; *SlCNAL1* - loss of bilateral symmetry; *SlPHB/PHV-LIKE1* protoxylem at metaxylem position; *SlVND6* - ectopic secondary cell wall deposition in other cell types) relative to the wild-type control (tomato transformed with *R. rhizogenes* with no plasmid) were determined with a logistical regression method using Generalized Linear Model (GLM) in R Studio software (Version 1.2.5001). The output from the GLM model was then used to determine an odds ratio for each independent line. Analysis was done on 3 independent lines for each overexpression construct with a minimum of 12 biological replicates per line. The results of all statistical tests performed are reported in Table S3.

Identifying transposase hypersensitive sites

A flow chart describing all steps of transposase hypersensitive site (THS) identification and analyses is found in https://github.com/plant-plasticity/tomato-root-atlas-2020/blob/master/Figures/Fig_S20_ATACseq_flowchart_with_legend.pdf). GFP expression patterns of stable transgenic lines were largely similar to that observed for hairy roots (Figure S1; Table S1). For each sample, 40-bp PE sequencing reads were trimmed using CutAdapt 2.0 and parameters for Nextera libraries (Martin, 2011). Trimmed reads were mapped using BWA-mem (Li and Durbin, 2009) software with default parameters to SL3.0 (https://www.ncbi.nlm.nih.gov/assembly/GCF_000188115.4/). Aligned sam files were converted to bam format using Samtools 1.6 (Li et al., 2009), sorted and filtered to retain only reads that had a mapping quality score of 2 or higher, and filtered to retain only reads that mapped to true nuclear chromosomes.

The tomato genome is repeat-rich (Bolger et al., 2014), and thus to account for mis-annotation of repeats as well as unknown copy number variation, we used highly conservative methods described for human DNaseI hypersensitive site sequencing to remove high-depth sequencing regions (Pickrell et al., 2011). Genomic DNA-based ATAC-seq libraries from 1 cm root tips were sequenced on the NextSeq 500 at the University of Georgia Genomics and Bioinformatics Core to obtain 36-bp paired-end reads (Reynoso et al., 2019). After mapping with Bowtie2 (Langmead and Salzberg, 2012) to SL3.0, the number of reads mapping to each position in the genome was determined. Next, the number of reads within 150-bp sliding windows (step size 20-bp) was counted and plotted in a histogram (Figure S6A). The top 0.1% most-accessible windows were then identified, merged and removed from cell type ATAC-seq sample bam files. Figure S6B demonstrates the distribution of sizes for these high sequencing depth regions. Masked bam files were then sub-sampled to a final count of 25 million reads.

In order to determine the best window size for peak calling, we took a deeply sequenced sample (SlWOX5_O08) and called peaks using three different window size parameters relative to increasing sizes of randomly sampled reads (Figures S6C and S6D). From

these, we determined that a 10-kb window size (the HOMER default) led to an asymptote at ~25 million reads. Peak calling was thus performed using the "Findpeaks" function of the HOMER 4.9 package (Heinz et al., 2010) with the parameters "-size 150," "-minDist 150" "-region" and "-regionRes 1." These regions are hereby referred to as transposase hypersensitive sites (THSs).

Independent of peak calling, "per base" bed files were also created. Specifically, the number of aligned reads within a bam file, or cut counts, were tallied at each position within the tomato genome. Any position with zero cut counts was discarded. Results were reported in standard bed file format. For visualization of data within a genome browser, bigWig files were also created from the sub-sampled with Deeptools 3.1.0 (Ramírez et al., 2014), with the parameters "–binSize 20," "–normalizeUsing RPGC," "–effectiveGenomeSize 807224664," and "–extendReads."

To find replicable THSs across a minimum of three, or a maximum of four biological replicates within a cell type, THSs from the replicates were merged into master replicate THS file using Bedtools 2.27 "merge" (Quinlan and Hall, 2010). In order to ensure that replicates were similar in terms of cut counts, we performed pairwise comparison of cut counts between cell types (https://github.com/plant-plasticity/tomato-root-atlas-2020/blob/master/Figures/Figure_S23_scatter_plot_replicates_repUnion_THSs_ALL_110520_v2_with_legend.pdf). Next, for each replicate, the number of cut counts within each region in the master replicate THS file were counted using BEDOPS 2.4.33 'bedmap,' with the replicate perbase bed file as the map file and the master replicate THS bed file as the reference (Neph et al., 2012). The coefficient of variation was then calculated for each THS across the replicates and the top 15% most variable THSs were removed from further analysis (Figures S6E and S6F). THSs below this 15% threshold are thought to be constitutive, non-variable THSs (Alexandre et al., 2018). 108,335 reproducible transposase hypersensitive sites (THSs) were identified across cell types, with more than half (66%) found in intergenic regions distal to the transcription start site (TSS) as previously described (Maher et al., 2018) (Figure S6G; Table S4). After repTHSs from each cell type were identified, repTHSs were merged into a master union THS bed file (uTHS bed file) using bedtools "merge." These uTHS regions were then used for downstream analysis of motif enrichment. Please see Table S4 for a summary of ATAC-seq data

Motif enrichment and TF networks

Motif database construction

Motif files were downloaded from CisBP for Weirauch, DAP-seq, Franco-Zorilla, and Sullivan motif datasets (Franco-Zorrilla and Solano, 2017; O'Malley et al., 2016; Sullivan et al., 2014; Weirauch et al., 2014). If a motif from the protein binding array studies overlapped with the DAP-seq database, it was discarded.

1-kb promoter network construction

1-kb upstream sequences of the TSS for each group of cell type-enriched genes were identified. Next, these sequences were used to perform motif enrichment with our custom motif database using Meme Suite AME (McLeay and Bailey, 2010), with the parameters "–scoring avg," "–method fisher," "–hit-lo-fraction 0.25," "–evalue-report-threshold 2000," "–control," "–shuffle–," and "–kmer 2." Next, the motif enrichment files for all cell types were converted to a matrix file where each row represents a transcription factor motif and each column represents the adjusted p value for that motif in a given cell type. Motifs were then filtered for ones that were significantly enriched in at least one cell type (padj > = 0.01). The matrix file was then split by motif family and adjusted p values were visualized in R 3.6 (https://www.R-project.org/) using Pheatmap (https://cran.r-project.org/web/packages/pheatmap/index.html).

uTHS promoter network construction

For each cell type-specific group of genes, uTHSs were identified that were 4-kb upstream of TTS, overlapping genic regions, or 1-kb downstream of the TTS. This was done using the bedtools "closest" tool, with the parameter "-D" and the uTHS files and the bed files for the genic locations for the cell type-specific genes. Fasta sequences for these regions were obtained using bedtools "getfasta." Next, motif enrichment was performed using Meme suite AME using the same parameters as the 1-kb upstream regions. Motif filtering and heatmap creation were performed as they were for the 1-kb upstream regions.

Cell type-unique network construction

To identify unique cell type functions and their underlying regulation, we also constructed unique cell type networks (Table S4). Transcription factor motifs that were significant and unique to each cell type were identified separately for uTHSs and 1-kb promoters. Next, we filtered the unique transcription factor motifs for positively correlated expressologs in tomato and whether they were expressed in the cell type of interest (TPM > = 1). After identification of unique expressologs, the union of unique transcription factors was taken between the 1-kb promoters and uTHS networks. These union networks comprising transcription factor motifs, as well as their targets, were then visualized with Cytoscape 3.7.1. (Shannon et al., 2003). Please see Table S4 for unique cell type network Cytoscape files.

Nitrogen network overlap

To test for enrichment of the exodermis-inferred network with *Arabidopsis* nitrogen-associated transcriptional regulatory network (Gaudinier et al., 2018), we filtered the expressolog list for positively correlated expressologs (cor > 0). The *Arabidopsis* nitrogen network contains a total of 429 genes. Of these, 362 have at least one positively correlated expressolog in *S. lycopersicum*. A total of 301 genes have an expressolog for both the TF and its target promoter in a TF/promoter interaction. We calculated if the overlap between the *S. lycopersicum* exodermis-inferred network genes and the expressologs of the orthologous nitrogen network genes in tomato was greater than expected by chance using the *fisher.test()* function in R with *alternative = "greater"* (Table S3).

A multi-species analysis of root cell type-atlases

Analysis overview can be found at https://github.com/plant-plasticity/tomato-root-atlas-2020.

Arabidopsis microarray data

.CEL files containing data resulting from translatome profiles of *Arabidopsis* root tips expressing FLAG-tagged cell type populations marked by endodermis (*AtSCR*), vasculature (*AtSHR*) and whole root (*35S*) promoters, as well as from translatome profiles of the meristematic zone (*AtRPL11C*) and meristematic cortex (*AtCO2*) marker lines, were downloaded from GEO (GSE14493) (Mustroph et al., 2009). The raw files were reanalyzed with the limma package, using default parameters (Ritchie et al., 2015) and normalized $\log_2$ intensity values can be found in Table S5.

Rice RNA-seq data processing and analysis

Rice data were processed as described above for tomato RNA-seq data processing and analysis with the following modifications: trimmed reads (on average 87% of the raw reads) were pseudo-aligned to IRGSP-1.0 transcriptome (cDNA, https://rapdb.dna.affrc.go.jp/index.html) using Kallisto (v0.43.1) (Bray et al., 2016) to obtain count estimates and transcript per million (TPM) values. Splice variants were summed to assess transcript values. On average 71% of the trimmed reads were aligned to the rice transcriptome. As a quality control we used STAR (Turner, 2012) to map the entire genome (including organelles), with default parameters. This approach resulted in additional mapping of 17% of the reads, which include expressed transposons or organellar transcripts that are beyond the scope of this study (Table S5). To validate our approach, we examined the expression patterns of core developmental cell type genes in the translatome of rice marker lines, as done for the tomato data (Table S5). Rice data is found on a gene-by-gene basis at: http://bar.utoronto.ca/eplant_rice/ via the Tissue eFP link.

Sample integration and expression clustering

Comparisons of transcript abundance were conducted for four homologous cell types/tissues (meristematic cortex, endodermis and quiescent center, vasculature and meristematic zone; Figures S1A, S1E, and S1F) within and between species (Figures S5A–S5C). We first examined the clustering of the biological replicates within each species separately (Figure S5A). Next, we explored expression similarities of homologous cell types/tissues between species (Figures 5B, S5B, and S5C). Since genes that undergo duplication events rapidly diverge in their expression profiles (Chung et al., 2006; Gu et al., 2005) three different orthology maps were generated, two maps based on sequence similarity and one based on sequence similarity coupled with expression correlation (i.e., "expressologs") (Table S5). The first orthology map includes 2,642 1:1:1 orthologs based on sequence homology, using Phytozome v12 gene families. Phytozome predicted gene families were generated using genome sequence data from 57 plant species. In Phytozome, the relationships between genes and species are determined by InParanoid, which uses an all-versus-all BLAST alignment of pairwise proteomes to identify orthology groups (Sonnhammer and Östlund, 2015). Phytozome uses an *S. lycopersicum* ITAG2.4 annotation, while data for all other analyses in Figures 1, 2, 3, and4 are from the ITAG3.2 genome. Hence, genes annotated in ITAG3.2 that are absent from ITAG2.4 were assigned to a gene family based on a blastp search against *A. thaliana* cDNAs (-max_target_seqs 1), with an E value cutoff of < 0.01). To identify 1:1:1 orthologs, only predicted gene families with one gene from each species were included (Figure 5B; Table S5). The second orthology map includes 3,505 1:1 orthologs, based on sequence homology to *Arabidopsis*. This map takes advantage of the plant-specific MapMan tool, which was originally developed for *Arabidopsis*, and currently supports more than 80 plant species (https://mapman.gabipd.org/home) (Thimm et al., 2004). The freely available MapMan annotation files of tomato and rice were parsed to include only 1:1 orthologs that are present in both files (Figure S5B; Table S5). Finally, the third orthology map consists of 1,771 *Arabidopsis* and rice expressologs of tomato with an expression correlation coefficient > 0.6 (Figure S5C; Table S5). "Expressologs" are determined using an approach to resolve orthologs by predicting putative functional orthology (i.e., expressologs). This map was constructed using sequence homology, based on OrthoMCL (Li et al., 2003), and complemented by published expression profile similarity to refine ortholog predictions as described in Patel et al. (2012).

As previously described, analyses of gene expression variation between species must take into account confounding factors (Gilad and Mizrahi-Man, 2015). Thus, we next considered how to address differences in experimental design between tomato, rice and *Arabidopsis*, and the fact that *i*) translatome samples were obtained from two expression platforms (i.e., RNA-seq for rice and tomato and microarray for *Arabidopsis*) and thus possess distinct dynamic ranges (Figure S5J), and *ii*) data obtained from each species were collected and processed in a different laboratory, which drives the clustering of samples (Figure S5K). We accounted for these issues by applying the functions normalizeBetweenArrays() and removeBatchEffect(), from the limma package, which were used to quantile normalize $\log_2$ transformed expression values, across samples of homologous cell types and 35S (Figure S5L) and to correct for the laboratory effect (Figure 5B), respectively (Ritchie et al., 2015). Since species and laboratory are completely confounded, by correcting for the batch effect we also removed the contribution of the species to gene expression variation, hence we can only assess the contribution of the tissues. Similarities between cell types/tissues were assessed with principal component analysis (PCA) using the function 'prcomp' in R. PCA plots were generated with the ggplot2 package (Wickham, 2009).

Root Cell Type TRAP-expressologs

Cell type- or tissue-resolution TRAP data can be utilized to define "expressologs" based on similarity of expression variation across homologous root cell types. Ortholog annotations for tomato, *Arabidopsis* and rice were determined as described in Patel et al. (2012) with the following modifications: (i) putative gene families that include at least two of the three species were retrieved from the ITAG3.2-updated Phytozome v12 gene family file (described above for the first orthology map); (ii) within each gene family, the Pearson correlation coefficient was calculated for each ortholog pair using the TRAP expression values of homologous cell types and

tissues. Tomato and *Arabidopsis* included eight homologous cell types and tissues (EP, COR, MCO, EN-QC, V, PH, MZ and 35S), tomato and rice included six homologous cell types and tissues (MCO, EN-QC, V, MZ, QC and 35S) and *Arabidopsis* and rice included five homologous cell types and tissues (MCO, EN-QC, V, MZ and 35S). (iii) The correlation matrices were reciprocally parsed to include only the best matching expressolog pairs using each species as a reference (e.g., maximum correlation between *Arabidopsis* to tomato and tomato to *Arabidopsis*, based on *Arabidopsis* as a reference species). (iv) To identify high confidence expressologs and define ortholog annotations for the cell type-enriched genes, only expressolog pairs with a positive correlation and a reciprocal match between *Arabidopsis* and tomato and *Arabidopsis* and rice were considered (Table S5). These filtering criteria resulted in the identification of 6,059 expressologs between *Arabidopsis* and rice, and 7,295 expressologs between *Arabidopsis* and tomato. To detect conserved expressologs, we selected only positively correlated expressologs that maintain the same relationship among the three species, independently of the reference species. To this end, high confidence expressologs among the three species were identified, using each species as a reference. This analysis resulted in identification of 6,293, 6,470 and 6,516 expressologs based on tomato, *Arabidopsis* and rice as a reference species, respectively. Next, the three datasets were intersected and expressologs with negative expression correlations were excluded, resulting in the identification of 1,555 expressologs that have both identical expressolog relationships independent of the reference species and positive expression correlations (referred to as consensus expressologs) (Table S5). Clustering of expression profiles of homologous cell types, based on the consensus expressologs, was done following quantile normalization and batch effect correction of $\log_2$ expression values, as described for the sample integration and clustering of expression profiles across-species (Figure S5D).

ANOVA to identify conserved cell type and tissue-specific expressologs

The clustering of the consensus expressologs based on tissue identity suggests that some of these genes have conserved tissue-specific patterns of expression (Figure S5D). To further explore these expression patterns and to identify consensus expressologs with conserved cell type and tissue-enriched expression we used an ANOVA. Expression values of MCO, EN-QC, V, MZ and 35S were processed for each species separately. For tomato and rice, upper quantile-normalized TPM values were filtered to remove genes with low expression (TPM $\leq$ 2), followed by adding a prior count of 3 and $\log_2$ transformation to reduce the contribution of low-abundance genes. Tomato data were further corrected for differences in sequencing date as described for the RNA-seq quality control and differential expression. For *Arabidopsis* we used normalized $\log_2$ intensity values. For each cell type and tissue in each species, we calculated the mean gene expression, if up to three biological replicates existed, or median gene expression, if four biological replicates existed. Next, the three datasets were combined based on the 1,555 consensus expressologs. The 15 sample mean/median values (MCO, EN-QC, V, 35S) were quantile normalized and corrected for the batch effect arising from the different laboratories, using the functions normalizeBetweenArrays() and removeBatchEffect() from the limma package, respectively, as described for the sample integration and clustering of expression profiles. To detect genes with conserved cell type and tissue specific expression we conducted subsequent analyses with MCO, EN-QC, V and MZ translatome data from each species. The R Stats functions lm(), aov() and the function HSD.test(), from the agricolae package (de Mendiburu, 2019), were used to fit a linear model, to test the effect of the tissue on gene expression and to identify the cell types or tissues with a significant effect, respectively. The consensus expressologs with the top 15% F-values ($\geq$6.6) were further filtered to include genes with a conserved enriched or depleted expression in one cell type, based on a Tukey test (p value $\leq$ 0.1) (e.g., conserved high expression in the MZ compared with the other three cell types). Finally, these genes were filtered against constitutively expressed genes (CEGs) within each species, as described below, resulting in the detection of 139 conserved cell type-specific expressologs (Table S5). Thirty-seven of these genes showed conserved cell type/tissue enriched expression among the three species (Figure 5C).

Detection of cell type and tissue-enriched genes and ontology terms across species

To allow a balanced comparison of CTEGs across species we used the same pipeline as described for the detection of cell type- or tissue-enriched genes within tomato (contrasts and parameters are specified in Table S5). Orthologs were resolved using the high confidence expressologs between *Arabidopsis* and the two other species, as described above for the Root Cell Type TRAP-expressologs (Table S5). Enrichment of GO and MapMan ontology terms of cell type-enriched genes were determined for each species as described above (Table S5). GO annotations were downloaded for the TAIR10 genome assembly (Arabidopsis.org) and retrieved from Ensembl with the biomaRt package (Durinck et al., 2009), for *Arabidopsis* and rice, respectively. Overlapping ontology terms among homologous cell types were visualized using Circos (Krzywinski et al., 2009) based plots, which also included 3-way overlaps, and their enrichment was evaluated by Fisher's exact test using fisher.test() function in R (Figure 5D; Table S5).

Additional analyses to confirm meristem similarity

Finally, we examined if the similarity observed in the homologous translatomes is specific to the MZ or rather a common feature with other meristematic tissues. To this end we substituted one meristematic cell population with another (i.e., MZ with MCO) and examined the number of overlapping expressologs and enriched ontology terms. This analysis resulted in a decline in the number of overlapping features, regardless of the species tested. For example, when the meristematic zone of *Arabidopsis* is replaced with its meristematic cortex the number of overlapping enriched expressologs and MapMan terms within *Arabidopsis* cell populations decreases by a factor of two (4.4% and 16.5% compared with 2.2% and 7.5% overlap, respectively).

Constitutively expressed genes (CEGs)

To identify CEGs we used a fold change difference < 1.5 between the maximum and minimum TPM or intensity values of each gene across the five homologous cell types/tissues (i.e., MCO, EN-QC, V, MZ and 35S), together with a cutoff of a per gene median

expression > median expression of each species. These filtering criteria resulted in detection of 308, 1,154 and 1,523 CEGs in tomato, *Arabidopsis* and rice, respectively (Table S5). Orthologs were resolved using the high confidence expressologs between *Arabidopsis* and the two other species, as described above for the Root Cell Type TRAP-expressologs (Table S5). Enrichment of gene and MapMan ontology terms of CEGs were determined for each species as described above for the detection of cell type and tissue-enriched genes and ontologies across species (Table S5). Assessment of the enrichment of the overlaps between the ontology terms of the CEGs compared with the CTEGs was carried out with a Fisher's exact test using fisher.test() function in R.

CellPress
OPEN ACCESS

Supplemental figures

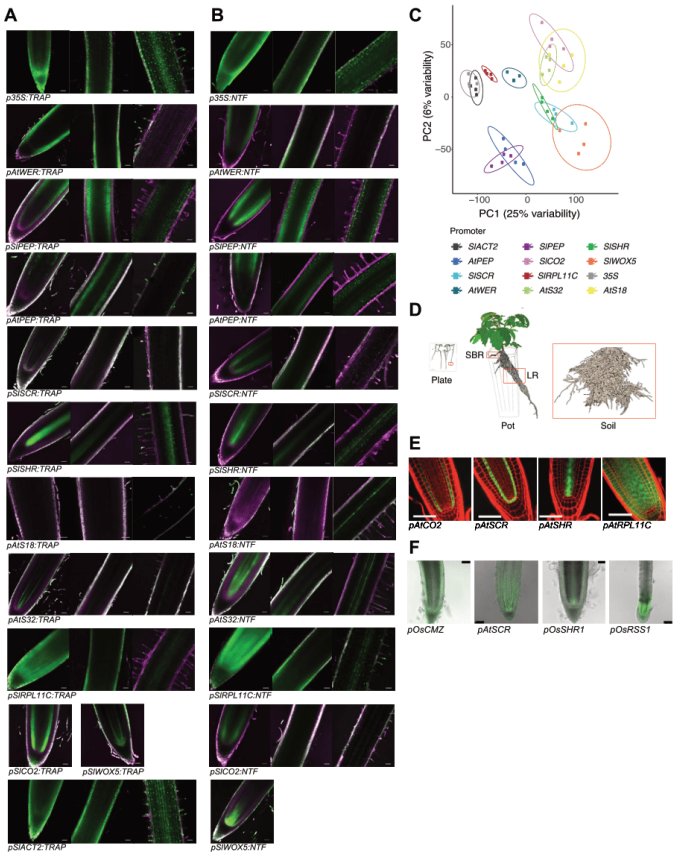


Figure S1. GFP expression in the TRAP and INTACT (nuclear tagging fusion [NTF]) lines and reproducibility of translatome biological replicates, related to Figure 1 and Table S1

(A-B) GFP expression for tomato promoter:TRAP marker lines (**A**) and promoter:NTF (nuclear tagging fusion) marker lines (**B**). The three panels represent the three root developmental zones; meristem, elongation zone and maturation zone. GFP signal is represented in green, autofluorescence in magenta, and the overlay of the two in white. Scale bars = 50 μm. (**C**) Principal component (PC) analysis of tomato marker-line derived translatomes. Ribosome-associated transcript abundance after normalization to library size and batch effect correction. Each sample is indicated by a dot and colored by the marker-line. (**D**) Line drawings of tomato root systems grown in the three growth set-ups; plate, pot and field. The drawings are in proportion to each other, and for pot set-up the drawing represents a washed rootball. Red squares indicate the sampled material from each set-up. SBR: shoot-borne root, LR; lateral root. **(E-F)** Expression patterns of GFP (green color) in the *Arabidopsis* (**E**) and rice (**F**) TRAP marker lines selected for the multi-species analysis. Red color denotes propidium iodide staining for *Arabidopsis*. *Arabidopsis* data is adapted from Mustroph et al. (2009). Scale bars represent 50μm for *Arabidopsis* (**E**) and 100μm for rice (**F**).

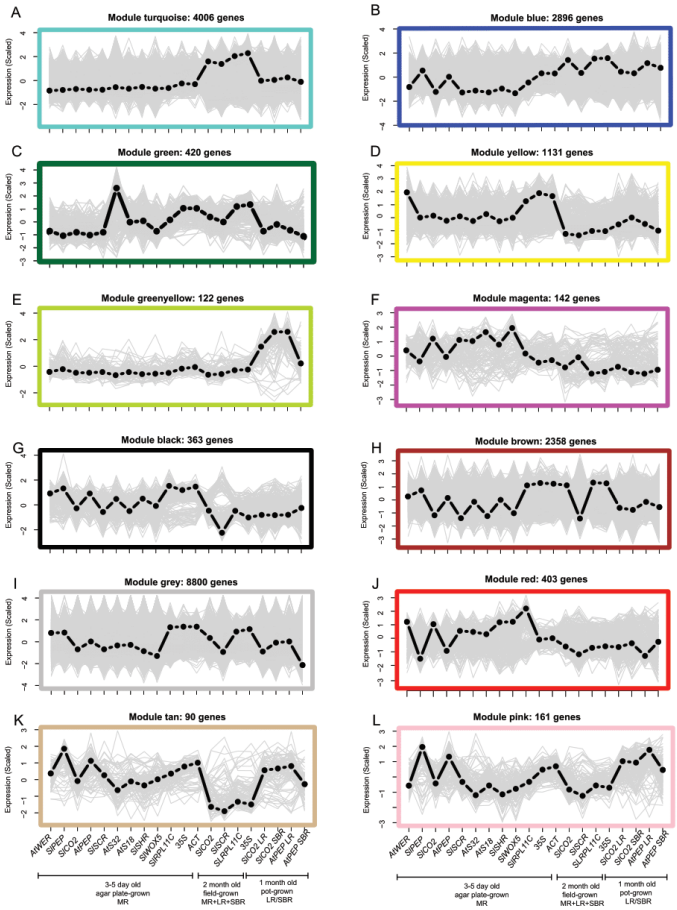
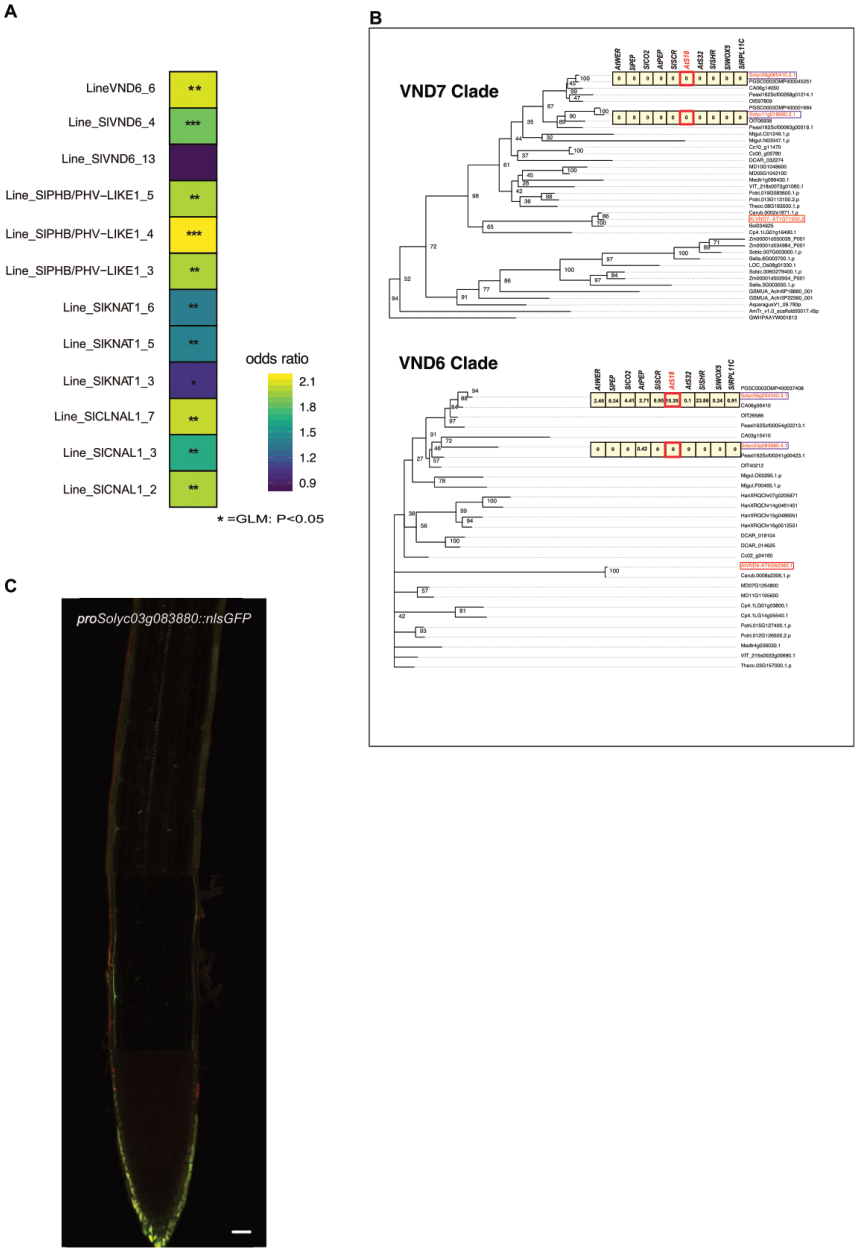


Figure S2. Expression profiles of WGCNA co-expression modules not shown in Figures 2D and 2E, related to Figure 2 and Table S2

(A) Field enriched co-expression module (**B**) Module of genes co-expressed in more typical cultivation conditions (pot- and field-grown plants). (**C**) Phloem and vascular initials co-expression module. (**D**) Whole root tissue co-expression module. (**E**) Module of genes co-expressed in lateral and shoot-borne roots of pot-grown plants. (**F**) Plate-enriched co-expression module. (**G**) Whole root tissue co-expression module. (**H**) General root tissue co-expression module of plants grown in sterile agar plates and the field. (**I**) Module of genes not assigned to any co-expression group. (**J**) Meristematic zone-enriched module. (**K**) Module with enrichment in the general cortex within the primary root. (**L**) Module with enrichment in the general cortex and inner cortex in primary and lateral roots. WGCNA co-expression modules with scaled expression values (y axis) across translatome profiles derived from different (i) promoters (*AtWER* = epidermis and lateral root cap; *SlPEP* = exodermis and cortex; *SlCO2* = meristematic inner cortex; *AtPEP* = inner cortex; *SlSCR* = endodermis and quiescent center; *AtS32* = phloem and vascular initials; *AtS18* = xylem and epidermis; *SlSHR* = vasculature; *SlWOX5* = quiescent center, vascular initials and meristematic pericycle; *SlRPL11C* = meristematic zone; 35S = nearly constitutive promoter; *ACT* = constitutive promoter; (ii) conditions (three-five day old plants grown on sterile agar plates in a growth chamber; two month old plants grown in the field; one month old plants grown in the growth chamber) and (iii) individual root types (MR - main root, LR - lateral roots, SBR - shoot-borne roots). Black dotted line = eigengene expression profile. The maximum peak of expression within the module is indicated by black font on top of the eigengene expression line. Grey line = expression values of all genes within the module. Most of the genes in these modules were positively correlated to the eigengene.

CellPress
OPEN ACCESS



(legend on next page)

Figure S3. Conservation of xylem regulation between *Arabidopsis* and tomato, related to Figure 3 and Table S3

(**A**) Quantification of abnormal xylem phenotypes in hairy root overexpression lines. Heatmap of $\log_2$ odds ratio of abnormal xylem phenotypes in 3 independent transgenic lines of *SlVND6, SlKNAT1, SlPHB/PHV-LIKE1* and *SlCNAL1*. n = ~15. See Table S3 for all odds ratios and p values. (**B**) Phylogenetic tree showing VND6 and VND7 clades only. Putative orthologs of *AtVND6* and *AtVND7* in *Solanum lycopersicum* are highlighted in red and purple. Numbers in boxes represent median normalized TPM from our TRAP-RNA-seq dataset in each cell type. Legend: AmTr: *Amborella trichopoda*, AT: *Arabidopsis thaliana*, Asparagus: *Asparagus officinalis*, Azfi: *Azolla filiculoides*, Bol: *Brassica oleracea*, Carub: *Capsella rubella*, CA: *Capsicum annuum*, Cc: *Coffea canephora*, Cp: *Cucurbita pepo*, DCAR: *Daucus carota*, Gb: *Ginkgo biloba*, HanXRQ: *Helianthus annuus*, MD: *Malus domestica*, Mapoly: *Marchantia polymorpha*, Medtr: *Medicago truncatula*, Migut: *Mimulus guttatus*, GSMUA: *Musa acuminata*, OIT: *Nicotiana attenuata*, GWHPAAYW: *Nymphaea colorata*, LOC_Os: *Oryza sativa japonica*, Peaxi: *Petunia axillaris*, Pp: *Physcomitrella patens*, MA: *Picea abies*, Potri: *Populus trichocarpa*, Semoe: *Selaginella moellendorffii*, Seita: *Setaria italica*, Solyc: *Solanum lycopersicum*, PGSC: *Solanum tuberosum*, Sobic: *Sorghum bicolor*, Thecc: *Theobroma cacao*, VIT: *Vitis vinifera*, Zm: *Zea mays*. (**C**) Confocal image showing lack of GFP expression in vascular tissue in *Solyc03g083880pro*::nlsGFP reporter line. Green signal in the cell wall represents autofluorescence. Red signal represents TagRFP (membrane-tagged RFP). Scale bar: 50 μm

CellPress
OPEN ACCESS

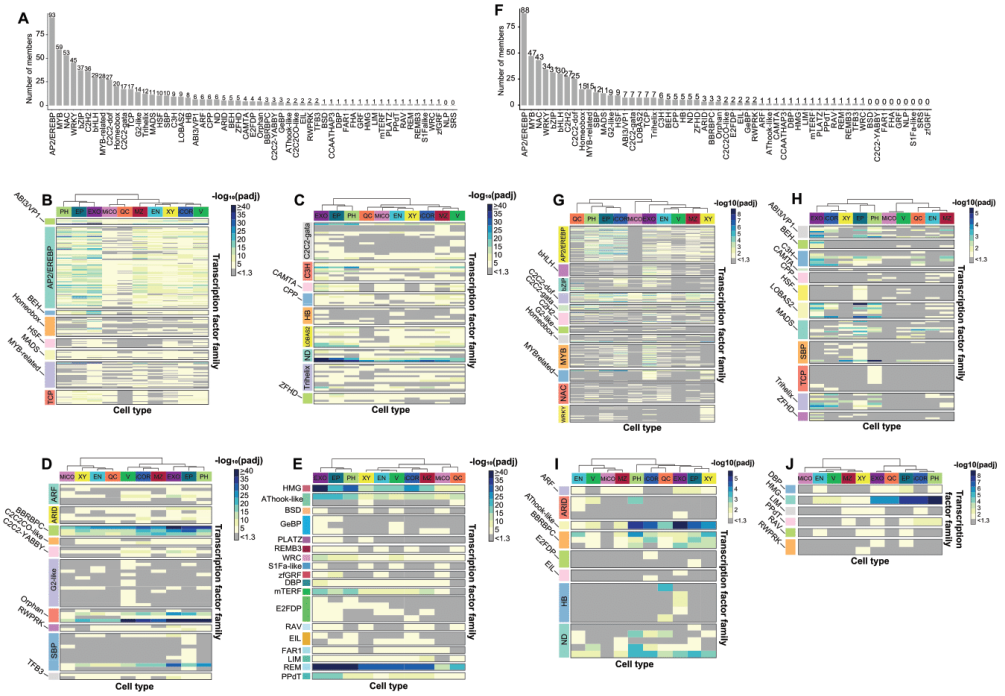


Figure S4. Transcription factor motifs enriched in 1-kb promoters and accessible regions near cell-type-enriched genes, related to Figure 4 and Table S4

1-kb promoters, transposase accessible regions 4-kb upstream of the transcription start site, or 1-kb downstream of the transcription termination site of cell type-enriched genes were used to perform motif enrichment. (**A-E**) Motif enrichment using 1-kb promoters of cell type-enriched genes. (**A**) Histogram demonstrating the number of identified transcription factor motifs in tomato. (**B-E**) All trees are hierarchically clustered to indicate similarity in enrichment across cell types. -$\log_{10}$ FDR adjusted p values are indicated according to the heatmap scale in the right part of the figure. (**B**) ABI3/VP1; BEH; Homeobox; Heat Shock Factor, MADS, and MYB-related transcription factor motif-enrichment. (**C**) Trihelix, C3H, CAMTA, CPP, Homeobox, LOB-AS2, C2C2-GATA and ZF-HD transcription factor motif enrichment. (**D**) ARF, ARID, BBRC/BPC; C2C2/CO-like; C2C2-YABBY, B2-like, Orphan, RWPRK, SBP and TFB3 transcription factor motif enrichment. (**E**) HMG, AThook-like, BSD, GeBP, PLATZ, REMB3, WRC, S1Fa-like, zfGRF, DBP, mTERF, E2FDP, EIL, FAR1, RAV, REM, PPdT, LIM transcription factor motif enrichment. (**F-J**) Motif enrichment using transposase accessible regions 4-kb upstream of the transcription start site or 1-kb downstream of the transcription termination site of cell type-enriched genes. (**F**) Histogram demonstrating the number of identified transcription factor motifs in tomato. (**G-J**) All trees are hierarchically clustered to indicate similarity in enrichment across cell types. -$\log_{10}$ FDR adjusted p values are indicated according to the heatmap scale in the right part of the figure. (**G**) AP2/EREBP, bHLH, bZIP, C2C2-dof, C2C2-gata, C2H2, G2-like, MYB, MYB-related, NAC and WRKY transcription factor motif-enrichment. (**H**) ABI3/VP1, BEH, C3H, CAMTA, CPP, HSF, LOBAS2, MADS, SBP, TCP, Trihelix and ZFHD transcription factor motif enrichment. (**I**) ARF, ARID, AThook-like, BBRBPC, E2FDP, EIL, HB and ND transcription factor motif enrichment. (**J**) DBP, HMG, LIM, PPdT, RAV, RWPRK, and zfGRF transcription factor motif enrichment. iCOR = cortex; EN = endodermis; EP = epidermis; EXO = exodermis; MiCO = meristematic inner cortex; MZ = meristematic zone; PH = phloem; V = vasculature; QC = quiescent center; XY = xylem.

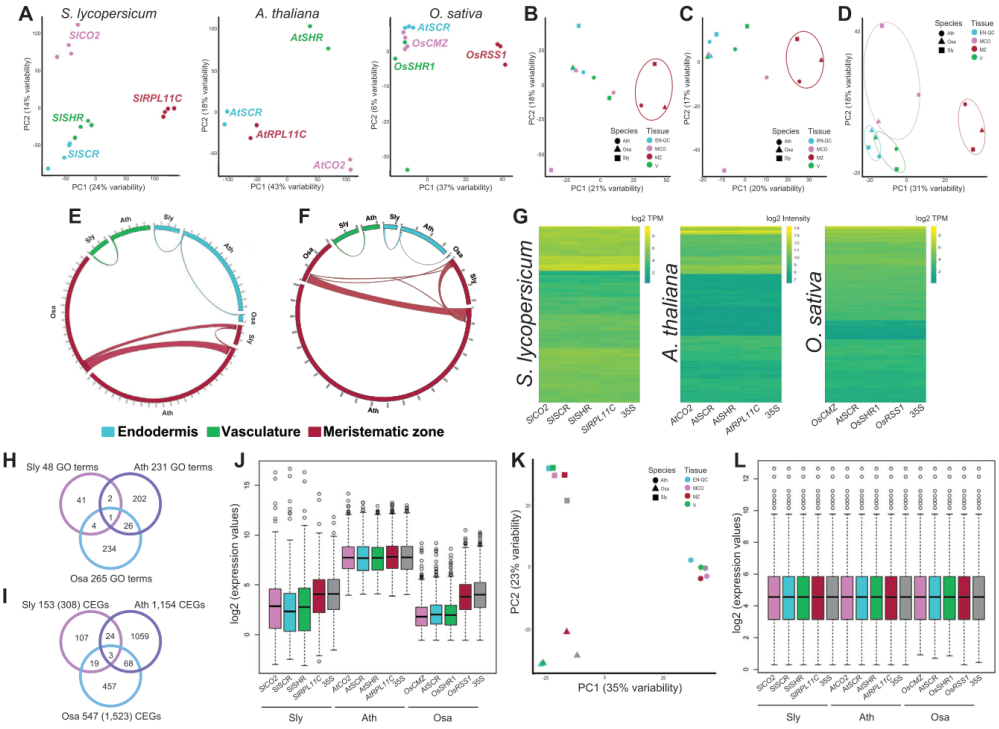


Figure S5. Multi-species analyses demonstrate similar translatome profiles of the meristematic zone compared with other cell populations and functional conservation of CEGs across species, related to Figure 5 and Table S5

(**A**) Clustering of cell population expression profiles based on top 5% most variable genes within each individual species (tomato, *Arabidopsis*, rice) using principal component analysis (PCA). (**B**) and (**C**) Clustering of cell population expression profiles between *Arabidopsis* (circle), rice (triangle) and tomato (square) using two independently derived orthology maps. (**B**) PCA plot of cell population expression of 3,505 1:1 orthologs, based on sequence homology to *Arabidopsis*, as determined by MapMan annotation files of tomato and rice. (**C**) PCA plot of cell population expression of 1,771 *Arabidopsis* and rice expressologs of tomato with an expression correlation coefficient > 0.6. (**D**) A Principal Component (PC) analysis of the expression of 1,550 consensus root TRAP expressolog between *Arabidopsis* (circle), rice (triangle) and tomato (square). Consensus expressologs have identical expressolog relationships independent of the reference species and positive expression correlations. (**E**) A Circos plot indicating overlapping GO terms of homologous cell type population between species. The width of the ribbon is proportional to the number of common terms. Numbers in the circle represent the number of common terms. (**F**) Circos-based plot indicating overlapping expressologs of homologous cell/tissue type enriched genes (CTEGs) between species. Ontology was determined based on 7,295 tomato and 6,059 rice root TRAP expressologs that have a reciprocal match and a positive expression correlation with *Arabidopsis* as a reference species. The width of the ribbon is proportional to the number of common expressologs. Numbers in the circle represent the number of expressologs within each group. (**G**) Expression patterns of constitutively expressed genes (CEGs) within each species. (**H**) Venn diagram of common and unique enriched GO terms. (**I**) Venn diagram of common and unique *Arabidopsis* root TRAP expressologs. Orthology was determined based on 7,295 tomato and 6,059 rice root TRAP expressologs that have a reciprocal match and a positive expression correlation with *Arabidopsis*. Numbers in parenthesis indicate the original number of CEGs detected within tomato and rice. (**J**) Boxplots of $\log_2$ expression values of homologous cell type populations before quantile normalization demonstrate different dynamic range of translatome data (microarray versus sequencing). Expression data for *Arabidopsis* are normalized $\log_2$ intensities and filtered $\log_2$ counts per million for rice and tomato. Tomato data were also corrected for the differences observed due to different sequencing dates. (**K**) Clustering of cell type and tissue expression profiles of 2,642 1:1:1 orthologs between *Arabidopsis* (circle), rice (triangle) and tomato (square) using principal component (PC) analysis without batch effect correction. (**L**) Boxplots of $\log_2$ expression values after quantile normalization. Ath = *Arabidopsis thaliana*; Osa = *Oryza sativa*: Sly = *Solanum lycopersicum;* EN+QC = endodermis and quiescent center; MCO = meristematic cortex; MZ = meristematic zone; V = vasculature.

CellPress
OPEN ACCESS

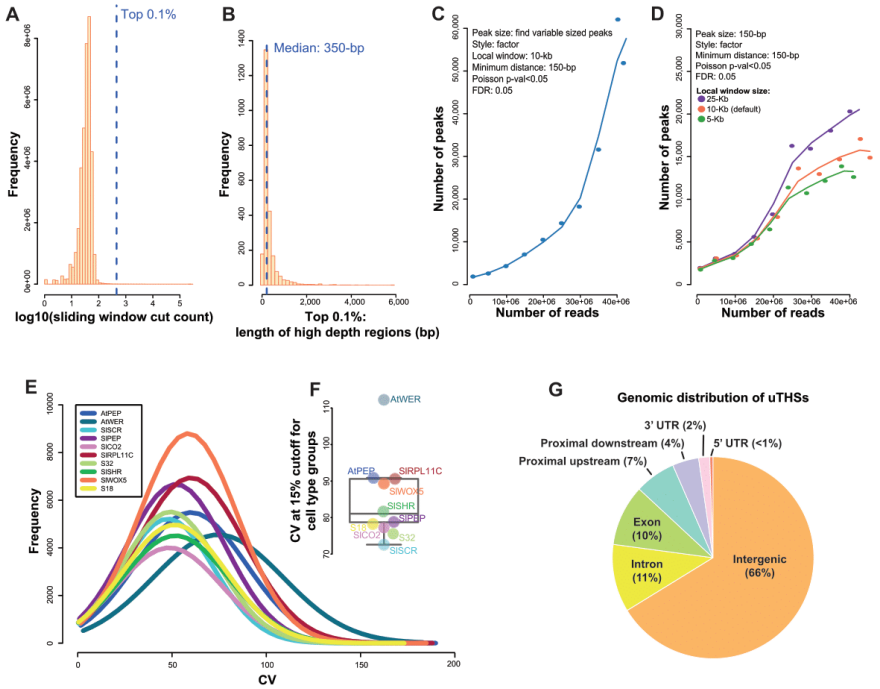


Figure S6. Summary of analysis methods for ATAC-seq, related to Figure 4 and Table S4

(**A-B**) Identification of high depth sequencing regions. (**A**) Cut counts from genomic DNA-based ATAC-seq were tallied across 150-bp sliding windows (step size 20-bp). x axis, $\log_{10}$ number of reads in window. y axis: frequency. The blue dashed line represents the top 0.1% most accessible windows. (**B**) x axis: length of top 0.1% high depth sequencing regions. y axis: frequency. This graph demonstrates the distribution of sizes for these high sequencing depth regions. Blue dashed line represents median high depth sequencing region size. (**C-D**) Choice of window size parameter for ATAC peak calling. (**C**) Peaks were called with increasing numbers of sub-sampled reads from the sample SlWOX5_O08. Here, the HOMER findPeaks parameters "-style factor," "-minDist 150," "-region" and "-regionRes 1." x axis: number of reads used to call peaks. y axis: number of peaks discovered. (**D**) Using 25 million sub-sampled reads from sample SlWOX5_O08, peaks were identified with HOMER using three different window sizes as well as the parameters "-size 150," "-minDist 150" "-region" and "-regionRes 1." x axis: number of reads used to call peaks. y axis: number of peaks discovered. (**E-F**) Threshold for identification of 15% most variable THSs for removal. THSs discovered within a marker line were merged and then used to tally the number of cut counts at that THS in each replicate. The coefficient of variation (CV) for cut counts was calculated at each replicate THS across all replicates. (**E**) x axis: CV across the replicate THSs for each cell type (see color legend) prior to filtering. (**F**) Boxplot of CV cutoff values for the top 15% most variable replicate THSs for each marker line (STAR Methods) to identify THSs with reduced variability across a marker line. (**G**) Genomic distribution of uTHSs. uTHSs (as defined by Maher et al., 2018) were found as described in STAR Methods. Proximal upstream = 2-kb upstream of transcription start site (TSS). Proximal downstream = 1-kb downstream of transcription termination site (TTS). Intergenic = more than 2-kb upstream of TSS or more than 1-kb downstream of the TTS.

Chapter II
Compositional and Transcriptional Alterations in the *Arabidopsis* Inflorescence Stem Mediated by *VND* Transcription Factors

ABSTRACT

This chapter focuses on investigating the composition of the cell wall in vascular tissue of the *Arabidopsis* inflorescence stem using *vnd6* and *vnd7* loss-of-function mutants. Previous reports have indicated that knockout mutations for these transcription factors (TFs) do not produce any noticeable phenotypic changes in xylem anatomy, morphology, or growth. Despite this, we hypothesized that since *VND6* and *VND7* target a broad range of major cell wall biosynthetic enzymes responsible for the biosynthesis of pectin, cellulose, hemicellulose, and lignin, they likely also play a role in controlling the composition of the xylem secondary cell wall (SCW). To investigate this, we employed glycome profiling in a tissue-specific manner using Laser Capture Microdissection (LCM) to dissect vessel elements and interfascicular fiber (IF) cells in the inflorescence stem. We detected perturbations in xylem and IF cell wall composition in *vnd6* and *vnd7* mutant backgrounds, which were not detectable at the resolution previously studied. We also integrated this glycomics analysis with transcriptional profiling of *vnd6* and *vnd7* single and double mutants to identify a group of candidate cell wall-associated genes that are direct targets of *VND6* and *VND7* that are misexpressed in the mutant backgrounds. We believe that these identified genes partially contribute to the observed global alterations in SCW composition in the mutant backgrounds.

INTRODUCTION

The plant cell wall is categorized into two groups: primary and secondary cell walls (SCW). In specialized cell types such as xylem, interfascicular fibers and anthers, a thick and rigid SCW is deposited inside the primary walls. The major components of the SCW are cellulose, hemicellulose (xylan and glucomannan) and lignin[1,2]. In the *Arabidopsis* inflorescence stem, there are two secondary cell wall-forming cell types: vessel elements and fibers (xylary or interfascicular). Transcriptional regulation of SCW development in *Arabidopsis* is controlled by a group of genes knows as *Secondary Wall NAC (SWN)* domain TFs, including *NAC SECONDARY WALL THICKENING PROMOTING FACTOR1 (NST1), NST2, SECONDARY WALL ASSOCIATED NAC DOMAIN PROTEIN1 (SND1)*, also called *NST3, VASCULAR-RELATED NAC-DOMAIN6(VND6) ,* and *VND7*[3,4]. In *Arabidopsis, MYB46/MYB83* are *SWN*-activated TFs that act as second-level master switches capable of activating the entire secondary wall biosynthetic program[5,6]. Additionally, *MYB58* and *MYB63,* which are directly regulated by *MYB46/MYB83*, have been shown to specifically regulate the genes involved in lignin biosynthesis[7]. Based on available evidence, the regulation of secondary wall biosynthesis involves a multi-leveled feed-forward loop regulatory structure that activates the transcription of secondary wall biosynthetic genes responsible for cellulose, xylan, and lignin[8]. *SND1* and it's close homolog *NST1* act redundantly as regulators of IF SCW development in *Arabidopsis*, while *VND6/7* along with other *VND* TFs (*VND1-5*) are the main SCW regulators in root xylem vessels[4,9]. *VND6* and *VND7* are additionally considered master regulators of SCW development in the inflorescence stem as they can drive ectopic xylem vessels and deposition of all three major secondary wall components,

including lignin, xylan and cellulose in other cell types[3,9,10]. Despite this, only dominant repression lines of *VND6* and *7,* and not loss-of-function mutant alleles have produced detectable xylem phenotypes in the root or shoot[9,11]. Thus, regulation of this process is highly redundant and combinatorial. However, no comprehensive analysis of SCW composition of inflorescence stem xylem vessels or interfascicular fibers in these mutants have been reported previously. We therefore aimed to further investigate *VND6/7*-dependent transcriptional and compositional variations in the *vnd6* and *vnd7* mutants in the inflorescence stem. Our results establish that *vnd6* and *vnd7* mutants indeed have altered SCW composition in xylem vessel elements and interfascicular fiber cells in the stem. Furthermore, the expression patterns of several cell-wall associated genes that are likely direct targets of these TFs are disrupted in the mutants and are candidates for control of these specific classes of SCW components.

RESULTS

In the *Arabidopsis* root, *VND6* and *VND7* single and double mutants have no xylem phenotypes. To investigate whether the mutant alleles of *VND6* and *VND7* TFs have any impact on lignin content in the inflorescence stem, cross-sections from 1cm above the rosette in Col-0*, vnd6* (*GABI_567F08*)*, vnd7* (*SALK_115812*) and double *vnd6/7* mutant lines were isolated and stained with Basic Fuchsin, a stain commonly used to visualize lignin in the SCWs, along with Direct yellow-96, a cellulose stain. Consistent with previous reports on the root, no detectable morphological differences in the xylem or interfascicular fiber cells was observed, and the fluorescent signal intensities for lignin measured from xylem vessel elements was comparable across all genotypes (Figure 1).

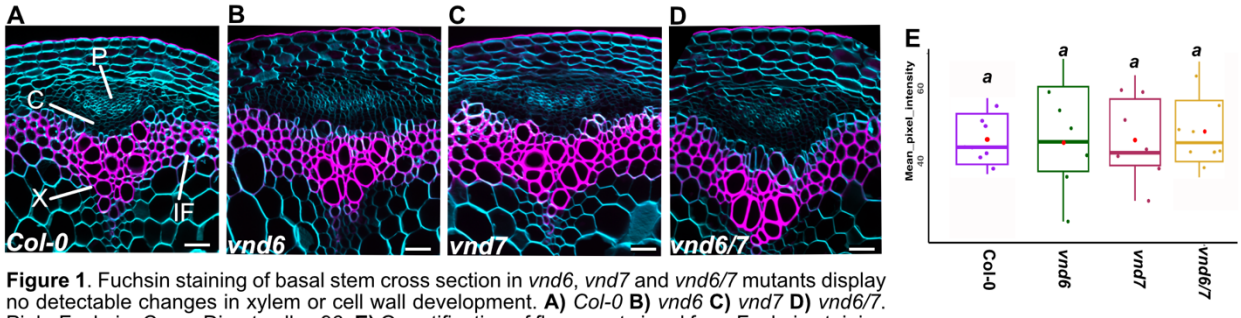


Figure 1. Fuchsin staining of basal stem cross section in *vnd6*, *vnd7* and *vnd6/7* mutants display no detectable changes in xylem or cell wall development. **A)** *Col-0* **B)** *vnd6* **C)** *vnd7* **D)** *vnd6/7*. Pink=Fuchsin, Cyan=Direct yellow96. **E)** Quantification of florescent signal from Fuchsin staining measured from xylem cells across 3 biological/2 section each. Statistical significance was determined by ANOVA with a post-hoc Tukey HSD test. X=xylem, C=cambium, P=phloem, IF= interfascicular fibers. Scale bar=20μm

A Method for Cell Type and Tissue-Resolution Cell Wall Composition Analysis

Although classical histochemical staining did not reveal any phenotypes for *vnd6* and 7 mutants in root or shoot tissue, it is possible that there are alterations in the structure and abundance of individual epitopes within the SCW that cannot be detected through histochemical techniques. Thus, to study the carbohydrate composition of xylem vessels and interfascicular fibers (IF) cells, two cell types in the stem that contain secondary cell walls, we performed glycome profiling on *Arabidopsis* Col-0 and *vnd6/7* mutants. To obtain these cell populations, we fixed a 1 cm section of the inflorescence stem from the apical and basal regions and embedded them in wax. We then transversely sectioned the tissue and used Laser Capture Microdissection (LCM) to isolate the xylem cells and IF from each section. This method allowed us to precisely collect and analyze the desired cell populations (Figures 2 & 3). To accurately perform LCM, it is essential to fix the tissue before sectioning to preserve the structural characteristics that distinguish different cell types and to ensure that the tissue maintains its original form allowing us to accurately

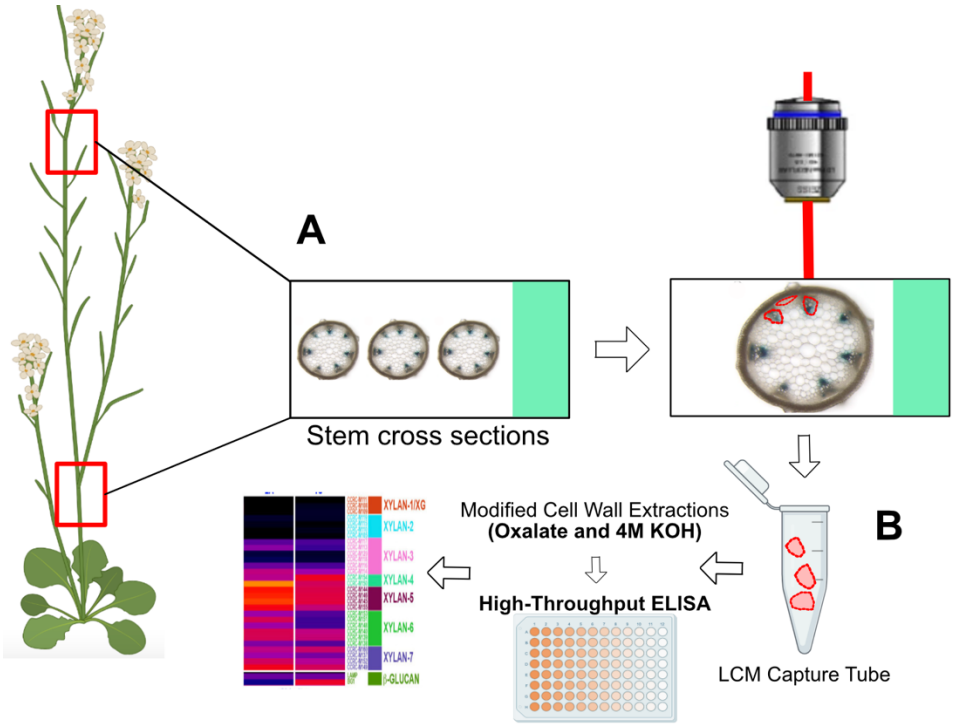


Figure 2. Workflow for tissue-specific LCM in Arabidopsis stem. **A)** inflorescence stem tissue was fixed, embedded and sectioned with a microtome. VB or IF were cut by LCM from each section and collected in a sample capture tube **B)** A modified extraction method was applied to collect fractions with for primary and secondary cell walls. Fractions were subjected to a high-throughput ELISA-based assay to generate a glycome profile displayed by a heat map for absorbance values for each mAb (representative epitopes).

identify and isolate specific cell populations. In addition, maintaining the integrity of the cell wall modifications and epitope structures during tissue preparation is crucial in glycome profiling. Therefore, we employed cold acetone fixation rather than a traditional fixative like formalin-acetic acid-ethanol (FAA). It is important to use a fixative that does not cause hydrolysis of sugars, as this can cleave many cell wall modifications such as fucose and xylose, potentially altering the epitope structures. By using cold acetone fixation, we were able to preserve the structural and biochemical features of the tissue[12]. To conduct a standard glycome profiling experiment, it is typically required to use a significant amount of whole organ cell wall residues, ideally a few hundred milligrams. These residues are then subjected to

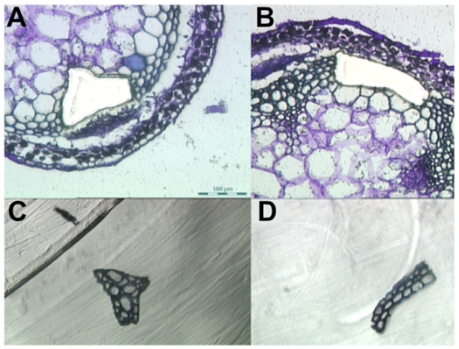


Figure 3. Xylem vessel elements **(A)** and Interfascicular fibers **(B)** in stem cross sections subjected to LCM and harvested tissue in capture tube**(C,D).**

sequential extractions using oxalate and/or carbonate to release pectin, followed by 1M and 4M KOH to extract hemicellulose and deeper embedded pectin glycans, and finally, chlorite to remove lignin (if present in notable quantities as in woody tissues)[13]. However, it is not feasible or reproducible to apply the full range of extractions to samples collected by LCM due to the small amounts of starting material. To overcome this challenge and recover sufficient amount of cell wall material, we reduced the number of extractions to only that of oxalate and 4M KOH fractions, which represent primary and secondary cell wall glycans respectively. This reduction in extractions reduces pipetting, washes, and transfers, helping to minimize the loss of the limited starting material. Standard glycome profiling involves performing total sugar estimation for each extract to ensure equal amounts are loaded for each sample. However, this assay is not possible with LCM-collected samples due to the low starting material. To normalize the input and accurately compare the samples, we harvested an equal area (800,000 μm^2 in total) from sections of equal thickness for each of 3 biological replicates to ensure equal loading of cell wall residues in the downstream ELISA assays (Figure 3). It is also important to acknowledge that this platform does not offer absolute quantification of cell wall epitopes, but rather serves as a qualitative indicator of abundance, as the extraction of cell wall material may be incomplete and not all tissue types may exhibit the same level of extractability[14]. Nonetheless, despite these limitations, this approach has proved to be a useful tool in comparative analyses of plant cell walls[15]. Taken together, these adapted methods allowed us to preserve the structural and biochemical features of the tissue and overcome the challenges of working with limited starting material in glycome profiling with LCM-collected samples to accurately compare the carbohydrate composition of the samples.

Glycomic Analysis of *vnd6* & *vnd7* Mutants Reveals Alterations in Cell Wall Epitope Profiles in the Inflorescence Stem

A. Xylem Vessel & Interfascicular Fiber Analysis of Cell Wall Composition:

Oxalate Extraction

By utilizing our optimized sample collection and glycome profiling method, we conducted a detailed analysis of xylem vessel and IF cell wall carbohydrate composition in *Arabidopsis thaliana* Col-0, *vnd6*, and *vnd7* mutants. Ammonium Oxalate extraction targets primary cell wall components including pectin and other loosely bound polymers. Pectin is a matrix polysaccharide in primary cell wall, but it is also present in smaller quantities in cells with secondary cell walls[16]. Another primary cell wall polymer is xyloglucan, which is similar to pectin in that it is also a matrix polysaccharide. Xyloglucan plays a crucial role in linking cellulose microfibrils in the primary cell wall of plants. unlike pectin and xyloglucan, xylan is mainly a SCW component in dicots, but it is also found in small amounts in primary walls, some of which are specific to primary walls. Biochemical information regarding xylan molecules in the primary cell walls of *Arabidopsis* is extremely limited[4].

Our findings revealed that while there were no notable changes or trends in the composition of the xylem cell walls in the apical compared to basal sample of the Col-0 stem, a slight decrease in the levels of xyloglucan and xylan epitopes was observed in the basal sections relative to the apical for the IF samples. (Figure 4 – red box). This decrease is consistent with the developmental trajectory of IF cells, which differentiate from mostly parenchyma cells with primary cell walls in the apical region to fully differentiated fiber cells with thick secondary cell walls. This transition from primary to

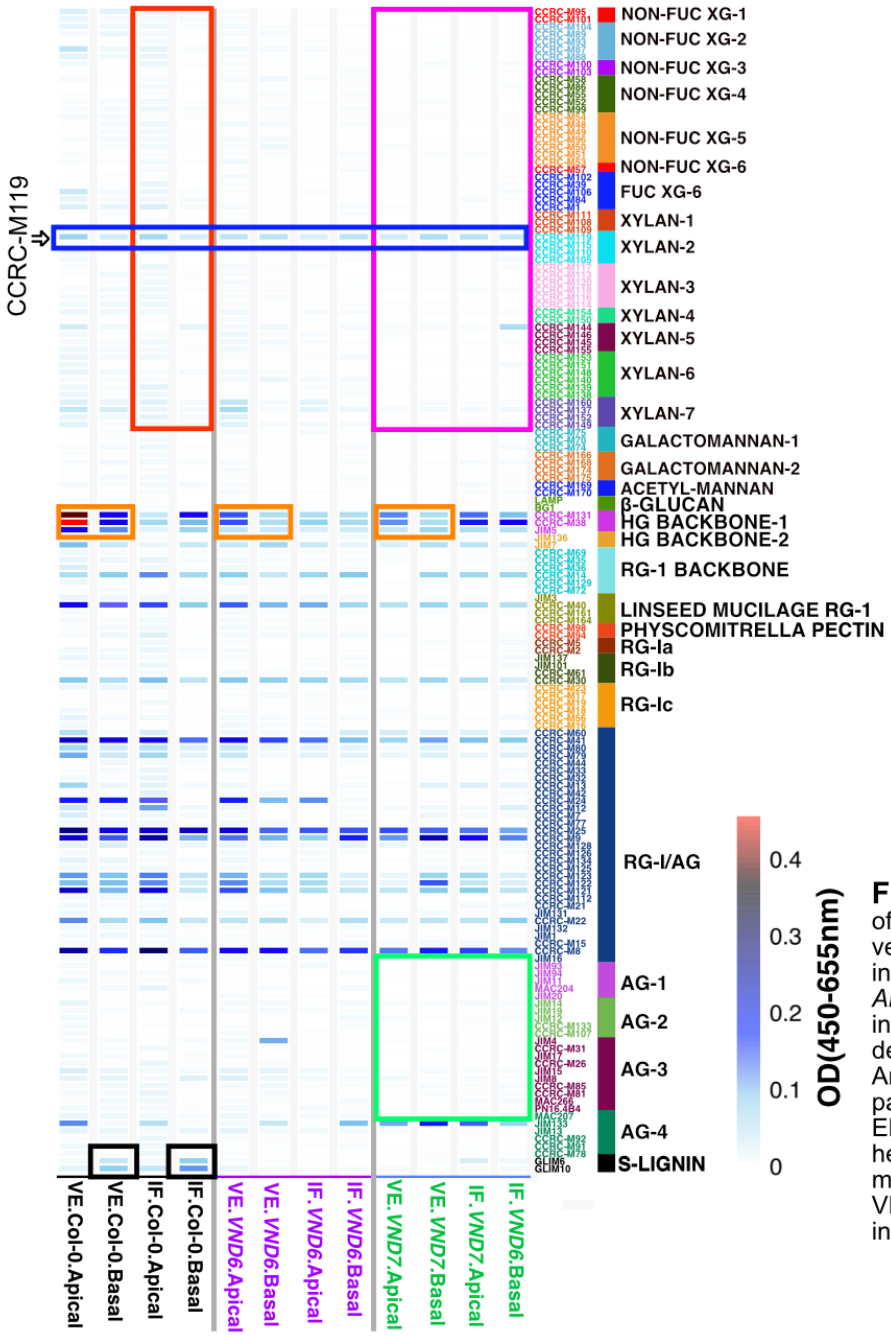


Figure 4. Glycome profiling of AO extracted cell walls from vessel elements and interfascicular fibers from *Arabidopsis thaliana* inflorescence stem at two developmental stages. Antibody binding strength for a panel of 155 mAbs used in the ELISA assay is visualized by a heat map. Color intensity is mean of 3 biological replicates. VE = vessel elements; IF = interfasciualr fiber.

secondary cell walls can result in reduced extractability with ammonium oxalate, which targets primary wall components, or it can lead to less efficient extraction of loosely incorporated cell wall components, such as xyloglucan/xylan, due to their increased cross-linking with secondary cell wall polymers in the basal section. There is also a general trend for reduced levels of the majority xyloglucan/xylan glycans in the xylem and

IFs of both apical and basal sections for the *vnd7* mutant compared to Col-0, (Figure 4 – pink box). An exception to this generality is that the xylan-2 epitope (CCRM-M119) was detected at relatively high levels in all genotypes and at all stages of development in the oxalate fraction, which is mainly rich in primary cell wall residues. In contrast, when the *Arabidopsis* stem is analyzed at the whole organ level on the same platform, this epitope does not seem to be detected in the oxalate extract[17]. This difference in detection in whole organ and LCM-based glycome profiling may be attributed to the tissue-specific sampling in this study, which has increased sensitivity to determine cellular resolution of epitopes. Our results suggest that CCRC-M119 may be an uncharacterized primary cell wall-specific xylan polymer that can be investigated further.

Pectin deposition is higher in apical compared to basal sections from xylem vessels; and differences are also observed in the *vnd6* and *vnd7* mutants relative to wild type (Figure 4 - orange box), as observed via monoclonal antibodies (mAbs) for the HG backbone-1 (homogalacturonan). This clade specifically recognizes low or fully de-esterified HG epitopes[18]. Additionally, the *vnd7* mutant exhibited a lack of detection for AG1-3 epitopes compared to Col-0 (Figure 4 - green box). The reduced abundance of pectin polymers in the mutant lines suggests that *VND6* and *VND7* may have a role in pectin biosynthesis in vessels and fibers. S-lignin (GLIM-6 and GLIM-10) abundance is also clearly reduced in *VND* mutants in the basal samples(Figure 4 – black box). This is not surprising as *VND6* and *VND7* are the master regulators of secondary cell wall development in vessel elements in the stem[11], and their impaired function is expected to impact lignin, which is a major component of secondary cell walls.

B. Xylem Vessel & Interfascicular Fiber Analysis of Cell Wall Composition: 4M KOH Extraction

Extraction with 4 M KOH was selected for its ability to yield a greater quantity and variety of alkaline-soluble glycans from the cell wall, which can provide valuable insights into the glycans of the secondary cell wall. This extraction effectively isolates hemicelluloses and tightly integrated pectic polysaccharides, although our results indicate it can also include some lignin polymers (Figure 4 – black box). The analysis of 4M KOH-extracted material from Col-0 and *vnd6* and *vnd7* mutants revealed distinct signatures in epitope abundance. First, no difference was observed in epitope abundance patterns for xyloglucan and xylan epitope profiles in xylem vessel and inflorescence fiber cells in both apical and basal sections (Figure 5 – blue box). This is contrary to the findings for IF cells in the oxalate extract (Figure 4 – red box), where a slight shift in the abundance of xyloglucan and xylan epitopes was observed during stem development. The lack of difference in detected glycans in vessel elements in both extracts could be attributed to the fact that the cells were fully differentiated at both zones from which they were harvested, leading to similar composition. Additionally, compared to fibers, vessels tend to have a more advanced secondary cell wall formation at the same developmental zone, as well as a more homogeneous distribution of xylan labeling and patterns during maturation[19]. This could contribute to the pattern observed in our glycomic data. The most striking change observed in this extract is a severe reduction in signal from xyloglucans and xylan1/2/3 epitopes in both mutants, indicating xyloglucan and xylan synthesis is *VND6/7*-dependent in both cell types (Figure 5 – red box).

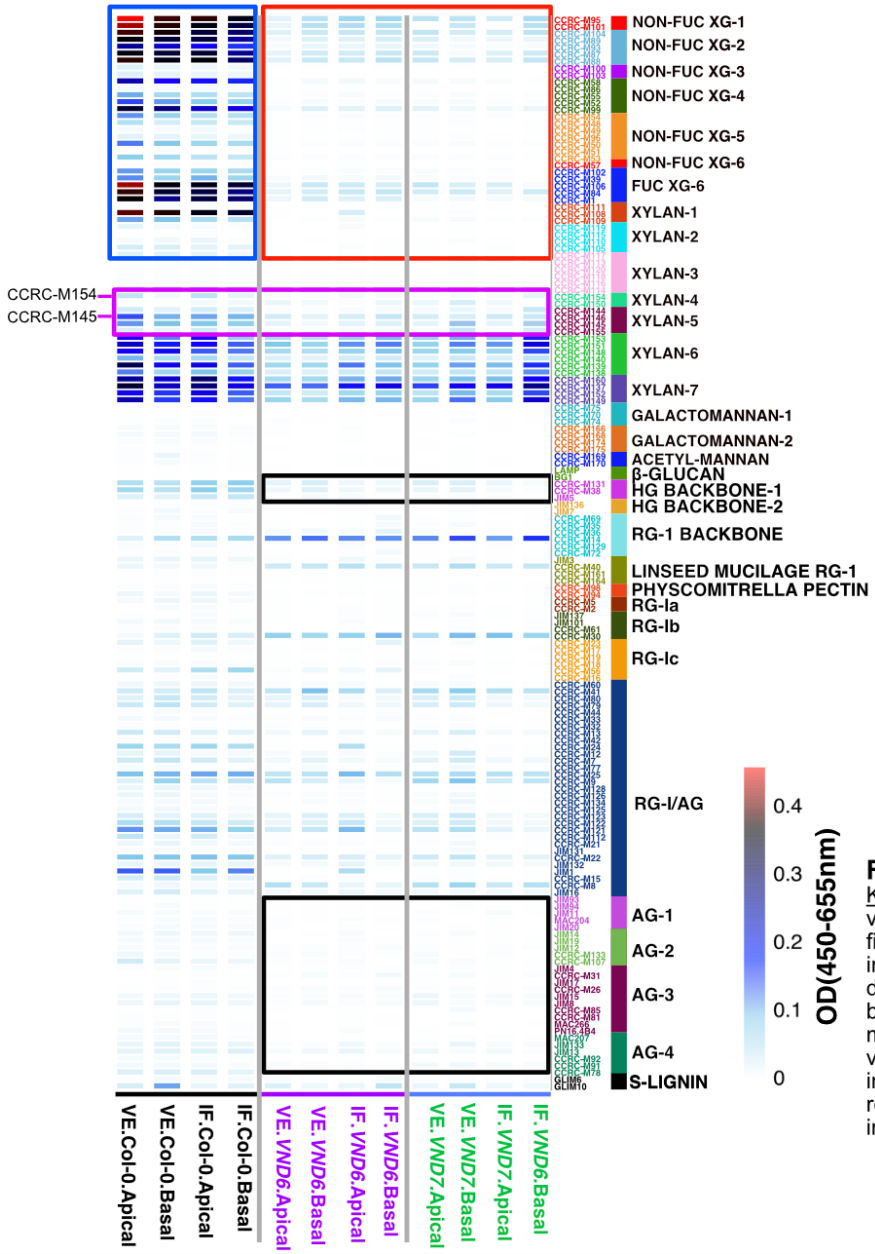


Figure 5. Glycome profiling of 4M KOH extracted cell walls from vascular bundles and interfascicular fibers from *Arabidopsis thaliana* inflorescence stem at two developmental stages. Antibody binding strength for a panel of 155 mAbs used in the ELISA assay is visualized by a heat map. Color intensity is mean of 3 biologcal replicates. VE = vessel elements; IF = interfasciualr fiber.

The mAbs of xylan 4/5 recognize xylan side-chains where xylose monomers are decorated with various groups, including arabinofuranose (Araf), glucuronic acid (GlcA), and [4-O-Methyl]Glucuronic acid ([Me]GlcA) groups[20]. In Col-0, the basal vessel elements and IF cells have lower levels of substituted xylan (xylan 4 – CCRC-M154) compared to the apical, and a similar trend is observed for CCRC-M145 (xylan 5) in vessel elements only (Figure 5, pink box). Therefore, these data reveal that there can be distinct

differences in the temporal and spatial distribution of xylan labeling among different cell types in *Arabidopsis* stem[19]. The *vnd6* mutant background shows a reduction in the abundance of these two epitopes in both apical vessel elements and IF cells (Figure 5, pink box), suggesting a VND-dependent alteration in side chain decorations of xylan polymers in the secondary cell wall possibly due to the absence of specific side chain moieties.

In contrast to Col-0, the *vnd6* and *vnd7* mutants displayed a decrease in the abundance of pectin epitopes from the AG-1,2,3,4 as well as HG-backbone1 clade in both the IF and VE (Figure 5 – black box). A similar trend was also observed in the oxalate extract, where AG1,2,3 showed a reduction in abundance in both xylem vessels and IF cells (more notable in the *vnd7* mutant) and HG-1 abundance was decreased in xylem vessels. These results further suggest that pectin deposition could be regulated by *VND6* and *VND7*, although the data presented here needs to be further validated by immunohistochemistry with individual monoclonal antibodies.

Immunohistochemistry Uncovers *VND6/7*-Dependent Patterns of Glycan Deposition in *Arabidopsis* Inflorescence Stem

Although glycome profiling has highlighted global differences in several cell wall glycans between Col-0 and *vnd6* and *vnd7* mutants, immunohistochemistry with individual epitopes can validate the glycome data acquired from LCM, which was qualitative in nature, and will confirm if differences in cell type specificity occur in *vivo*. The glycome analysis generally revealed similar phenotype in epitope abundance in the vessel elements and IF cells relative to Col-0. Therefore, we decided to focus on evaluating xylem vessel elements in immunolabelling experiments. We selected a set of 11 epitopes

representative of glycan clades including pectin, xyloglucan, xylan, and lignin which had different accumulation patterns in the AO and 4MKOH fractions (Figures 4&5). Of these, probing with 6 out of the 11 mAbs (Table 1) displayed less background noise and significant differences between genotypes (Figure 6).

Antibody	Clade
M-1	Fucosylated xyloglucan | **XG**
M-26	AG-3 (Arabinogalactan-3) | **Pectin**
M-95	Non-Fucosylated xyloglucan-1 | **XG**
M160	Xylan-7 | **Xylan**
GLIM-6	S-lignin
GLIM-10	S-lignin

Table.1 - Selected antibodies for immunolabeling

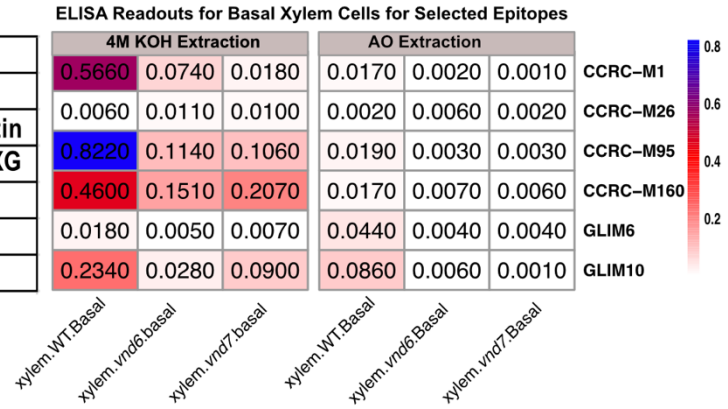


Figure 6. List of selected mAbs for immunolabelling. Heatmap on the right indicates the raw ELISA readings from AO and 4M KOH extractions (mean of 3 biologocal replicates) in glycome profiling. XG = xyloglucan.

Immunolabeling was performed on 40µm thick sections obtained from the basal region of the inflorescence stem (12-14 cm), except for S-lignin antibodies which were also tested on apical segments. Fluorescence intensity was measured from a pre-defined area in xylem vessel elements within the vascular bundle across a minimum of three sections per biological replicate. In addition to the single mutants, the *vnd6/7* double mutant was also evaluated in these experiments.

Immunohistochemical analysis of basal sections using CCRC-M95 revealed binding signals in cambium, phloem, and xylem cells, indicating a lack of specificity to cells with secondary cell walls. However, the staining exhibited a preference for developing xylem cells over metaxylem cells. In both single and double *vnd6/7* mutant lines, the density and intensity of the staining were significantly decreased compared to the wild type (Figure 7a).

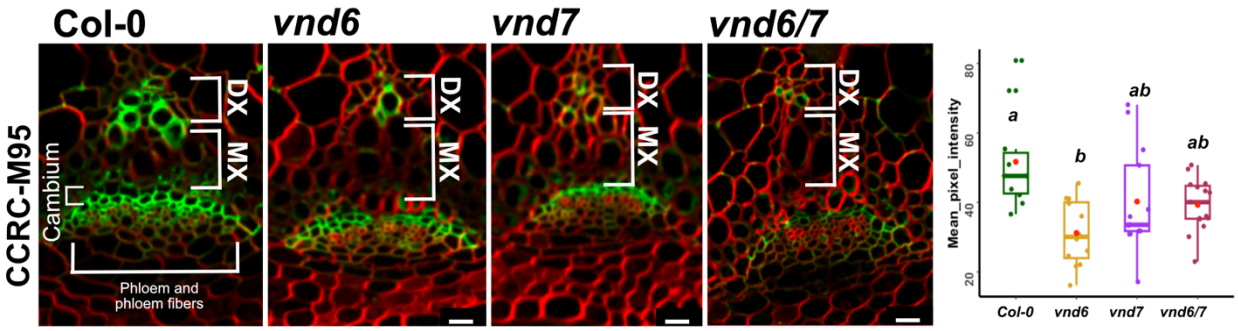


Figure 7a. Immunolabeling with CCRC-M95 mAbs for *in situ* visualization in Col-0 and *vnd6*, *vnd7* and *vnd6/7* mutants in *Arabidopsis* inflorescence stem. Section taken from 1cm from the base of the inflorescence unless indicated otherwise on the image. Image is taken zoomed-in on the vascular bundles. Red= Congo red counter stain; Green=Alexa Fluor 488 secondary antibody. Scale bar = 20µm. Signal intensity is quantified using a predefined area from 3 biological replicates/2 sections per sample. DX = developing xylem; MX = metaxylem.

The CCRC-M1 mAb, which targets fucosylated xyloglucan backbones, showed a comparable binding pattern to CCRC-M95. The labeling intensity decreased in both single and double mutants in comparison to Col-0.The reduction in labeling intensity was more pronounced in the double mutant line (Figure 7b) .

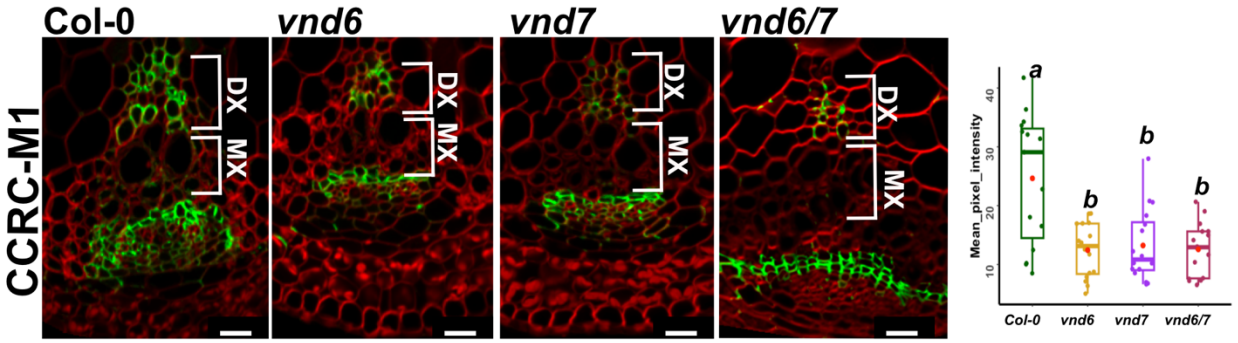


Figure 7b. Immunolabeling with CCRC-M1 mAbs for *in situ* visualization in Col-0 and *vnd6*, *vnd7* and *vnd6/7* mutants in *Arabidopsis* inflorescence stem. Section taken from 1cm from the base of the inflorescence unless indicated otherwise on the image. Image is taken zoomed-in on the vascular bundles. Red= Congo red counter stain; Green=Alexa Fluor 488 secondary antibody. Scale bar = 20µm. Signal intensity is quantified using a predefined area from 3 biological replicates/2 sections per sample. Statistical significance was determined by ANOVA with a post-hoc Tukey HSD testDX = developing xylem; MX = metaxylem.

The CCRC-M160 mAb, which targets the xylan-7 clade, exhibits a binding pattern that differs slightly from xyloglucans. Specifically, it shows a stronger affinity for mature xylem cells and appears to extend into the interfascicular fiber region as well. However, all mutants exhibit a significant reduction in binding compared to Col-0, with no discernible difference observed between the single and double mutants (Figure 7c).

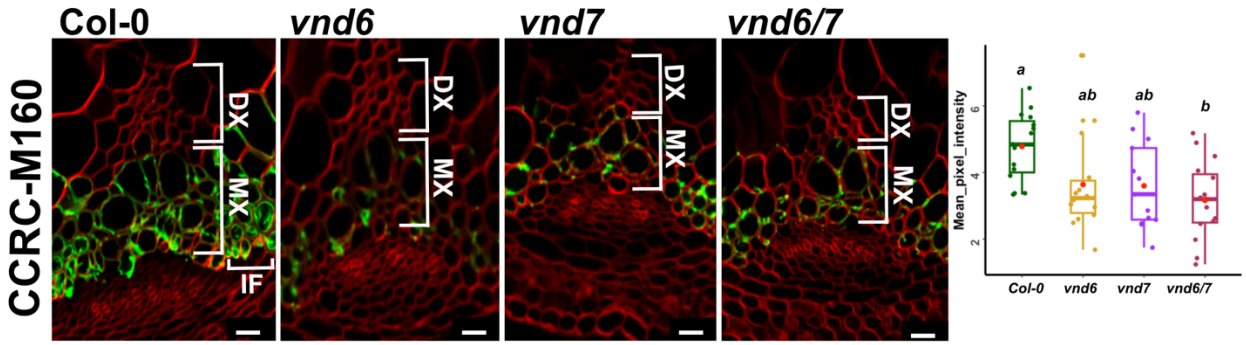


Figure 7c. Immunolabeling with CCRC-M160 mAbs for *in situ* visualization in Col-0 and *vnd6*, *vnd7* and *vnd6/7* mutants in *Arabidopsis* inflorescence stem. Section taken from 1cm from the base of the inflorescence unless indicated otherwise on the image. Image is taken zoomed-in on the vascular bundles. Red= Congo red counter stain; Green=Alexa Fluor 488 secondary antibody. Scale bar = 20µm. Signal intensity is quantified using a predefined area from 3 biological replicates/2 sections per sample. Statistical significance was determined by ANOVA with a post-hoc Tukey HSD test. DX = developing xylem; MX = metaxylem.

The pectin-targeted antibody CCRC-M26 (arabinogalactan: polymers of d-galactose and l-arabinose) specifically bound to both xylem elements and IF cells in the stem cross-section. Binding intensity did not differ in single mutant lines compared to the wild type. However, in the double mutant, there was a significant reduction in binding of this epitope (Figure 7d).

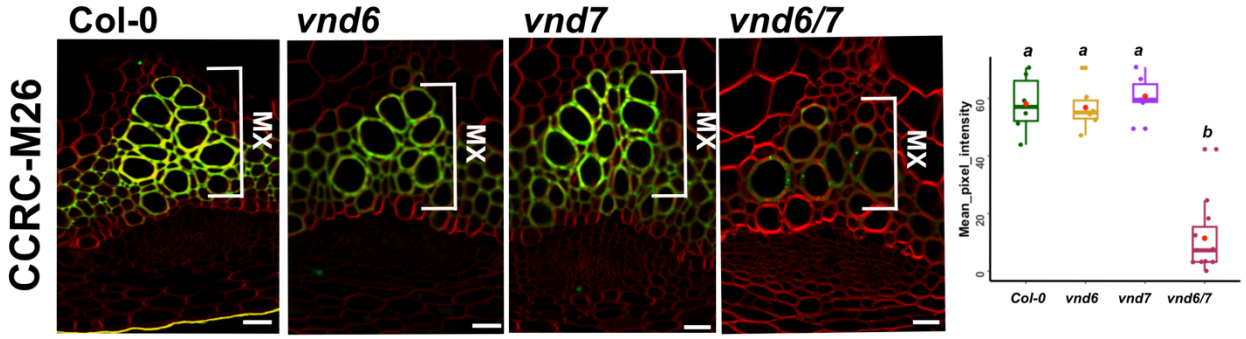


Figure 7d. Immunolabeling with CCRC-M26 mAbs for *in situ* visualization in Col-0 and *vnd6*, *vnd7* and *vnd6/7* mutants in *Arabidopsis* inflorescence stem. Section taken from 1cm from the base of the inflorescence unless indicated otherwise on the image. Image is taken zoomed-in on the vascular bundles. Red= Congo red counter stain; Green=Alexa Fluor 488 secondary antibody. Scale bar = 20µm. Signal intensity is quantified using a predefined area from 3 biological replicates/2 sections per sample. Statistical significance was determined by ANOVA with a post-hoc Tukey HSD test. DX = developing xylem; MX = metaxylem.

In apical stem sections of *VND* single and double mutant lines, both GLIM6 and GLIM10 mAbs that recognize S-lignin epitopes exhibited differential binding in developing xylem cells. The signal intensity was significantly reduced in all mutant lines. Similarly, GLIM-10 also showed a reduction in signal intensity in the apical sections across all mutant backgrounds, and there was a significant difference in binding intensity in the basal

sections. The labeling intensity in basal sections of GLIM-6 was comparable to the wild type (Figure 7e). Taken together, by analyzing a limited number of antibodies *in vivo*, we found that the differences in epitope abundance observed in *VND* mutants for select

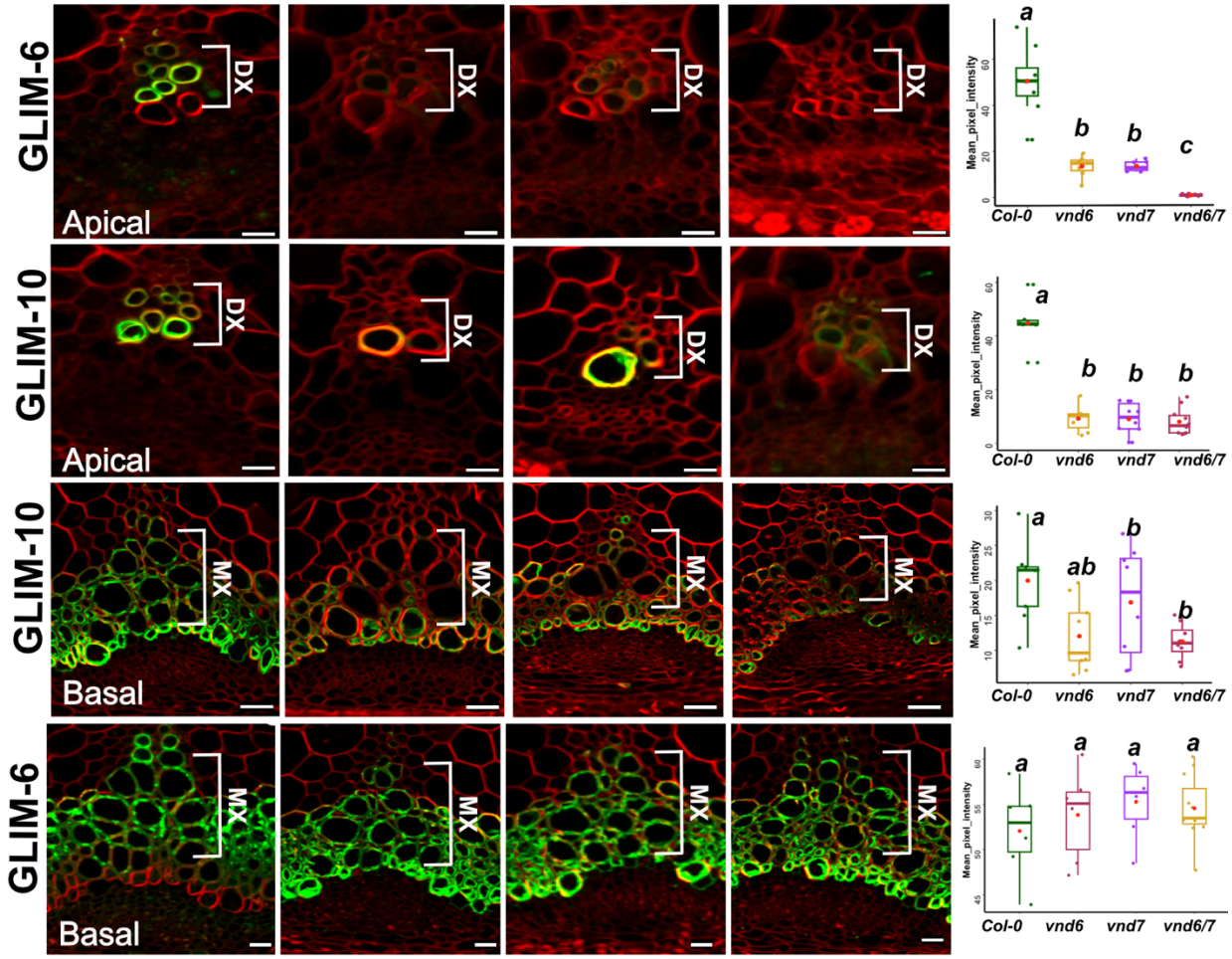


Figure 7E. Immunolabeling with GLIM -6 and GLIM-10 mAbs for *in situ* visualization of S-lignin in Col-0 and *vnd6*, *vnd7* and *vnd6/7* mutants in *Arabidopsis* inflorescence stem. Section taken from 1cm from the base and apex of the inflorescence stem. Image is taken zoomed-in on the vascular bundles. Red= Congo red counter stain; Green=Alexa Fluor 488 secondary antibody. Scale bar = 20µm. Signal intensity is quantified by using a predefined area from 3 biological replicates/2 sections per sample. Statistical significance was determined by ANOVA with a post-hoc Tukey HSD test. DX = developing xylem; MX = metaxylem.

antibodies from the xyloglucan, xylan, and pectin clades are consistent with glycome data, and these differences were not attributed to technical artifacts or possible changes in epitope structure that could occur during the chemical extraction of cell walls.

Transcriptome Profiling of *vnd6 and vnd7* and *vnd6/7* Apical and Basal Inflorescence Stem

To link our observations from glycome profiling and immunolabelling to an underlying molecular mechanism, we conducted transcriptome profiling from 1-2 cm apical/basal segments of elongated inflorescence stems in the *vnd6*, *vnd7*, and *vnd6/7* mutant lines. This sampling corresponds to the same developmental zones assayed in glycome profiling and immunolabelling studies. Multidimensional scaling (MDS) analysis revealed that developmental stage is a major factor contributing to transcriptional variation of these

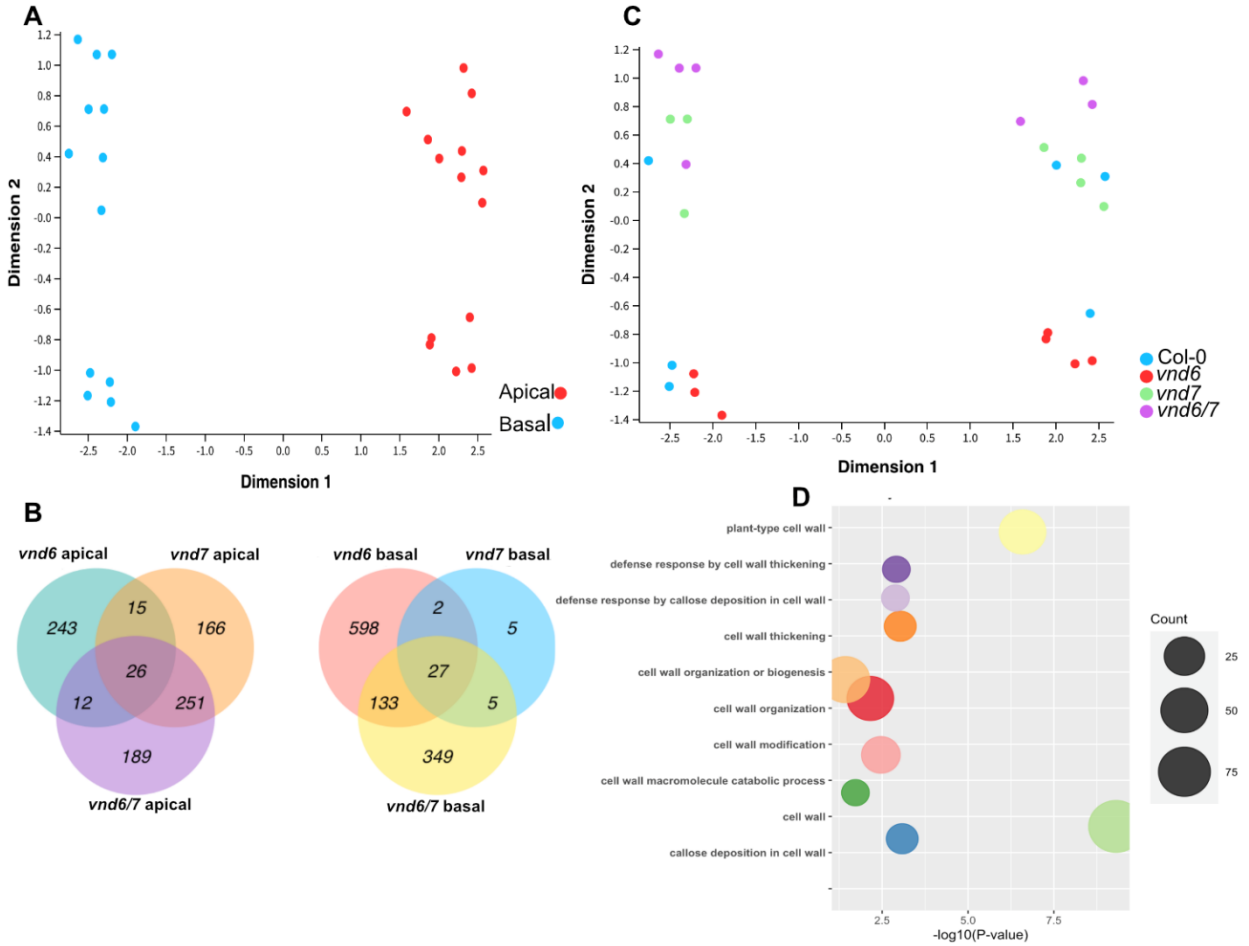


Figure 8. Transcriptome profiles in *vnd6,vnd7* and *vnd6/7* background. **A**) MDS analysis of Col-0 and mutants is derived by developmental stage - Red = basal; Blue= apical samples. **B**) MDS analysis of Col-0,*vnd6,vnd7* and *vnd6/7* by genotype from apical and basal segments **C**) Venn diagram indicating the number of unique and overlapping DEGs (adj.P.Val < 0.1; log(FC)>1) in single and double mutants in apial and basal samples. **D**) Go enrichment analysis from combined DEGs in *vnd6/7* apical and basal samples (p.value <0.05 ;Fisher exact test).

samples, as reflected by apical and basal samples grouping together and separating on

dimensions 1 and 2 (Figure 8A). Moreover, the mutant alleles exhibit distinct transcriptomes from each other, with the double mutant displaying the most divergent profile compared to the wild type (Figure 8B). Differentially Expressed Genes (DEG) were identified in the single and double mutants relative to wild type (Figure 8C - adj.P.Val < 0.1; log(FC)>1). For an overview on molecular processes enriched in mutant alleles, non-overlapping DEGs in all 3 mutant backgrounds relative to wild type from apical and basal samples (1822 genes) were compiled and gene ontology enrichment analysis was performed. Notably, this gene set showed significant enrichment for defense response-related genes (P.val<0.05). It is well known that cell wall mutants are associated with hypersensitive response to stimuli[21,22]. Further, other terms including cell wall biogenesis and organization, cell wall thickening, cell wall modification and genes related to cell wall polysaccharide metabolic processes were also enriched (Figure 8D).

Transcriptional Profile of Cell Wall-Associated DEGs in *vnd6 and vnd7* Single and Double Mutants

VND6 and *VND7* are crucial in the direct and indirect regulation of many of the cell wall biosynthetic enzymes. Any changes in the expression of these transcription factors would likely have an impact on the expression of their downstream target genes. To examine this further, a comprehensive gene list was compiled including xylem and SCW regulators, major primary and secondary cell wall biosynthetic pathways as well as their transcriptional regulators (Supplemental Table 1). This gene list was then intersected with DEGs from apical and basal samples in all 3 mutant backgrounds, resulting in the identification of a total of 136 cell wall-associated DEGs. Of these 136 DEGs, 40 were specific to the apical inflorescence stem, 77 were specific to the basal inflorescence stem,

and 19 were common in both groups. The common genes include those associated with cell wall modifications, regulation of vascular development, and the S-lignin biosynthesis pathway (Supplemental Table 1). The common DEGs have a highly similar expression pattern across all genotypes and developmental zones, suggesting a consistent regulatory response. Among them, *XTH4* is involved in xyloglucan biogenesis and is reported to have positive effects on SCW thickening[23]. *XTH4* displays relatively increased expression in all 3 mutant backgrounds in both the apical and basal zones.

VND6 and *VND7* regulate expression of their downstream targets by binding to an 11-bp tracheary element-specific *cis*-element (TERE) located in the promoters of their targets[24]. Therefore, to further narrow down the candidate genes impacting the observed cell wall composition phenotypes, we looked to see if any if any of the 136 cell wall-associated DEGs had a TE-element in their promoters within their 500bp upstream of their transcription start site. The TERE sequence was found in the upstream region of 37 out of the 136 DEGs at a significantly higher percentage than all genes in the *Arabidopsis* genome (Supplemental Table 2 - E-threshold 1.0E-4). Additionally, the DEGs were also mined for the *VND*6 binding motif[25]. The promoters of 58 DEGs were found to have a *VND6* binding motif (including *XTH4* (Supplemental Table 2)). Thirty-six genes had a TERE binding element and fourteen genes were found to include both TERE and *VND6* motifs (Supplemental Table 2). The CW-associated DEGs were also intersected with a vessel element-specific gene set in the *Arabidopsis* inflorescence stem[26] as well as *VND6* and *VND7* direct targets[27,28] (Supplemental table 3). . Although only three genes were identified as *VND6* or *VND7* direct targets, 2 of these have unknown function, and deserve further study. The third gene, GLAS2, is associated with β-1,4-galactan

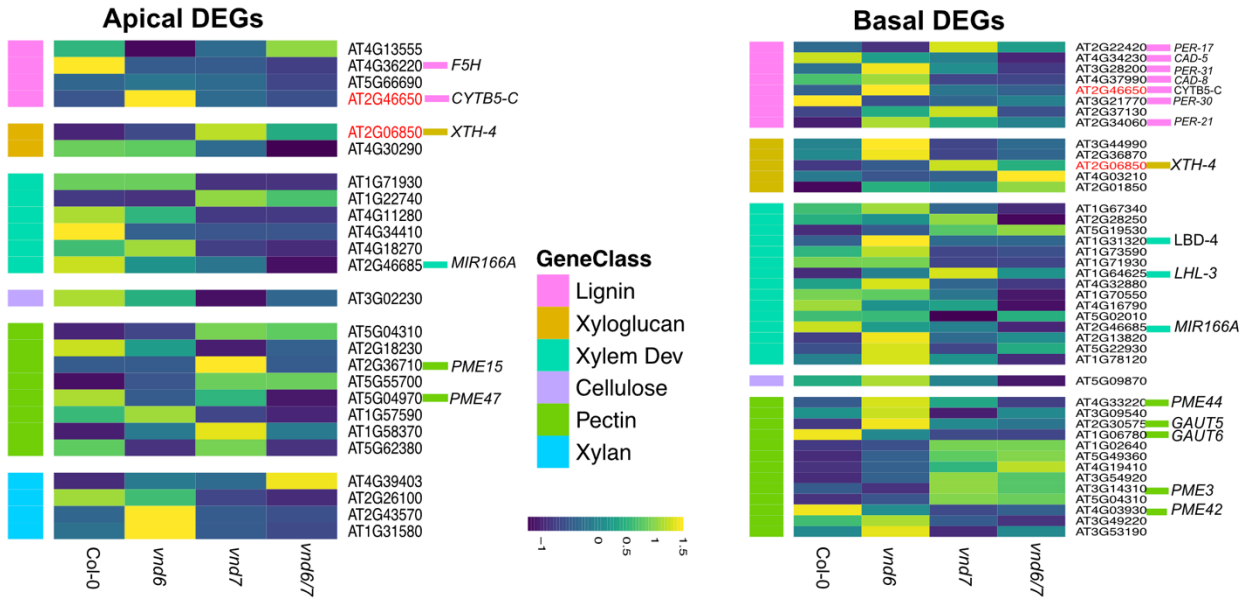


Figure 9. Select Combined DEGs corresponding to major classes of cell wall polysaccharides from **A**) apical and **B**) basal samples in *vnd6,vnd7* and *vnd6/7* lines. Color intensity is row normalized z-scores. Rows are separated based on their gene class. Common DEGs in apical and basal regions are highlighted in red.

synthesis, generally found in side chains of rhamnogalacturonan I, a major component of pectin polymer[29].

Interestingly, in our data, genes differentially expressed in a *VND6/7*-dependent manner in both apical and basal zones included those annotated as pectin-related. This is noteworthy as the role of pectin in SCW structure is considered to be minimal. Overall, the observed expression patterns within each polymer class in the apical or basal zone, both in single and double mutants, does not reveal any apparent trends within the different categories of cell wall genes. This further highlighting the complexity of gene expression and regulation in this context. Nonetheless, these expression profiles underscore a regulatory shift in expression of *VND6* and 7 cell wall-related genes in their respective mutant backgrounds.

DISCUSSION

VND6 and *VND7* along with 5 other members of *VND* TF family are expressed in inflorescence stem vessel elements[11]. Overexpression of these two TFs, as well as other *VND*s can induce ectopic SCW deposition and activate many of the downstream genes involved in biosynthesis of SCW polymers including cellulose, hemicellulose, and lignin. Single knock out mutant alleles of *vnd6* or *vnd7* show no obvious anatomical or developmental phenotypes in xylem vessels within the inflorescence stem or root[9,30]. Our observation in the inflorescence stem basal section using a general florescent lignin stain (fuchsin), align with previous studies in the roots, indicating no observable anatomical differences or loss of lignin deposition in the single *vnd6, vnd7,* or *vnd6/7* double mutants (Figure 1). Due to the high heterogeneity of cell wall composition in the inflorescence stem and the dynamic nature of glycan epitopes throughout inflorescence stem development, we utilized a glycome profiling platform in combination with LCM to obtain a comprehensive and high-resolution cell wall composition profile for vessel elements and interfascicular fiber cells in single *vnd6* and *vnd7* and their double mutant at two developmental stages(Figure 2). In contrast to previous reports indicating a lack of phenotype in the *vnd6* and *vnd7* mutants our approach demonstrates subtle alterations in the compositional profile and abundance of multiple classes of cell wall polymers in *vnd6* and *vnd7* mutants relative to the wildtype (Figures 4 and 5).

A small subset of mAbs were selected for *in situ* immunohistochemistry validation based on differences observed in glycome profiles relative to the wildtype (Figure 6). More than 50% of selected epitopes, covering major cell wall polymers including pectin, xyloglucan, xylan and lignin, mirrored the results from glycome profiling. It should not be surprising

that some epitopes did not produce overlapping results with glycome data as the two methods employ vastly different techniques and treatments to assess cell wall constituents. Immunolabeling not only reveals abundance of cell wall epitopes through signal intensity but also can provide valuable information on the spatiotemporal distribution of cell wall epitopes. In basal stem sections, CCRC-M1 and M-95 (XG-clade) are mainly labeled in developing xylem cells, whereas xylan-7 mAb signal is observed in mature vessel elements (Figure 7). This pattern corresponds with the fact that xylan is the major constituent of cells with already developed secondary cell walls, whereas xyloglucan is more prevalent in primary walls and hence present in developing xylem cells. The S-lignin antibody (GLIM-10) also provides *VND*-dependent temporal information on lignin deposition, interestingly, showing a greater reduction in lignin content in all mutants in the apical sections than basal sections with more developed xylem (Figure 7). While fuchsin is a general lignin stain, this demonstrates that differences in lignin composition can be revealed by this higher resolution biochemical analysis. Overall, these results emphasize the importance of complementary approaches when examining cell wall composition and underscores *VND*-mediated glycan and lignin remodeling in SCW-forming tissues in the stem.

Transcriptome profiling of the stem identified a group of DEGs associated with the regulation and biosynthesis of the cell wall in the mutant backgrounds. Many of these DEGs are known cell wall regulators and biosynthetic enzymes, making them potential candidate genes that could partially account for the observed alterations in cell wall composition. Among the xylem developmental regulators, several well-studied genes were identified as being downregulated in the *vnd6* mutant within the basal section,

including *MICRO-RNA166A* (*MIR166A*), which targets multiple *HD-ZIP III* transcription factors crucial for vascular differentiation and patterning[31]. Additionally, *LOB DOMAIN-CONTAINING PROTEIN 4* (*LBD4*), responsible for regulating the boundary and organization of vascular bundles, is also downregulated[32]. Furthermore, *LONESOME HIGHWAY LIKE 3* (*LHL3*), known for its significant role in early vascular development, was also found to be downregulated (Figure 9)[33].

Regarding the pectin polymer, expression data has revealed two DEGs in the *vnd6* mutant: *AT2G30575 (GAUT5)* and *AT1G06780 (GAUT6)*, which are homologs of each other. These genes are known to be involved in the biosynthesis of homogalacturonan (HG), a major cell wall pectic glycan. *GAUT6* is downregulated in all mutant backgrounds, which corresponds with the glycome data showing a reduced abundance of epitopes in the HG-backbone 1 and 2 clades in the mutant backgrounds (Figure 4, orange box; Figure 5, black box). Furthermore, it contains a TERE element and a *VND6* binding motif, suggesting it may potentially be a direct target of *VND6* or *VND7* (Supplemental table 2). Among other notable DEGs in the pectin category, there are five pectin methylesterases (PMEs) responsible for modifications of the pectin network including demethylesterification and acetylation. These enzymes are also known to play a role in maintaining cell wall integrity[34]. Additionally, *BETA-D-XYLOSIDASE* (BXL1), another pectin modifier, is among the common DEGs (supplemental table 1) and is associated with tissues undergoing secondary cell wall thickening[35]. These findings highlight the important and previously unknown role played by *VND6* and *VND7* in regulating the expression of genes involved in pectin biosynthesis and modification.

We identified 12 DEGs that overlapped with the lignin-associated processes (Figure 9). Among the apical DEGs, *AT4G36220*, which encodes for *F5H* (*FERULIC ACID 5-HYDROXYLASE)* is the rate-limiting enzyme for converting guaiacyl monolignol (G-monolignol) to syringyl monolignol (S-monolignol) in the lignin pathway[36]. This gene is downregulated in all mutant backgrounds (Figure 9). This finding is consistent with our observations from glycome profiling (Figure 4 - black box) and immunohistochemistry (Figure 7), both of which indicate a reduction in the S-lignin polymer in all mutant lines (Figure 4 - black box; Figure 7). In addition, The *F5H* gene has a putative *VND6* binding motif in its promoter region, suggesting it may be a target of *VND6* and could play a role in the S-lignin phenotype observed in vessel elements of the inflorescence stem in the mutants. Further, among common DEGs in apical and basal zones, *CYB-C5* is known to acts as a component that contributes to lignin biosynthesis, especially S-lignin formation. Interestingly, this gene is upregulated in the vnd6 background. It is possible that *VND6* acts as repressor of this gene. There are also several *PEROXIDASES* (*PER17,19,20,31*) and *CINNAMYL ALCOHOL DEHYDROGENASE* (*CAD5&8*) among DEGs within the basal section in the lignin category. Both *CAD* genes are downregulated while peroxidases have a mixed expression pattern. The combination of described DEGs associated with lignin could contribute to the lignin phenotype observed in the mutants.

To conclude, this study presents a novel approach for detecting glycan changes with tissue and cell type-resolution. Our findings reveal previously undetectable cell wall composition alterations in *vnd6* and *vnd7* mutants, which were considered to have no mutant phenotype. By integrating transcriptomics and cell wall composition data, we demonstrate that compromised function of *VND6* and *VND7* impacts multiple groups of

cell wall biosynthetic enzyme, resulting in perturbations in their expression and altered deposition of major cell wall constituents in vessel elements and fiber cells in the *Arabidopsis* inflorescence stem.

Materials and Methods

Material Growth for LCM, immunolabelling and RNA-Seq

Arabidopsis thaliana seeds (Col-0, T-DNA mutants *vnd6 (*GABI_567_F08*), vnd7 (*SALK_115821) and *vnd6/7* were sterilized with 50% bleach for 9 minutes followed by extensive washes with water. Seeds were stratified at 4°C for 48 hours, following germination on full strength MS medium with 1% Bactoagar. Seven days post germination, the seedlings were transplanted into soil in a growth chamber and grown at 22°C under long-day conditions (16 hours of light and 8 hours of dark). The inflorescence stems were allowed to elongate to a height of 12-14 cm, at which point 1cm sections were harvested from either from the apex or base for further analysis.

Histology on stem cross section – Lignin and cellulose imaging

To obtain 40 µm sections from the 12-14cm elongated inflorescence stem, we took a segment 1cm above the rosette and embedded it in 6% agar. These sections were then stained with 0.2% Basic Fuchsin dye in ethanol for 20 minutes, followed by rinsing with DI water and a subsequent staining with 0.5% Direct Yellow-96 (dissolved in water). The sections were imaged using a Confocal Laser Scanning microscope(Zeiss LSM700) equipped with a 20X objective lens. Basic Fuchsin was excited at 550-561 nm and detected at 570-650 nm. Direct Yellow 96 was excited at 488 nm and detected at 519 nm.

Histology and Sectioning for LCM

To prepare the samples, 1 cm segments were taken from the basal and apical parts of the inflorescence stem (12-14 cm). These segments were then placed in cold acetone for 15 minutes under a vacuum. The fixative was replaced with fresh cold acetone and the samples were rotated at 40 rpm at 4°C overnight. The samples were then treated with different ratios of acetone: histoclear (3:1, 1:1, and 1:3) for one hour each, ending with a final treatment of pure histoclear with a few grams of paraplast chips (Sigma - P3558) added. The samples were rotated at 40 rpm overnight at room temperature (RT) and then infiltrated with paraplast at 57°C for 2-3 days, changing the wax solution twice daily. Finally, the tissue was embedded in wax using tissue blocks. To prepare the sections for LCM, thin 11um strips of transverse stem sections were cut. These strips were then placed on PEN membrane glass slides(LCM0522) mounted in dH2O at 55°C using a hot plate. The slides were further dried at 42°C for 30-40 minutes. Before being examined, the slides were treated with histoclear and ethanol for 5 minutes each to remove the paraffin. The sections were then stained with toluidine blue (0.1%w/v in dH2O) for 2 minutes and dehydrated using 70% ethanol for observation under a microscope and LCM.

Laser Capture Microdissection

To collect the IF and xylem cells from the strips of sections on the slides, a Leica LMD7000 laser system was used. The areas to be harvested were outlined using the freehand tool. The tissue regions were then cut and collected using a laser beam with a diameter of 9 micrometers and pressure; laser pulse duration of 500 microseconds. The collected tissue was captured in Adhesive Cap 500 clear tubes containing 50ul of dH2O. A total area of 800 um^2 was collected for each sample.

Glycome Profiling

To perform glycome profiling, cell walls were extracted from laser-dissected material using a modified version of the method described in Pattathil et al. (2012). The samples were treated with 50mM ammonium oxalate, rotated for 24 hours, and centrifuged for 20 minutes. The supernatant was collected, and the resulting pellet was reconstituted with 4M KOH and rotated for an additional 24 hours. The KOH extracts were then neutralized with glacial acetic acid. Both the oxalate and KOH extracts were analyzed using a high-throughput enzyme-linked immunoabsorbent assay (ELISA) with a set of plant cell wall glycan-specific monoclonal antibodies (mAbs). A comprehensive list of the mAbs used can be found in monoclonal antibody database (http://www.wallmabdb.net). This comprehensive database offers detailed information about each monoclonal antibody (mAb), including the immunogen, antibody isotype, epitope structure (to the extent it is known), supplier details, and relevant literature citations.

Immunolabelling: Fixation

To maintain the integrity of the samples and antigenicity, small pieces of stem measuring 0.5 cubic centimeters were incubated overnight in a solution of 4% paraformaldehyde in PEM (50 mM piperazine-N,N'-bis[2-ethanesulphonic acid], 5 mM EGTA, 5 mM MgSO4; pH 6.9) buffer at 4°C. The fixation process was stopped by washing the samples three times with PEM buffer for 5 minutes each and once with deionized water for 5 minutes. The samples were then embedded in agarose.

Immunolabelling: Agarose-Embedding and Vibratome Sectioning

To embed the tissue, freshly made 7% agarose was placed in conical centrifuge tubes and allowed to cool for 1 minute. As soon as the agarose began to solidify, the plant

material was inserted into the gel matrix. Once the agarose had completely solidified, the bottom of the tube was cut off and the cylinder of agarose containing the tissue was removed. The cylinder was then attached to a specimen disk using crazy glue. The specimen disks were mounted on a buffer tray. Thin sections of 40 micrometers were cut from the tissue embedded in agarose blocks using a vibratome (Leica VT1000 S). These sections were collected and stored in distilled water at 4°C.

Immunolabelling

Sections were placed on glass slides and outlined with an Aqua Barrier Pen. The sections were treated with 15ul of 3% non-fat dry milk in PBS for 30 minutes at RT to block non-specific binding. The slides were incubated in a Petri dish containing soaked tissue wipes to create a humid chamber. After washing the sections with PBS for 5 minutes, they were incubated with 10-15 ul of primary antibody diluted 1:10 in PBS for 45 minutes at RT. The sections were washed with PBS three times, each for 5 minutes, before being incubated with a secondary antibody (Alexa Fluor 488 anti-mouse IgG (H + L),Invitrogen, A11001 - diluted 1:100 in PBS) for 1 hour at room temperature in the dark. The samples were extensively washed and stained for 1 minute in the dark with 0.1% Congo Red in dH_2O, a general cellulose binding stain, for 1-2 mins. The sections were rinsed with dH_20 one last time for 1 minute before being covered with a cover slip and ready for microscopy. As a negative control, sections that were not incubated with the primary antibody were also examined to detect any natural primary fluorescence present in the plant material. Alexa Fluor 488 excitation was with 488 nm laser and an emission peak at 496 nm - Congo Red was excited with a 560 nm laser, and emission collected at 640 nm.

RNA-Seq Library Preparation and Processing

Stem tissue (100mg – pooled from 1cm segments from apical or basal portion) was manually ground in liquid nitrogen and RNA was extracted with RNeasy Plant Mini Qiagen Kit (cat. nos. 74903) per manufacturers guidelines . DNAse treatment was applied on the column. RNA-seq libraries were synthesized with QuantSeq 3′ mRNA-Seq Library Prep Kit FWD (Lexogen- Cat no. 015 QuantSeq FWD 3' mRNA-Seq Library Prep Kit – with single indexing) according to the manufacturer's protocol. Libraries were sequenced at Novogene with NovaSeq 6000 S4 platform with 5 biological replicates per sample. Sequences were trimmed, and filtered and trimmed using FastQC v.0.11.7 ((https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Gene counts were obtained by mapping trimmed reads to the *Arabidopsis thaliana* genome (TAIR10) using STAR aligner v2.7.0f (Dobin et al, 2013) with options: --outFilterMultimapNmax 20 \ --alignSJoverhangMin 8 --alignIntronMin 20 --sjdbGTFtagExonParentGene gene_id --alignIntronMax 10000 --bamRemoveDuplicatesType UniqueIdentical --outFilterMismatchNmax 20 --outSAMtype BAM SortedByCoordinat --quantMode TranscriptomeSAM GeneCounts. Raw RNA-seq read counts were filtered to remove genes with zero counts across all samples. Counts were normalized and counts per million (CPM) were calculated using edgeR (Robinson et al., 2009) and limma-voom (Ritchie et al., 2015) Bioconductor packages in R. Differential expression analysis was performed in edgeR 640 (v3.34.1) - with adj.P.Val < 0.05 and logFC > 1.

GO Enrichment Analysis

Analysis was conducted on AgriGo by submitting the list of AGIs of unique genes significantly differentially expressed in all mutant backgrounds in apical and basal - The complete biological process GO annotation dataset for *Arabidopsis thaliana* was used for

testing statistical overrepresentation with options: significance level of 0.05 and Fisher's exact.

Motif Enrichment

We used FIMO (Charles E. Grant, Timothy L. Bailey, and William Stafford Noble, 2011) to identify sequences with TERE *cis*-element (CTTGAAAGCAA) or *VND6* binding motif (O'Malley *et* al., 2016) in the 500 bp region upstream of the cell wall-associated DEGs. MEME The parameters were set as follows: fimo --oc . --verbosity 1 --thresh 1.0E-4. meme/5.4.1 module was used to covert the TERE sequence to FIMO compatible (.meme) motif input file used with MEME suite.

Supplemental Table 1 - Common CW-asssociated DEGs in the apical and basal *vnd6,vnd7 and vnd6/7*	
gene_id	Name
AT1G03870	FLA9
AT1G31580	ECS1
AT1G72416	NA
AT1G78830	NA
AT5G04310	NA
AT5G15500	NA
AT5G49360	BXL1
AT5G51550	EXL3
AT5G62380	NAC101
AT3G13750	BGAL1
AT3G25700	NA
AT3G52840	BGAL2
AT2G06850	XTH4
AT2G15880	PEX3
AT2G46650	CYTB5-C
AT2G46685	MIR166A
AT4G18270	TRANS11
AT4G29310	DUF1005 family protein

Supplemental Table 2 - CW-associated DEGs with TERE or *VND6* binding motif					
***VND6* binding motif**	**Gene Name**	***VND6* binding motif**	**Gene Name**	**TERE element**	**Gene Name**
AT4G36220	CYP84A1	AT5G56870	BGAL4	AT2G28250	NCRK
AT5G66690	UGT72E2	AT2G06850	XTH4	AT5G19530	ACL5
AT2G43570	CHI	AT5G22930	NA	AT1G70550	NA
AT1G26380	FOX1	AT1G78120	TPR12	AT3G09540	NA
AT1G73750	NA	AT1G02640	BXL2	AT2G30575	GAUT5
AT2G06850	XTH4	AT5G49360	BXL1	AT1G06780	GAUT6
AT1G22740	RABG3B	AT4G19410	NA	AT2G46650	CYTB5-C
AT3G02230	RGP1	AT5G04310	NA	AT1G04980	PDIL2-2
AT2G18230	PPA2	AT4G03930	PME42	AT2G02100	PDF2.2
AT2G36710	PME15	AT3G49220	NA	AT3G25700	NA
AT5G55700	BAM4	AT3G13750	BGAL1	AT5G56870	BGAL4
AT1G10070	BCAT2	AT1G02630	ETN8	AT4G03210	XTH9
AT3G13750	BGAL1	AT3G06070	NA	AT3G54920	PMR6
AT4G39403	PLS	AT5G61130	PDCB1	AT3G53190	NA
AT5G04970	PME47	AT2G37130	PER21	AT5G61130	PDCB1
AT5G49360	BXL1	AT1G80110	PP2B11	AT2G34060	PER19
AT5G40210	NA	AT3G12500	CHI-B	AT1G28110	SCPL45
AT2G20750	EXPB1	AT3G52840	BGAL2	AT2G46330	AGP16
AT2G28250	NCRK	AT4G36120	FPP5	AT3G52840	BGAL2
AT5G19530	ACL5	AT1G66830	NA	AT4G36120	FPP5
AT1G73590	PIN1	AT5G55350	AT4	AT4G18270	TRANS11
AT1G71930	NAC030	AT5G47530	NA	AT2G46650	CYTB5-C
AT1G64625	BHLH157	AT1G06780	GAUT6	AT4G04745	NA
AT4G32880	ATHB-8	AT1G75500	WAT1	AT1G22740	RABG3B
AT1G70550	NA	AT1G80120	NA	AT3G02230	RGP1
AT4G16790	NA	AT5G61290	NA	AT1G58370	RXF12
AT4G33220	PME44	AT5G37790	NA	AT2G36710	PME15
AT2G30575	GAUT5	AT5G50150	NA	AT4G18270	TRANS11
AT5G02010	ROPGEF7	AT1G04980	PDIL2-2	AT3G52840	BGAL2
				AT1G45010	NA
				AT2G28790	NA
				AT3G07320	NA
				AT3G25700	NA
				AT2G26100	B3GALT12
				AT1G57590	PAE2
				AT5G53590	NA

Supplemental Table 3	
CW-associated DEGs identified as direct targets of *VND6* or *VND7*	
AT5G53590	NA
AT5G44670	GALS2
AT1G70550	NA
CW-associated DEGs with vessel element-specific expression	
AT1G10070	BCAT2
AT1G28110	NA
AT1G58370	RXF12
AT1G70550	NA
AT5G19530	ACL5
AT5G22930	NA
AT5G49800	NA
AT5G50150	NA
AT5G56870	BGAL4

Supplemental Table 4 - Direct Targets of VND7
https://doi.org/10.1093/mp/ssq062

AGI Code	Gene Annotation
At1g65570	polygalacturonase, putative / pectinase, putative
At5g40020	pathogenesis-related thaumatin family protein
At1g58300	heme oxygenase, putative
At1g26820	ribonuclease 3 (RNS3)
At3g52830	hypothetical protein
At5g02010	expressed protein
At1g11190	bifunctional nuclease (BFN1)
At4g08160	glycosyl hydrolase family 10 protein / carbohydrate-binding
At5g07800	flavin-containing monooxygenase family protein / FMO fam
At2g34790	FAD-binding domain-containing protein
At4g00220	LOB domain protein 30 / lateral organ boundaries domain p
At4g14940	copper amine oxidase, putative
At3g55720	expressed protein
At5g01930	(1-4)-beta-mannan endohydrolase, putative
At5g19870	expressed protein
At2g46760	FAD-binding domain-containing protein
At3g03540	phosphoesterase family protein /// phosphoesterase family
At3g52820	purple acid phosphatase (PAP22)
At5g60660	major intrinsic family protein / MIP family protein
At2g38320	expressed protein
At1g10810	aldo/keto reductase family protein
At2g40470	LOB domain protein 15 / lateral organ boundaries domain p
At3g47400	pectinesterase family protein
At5g17600	zinc finger (C3HC4-type RING finger) family protein
At1g20850	cysteine endopeptidase, papain-type (XCP2)
At3g26125	cytochrome P450, putative
At3g52900	expressed protein
At5g66390	peroxidase 72 (PER72) (P72) (PRXR8)
At1g70500	polygalacturonase, putative / pectinase, putative
At1g13470	expressed protein
At3g21550	expressed protein
At5g38610	invertase/pectin methylesterase inhibitor family protein
At5g56540	arabinogalactan-protein (AGP14)
At1g77330	1-aminocyclopropane-1-carboxylate oxidase, putative / AC
At4g35350	cysteine endopeptidase, papain-type (XCP1)
At4g37445	expressed protein
At3g50220	expressed protein
At3g10590	myb family transcription factor
At5g60530	late embryogenesis abundant protein-related / LEA protein-
At4g18425	expressed protein
At3g52620	hypothetical protein
At1g70550	expressed protein
At1g63650	basic helix-loop-helix (bHLH) family protein
At1g55180	phospholipase D, putative (PLDEPSILON)
At5g04200	latex-abundant protein, putative (AMC9) / caspase family pr
At2g31980	cysteine proteinase inhibitor-related
At1g55760	BTB/POZ domain-containing protein
At1g43790	expressed protein
At1g19230	respiratory burst oxidase protein E (RbohE) / NADPH oxida
At1g21810	expressed protein
At2g27920	serine carboxypeptidase S10 family protein
At3g45280	syntaxin 72 (SYP72)
At2g14620	xyloglucan:xyloglucosyl transferase, putative / xyloglucan e
At1g10800	hypothetical protein
At3g20960	cytochrome P450 family protein
At5g12870	myb family transcription factor (MYB46)
At4g16620	integral membrane family protein / nodulin MtN21-related
At2g14095	expressed protein
At4g32890	zinc finger (GATA type) family protein
At4g17220	expressed protein
At1g04490	expressed protein
At1g13920	remorin family protein
At3g13672	seven in absentia (SINA) family protein
At2g32610	cellulose synthase family protein
At4g10390	protein kinase family protein
At5g03260	laccase, putative / diphenol oxidase, putative
At2g44260	expressed protein
At3g27200	plastocyanin-like domain-containing protein
At4g24430	expressed protein
At1g78110	expressed protein
At1g23460	polygalacturonase, putative / pectinase, putative
At1g54790	GDSL-motif lipase/hydrolase family protein
At1g68500	expressed protein
At3g26190	cytochrome P450 71B21, putative (CYP71B21)
At4g34320	expressed protein
At2g38120	amino acid permease, putative (AUX1)
At1g16440	protein kinase, putative
At3g14360	lipase class 3 family protein
At5g45320	expressed protein
At4g36430	peroxidase, putative
At3g07510	expressed protein
At5g03510	zinc finger (C2H2 type) family protein
At5g48500	expressed protein
At5g60720	expressed protein
At2g44000	expressed protein
At2g43840	UDP-glucoronosyl/UDP-glucosyl transferase family protein
At4g23110	hypothetical protein
At3g49190	condensation domain-containing protein
At1g35780	expressed protein
At5g23020	2-isopropylmalate synthase 2 (IMS2)
At5g54670	kinesin-like protein C (KATC)
At5g57260	cytochrome P450 71B10
At5g15050	glycosyltransferase family 14 protein / core-2/I-branching e
At1g58070	expressed protein
At1g15960	NRAMP metal ion transporter 6, putative (NRAMP6)
At1g78340	glutathione S-transferase, putative
At1g53660	phosphate translocator-related
At3g16920	glycoside hydrolase family 19 protein
At5g34940	glycosyl hydrolase family 79 N-terminal domain-containing
At5g05270	chalcone-flavanone isomerase family protein
At5g14510	armadillo/beta-catenin repeat family protein
At4g36450	mitogen-activated protein kinase, putative / MAPK, putative
At2g03740	late embryogenesis abundant domain-containing protein / L
At2g24762	expressed protein
At3g53380	lectin protein kinase family protein
At3g54040	photoassimilate-responsive protein-related
At2g14890	arabinogalactan-protein (AGP9)
At2g15680	calmodulin-related protein, putative
At1g49660	expressed protein
At4g26140	beta-galactosidase, putative / lactase, putative
At5g12950	expressed protein /// expressed protein
At4g16790	hydroxyproline-rich glycoprotein family protein
At2g18720	eukaryotic translation initiation factor 2 subunit 3, putative /
At1g06020	pfkB-type carbohydrate kinase family protein
At3g19380	U-box domain-containing protein
At3g62160	transferase family protein
At1g30900	vacuolar sorting receptor, putative
At1g21340	Dof-type zinc finger domain-containing protein
At1g70670	caleosin-related family protein /// caleosin-related family pro
At3g52370	beta-lg-H3 domain-containing protein / fasciclin domain-con
At3g56240	copper homeostasis factor / copper chaperone (CCH) (ATX
At5g59845	gibberellin-regulated family protein
At5g26330	plastocyanin-like domain-containing protein / mavicyanin, p
At2g36650	expressed protein
At1g24600	expressed protein
At1g65860	flavin-containing monooxygenase family protein / FMO fam
At2g33560	spindle checkpoint protein-related
At5g39350	pentatricopeptide (PPR) repeat-containing protein
At5g25470	expressed protein
At1g72180	leucine-rich repeat transmembrane protein kinase, putative
At2g33570	expressed protein
At4g12430	trehalose-6-phosphate phosphatase, putative
At2g32510	protein kinase family protein
At2g27130	protease inhibitor/seed storage/lipid transfer protein (LTP) f
At1g27440	exostosin family protein
At2g21550	bifunctional dihydrofolate reductase-thymidylate synthase,
At1g71695	peroxidase 12 (PER12) (P12) (PRXR6)
At1g06120	fatty acid desaturase family protein /// fatty acid desaturase
At4g23690	disease resistance-responsive family protein / diligent famil
At4g34950	nodulin family protein
At3g54520	hypothetical protein
At5g43700	auxin-responsive protein / indoleacetic acid-induced protein
At2g04780	fasciclin-like arabinogalactan-protein (FLA7)
At5g51890	peroxidase-related
At1g28230	purine permease (PUP1)
At2g19780	leucine-rich repeat family protein / extensin family protein
At1g29240	expressed protein
At3g23110	disease resistance family protein /// disease resistance fam
At4g26890	protein kinase family protein
At5g38400	hypothetical protein
At2g04850	auxin-responsive protein-related
At1g59850	expressed protein
At1g43580	expressed protein
At1g70690	kinase-related
At3g16460	jacalin lectin family protein
At4g28650	leucine-rich repeat transmembrane protein kinase, putative
At2g32680	disease resistance family protein
At1g68700	hypothetical protein
At3g01070	plastocyanin-like domain-containing protein
At4g11470	protein kinase family protein
At4g12910	serine carboxypeptidase S10 family protein
At4g28140	AP2 domain-containing transcription factor, putative
At5g07080	transferase family protein
At5g15180	peroxidase, putative
At5g44670	expressed protein
At2g43060	expressed protein
At3g53350	myosin heavy chain-related
At5g01360	expressed protein
At5g47820	kinesin-like protein (FRA1)
At5g62180	expressed protein
At5g64340	expressed protein
At5g27930	protein phosphatase 2C, putative / PP2C, putative
At5g04390	zinc finger (C2H2 type) family protein
At2g38480	integral membrane protein, putative
At2g17710	expressed protein
At1g79030	DNAJ heat shock N-terminal domain-containing protein / S-
At3g14350	leucine-rich repeat transmembrane protein kinase, putative
At3g53040	late embryogenesis abundant protein, putative / LEA protei
At3g56060	glucose-methanol-choline (GMC) oxidoreductase family pro
At5g14230	ankyrin repeat family protein
At5g18970	AWPM-19-like membrane family protein
At5g47000	peroxidase, putative
At2g28760	NAD-dependent epimerase/dehydratase family protein
At1g63020	DNA-directed RNA polymerase alpha subunit family protein
At4g04460	aspartyl protease family protein
At3g63050	expressed protein
At5g18460	expressed protein
At5g45020	expressed protein
At5g54130	calcium-binding EF hand family protein
At5g16230	acyl-(acyl-carrier-protein) desaturase, putative / stearoyl-AC
At4g36540	basic helix-loop-helix (bHLH) family protein
At2g22270	expressed protein
At3g10740	glycosyl hydrolase family protein 51
At3g13100	ABC transporter family protein
At4g08250	scarecrow transcription factor family protein
At4g29680	type I phosphodiesterase/nucleotide pyrophosphatase fami
At3g59110	protein kinase family protein
At5g50170	C2 domain-containing protein / GRAM domain-containing p
At5g15740	expressed protein
At1g75730	expressed protein
At3g05270	expressed protein
At3g44450	expressed protein
At5g55950	transporter-related
At5g63930	leucine-rich repeat transmembrane protein kinase, putative
At5g67440	phototropic-responsive NPH3 family protein
At1g51940	protein kinase family protein / peptidoglycan-binding LysM
At2g32940	PAZ domain-containing protein / piwi domain-containing pro
At2g46225	expressed protein
At2g45040	matrix metalloproteinase
At2g36020	abscisic acid-responsive HVA22 family protein
At1g30440	phototropic-responsive NPH3 family protein
At3g01860	expressed protein
At3g14340	expressed protein
At4g01680	myb family transcription factor (MYB55)
At1g23030	armadillo/beta-catenin repeat family protein / U-box domain
At1g54510	protein kinase family protein
At1g50010	tubulin alpha-2/alpha-4 chain (TUA2)
At1g01240	expressed protein
At3g17110	hypothetical protein
At3g22070	proline-rich family protein
At4g09350	DNAJ heat shock N-terminal domain-containing protein
At3g45010	serine carboxypeptidase III, putative
At3g53620	inorganic pyrophosphatase, putative (soluble) / pyrophosph
At5g62960	expressed protein
At5g62880	Rac-like GTP-binding protein (ARAC10)
At5g64530	no apical meristem (NAM) family protein
At5g04840	bZIP protein
At2g34340	expressed protein
At2g21300	kinesin motor family protein
At1g48260	CBL-interacting protein kinase 17 (CIPK17)
At1g63520	expressed protein
At2g41770	expressed protein
At1g69680	expressed protein
At3g23340	casein kinase, putative
At3g24460	TMS membrane family protein / tumour differentially expres
At4g25780	pathogenesis-related protein, putative
At5g37680	ADP-ribosylation factor, putative
At5g48910	pentatricopeptide (PPR) repeat-containing protein
At5g15940	short-chain dehydrogenase/reductase (SDR) family protein
At2g37670	WD-40 repeat family protein
At1g79110	expressed protein
At1g76020	expressed protein
At1g32760	glutaredoxin family protein
At2g41800	expressed protein
At1g74440	expressed protein
At1g13310	expressed protein
At4g37740	expressed protein
At3g61960	protein kinase family protein
At5g59290	UDP-glucuronic acid decarboxylase (UXS3)
At5g63800	glycosyl hydrolase family 35 protein
At5g25180	cytochrome P450 71B14, putative (CYP71B14)
At5g27950	kinesin motor protein-related
At2g28200	zinc finger (C2H2 type) family protein
At3g45850	kinesin motor protein-related
At3g49070	expressed protein
At3g53280	cytochrome P450 71B5 (CYP71B5)
At5g49900	expressed protein
At2g28470	beta-galactosidase, putative / lactase, putative
At1g26920	expressed protein
At1g16500	expressed protein
At1g74690	calmodulin-binding family protein
At3g11600	expressed protein
At1g33265	expressed protein
At4g18050	ABC transporter family protein
At3g62020	germin-like protein (GLP10)
At5g03280	ethylene-insensitive 2 (EIN2)
At5g05820	phosphate translocator-related
At5g22910	cation/hydrogen exchanger, putative (CHX9)
At2g25440	leucine-rich repeat family protein
At2g21970	stress enhanced protein 2 (SEP2)
At2g38820	expressed protein
At1g28470	no apical meristem (NAM) family protein
At1g20180	hypothetical protein
At4g03450	ankyrin repeat family protein
At4g35550	homeobox-leucine zipper protein (HB-2) / HD-ZIP protein
At5g53590	auxin-responsive family protein

Supplemental Table 5 - Direct Targets of VND6
https://doi.org/10.1105/tpc.110.075036

At5g12870	MYB46
At3g08500	MYB83
At1g63910	MYB103
At5g18090	transcriptional factor B3 family protein
At5g44030	CESA4/IRX5
At5g17420	CESA7/IRX3
At4g18780	CESA8/IRX1
At2g32610	ATCSLB01 (Cellulose synthase-like B1)
At5g53340	galactosyltransferase family protein
At5g15470	GAUT14 (Galacturonosyltransferase 14)
At5g03260	LAC11 (laccase 11)
At2g47670	invertase/pectin methylesterase inhibitor
At3g47400	pectinesterase family protein
At4g14940	ATAO1 (Arabidosis thaliana amine oxidase 1)
At2g34790	EDA28/MEE23
At3g16920	chitinase
At1g32860	glycosyl hydrolase family 17 protein
At5g59290	UXS3
At2g28760	UXS6; catalytic
At1g50010	TUA2 (tubulin alpha-2 chain)
At1g04820	TUA4 (tubulin alpha-4 chain)
At4g14960	TUA6 (tubulin alpha-6 chiain)
At5g23860	TUB8 (tubulin beta-8)
At5g27950	kinesin motor protein-related
At3g53350	MIDD1
At5g54670	ATK3 (ARABIDOPSIS THALIANA KINESIN 3)
At5g65270	AtRABA4a
At4g09720	AtRABG3a; GTP binding
At2g38360	prenylated rab acceptor (PRA1) family protein
At3g24840	SEC14 cytosolic factor
At4g18640	MRH1 (morphogenesis of root hair 1)
At1g10850	ATP binding / protein serine/threonine kinase
At3g52820	ATPAP22/PAP22 (purple acid phosphatase 22)
At1g63120	ATRBL2
At2g17760	aspartyl protease family protein
At1g11190	BFN1 (BIFUNCTIONAL NUCLEASE I)
At4g29680	phosphodiesterase/nucleotide pyrophosphatase
At3g45010	SCPL48 (serine carboxypeptidase-like 48)
At2g27920	SCPL51; serine carboxypeptidase
At5g48940	leucine-rich repeat protein kinase
At1g60810	ACLA-2 (ATP-citrate lyase A-2)
At5g49460	ACLB-2 (ATP-citrate lyase B-2)
At4g17890	AGD8 (ARF-GAP DOMAIN 8); DNA binding
At5g14510	armadillo/beta-catenin repeat family protein
At1g61350	armadillo/beta-catenin repeat family protein
At5g62180	ATCXE20
At1g19940	ATGH9B5
At5g45940	ATNUDT11
At1g26570	ATUGD1/UGD1
At5g26220	ChaC-like family protein

At2g35890	CPK25 (calcium-dependent protein kinase 25)
At1g61667	expressed protein
At5g35730	EXS family protein
At1g06120	fatty acid desaturase family protein
At5g07800	flavin-containing monooxygenase family protein
At5g16570	GLN1;4 (Glutamine synthetase 1;4)
At3g56060	glucose-methanol-choline oxidoreductase
At1g73160	glycosyl transferase family 1 protein
At1g23040	hydroxyproline-rich glycoprotein family protein
At2g38480	integral membrane protein, putative
At3g05280	integral membrane Yip1 family protein
At4g00820	IQD17 (IQ-domain 17); calmodulin binding
At1g18840	IQD30; calmodulin binding
At1g75280	isoflavone reductase, putative
At1g70690	kinase-related
At1g16860	merozoite surface protein-related
At5g20360	octicosapeptide/Phox domain-containing
At4g18050	PGP9 ATPase
At4g33330	PGSIP3
At1g53660	phosphate translocator-related
At1g06470	phosphate translocator-related
At5g60660	PIP2;4/PIP2F
At3g22070	proline-rich family protein
At1g09440	protein kinase family protein
At3g51470	protein phosphatase 2C
At5g55550	RNA recognition motif (RRM)-containing protein
At2g37540	short-chain dehydrogenase
At2g02480	STI (STICHEL)DNA-directed DNA polymerase
At3g54690	sugar isomerase (SIS) domain-containing protein
At1g76670	transporter-related
At5g15490	UDP-glucose 6-dehydrogenase, putative
At3g29360	UDP-glucose 6-dehydrogenase, putative
At5g45020	unknown protein
At5g61340	unknown protein
At1g33800	unknown protein
At1g69030	unknown protein
At1g64980	unknown protein
At3g06035	unknown protein
At4g18425	unknown protein
At1g76250	unknown protein
At1g67060	unknown protein
At1g62045	unknown protein
At1g52910	unknown protein
At4g01240	unknown protein
At4g09990	unknown protein
At5g67600	unknown protein
At2g28870	unknown protein
At5g17600	zinc finger (C3HC4-type RING finger)

1. Caño-Delgado, A., Penfield, S., Smith, C., Catley, M. & Bevan, M. Reduced cellulose synthesis invokes lignification and defense responses in Arabidopsis thaliana. *Plant J* **34**, 351–362 (2003).
2. Mitsuda, N. & Ohme-Takagi, M. Functional Analysis of Transcription Factors in Arabidopsis. *Plant and Cell Physiology* **50**, 1232–1248 (2009).
3. Zhong, R., Demura, T. & Ye, Z.-H. SND1, a NAC Domain Transcription Factor, Is a Key Regulator of Secondary Wall Synthesis in Fibers of Arabidopsis. *Plant Cell* **18**, 3158–3170 (2006).
4. Zhong, R., Richardson, E. A. & Ye, Z.-H. Two NAC domain transcription factors, SND1 and NST1, function redundantly in regulation of secondary wall synthesis in fibers of Arabidopsis. *Planta* **225**, 1603–1611 (2007).
5. Kim, W.-C. *et al.* MYB46 directly regulates the gene expression of secondary wall-associated cellulose synthases in Arabidopsis. *The Plant Journal* **73**, 26–36 (2013).
6. McCarthy, R. L., Zhong, R. & Ye, Z.-H. MYB83 Is a Direct Target of SND1 and Acts Redundantly with MYB46 in the Regulation of Secondary Cell Wall Biosynthesis in Arabidopsis. *Plant and Cell Physiology* **50**, 1950–1964 (2009).
7. Zhou, J., Lee, C., Zhong, R. & Ye, Z.-H. MYB58 and MYB63 Are Transcriptional Activators of the Lignin Biosynthetic Pathway during Secondary Cell Wall Formation in Arabidopsis. *The Plant Cell* **21**, 248–266 (2009).
8. Taylor-Teeples, M. *et al.* An Arabidopsis gene regulatory network for secondary cell wall synthesis. *Nature* **517**, 571–575 (2015).

9. Kubo, M. *et al.* Transcription switches for protoxylem and metaxylem vessel formation. *Genes Dev.* **19**, 1855–1860 (2005).

10. Yamaguchi, M., Kubo, M., Fukuda, H. & Demura, T. 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 (2008).

11. Zhou, J., Zhong, R. & Ye, Z.-H. Arabidopsis NAC Domain Proteins, VND1 to VND5, Are Transcriptional Regulators of Secondary Wall Biosynthesis in Vessels. *PLOS ONE* **9**, e105726 (2014).

12. Gautam, V., Singh, A., Singh, S. & Sarkar, A. K. An Efficient LCM-Based Method for Tissue Specific Expression Analysis of Genes and miRNAs. *Sci Rep* **6**, 21577 (2016).

13. Avci, U., Pattathil, S. & Hahn, M. G. Immunological Approaches to Plant Cell Wall and Biomass Characterization: Immunolocalization of Glycan Epitopes. in *Biomass Conversion: Methods and Protocols* (ed. Himmel, M. E.) 73–82 (Humana Press, 2012). doi:10.1007/978-1-61779-956-3_7.

14. Pattathil, S. *et al.* A comprehensive toolkit of plant cell wall glycan-directed monoclonal antibodies. *Plant Physiol* **153**, 514–525 (2010).

15. Popper, Z. A. *et al.* Evolution and diversity of plant cell walls: from algae to flowering plants. *Annu Rev Plant Biol* **62**, 567–590 (2011).

16. Hao, Z. *et al.* Loss of Arabidopsis GAUT12/IRX8 causes anther indehiscence and leads to reduced G lignin associated with altered matrix polysaccharide deposition. *Frontiers in Plant Science* **5**, (2014).

17. Peralta, A. G., Venkatachalam, S., Stone, S. C. & Pattathil, S. Xylan epitope profiling: an enhanced approach to study organ development-dependent changes in xylan structure, biosynthesis, and deposition in plant cell walls. *Biotechnology for Biofuels* **10**, 245 (2017).

18. Xu, C., Zhao, L., Pan, X. & Šamaj, J. Developmental Localization and Methylesterification of Pectin Epitopes during Somatic Embryogenesis of Banana (Musa spp. AAA). *PLoS One* **6**, e22992 (2011).

19. Kim, J. S. & Daniel, G. Immunolocalization of hemicelluloses in Arabidopsis thaliana stem. Part I: temporal and spatial distribution of xylans. *Planta* **236**, 1275–1288 (2012).

20. Schmidt, D., Schuhmacher, F., Geissner, A., Seeberger, P. H. & Pfrengle, F. Automated synthesis of arabinoxylan-oligosaccharides enables characterization of antibodies that recognize plant cell wall glycans. *Chemistry* **21**, 5709–5713 (2015).

21. Hernández-Blanco, C. *et al.* Impairment of cellulose synthases required for Arabidopsis secondary cell wall formation enhances disease resistance. *Plant Cell* **19**, 890–903 (2007).

22. Molina, A. *et al.* Arabidopsis cell wall composition determines disease resistance specificity and fitness. *Proc Natl Acad Sci U S A* **118**, e2010243118 (2021).

23. Kushwah, S. *et al.* Arabidopsis XTH4 and XTH9 Contribute to Wood Cell Expansion and Secondary Wall Formation1 [OPEN]. *Plant Physiology* **182**, 1946–1965 (2020).

24. Pyo, H., Demura, T. & Fukuda, H. TERE; a novel cis-element responsible for a coordinated expression of genes related to programmed cell death and secondary

wall formation during differentiation of tracheary elements. *Plant J* **51**, 955–965 (2007).

25. O'Malley, R. C. *et al.* Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. *Cell* **165**, 1280–1292 (2016).

26. Shi, D. *et al.* Tissue-specific transcriptome profiling of the Arabidopsis inflorescence stem reveals local cellular signatures. *The Plant Cell* **33**, 200–223 (2021).

27. Zhong, R., Lee, C. & Ye, Z.-H. Global Analysis of Direct Targets of Secondary Wall NAC Master Switches in Arabidopsis. *Molecular Plant* **3**, 1087–1103 (2010).

28. Ohashi-Ito, K., Oda, Y. & Fukuda, H. Arabidopsis VASCULAR-RELATED NAC-DOMAIN6 Directly Regulates the Genes That Govern Programmed Cell Death and Secondary Wall Formation during Xylem Differentiation[C][W]. *Plant Cell* **22**, 3461–3473 (2010).

29. Liwanag, A. J. M. *et al.* Pectin Biosynthesis: GALS1 in Arabidopsis thaliana Is a β-1,4-Galactan β-1,4-Galactosyltransferase. *The Plant Cell* **24**, 5024–5036 (2012).

30. Arabidopsis NAC Domain Proteins, VND1 to VND5, Are Transcriptional Regulators of Secondary Wall Biosynthesis in Vessels | PLOS ONE. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0105726.

31. Du, Q. & Wang, H. The role of HD-ZIP III transcription factors and miR165/166 in vascular development and secondary cell wall formation. *Plant Signal Behav* **10**, e1078955 (2015).

32. Smit, M. E. *et al.* A PXY-Mediated Transcriptional Network Integrates Signaling Mechanisms to Control Vascular Development in Arabidopsis[OPEN]. *Plant Cell* **32**, 319–335 (2020).

33. Ohashi-Ito, K., Matsukawa, M. & Fukuda, H. An Atypical bHLH Transcription Factor Regulates Early Xylem Development Downstream of Auxin. *Plant and Cell Physiology* **54**, 398–405 (2013).

34. Wu, H.-C., Bulgakov, V. P. & Jinn, T.-L. Pectin Methylesterases: Cell Wall Remodeling Proteins Are Required for Plant Response to Heat Stress. *Front. Plant Sci.* **9**, (2018).

35. Guzha, A. *et al.* Cell wall-localized BETA-XYLOSIDASE4 contributes to immunity of Arabidopsis against Botrytis cinerea. *Plant Physiology* **189**, 1794–1813 (2022).

36. Jiang, W., Zeng, Q., Jiang, Y., Gai, Y. & Jiang, X. Molecular and functional characterization of ferulate-5-hydroxylase in Populus tomentosa. *J. Plant Biochem. Biotechnol.* **30**, 92–98 (2021).

CHAPTER III

Identification and Characterization of Putative Vascular NAC Domain Transcription Factors in Tomato Root

ABSTRACT

The NAC family of plant-specific Transcription Factors (TFs) play a significant role in regulating various complex biological processes in plants including root formation, flower maturation and the development of meristems[1] . A subgroup of these TFs, known as *VASCULAR NAC DOMAIN* (*VND*) TFs have been identified as key regulators of xylem cell differentiation and secondary cell wall deposition (SCW) in the model species *Arabidopsis thaliana*[2]. However, there are only a handful of studies that have investigated the conservation and function of *VND* orthologs in other species[1,3]. To address this gap in knowledge, I conducted a study that employed hairy root transgenic lines to identify likely *VND* orthologs in tomato (*Solanum lycopersicum* cv. M82). I demonstrated that several tomato *VND* TFs were expressed in root vascular tissue and sufficient to induce ectopic xylem differentiation. Overall, the findings of this study highlight both the evolutionary conservation and unique differences present within the *VND* TF family in the context of xylem development between two evolutionary distant plant species.

INTRODUCTION

Xylem cells are crucial for plant survival, facilitating the transport of water and minerals from roots to aerial plant parts, while also contributing to plant adaptation to various stresses[4]. Given the importance of these specialized cell types, there is an increasing interest in understanding their development across diverse plant species. Using the model plant *Arabidopsis thaliana*, we have obtained a fundamental understanding of

xylem cell biogenesis and its associated gene regulatory network that produces the secondary cell wall (SCW) including enzymes involved in cellulose, hemicellulose and lignin bisynthesis[4]. In the *Arabidopsis* root, the differentiation of xylem vessels is in part controlled by *VASCULAR-RELATED NAC-DOMAIN1* (*VND1*) through *VND7* transcription factors[5]. In particular, *VND6* and *VND7* are considered the master regulators of xylem differentiation. When overexpressed, these two TFs are sufficient to drive xylem cell transdifferentiation as demonstrated by deposition of a helical, reticulate or pitted SCW in other root cell types[2]. In root tissue, *VND6* is expressed in metaxylem cell file in root differentiation zone, while *VND7* is expressed in protoxylem cells. Loss of function mutant alleles of *VND6* and *VND7* produce no visible xylem developmental abnormalities, likely due to functional redundancy with other *VND* TFs. Further, *VND*-associated xylem phenotypes in roots were only demonstrated by using the SRDX dominant repression system[2]. Previous studies in the *Arabidopsis* root have demonstrated that the other related *VND* transcription factors (*VND1* to *VND5*) are also xylem vessel or vascular-enriched and that their overexpression activates xylem transdifferentiation and SCW deposition[2,5]. In the Arabidopsis roots, *VND1-3* are preferentially expressed in procambial cells in the meristem, while *VND4-VND7* are expressed in the vessel elements[2]. similar to *VND6* and *7*, overexpression of *VND1* to *VND5* induce ectopic deposition of SCW by producing a helical or reticulate cell walls in the roots, root hypocotyl as well as mesophyll and epidermal cells in leaves[5,6]. It should be noted that the *VND1-5* overexpression phenotypes in roots were only observed when an inducible overexpression system was used whereas overexpression under a 35S promoter did not produce any ectopic SCW deposition[2,6.]

Within the root, *VND6* and *VND7* directly or indirectly regulate the expression of numerous *MYB* genes[2]. *MYB* TFs are the second-layer regulators of SCW deposition, including *AtMYB58/63/83/85*, which are lignin-specific TFs, and *AtMYB26/32/41/44/46/61/83/103*, which participate in regulating enzymes responsible for biosynthesis of other cell wall polymers including cellulose, hemicellulose and pectin biosynthesis[7]. Among these *MYB* TFs, *MYB46* and *MYB83* were identified as direct downstream targets of *VND6,7*. Additionally, they are sufficient to regulate SCW deposition and were shown to directly regulate the expression of numerous SCW associated genes. Dominant repression of their function results in reduced secondary wall thickening in vessel elemnts[8,9].

The conservation of the SCW regulatory network is understudied. The findings regarding *VND*s and their essential role in xylem differentiation in *Arabidopsis* make a strong case for functional analysis of these TFs in other species. Further, genome duplications and speciation have resulted in expansion of many TF families, and subsequent gene subfunctionalization or neofunctionalization can present underlying mechanisms through which gene function and regulatory pathways differ and evolve in different species[10]. The focus of this chapter is the characterization of select tomato *VND* orthologs that are putative regulators of xylem differentiation. Given the importance of tomato as a significant agricultural crop and its potential as a complementary model organism to *Arabidopsis*, investigating these orthologs is of great importance. Accurately identifying orthologous genes is essential in evaluating gene function and evolution of *VND*s in tomato. There are multiple approaches for determining orthology, including phylogenetic analysis to establish (1:1 or 1:many or many:1) patterns of orthology, comparisons of

gene expression patterns, localization, or loss-of-function phenotypes between *Arabidopsis* and tomato. This study aims to identify *VND* orthologs in tomato by utilizing both phylogenetic analysis and analysis of the expression patterns driven by tomato *VND* promoters. Moreover, the function of tomato *VND* TFs in xylem differentiation were examined by analyzing the effects of their perturbed expression in transgenic tomato hairy roots[11]. To this end, overexpression and CRISPR edited mutant hairy root lines were generated and assayed for xylem developmental abnormalities. Lastly, the tomato *VND6* ortholog was overexpressed in *Arabidopsis* to assess its functional conservation.

RESULTS

Identification of Putative *VND* Family Orthologs in Tomato

A phylogenetic tree was constructed using genetic information from *Arabidopsis VND1-7* (*AT2G18060, AT4G36160, AT5G66300*, *AT1G12260, AT1G62700, AT5G62380, AT1G71930*) TFs to identify their closest putative orthologs in tomato (Fig.1A). The resulting genes were grouped into four distinct groups: the *VND1,2,3* clade, *VND4/5* clade, *VND6* clade, and *VND7* clade. In the *VND1,2,3* clade, the orthology determined is (many:one – 3 Arabidopsis gene to 1 tomato gene), and a single tomato gene (*Solyc02g084350*) was identified as a *SlVND1,2,3-like* ortholog. Similarly in the *VND4/5* clade (many:1) the only gene identified is *SlVND4,5-like* (*Solyc08g079120*). For the *VND6* clade and *VND7* clade the relationship is (1:many) , as two tomato genes were identified for each clade, namely *SlVND6* (*Solyc06g034340*)*, SlVND6-1* (*Solyc03g083880*) and *SlVND7 (Solyc11g018660), SlVND7-1* (*Solyc06g065410*). Analysis of single cell data from tomato roots confirmed that transcripts of the identified tomato *VND* homologs are

enriched in xylem cells[12], providing further evidence of their potential roles in regulating xylem development and function in tomato (Fig.1B,C).

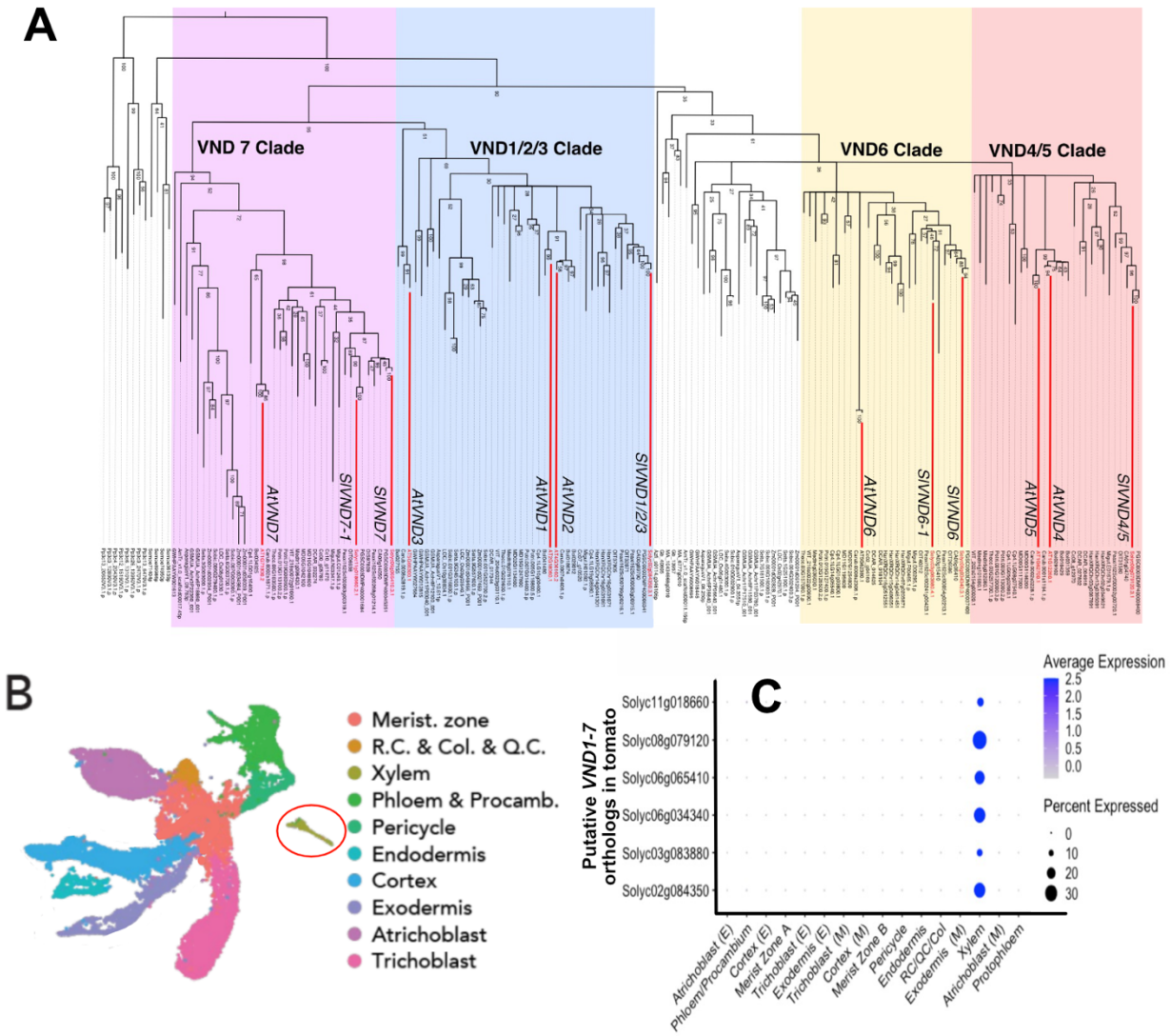
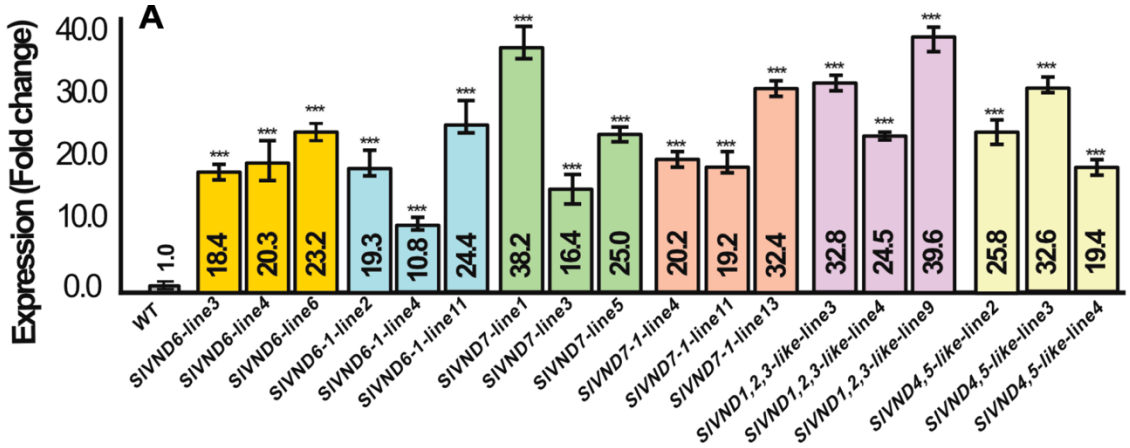


Figure 1. Phylogenetic and gene expression analyses of *VND*s in tomato**. A)** Phylogenetic analysis where 6 tomato genes are grouped with members of *AtVND1-7* genes. **B**)Uniform Manifold Approximation Projection (UMAP) displaying 10 clusters and their associated cell types identity in tomato root (Canto-Pastor, et al. 2022). The xylem cluster is circled in red **C**) Cell type or tissue-specific expression profiles for the putative *VND* orthologs visualized from single-cell dataset from 3cm of the tomato primary root tip.

Overexpression of *SlVND6, SlVND6-1* and *SlVND7* Causes Ectopic Deposition of Secondary Cell Walls in Tomato Hairy Roots

To investigate whether the identified *VND* genes act as transcriptional activators of secondary wall biosynthesis, I examined whether their overexpression in tomato roots could trigger the deposition of ectopic SCWs. I generated overexpression lines for each candidate gene. The full-length cDNAs of tomato *VND*s were cloned downstream of the cauliflower mosaic virus (*CaMV*) *35S* promoter in a binary vector, and these



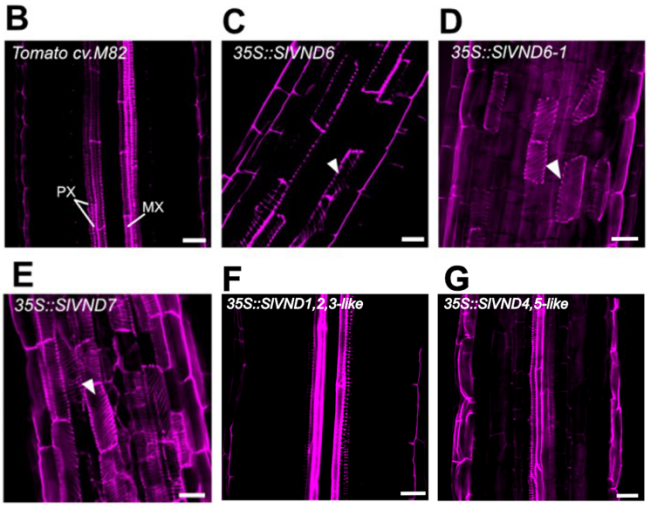


Figure 2. Overexpression of *SlVND* orthologs in hairy roots **A**) qRT-PCR overexpression measurements for wild type and *SlVNDs* in hairy root transgenic lines. (*=p<0.05; **=p<0.01;***=p<0.001) using a student's t-test – An absolute quantification method was used where expression values were adjusted to a tomato housekeeping gene(*Solyc07g025390)* and shown relative to WT values. **B)** Xylem cell development in a WT tomato root. **C-G**) Fuchsin staining of roots for lignin. Ectopic SCW deposition in tomato hairy roots in **C**) *SlVND6* **D**) *SlVND6-1* **E**) *SlVND7.* Lack of ectopic SCW in **F***) SlVND1,2,3-like* **G**) *SlVND4,5-like* PX=protoxylem; MX=metaxylem. White arrowheads indicate ectopic SCW. Scale bars, 20µm.

constructs were then introduced into tomato hairy roots (Supplemental Table 1). The increased expression level of corresponding genes in the overexpression lines was confirmed by quantitative PCR analysis (Fig.2A). Subsequent histological staining of transgenic lines with basic fuchsin to mark SCW lignin deposition, revealed that overexpression of three out of the six identified tomato *VND*s resulted in ectopic SCW

deposition. While overexpression of *Arabidopsis VND1-5* results in ectopic deposition of SCWs[11], overexpression of the putative orthologs to *VND1-5* in tomato hairy roots did not show any phenotype. However, similar to what was previously reported for *SlVND6* (described in Chapter I[11]), *SlVND6-1* and only one of the two *VND7* orthologs, *SlVND7*, were sufficient to produce ectopic secondary cell wall deposition (Fig. 2B-E, Table1).

***Arabidopsis* TF**	**Overexpression phenotype**	**Tomato Ortholog**	**Overexpression phenotype**
VND1	Ectopic SCW	*SlVND1,2,3-like;Solyc02g084350*	None
VND2	Ectopic SCW	*SlVND1,2,3-like;Solyc02g084350*	None
VND3	Ectopic SCW	*SlVND1,2,3-like;Solyc02g084350*	None
VND4	Ectopic SCW	*SlVND4,5-like;Solyc08g079120*	None
VND5	Ectopic SCW	*SlVND4,5-like;Solyc08g079120*	None
	Ectopic SCW	*SlVND6;Solyc06g034340*	Ectopic SCW
		SlVND6-1;Solyc03g083880	Ectopic SCW
	Ectopic SCW	*SlVND7;Solyc11g018660*	Ectopic SCW
		SlVND7-1;Solyc06g065410	None

Table1. Summary of *VND* TFs in *Arabidopsis* and tomato including their respective overexpression phenotypes in root tissue.

SlVND6 and *SlVND7* Display Vascular Specific Transcriptional Activity

The three TFs identified in previous section, *SlVND6, SlVND6-1* and *SlVND7* with the ability to induce ectopic SCW deposition are hypothesized to be key players in xylem

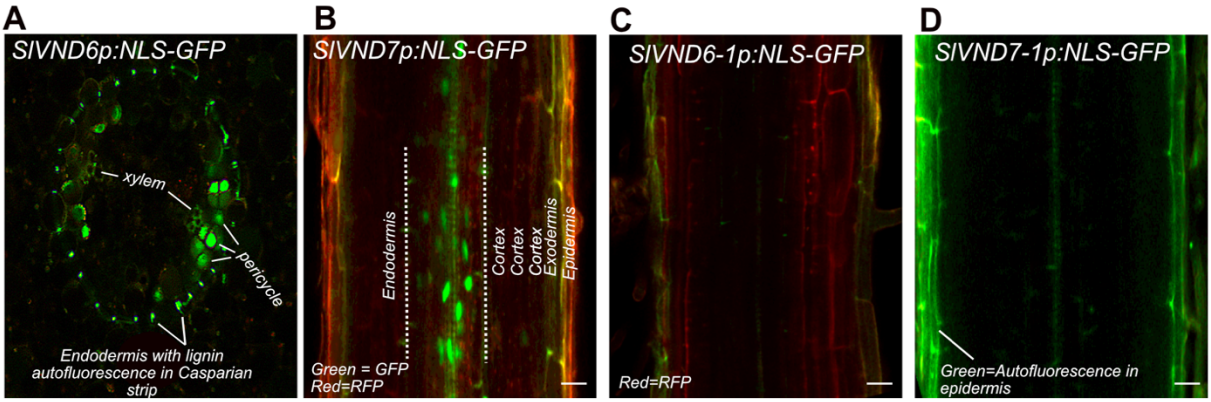


Figure 3. Transcriptional activity of *VND6* and *VND7* orthologs in tomato hairy roots **A**) *SlVND6* transcriptional fusion with GFP signal present in pericycle cells in a tomato cross section from the elongation zone **B)** Transcriptional fusion with GFP signal in vascular tissue in maturation zone. **C-D**) No GFP signal is observed in maturation zone for *SlVND6-1* and *SlVND7-1* reporter lines. Green signal in epidermis is autofluorescence. Constructs include a membrane RFP-Tag - Scale bars, 20µm.

differentiation in tomato roots. To further examine their transcriptional activity and validate the available data from single cell dataset, transcriptional GFP reporter hairy root lines were constructed to evaluate and validate their activity domain in tomato roots. NLS-GFP expression driven by the respective ~2kb promoters (Supplemental Table 1) of *SlVND6* and *SlVND7* was observed in the xylem pole pericycle and vascular tissue (maturation zone) respectively (Fig. 3A-B), while there was no GFP signal detected for *SlVND6-1* and *SlVND7-1* in 3 independent transgenic lines (Fig. 3C-D, Table 2).

***Arabidopsis* TF**	**Transcript Localization**	**Tomato Ortholog**	**Transcript Localization**
VND6	Metaxylem	*SlVND6*	Xylem pole pericycle
		SlVND6-1	Not Detected
VND7	Protoxylem	*SlVND7*	Vascular Tissue
		SlVND7-1	Not Detected

Table 2. Summary of *VND* TFs in *Arabidopsis* and tomato and their respective expression pattern in reporter lines in root tissue.

Heterologous Overexpression of *SlVND6* in *Arabidopsis* Heterologous expression is a widely adopted method to evaluate gene functional conservation between plant species[13]. Studying the behavior of proteins in a well-studied model organism can yield insights into their roles within complex pathways. Prior research has demonstrated that heterologous overexpression of various secondary cell wall transcription factors from other species can activate corresponding pathways in *Arabidopsis*[1,3]. To investigate this in tomato and further assess functional conservation of *SlVND6*, I conducted a complementary experiment to evaluate if *SlVND6* is sufficient to drive ectopic xylem cell differentiation in *Arabidopsis*. Indeed, the overexpression of *SlVND6* in *Arabidopsis* led to ectopic SCW, phenocopying the ectopic xylem SCW deposition observed in tomato (Fig. 4A,B).

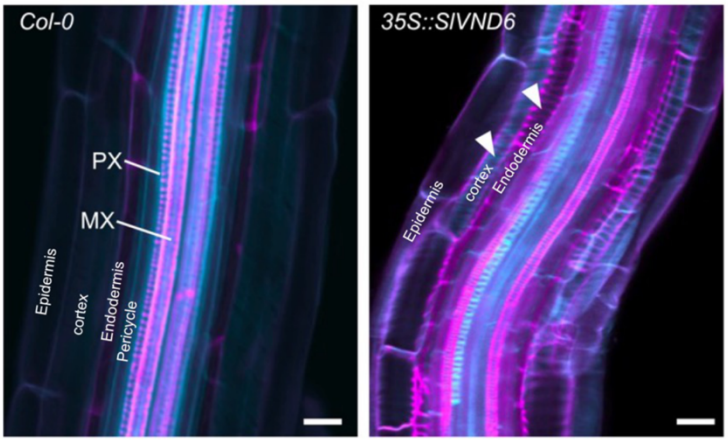


Figure 4. Heterologous overexpression of *SlVND6-1* in Arabidopsis. **A**) Xylem cell development in a WT *Arabidopsis* (Col-0) root elongation zone. **B**) Overexpression of *SlVND6-1* in *Arabidopsis.* Arrowheads denote ectopic SCW. PX= protoxylem; MX=metaxylem. Scale bars, 20µm.
Pink = Fuchsin staining for lignin; Cyan= Direct yellow96 staining for cellulose.

SlVND6 Functional Redundancy in Tomato Roots

The *VND* genes display overlapping expression patterns within the vascular tissues of *Arabidopsis* roots[2], suggesting potential functional overlapping roles. To explore this in tomato roots, I utilized the CRISPR Cas-9 system to generate a *slvnd6* mutant hairy root lines. Three independent CRISPR *slvnd6* edited lines were generated (Supplemental Table 1). Similar to the *Atvnd6* T-DNA mutant[2], no morphological abnormalities were detected in *Slvnd6* mutant lines (n=3, Supplemental Table 1) with respect to xylem development. Two protoxylem cell files and one metaxylem cell file is observed at 1cm

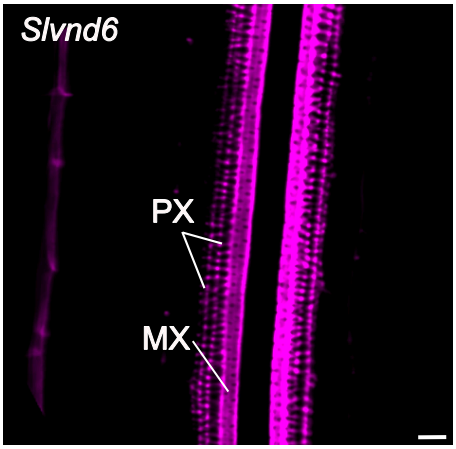


Figure 5. CRISPR edited *SlVND6.* Image taken at 1cm from the root tip with two protoxylem cell files and one metaxylem file.
Pink = Fuchsin staining for lignin. PX= protoxylem; MX=metaxylem.
Scale bar = 20µm

from the root tip(Fig. 5A), similar wild type tomato roots(Fig. 2B). Given the functional redundancy of *VND* genes in *Arabidopsis*[2], elucidating the hierarchical gene regulation involved in the development of various xylem cell types in tomato roots, requires combinatorial *VND* mutants.

DISCUSSION

In this sub-section of Chapter III, several homologs of *Arabidopsis VND* transcription factors were identified in tomato roots. The putative functional orthology of these tomato *VND* genes were characterized using overexpression studies in tomato roots, which provided evidence that *SlVND6, SlVND6-1,* and *SlVND7* but not *VND1,2,3-like or VND4,5-like* orthologs are indeed sufficient to determine xylem differentiation in tomato as marked by helical secondary cell wall deposition upon their overexpression. The ectopic deposition of secondary cell walls in tomato hairy roots overexpressing these genes suggests that they are sufficient to act as transcriptional activators of secondary wall biosynthesis. The conservation of *SlVND6* function across plant species was further demonstrated by its ability to induce xylem differentiation in *Arabidopsis* roots, as is observed for the *Arabidopsis VND6* transcription factor[2].

In *Arabidopsis*, when an inducible system is used , the overexpression of all *VND* TFs leads to ectopic secondary cell wall deposition in various root cell types characteristic of xylem cells. However, in tomato, overexpression of putative orthologs to At*VND1,2,3,4,5* (*SlVND1,2,3-like* and *SlVND4,5-like*) did not produce any phenotypic changes in vascular tissue in roots. This suggests that these genes may have evolved distinct functions in tomato compared to their *Arabidopsis* counterparts. It could be hypothesized that these

genes function redundantly to *SlVND6,6-1* and *SlVND7* and are expressed upon their loss of function. Another hypothesis is that *SlVND1,2,3-like* and *SlVND4,5-like* regulate xylem developmental plasticity. In line with this, a recent study highlighting *VND*-dependent xylem cell acclimation to stress has shown that under ABA treatment, only homologs of *VND1,2,3,4,5* are upregulated in tomato, while the *VND6* and *7* homologs had negligible upregulation[14]. In an analogous fashion, it could be hypothesized that *VND6* and *VND7* orthologs in tomato are responsible for the innate xylem developmental program, while other tomato *VND*s are only activated in response to a stimulus.

The overexpression of *SlVND7-1* in tomato roots did not lead to ectopic SCW, and transcriptional reporters revealed that this gene is not expressed in the roots. This is in contrast to single cell expression data available from the tomato root indicating that this transcript is present in xylem cells (Fig. 1C). Conversely, *SlVND6-1* was found to induce ectopic SCW deposition, but GFP expression was not detected in the roots of reporter lines although this gene has expression in xylem cells in data obtained from single cell data set[12]. It is important to note that the 2kb upstream region used in the reporter constructs may not be adequate to promote transcription or to recruit other necessary *trans*-factors. Furthermore, although hairy roots offer a convenient approach for examining promoter activity without the necessity of regenerating a whole plant, they possess certain limitations. According to reports, hairy roots have significantly higher sensitivity to auxin than normal roots[15], which can impact the expression profile of target genes. Hence, findings derived from this system should be re-evaluated and validated in stable transgenic tomato plants.

The absence of defects in xylem development in roots of the *Slvnd6* loss-of-function line suggests that *Slvnd6* is not essential for this process. This lack of phenotype may be due to the functional redundancy with other *VND* TFs, similar to what has been observed in *Arabidopsis*[2]. However, it is important to note that while histological staining did not reveal any visible morphological changes, it is possible that there could be modifications in the composition of secondary cell walls, similar to *vnd6* and *vnd7* mutants in the *Arabidopsis* inflorescence stem(as described in Chapter II). Such cell wall modifications are not detectable using the methods utilized in this study.

Taken together, this work provides a preliminary framework to assess the functional conservation of *VND* transcription factors in tomato roots.

MATERIALS AND METHODS

Overexpression Construct Design and Cloning

The coding sequence (CDS) for target genes was obtained from the Sol Genomics database (https://solgenomics.net - ITAG3.2). CDSs were amplified from tomato (*Solanum lycopersicum* cv. M82) cDNA. In brief, total RNA was isolated from 50mg of tomato root tissue using the Zymo-Direct-Zol RNA Miniprep Plus Kit (Zymo Research-catalog#R2071) according to manufacturer's instructions and treated with RNase-Free DNase (1unit/10µl). 1µg of DNAse-treated RNA was reverse-transcribed into cDNA using oligo(dT) primers and SuperScript III Reverse Transcriptase (SuperScript III First-Strand Synthesis System; Invitrogen) per kit instructions. Cloning primers were designed to PCR amplify the CDS without the stop codon(Supplemental Table 1). PCR products were purified from the agarose gel (QIAquick Gel Extraction kit. Catalog#28704) for

subsequent recombination and cloning. Purified cDNAs were introduced into the pENTR/D-Topo vector (Invitrogen). The resulting pENTR plasmids were then LR recombined (LR Clonase II Enzyme mix; Invitrogen) into the pGWB417 binary destination vector (Addgene plasmid #74811; http://n2t.net/addgene:74811; RRID:Addgene_74811) containing a 35S promoter driving the expression of the CDS. All constructs were confirmed by Sanger sequencing.

Transcriptional Reporter Construction and Imaging

Cloning primers were designed to amplify 2000 bp upstream of the translational start site of *SlVND6,SlVND7,SlVND6-1 and SlVND7-1* respectively (Supplemental Table 1), using the tomato reference genome annotation ITAG3.2 (https://solgenomics.net). The promoters were amplified from genomic DNA using Phusion DNA polymerase (New England Biolabs). Amplified fragments were cloned into pENTR5'TOPO (Invitrogen) and sequences were confirmed by Sanger sequencing. LR Clonase II Enzyme mix (Invitrogen) was used to clone the promoters upstream of a nlsGFP binary vector pMR99 which also contains a ubiquitously expressing plasma membrane marker TagRFP-LTI6b. The binary vectors were used for hairy root (*Rhizobium rhizogenes*) transformation as described below. Transgenic hairy root fluorescence was visualized using Confocal Laser Scanning Microscopy with a Zeiss Observer Z1 LSM700 (Zeiss) microscope (water immersion, 20X objective) with excitation at 488 nm and emission at 493–550 nm for GFP and excitation at 555 nm and emission at 560–800 nm for RFP. Images were taken at approximately 1 cm from the root tip.

Generating CRISPR Constructs in Hairy Roots

Guide RNAs were designed using the CRISPR-PLANT web tool (https://www.genome.arizona.edu/crispr/CRISPRsearch.html). The guide RNAs (Supplemental Table 1) were phosphorylated and ligated into pYPQ131-3 vectors and recombined into p278 via Gateway cloning. p278 vector containing all 3 gRNA expression cassettes was then recombined into pMR286 vector by Gateway cloning. pMR286 contains Cas9 and Kan resistance expression cassettes. The final CRISPR vector was introduced into *Rhizobium rhizogenes* for hairy root transformation. Primers for genotyping (Supplemental Table 1) were designed with Primer-BLAST software (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) and checked against Solanum lycopersicum genome for specificity.

Quantitative RT–PCR of Overexpression Lines

All quantitative RT-PCR primers were designed with Primer3Plus software (http://www.primer3plus.com/). Primers were designed to amplify a 100-150 bp region near the 3' end of each target TF coding sequence (Supplemental Table 1). qRT-PCR was performed by setting up a 20 µL PCR reaction containing 5ul of cDNA (100ng/reaction) and 200 nM of each primer (PCRBIO Taq DNA Polymerase/Mix- catalog no. PB10.11-05 and EvaGreen dye; PCRBIO catalog number 89138-982). qRT–PCR was performed in a Bio-RAD CFX384-Real Time System with the following thermal cycling conditions: 5 min at 95°C, followed by 40 cycles of 20s at 95°C, 20s at 60°C, and 20s at 72°C. To ensure that PCR products were unique, a melting-curve analysis was performed after the amplification step. The experiment was carried out on a minimum of three independent lines and three technical replicates for each overexpression line. To

determine the fold change of the overexpression line relative to the wild type control (tomato transformed with *R. rhizogenes* with no plasmid), an absolute quantification method was conducted by generating a standard curve for each primer set. Values were normalized to the Ct value of an endogenous control gene (*Solyc07g025390*). The qPCR data for each gene is shown as a relative expression with respect to a control hairy root sample to which an expression value of 1 was assigned. Standard error of the mean (SEM) was then calculated from the normalized expression for each sample represented in the graphs. P-values were determined by performing a simple t-test; subtracting Ct number of the target gene for 3 replicates from that of the reference gene, which provides ΔCt values for overexpression lines and the wild type control to be subject to a t.test.

Rhizobium (Agrobacterium) rhizogenes Transformation

Rhizobium rhizogenes (ATCC: Strain 15834) transformation followed the protocol previously described (Ron et al., 2014).

Histochemistry and Imaging of Xylem Phenotypes

Hairy root tissue was cleared for 4-5 days in ClearSee buffer (Ursache et al., 2018). The presence of SCW was determined through a series of staining and imaging procedures. For Basic Fuchsin staining, cleared roots were incubated for 24 hours in a 0.02% w/v solution of Basic Fuchsin in ClearSee, followed by a 1-2 hour wash in ClearSee buffer prior to imaging. In cases where Direct Yellow 96 was used as a general cell wall stain, roots were first stained with Basic Fuchsin, then washed twice for 15 minutes in ClearSee, and subsequently incubated in a 0.1% solution of Direct Yellow 96 dissolved in ClearSee for an hour. After two additional 15-minute washes in ClearSee, roots were mounted for imaging. For Congo Red staining, root sections were incubated with a 0.1% solution of

Congo Red dissolved in water for 5 minutes, rinsed twice with water, and then imaged. Confocal Laser Scanning microscopy was conducted using a Zeiss LSM700 confocal with a 20X objective. Basic Fuchsin was excited at 550-561 nm and detected at 570-650 nm, Direct Yellow 96 was excited at 488 nm and detected at 519 nm, while Congo Red was excited at 497 nm and detected at 600 nm.

Phylogenetic Tree Construction

The methods to construct the phylogenetic tree were as previously described in Kajala et al., 2021(Chapter I).

Construction of *Arabidopsis 35S::SlVND6* transgenic line

Arabidopsis plants (Col-0) were floral dipped in *Agrobacterium tumefaciens* strain GV3101 (Bechtold and Pelletier, 1998). Transgenic plants were selected for on Kanamycin(100 mg L^{-1}) plates and T3 homozygous plants were used for experiments.

Supplemental Table 1 - List of primers used in this study, sg RNAs and resulting mutations.

Overexpression Constructs		
gene_id	F-primer	R-primer
Solyc02g084350	CACCATGGATACAAACGAATCATGCG	TTTATCAAATAAACAAATGCCAATATCA
Solyc08g079120	CACCATGAATTCTTTTGCACATGTTC	CTTCCATATATTATCAAATTG
Solyc06g034340	CACCATGAATTCCTTGTCACACGTTCCT	TTTCCACAAATCAATTTGAGAAGTTGAAG
Solyc03g083880	CACCATGAATTCGTTTTCACAGATTCCT	ATCAACAAATAAAGAAGTGTACTTCC
Solyc06g065410	CACCATGGAGATGGACTCATGTGTAC	TAAATCTGGGAAGCACTCAAGC
Solyc11g018660	CACCATGGAAATAGAGTCATGTGTT	TAAATCAGGGTAACAATGTAGC
Transcriptional Reporters		
Solyc02g084350	CAATGACGTCAACAATGGG	TGTTGATCACTCTTTTATTTAATTTATAAC
Solyc08g079120	TTGACATCGATCGGAACTTCA	CTCTTTTCTTGATCCACAAAATCAAAAC
Solyc06g034340	TTATTTCTATTTCAACCAATTTTGC	CTTCTTTGATCAATCAAATGTACTACTATAAC
Solyc03g083880	TCAAACCTAACCCAAATAGTATCG	CTGTGAAAACGAATTCATTCTACATATA
Solyc06g065410	TTCAATTTTTCAACATGACTATGAA	TCCATCTCCATTCTTGATTTTATTG
Solyc11g018660	GACTCGCTATAGCCCAGCAC	CTATTTCCATCTTGATCAATCTAACAAATT
qPCR primers		
Solyc02g084350	CAGCTTCCACAGCTCGAA	AATTTGTCAAGTGCTCTCCA
Solyc08g079120	GACTTGAAACCAACGAGAATGC	TGGTCATCGTTATACCAACAT
Solyc06g034340	TACCAAACCCCACCTCAAAA	TGACCAATGCTGGTCGAATA
Solyc03g083880	TGGCAAGCACAAGGAAAGAA	ACCACCACCACCAACAAA
Solyc06g065410	TTACACAGTCAATTACAACAATG	TAAATCTGGGAAGCACTCAAGC
Solyc11g018660	GCTACAAATGATCATCAAGGCCA	AGAAGAAGAAGATGGTGCCTCC
gRNA		
Solyc06g034340	TCCTTGTCACACGTTCCTCC	Exon1; 54bp deletion@position594 - Premature stop codon
Solyc06g034340	CTAGAAGGATTGACCTCGAT	Exon1; 56bp deletion@position593 - Premature stop codon
Solyc06g034340	TGGGATCTTCAAGGTACTTA	Exon1; 55bp deletion @position593 - Premature stop codon
Genotyping primers for CRISPR lines		
Solyc06g034340	TGTTGTCCTCTTCATTTGGATCGT	TCACTCTGCTCTTCTGTCCCC

References

1. Grant, E. H., Fujino, T., Beers, E. P. & Brunner, A. M. Characterization of NAC domain transcription factors implicated in control of vascular cell differentiation in Arabidopsis and Populus. *Planta* **232**, 337–352 (2010).
2. Kubo, M. *et al.* Transcription switches for protoxylem and metaxylem vessel formation. *Genes Dev.* **19**, 1855–1860 (2005).
3. Laubscher, M. *et al.* Temporal analysis of Arabidopsis genes activated by Eucalyptus grandis NAC transcription factors associated with xylem fibre and vessel development. *Sci Rep* **8**, 10983 (2018).
4. Taylor-Teeples, M. *et al.* An Arabidopsis gene regulatory network for secondary cell wall synthesis. *Nature* **517**, 571–575 (2015).
5. Zhou, J., Zhong, R. & Ye, Z.-H. Arabidopsis NAC Domain Proteins, VND1 to VND5, Are Transcriptional Regulators of Secondary Wall Biosynthesis in Vessels. *PLOS ONE* **9**, e105726 (2014).
6. Endo, H. *et al.* Multiple Classes of Transcription Factors Regulate the Expression of VASCULAR-RELATED NAC-DOMAIN7, a Master Switch of Xylem Vessel Differentiation. *Plant and Cell Physiology* **56**, 242–254 (2015).
7. Zhong, R. & Ye, Z.-H. Transcriptional regulation of lignin biosynthesis. *Plant Signal Behav* **4**, 1028–1034 (2009).
8. Zhong, R., Richardson, E. A. & Ye, Z.-H. The MYB46 Transcription Factor Is a Direct Target of SND1 and Regulates Secondary Wall Biosynthesis in Arabidopsis. *The Plant Cell* **19**, 2776–2792 (2007).

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

10. Panchy, N., Lehti-Shiu, M. & Shiu, S.-H. Evolution of Gene Duplication in Plants. *Plant Physiology* **171**, 2294–2316 (2016).

11. Kajala, K. *et al.* Innovation, conservation, and repurposing of gene function in root cell type development. *Cell* **184**, 3333-3348.e19 (2021).

12. Cantó-Pastor, A. *et al.* A suberized exodermis is required for tomato drought tolerance. 2022.10.10.511665 Preprint at https://doi.org/10.1101/2022.10.10.511665 (2022).

13. Yesilirmak, F. & Sayers, Z. Heterelogous Expression of Plant Genes. *Int J Plant Genomics* **2009**, 296482 (2009).

14. Ramachandran, P. *et al.* Abscisic acid signaling activates distinct VND transcription factors to promote xylem differentiation in Arabidopsis. *Current Biology* **31**, 3153-3161.e5 (2021).

15. Shen, W. H., Petit, A., Guern, J. & Tempé, J. Hairy roots are more sensitive to auxin than normal roots. *Proc Natl Acad Sci U S A* **85**, 3417–3421 (1988).

Chapter IV

Investigating Hysteresis in Xylem Cell Fate Switching: Experimental Design, Considerations and Challenges

INTRODUCTION

Hysteresis is a phenomenon associated with bistable switching in biological systems. Bistability refers to a system that exists in two different stable states, where a small perturbation can trigger a switch from one state to another[1,2]. It was previously reported that xylem cell differentiation has features of a bistable switch system with characteristics of hysteresis[3]. The term "hysteretic" refers to the fact that the switch occurs only when the input signal reaches a certain threshold, and the system remains in the new state even after the input signal is removed[4]. Further, the system's output or state depends not only on its current input, but also on its previous inputs or the memory of the system[5]. Bistability in xylem development was revealed when *VND7* was used as a stimulus via an estradiol inducible system[3]. Induction of *VND7* causes cells in the root to switch from a non-xylem cell identity to a xylem cell identity in a bistable manner[3]. Given the bistable switch system in xylem development, we hypothesized that this switch behavior has features of hysteresis or memory of a prior state. Unlike most plant cells, xylem developmental process ends with terminal differentiation which consists of secondary cell wall (SCW) deposition followed by programmed cell death, ultimately forming long, hollow vessels[6]. Cells that undergo terminal differentiation cannot revert back, so to test “memory” in xylem development, I used an estradiol-inducible system to induce a short pulse of *VND7*, sufficient to induce *VND7* expression without having cells undergo terminal transdifferentiation.

The aim of this work was to compare roots pre-treated with different concentrations of the inducer to determine whether there is any evidence to support memory of their prior induction state. However, my initial data collection, which included the assessment of i) the threshold prior to terminal differentiation and VND7 protein stability and ii) evaluating estradiol stability within the cells, suggested that an alternative induction system is required to effectively test this hypothesis. The extended stability of estradiol within the cells over prolonged periods of time interfered with our experimental design, requiring the use of a different induction system.

Determining VND7 Protein Decay & Threshold Prior to Terminal Differentiation

Terminal differentiation is an irreversible state, thus, to test memory, an earlier time point prior to cell death had to be used in our experiments. In order to identify this time point, a time course experiment was conducted with inducible transgenic *Arabidopsis thaliana* seedlings expressing *VND7* fused to a Myc epitope tag (XVE:*VND7*:Myc) and treated with 20 μM estradiol (the concentration at which 100% conversion to xylem cell was observed[3]). Roots were stained with basic fuchsin to detect SCW deposition, and images were taken at 6, 8, 16, and 24 hours after estradiol induction for xylem cell transdifferentiation (Fig. 1A). VND7 levels are regulated post-transcriptionally by the 26s proteasome[7]. Therefore, at the same time points, roots were subjected to treatment with either MG-132 (a proteasome inhibitor) or a mock control at specific time intervals to

measure protein levels (Fig. 1B). VND7 protein levels were quantified using an antibody to the Myc epitope (Fig. 1B-C). Xylem cell transdifferentiation was first observed at 8 hours after estradiol application with minimal protein abundance. At the earlier time point, 6 hours after estradiol application, although the protein was translated, as confirmed by treatment with MG-132, VND7-Myc tagged protein was hardly detectable (Fig. 1B). After

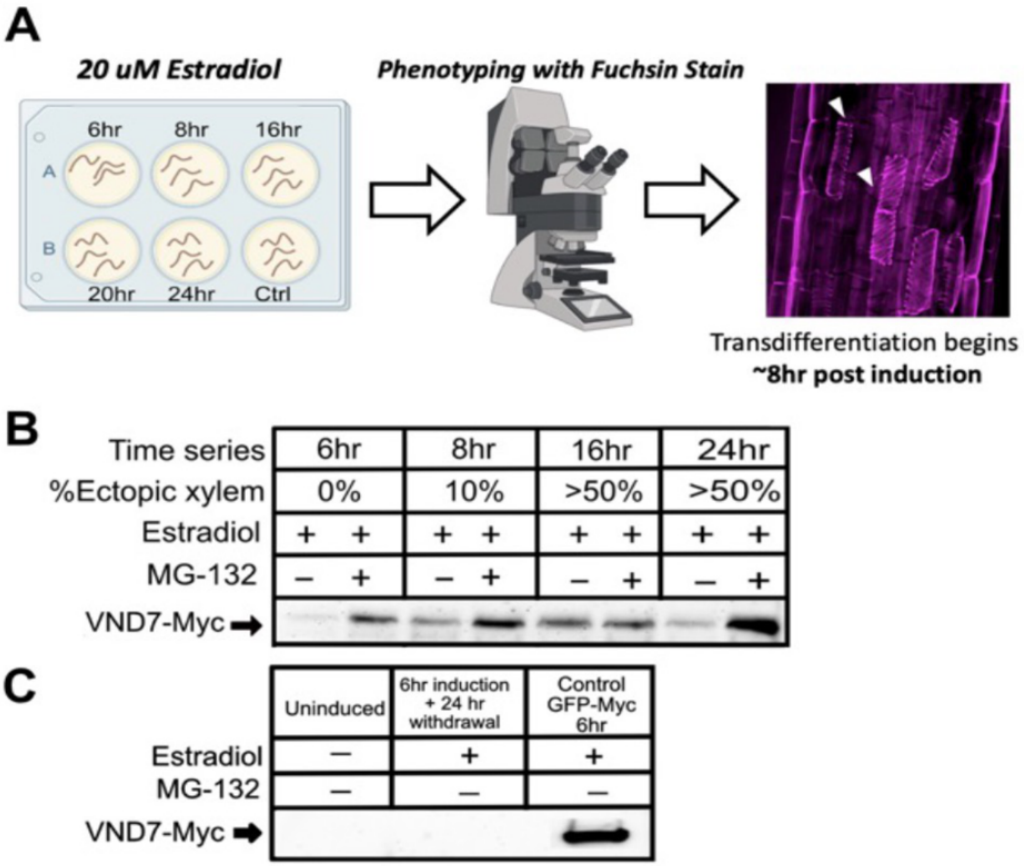


Figure 1. VND7 protein turnover. **(A)** Time series experiment to determine the time points at which xylem conversion can be detected in the inducible system. **(B)** Western blot analysis for the corresponding time series in A. **(C)** VND7 decay post-induction (columns 1 and 2), positive control from an estradiol-inducible GFP-MYC plant on far right.

24 hours of estradiol removal, VND7 protein was not detected (Fig. 1C). Therefore, the optimal protocol for inducing VND7 protein without xylem cell terminal differentiation comprises pre-treating plants with 20 μM estradiol for 6 hours followed by 24 hours on MS media.

Experimental Design to Generate a Dose Response Curve for Xylem Conversion

To evaluate hysteresis, we needed to examine the effect of prior induction on the dose-response curve of xylem cell conversion as a function of *VND7* expression in roots (Fig. 2D). Roots were pre-treated with either the *VND7* inducer (estradiol) or MS media alone (Fig. 2A), followed by measuring the conversion of xylem cells in response to varying concentrations of *VND7* expression (Fig. 2C). Xylem identity was quantified by histological staining for secondary cell walls. Pre-treatment with the *VND7* inducer was done for 6 hours with 20 μM estradiol (as determined in the prior section) where *VND7* was induced but xylem terminal transdifferentiation was not yet observed.

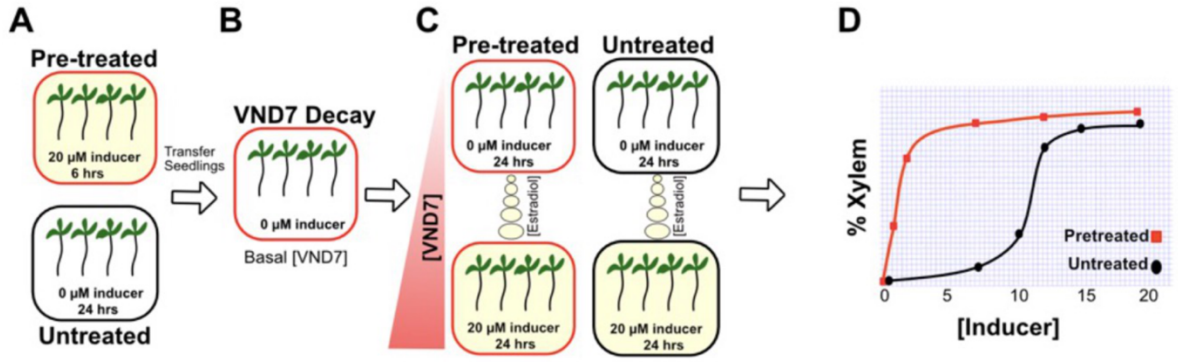


Figure.2 Experimental design for VND7-induced memory. (A) per-treatment of plates with the VND7 inducer. **(B)** Measuring the VND7 decay over time to determine basal VND7 concentration. **(C)** Challenging the plants with different doses of VND7 inducer. **(D)** Expected dose-response curves for pre-treated, untreated plants.

If memory of the prior exposure to the *VND7* inducer is still present, then the concentration of *VND7* needed to trigger the switch to xylem fate should be lower in plants that were pretreated compared to plants that did not undergo prior induction or exposure (Fig.1).

Evaluating Estradiol Stability in the System

To ensure the reliability of our experimental design to test memory of a prior state, it was essential to evaluate the stability of estradiol within the cells. To prevent prolonged levels

of *VND7* transcript following inducer pretreatment, it is crucial for the inducer to have a short half-life once it enters the cells. To monitor this in *Arabidopsis* inducible *VND7* lines, I conducted an experiment similar to that described in Fig. 2A. Specifically, seedlings were treated with 20uM of estradiol for 6 hours, which had previously been determined to be the time point threshold before terminal differentiation (Fig. 1A). At the end of the induction period, the seedlings were divided into two groups of 10 seedlings each (Fig. 3). The first group was immediately fixed, cleared, stained with Basic Fuchsin staining and finally imaged to be assessed for ectopic SCW detection in roots. As previously determined, no xylem transdifferentiation was observed at this time point. The second group was thoroughly rinsed and transferred to MS media alone, and imaged 24 hours later. The results showed that 80% of the seedlings exhibited trans-differentiation when imaged 24 hours later in the absence of inducer, suggesting estradiol stability and induction of *VND7* beyond our intended pretreatment duration.

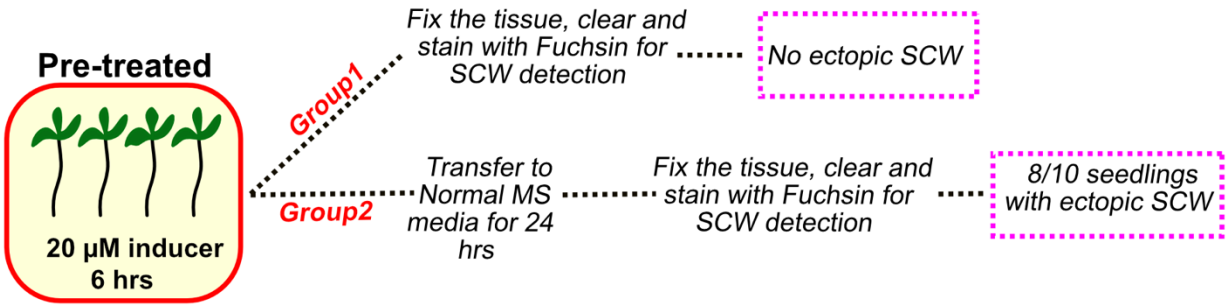


Figure. 3 Evaluating estradiol stability in *Arabidopsis* seedlings. Schematic summarizing the experimental work flow to determine SCW deposition in roots after induction with estradiol.

In order to gain a better understanding of the stability of estradiol and identify optimal experimental conditions, I utilized an inducible GFP construct (XVE::GFP:MYC) in a transient system in tobacco leaves (Fig. 4). The decision to use GFP was based on its ease of imaging, its documented half-life[8] and the added advantage of not needing fixation or clearing as compared to SCW imaging in roots. Estradiol-Inducible GFP

plasmid and 5 μM estradiol were infiltrated into 6 different tobacco leaves from 3 individual plants. Subsequently, GFP levels were monitored by confocal imaging. If the GFP signal persisted beyond its reported half-life of 18-20 hours[8], it would indicate the continued presence of estradiol and its ability to sustain continuous expression of GFP. GFP levels remained significantly high even 9 days after the induction, suggesting that estradiol was present to drive GFP expression and that it remained stable within the plant cells (Fig.4).

Figure 4. Estradiol stability in tobacco leaves was assayed by infiltrating the tobacco leaves and monitored overtime. The image on the right displays GFP levels in the leaves 9 days post induction. White arrowheads indicate GFP in the nucleus.

To confirm, compare, and contrast the results from the transient tobacco assay, I replicated that experiment with slight modifications, in six-day-old seedlings of *Arabidopsis* lines carrying GFP-inducible constructs (Fig. 5A). Initially, our experimental design proposed a concentration of 20uM of estradiol, but I replaced this with a much lower concentration of 5nM to avoid saturation of the system. Additionally, I reduced the induction time to a minimum of 30 minutes instead of 6hrs. Despite these efforts, as depicted in Fig. 5A, GFP was still detectable beyond its half-life of 18-20 hours[8], even with low estradiol concentrations and a very limited period of induction. In an effort to address the issue of estradiol stability, I also investigated the potential of an estradiol antagonist, 4HydroxyTamoxifen (4OH-T)[9]. Unfortunately, the results indicated that the plant cells were unable to either effectively absorb the chemical, or that it doesn't behave

as intended in plant cells. GFP signal persisted even after treatment with the 4OH-T compound(Fig. 5B).

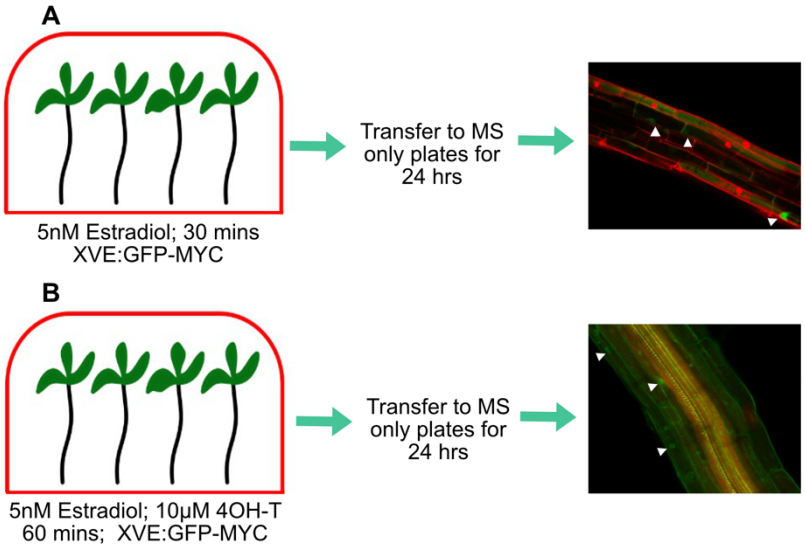


Figure 5. Evaluating estradiol stability. Inducible GFP plants treated with 5nM estradiol for 30 min (A) and 60 minutes (B). When applied for 60 minutes, plants were also treated with the estradiol inhibitor (4OH-T) for simultaneous induction and inhibition. White arrowheads indicate GFP expression.

Conclusion

The data presented in this section are essential pieces to be used in the future experiments to continue this work. However, it also suggests that developing a new induction system is necessary to examine *VND7*-dependent hysteresis due to stability of estradiol inside plant cells. Another option to explore is the use of a dexamethasone inducible system, although its stability in the system needs to be evaluated thoroughly. Alternatively, a more effective strategy could be the use of an induction system that features an "on" and "off" switch, such as the tetracycline-off and heat shock-on expression system[10]. This approach has been successfully tested in tobacco plants and seems to provide strict control over the target transgene's on/off state of expression[10]. The ability to precisely control expression states of *VND7* can facilitate the accurate examination of hysteresis in our system.

Materials and Methods

Construction of the Inducible Myc-Tagged *VND7* and GFP Lines

The plasmids p1R4-pG1090:XVE, p2R3a-nosT, and a gateway entry clone that contained AtVND7 fused to a 4xMyc epitope (or GFP fused to 4xMyc) were incorporated into the pCAM-kan-R4R3M binary vector using the MultiSite gateway cloning system. The final expression vector was confirmed by Sanger sequencing. This construct was introduced into *Agrobacterium tumefaciens* or *Rhizobium rhizogenes* as described in Chapter I.

VND7 Induction Experiments

XVE:*VND7*-Myc or XVE:GFP-Myc plants were grown vertically on sterile mesh placed on MS media plates. Five days after germination, seedlings were transferred, with the mesh, to MS media containing various concentrations of estradiol (dissolved in 100% ethanol). In cases where MG-132 was used, MG-132(M7449-200UL) was added at the concentration of 80μM[7] (dissolves in 100% ethanol) and incubated along with estradiol induction time series. 4OH-T was dissolves in DMSO and used at 10μM concentration.

Western Blot

To extract total protein, 100mg of flash frozen tissue was ground into a fine powder and then mixed with an equivalent volume of 4X BioRad Laemmli sample buffer (1610747) to create a buffer solution of approximately 100uL per 100mg of tissue. The mixture was then centrifuged at maximum speed for 2 minutes, and the resulting supernatant was boiled at 95C for 5 minutes. The protein concentration was determined using a Bradford Assay kit (BioRad – 5000201), and equal amounts of protein were loaded onto a 4–20% gradient gel (Mini-PROTEAN TGX; Bio-Rad) for separation. The separated proteins were

then transferred to a nitrocellulose membrane, which was blocked with 10% milk in PBST (1xPBS +0.1% Tween20) for 1 hour. The membrane was then incubated with a primary antibody (anti-Myc antibody, Sigma M4439 at 1:1000) at room temperature with shaking for 1 hour. After three 15-minute washes in PBST, the primary antibodies were detected using horseradish peroxidase (HRP)–conjugated Anti-mouse antibody at a 1:10,000 dilution (ThermoFisher; Catalog # 31430). Finally, the signals from the antibodies were detected using enhanced chemiluminescence (Amersham ECL Prime Blotting Detection Reagent) with a Bio-Rad ChemiDoc MP System (Bio-Rad).

References

1. Fukushige, T., Hawkins, M. G. & McGhee, J. D. The GATA-factor elt-2 is essential for formation of the Caenorhabditis elegans intestine. *Dev Biol* **198**, 286–302 (1998).
2. Lai, K., Robertson, M. J. & Schaffer, D. V. The sonic hedgehog signaling system as a bistable genetic switch. *Biophys J* **86**, 2748–2757 (2004).
3. Turco, G. M. *et al.* Molecular Mechanisms Driving Switch Behavior in Xylem Cell Differentiation. *Cell Reports* **28**, 342-351.e4 (2019).
4. Kramer, B. P. & Fussenegger, M. Hysteresis in a synthetic mammalian gene network. *Proc Natl Acad Sci U S A* **102**, 9517–9522 (2005).
5. Wang, L. *et al.* Bistable switches control memory and plasticity in cellular differentiation. *Proceedings of the National Academy of Sciences* **106**, 6638–6643 (2009).
6. De Rybel, B. *et al.* Plant development. Integration of growth and patterning during vascular tissue formation in Arabidopsis. *Science* **345**, 1255215 (2014).

7. Yamaguchi, M., Kubo, M., Fukuda, H. & Demura, T. 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 (2008).
8. Corish, P. & Tyler-Smith, C. Attenuation of green fluorescent protein half-life in mammalian cells. *Protein Engineering, Design and Selection* **12**, 1035–1040 (1999).
9. Characterization of tamoxifen and 4-hydroxytamoxifen glucuronidation by human UGT1A4 variants | Breast Cancer Research | Full Text. https://breast-cancer-research.biomedcentral.com/articles/10.1186/bcr1539.
10. Zhou, Y. *et al.* Establishment of a tetracycline-off and heat shock-on gene expression system in tobacco. *Journal of Integrative Agriculture* **16**, 1112–1119 (2017).