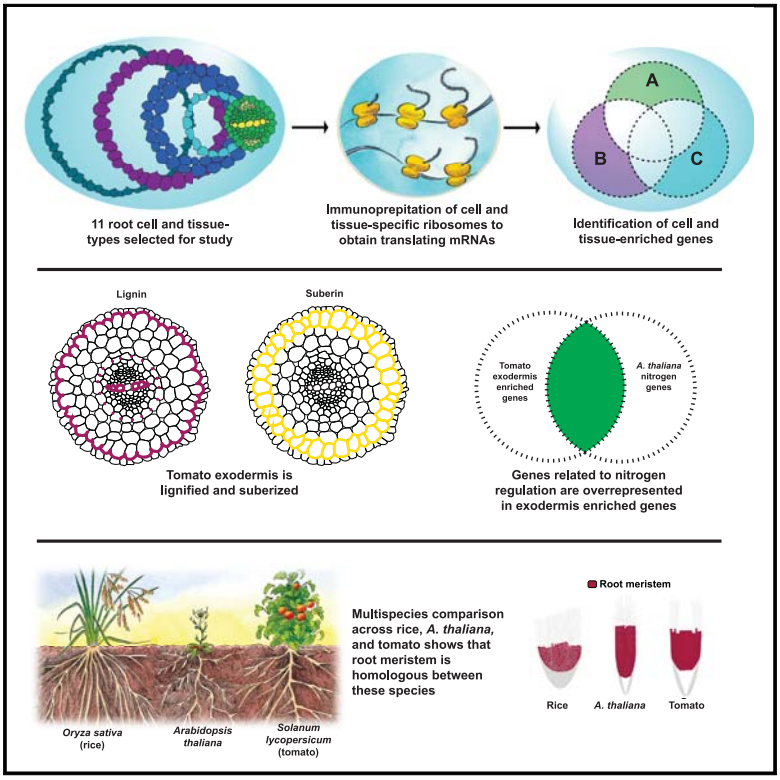
Cell

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*, 1–16
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

Please cite this article in press as: Kajala et al., Innovation, conservation, and repurposing of gene function in root cell type development, Cell (2021), https://doi.org/10.1016/j.cell.2021.04.024

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

Please cite this article in press as: Kajala et al., Innovation, conservation, and repurposing of gene function in root cell type development, Cell (2021), https://doi.org/10.1016/j.cell.2021.04.024

CellPress
OPEN ACCESS

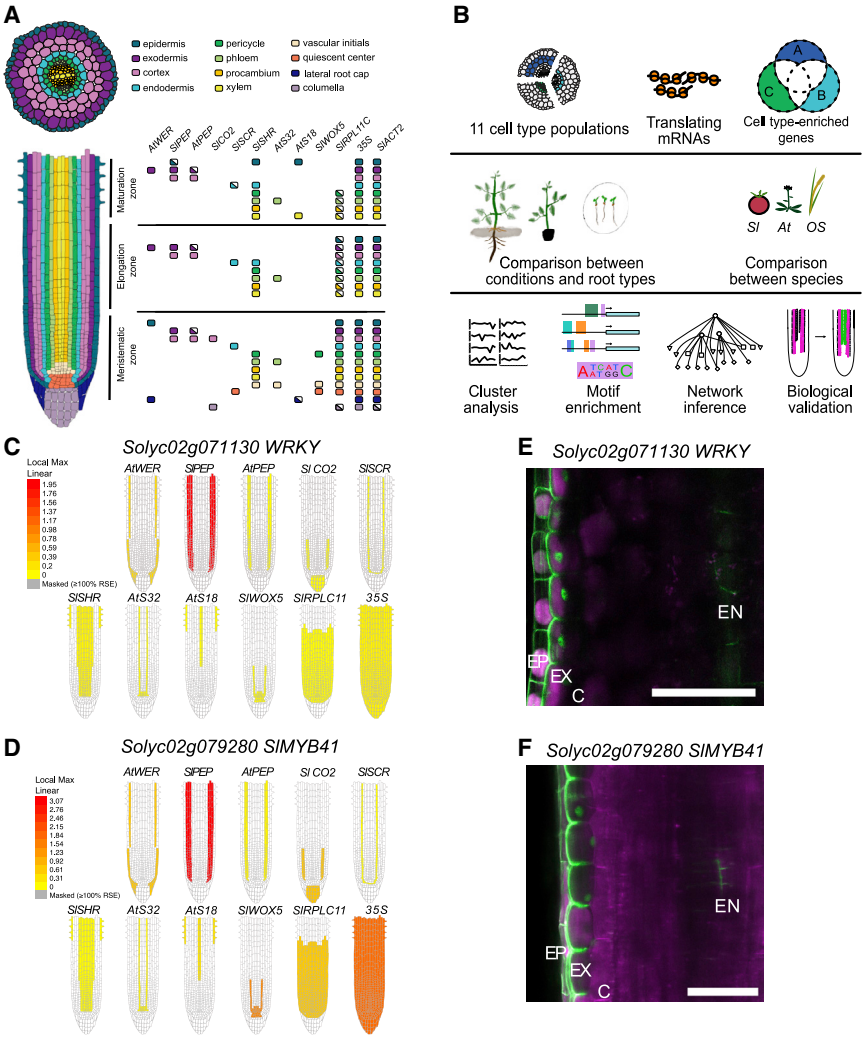


Figure 1. Tomato translatome root atlas at cell type resolution

(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

Please cite this article in press as: Kajala et al., Innovation, conservation, and repurposing of gene function in root cell type development, Cell (2021), https://doi.org/10.1016/j.cell.2021.04.024

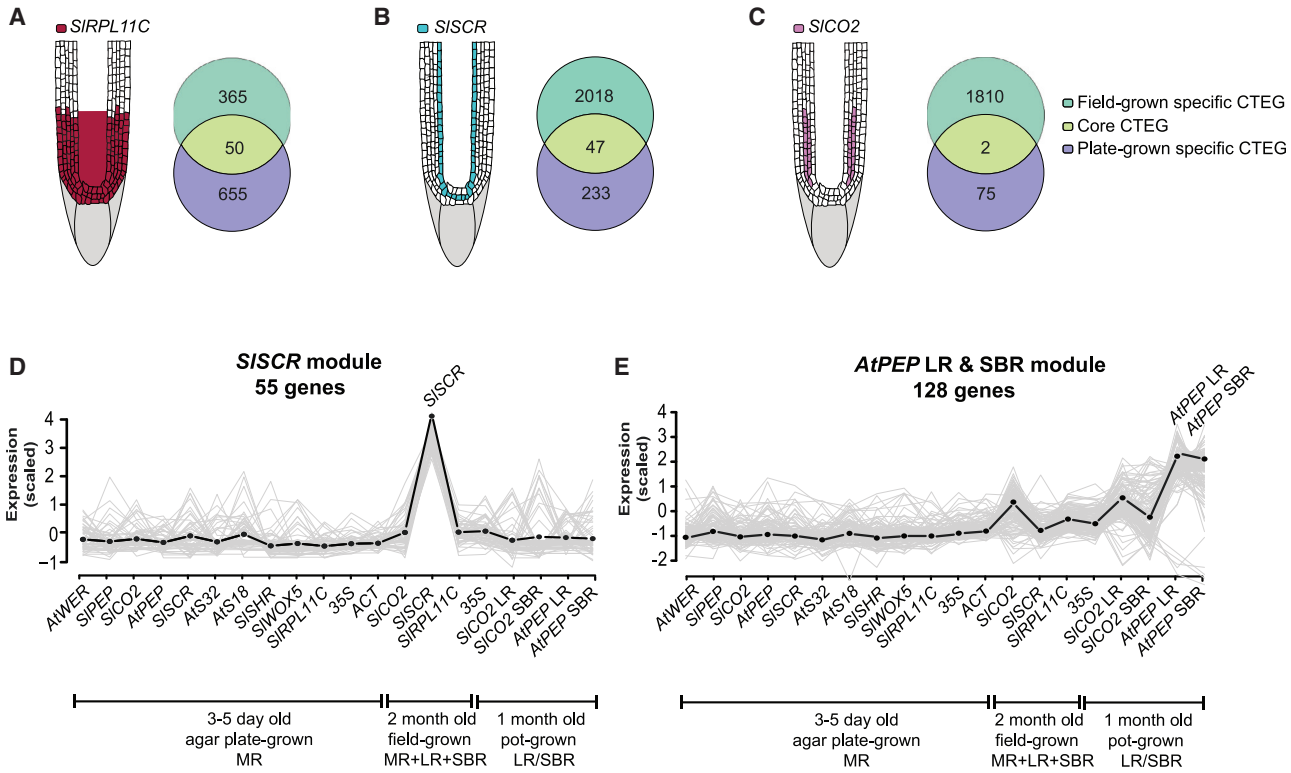


Figure 2. Cell type-enriched genes (CTEGs) across multiple developmental zones, conditions, and root types

(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 metaxylem cells mis-specified as protoxylem (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 protoxylem specification. Constitutive expression of miRNA-resistant versions of *SlCNAL1* and *SlPHB/PHV-LIKE1* was indeed sufficient to regulate protoxylem vessel identity and patterning within the vascular cylinder, with *SlPHB/PHV-LIKE1* recapitulating the phenotype of constitutive vascular *AtPHB* expression in *Arabidopsis* (Miyashima et al., 2011). 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

Please cite this article in press as: Kajala et al., Innovation, conservation, and repurposing of gene function in root cell type development, Cell (2021), https://doi.org/10.1016/j.cell.2021.04.024

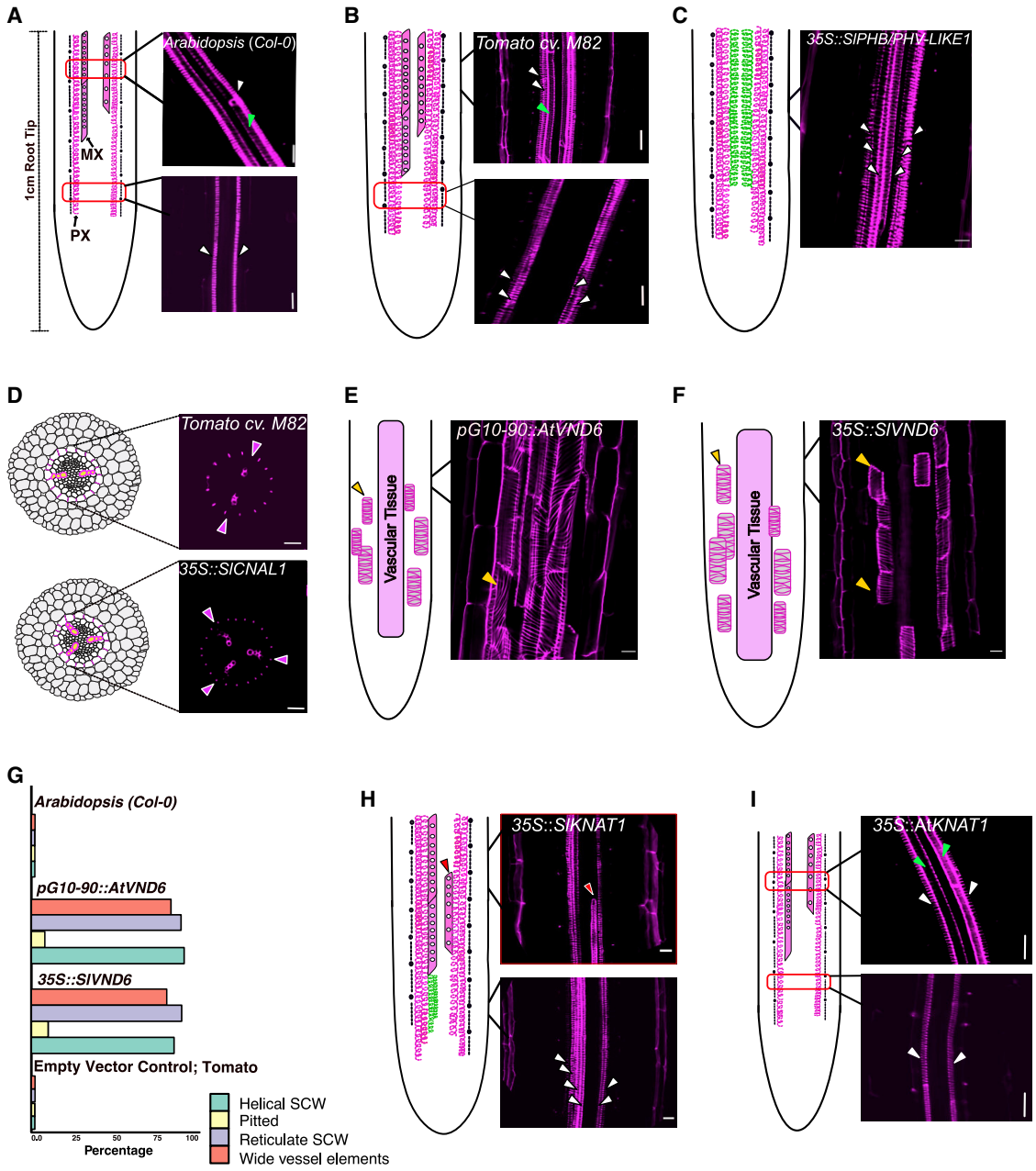


Figure 3. Identification of xylem vessel transcriptional regulators in tomato

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-LIKE1* 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::AtKNAT1* 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.

functional ortholog of *AtVND6* and assign it the name *SlVND6*. These experimental validations demonstrate likely 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

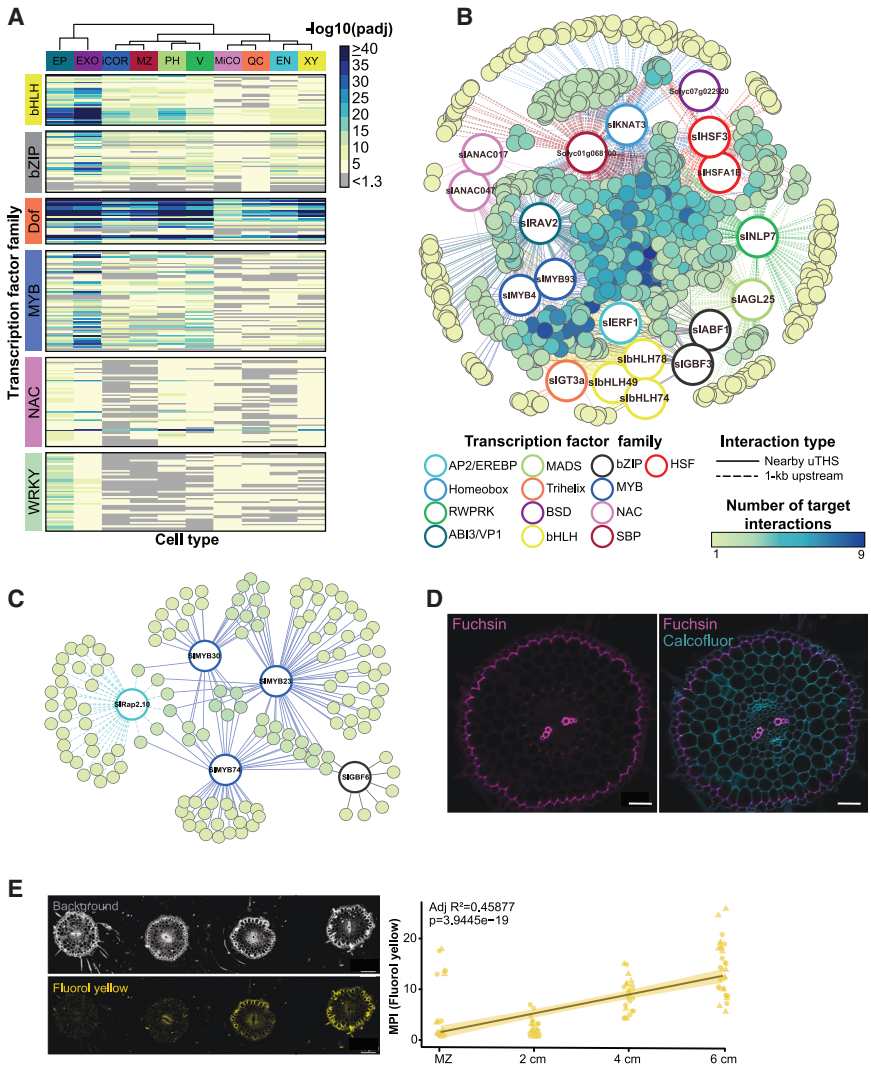


Figure 4. Inferred exodermis and inner cortex regulatory network and function

(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

Please cite this article in press as: Kajala et al., Innovation, conservation, and repurposing of gene function in root cell type development, Cell (2021), https://doi.org/10.1016/j.cell.2021.04.024

CellPress
OPEN ACCESS

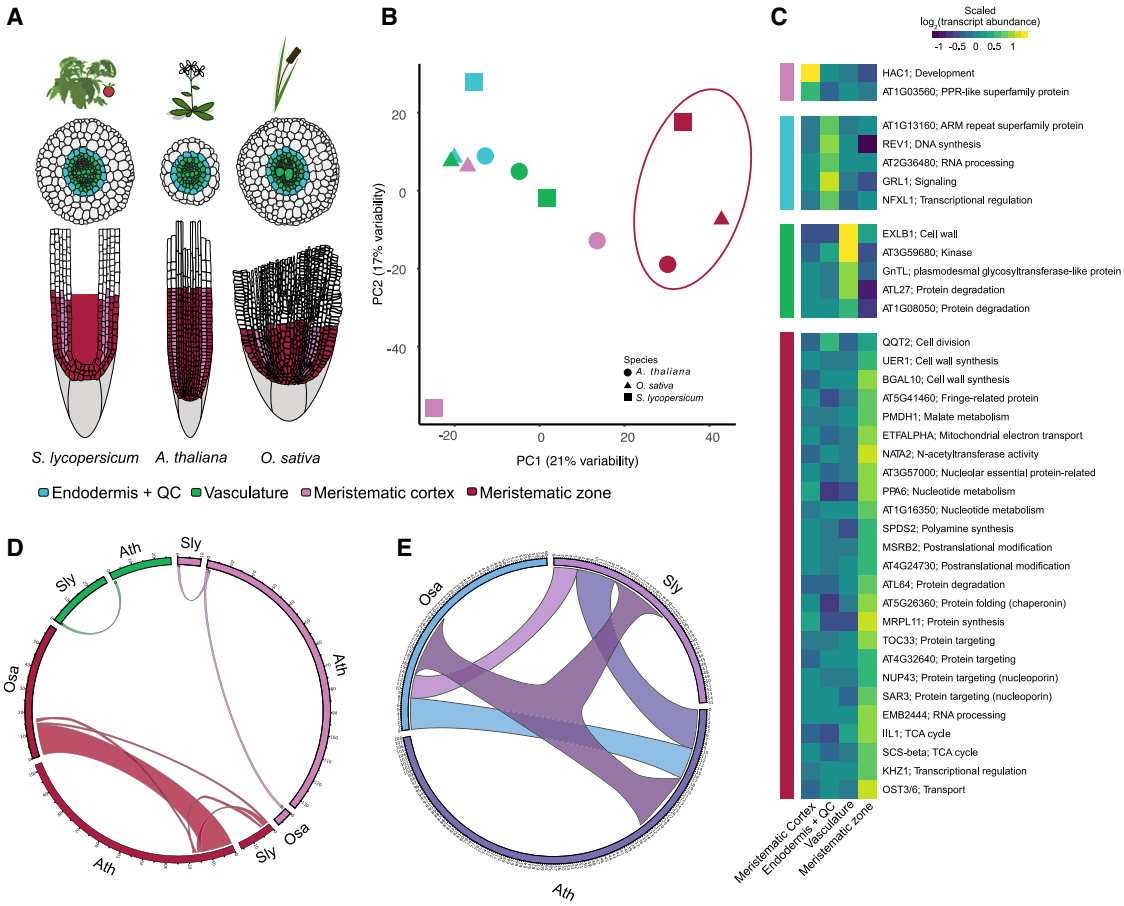


Figure 5. Homologous cell populations show limited conservation of gene expression

(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

Please cite this article in press as: Kajala et al., Innovation, conservation, and repurposing of gene function in root cell type development, Cell (2021), https://doi.org/10.1016/j.cell.2021.04.024

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.

Please cite this article in press as: Kajala et al., Innovation, conservation, and repurposing of gene function in root cell type development, Cell (2021), https://doi.org/10.1016/j.cell.2021.04.024

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*)

Please cite this article in press as: Kajala et al., Innovation, conservation, and repurposing of gene function in root cell type development, Cell (2021), https://doi.org/10.1016/j.cell.2021.04.024

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-Sl	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.

Please cite this article in press as: Kajala et al., Innovation, conservation, and repurposing of gene function in root cell type development, Cell (2021), https://doi.org/10.1016/j.cell.2021.04.024

CellPress
OPEN ACCESS

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 μmol/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 $\geq$ 2 and FDR adjusted P value (adj.P.Val) $\leq$ 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 $\geq$ 2, FDR adjusted P value $\leq$ 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

Please cite this article in press as: Kajala et al., Innovation, conservation, and repurposing of gene function in root cell type development, Cell (2021), https://doi.org/10.1016/j.cell.2021.04.024

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

Please cite this article in press as: Kajala et al., Innovation, conservation, and repurposing of gene function in root cell type development, Cell (2021), https://doi.org/10.1016/j.cell.2021.04.024

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).

Please cite this article in press as: Kajala et al., Innovation, conservation, and repurposing of gene function in root cell type development, Cell (2021), https://doi.org/10.1016/j.cell.2021.04.024

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

Please cite this article in press as: Kajala et al., Innovation, conservation, and repurposing of gene function in root cell type development, Cell (2021), https://doi.org/10.1016/j.cell.2021.04.024

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

Please cite this article in press as: Kajala et al., Innovation, conservation, and repurposing of gene function in root cell type development, Cell (2021), https://doi.org/10.1016/j.cell.2021.04.024

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

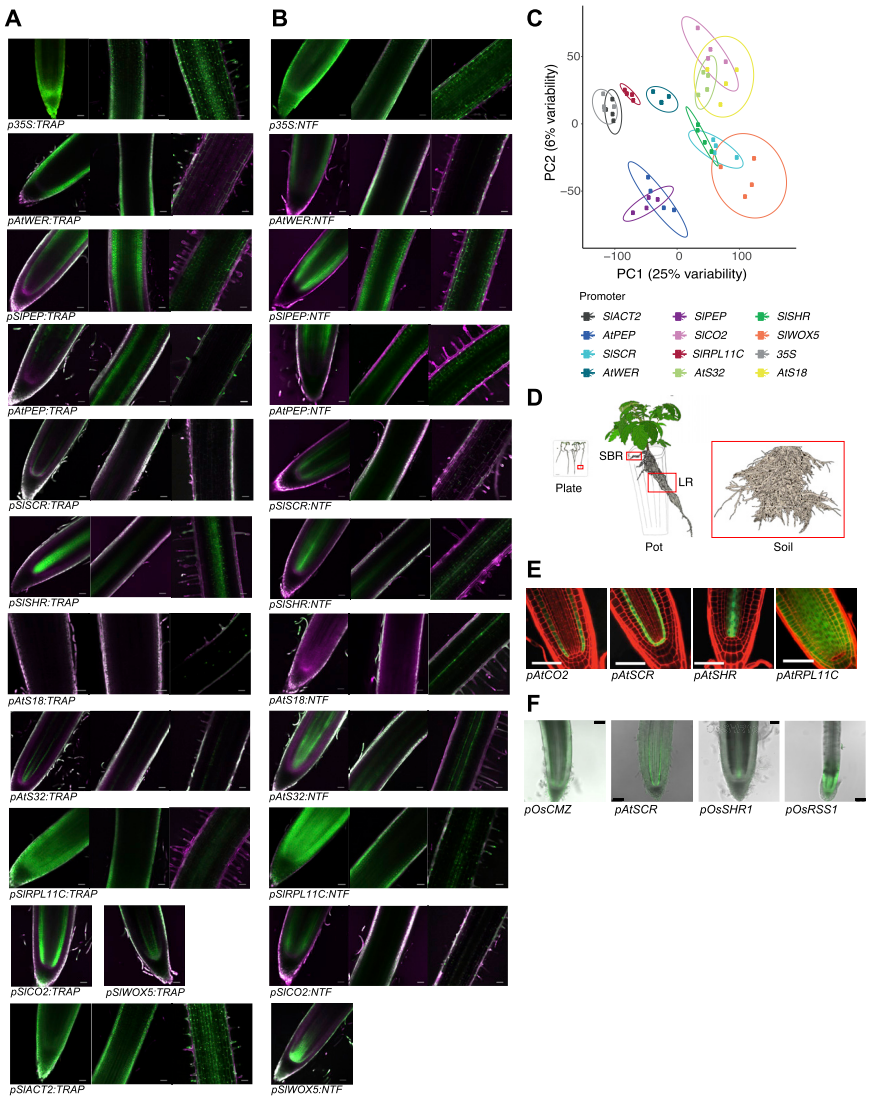


Figure S1. GFP expression in the TRAP and INTACT (nuclear tagging fusion [NTF]) lines and reproducibility of translatome biological replicates, related to Figure 1 and Table S1

(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**).

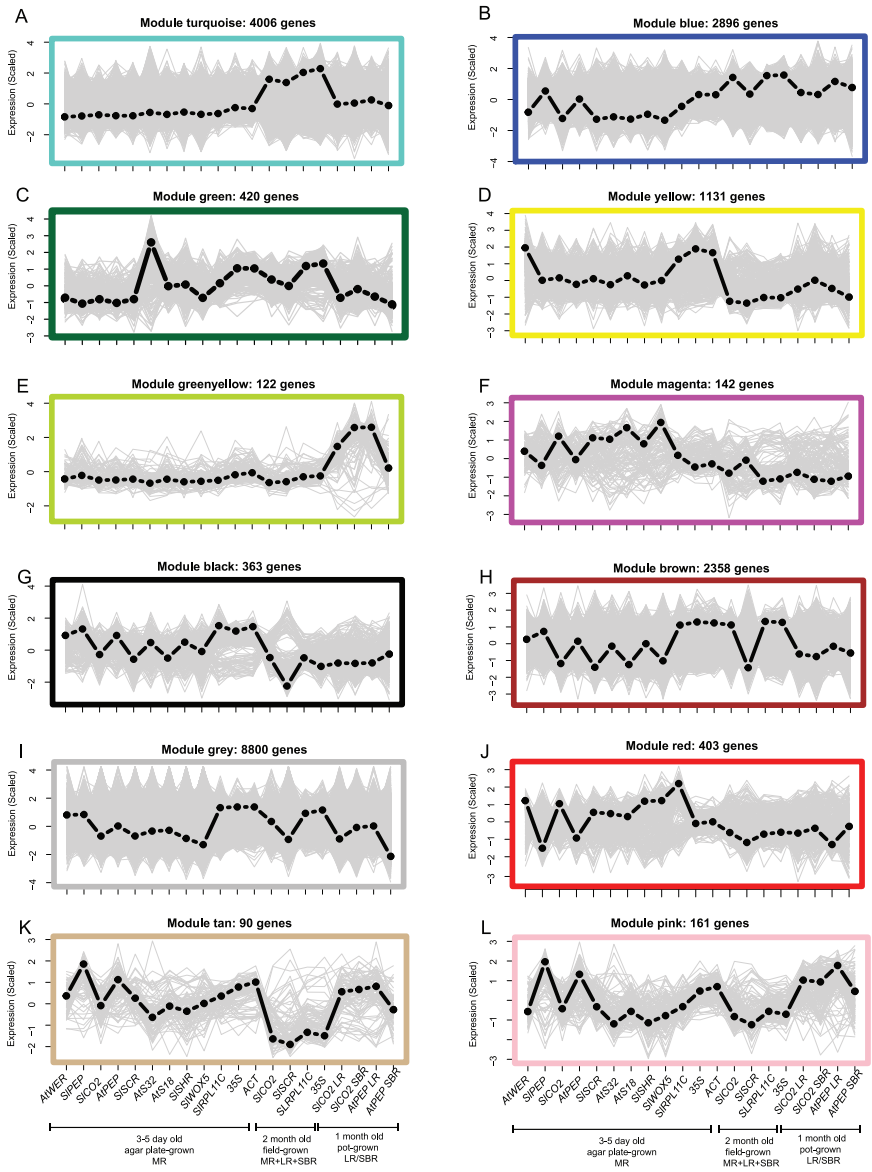
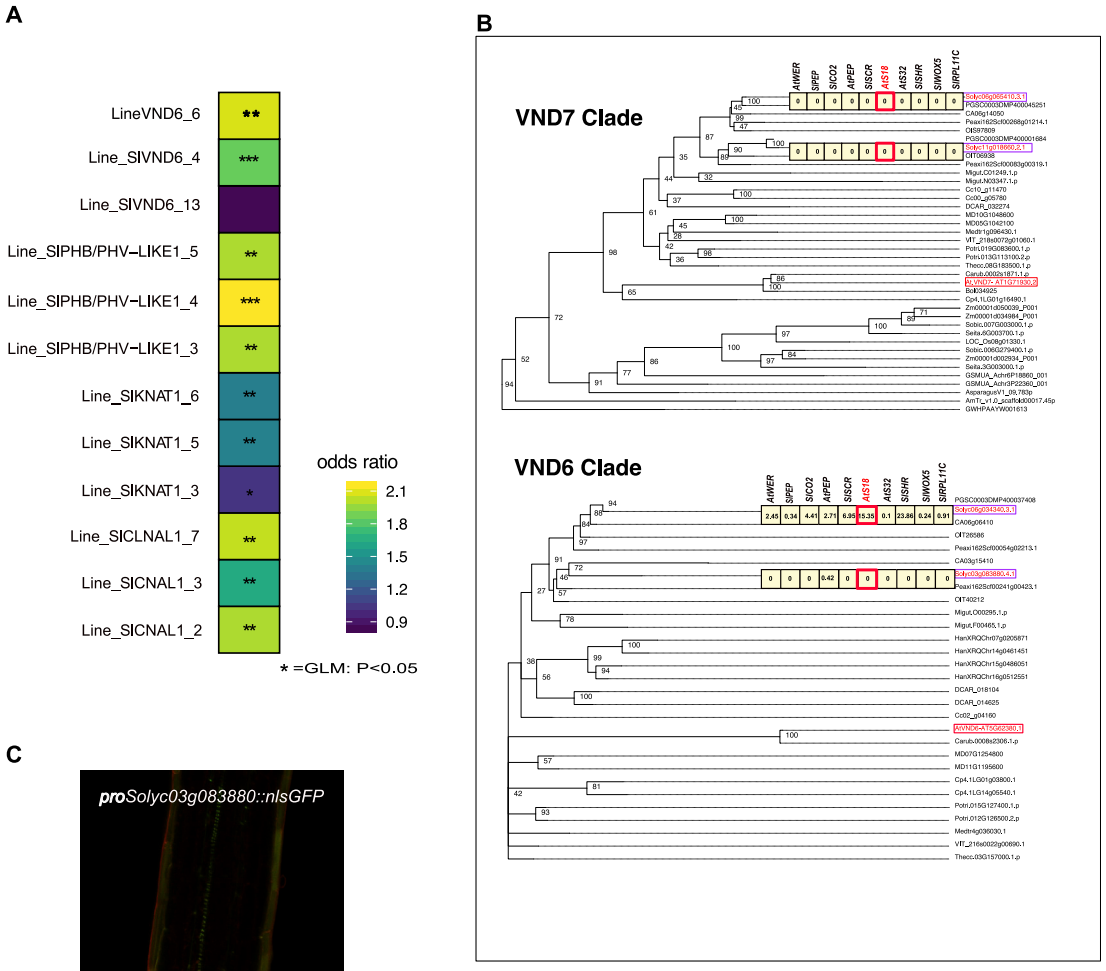


Figure S2. Expression profiles of WGCNA co-expression modules not shown in Figures 2D and 2E, related to Figure 2 and Table S2

(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.



(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

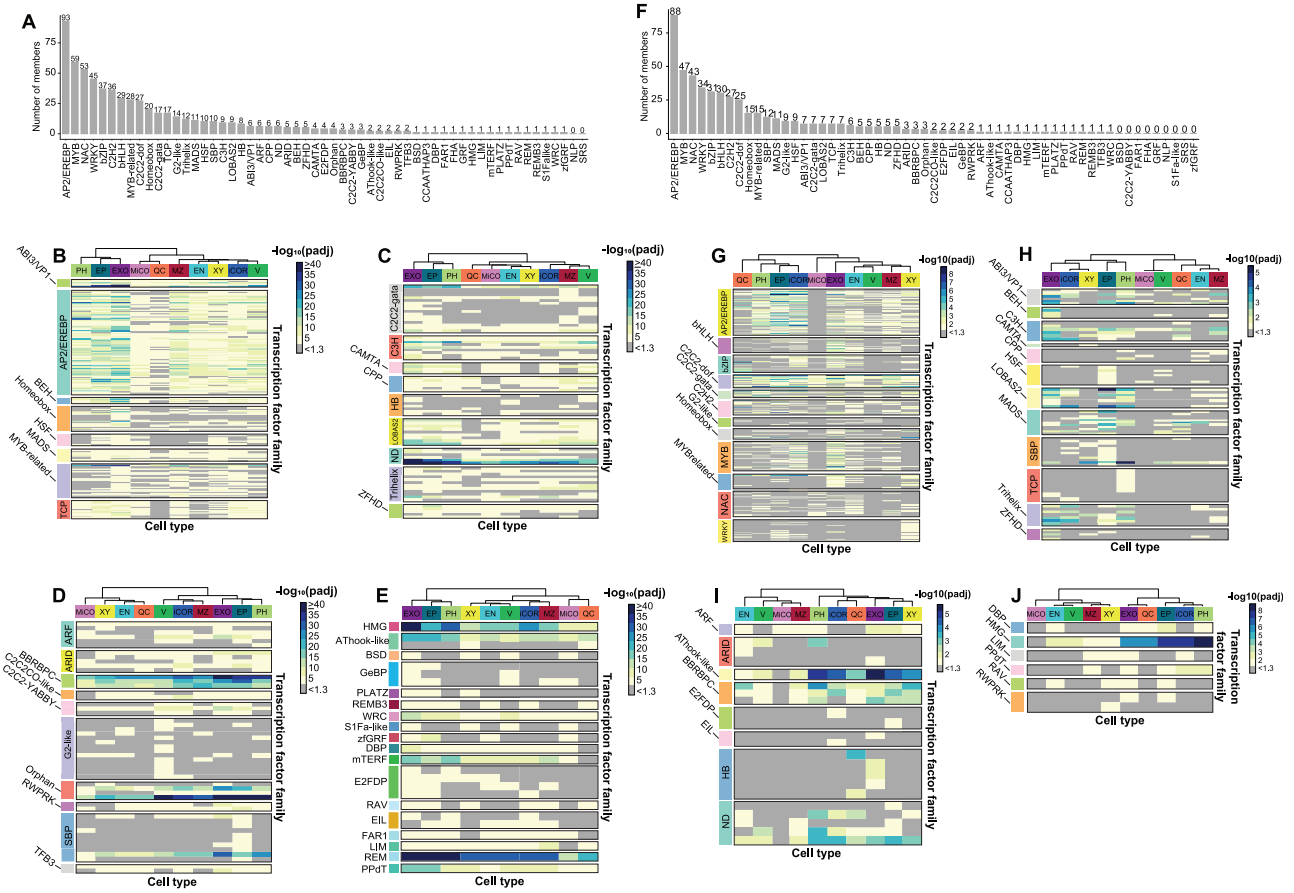


Figure S4. Transcription factor motifs enriched in 1-kb promoters and accessible regions near cell-type-enriched genes, related to Figure 4 and Table S4

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.

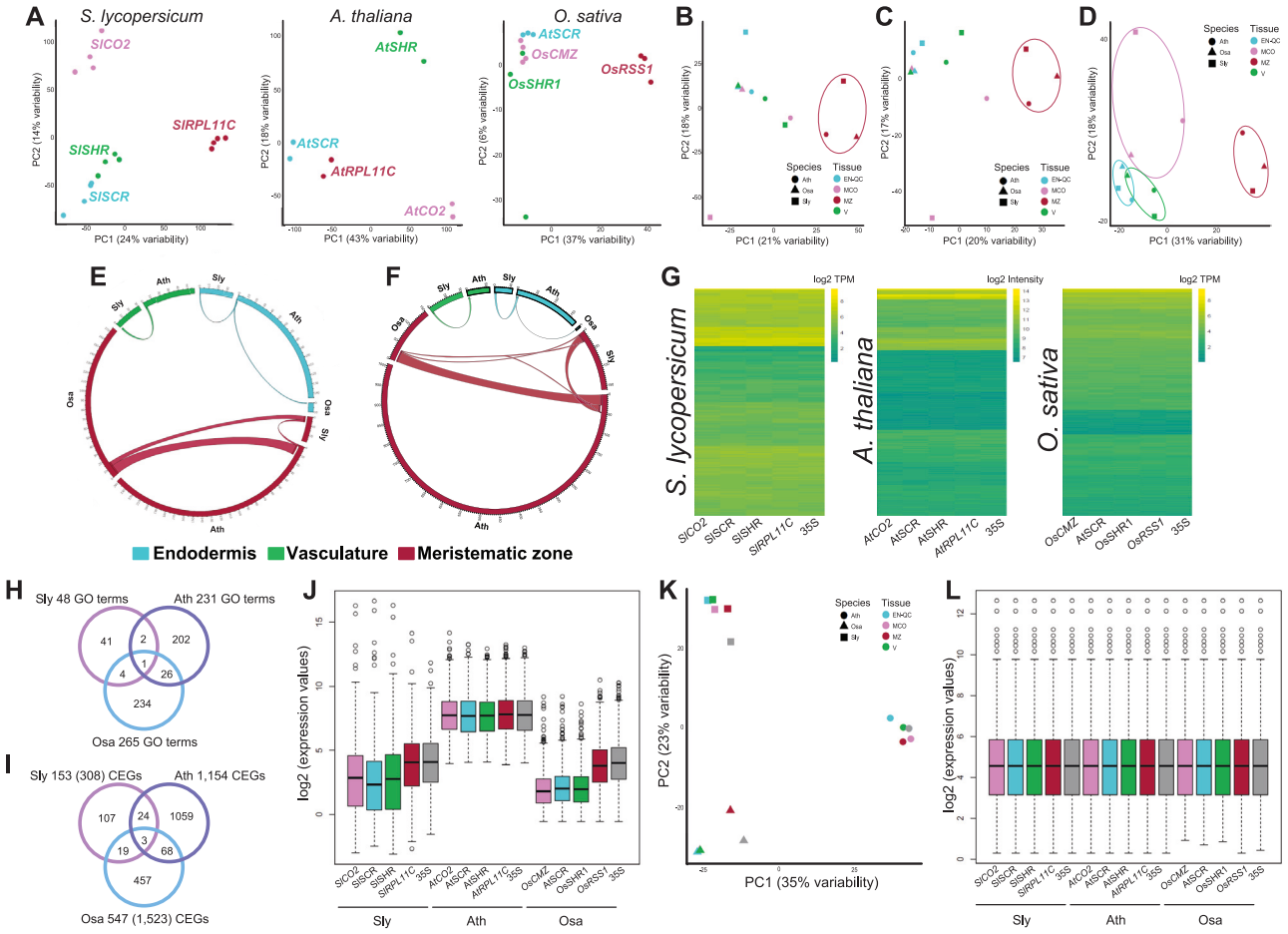


Figure S5. Multi-species analyses demonstrate similar translatome profiles of the meristematic zone compared with other cell populations and functional conservation of CEGs across species, related to Figure 5 and Table S5

(**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.

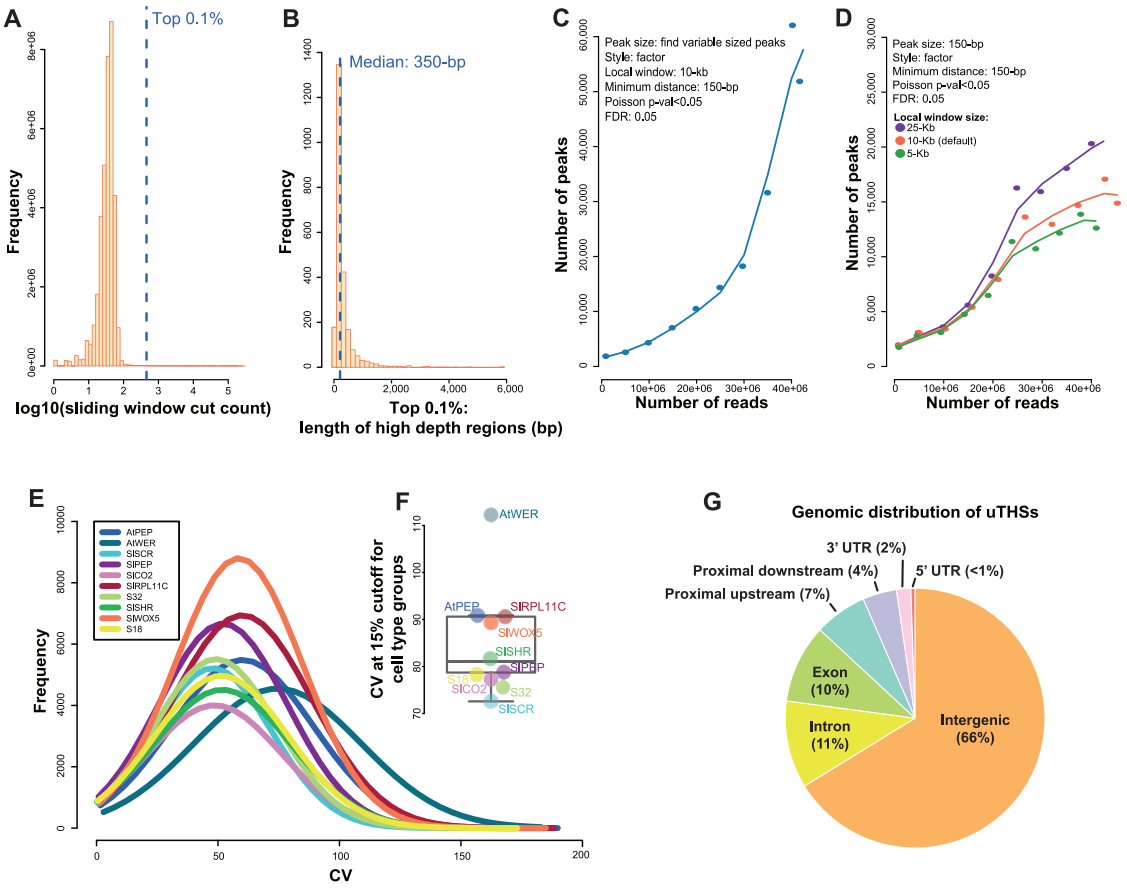


Figure S6. Summary of analysis methods for ATAC-seq, related to Figure 4 and Table S4

(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.