Cell-by-cell dissection of phloem development links a maturation gradient to cell specialization

Roszak, Pawel

2021-12-24

Roszak , P , Heo , J O , Blob , B , Toyokura , K , Sugiyama , Y , de Luis Balaguer , M A , Lau , W W Y , Hamey , F , Cirrone , J , Madej , E , Bouatta , A M , Wang , X , Guichard , M , Ursache , R , Tavares , H , Verstaen , K , Wendrich , J , Melnyk , C W , Oda , Y , Shasha , D , Ahnert , S E , Saeys , Y , De Rybel , B , Heidstra , R , Scheres , B , Grossmann , G , Mähönen , A P , Denninger , P , Göttgens , B , Sozzani , R , Birnbaum , K D & Helariutta , Y 2021 , ' Cell-by-cell dissection of phloem development links a maturation gradient to cell specialization ' , Science , vol. 374 , no. 6575 , eaba5531 . https://doi.org/10.1126/science.aba5531

http://hdl.handle.net/10138/343022
https://doi.org/10.1126/science.aba5531

unspecified
acceptedVersion

Title: Cell-by-cell dissection of phloem development links a maturation gradient to cell specialization.

Authors: Pawel Roszak[1]†, Jung-ok Heo[1,2]†, Bernhard Blob[1]†, Koichi Toyokura[1,3,4,5]†, Yuki Sugiyama[1,6]‡, Maria Angels de Luis Balaguer[7]‡, Winnie W. Y. Lau[8]‡, Fiona Hamey[8]‡, Jacopo Cirrone[9]‡, Ewelina Madej[10], Alida M. Bouatta[11], Xin Wang[2], Marjorie Guichard[12,13], Robertas Ursache[1], Hugo Tavares[1,14], Kevin Verstaen[15,16], Jos Wendrich[17,18], Charles W. Melnyk[19], Yoshihisa Oda[6,20], Dennis Shasha[9], Sebastian E. Ahnert[1,21,22], Yvan Saeys[15,16], Bert De Rybel[17,18], Renze Heidstra[23], Ben Scheres[23,24], Guido Grossmann[12,13], Ari Pekka Mähönen[2], Philipp Denninger[11], Berthold Göttgens[8], Rosangela Sozzani[7]*, Kenneth D. Birnbaum[25]*, Yrjö Helariutta[1,2]*

Affiliations:

[1]The Sainsbury Laboratory, University of Cambridge; Cambridge, UK.

[2]Institute of Biotechnology, HiLIFE/Organismal and Evolutionary Biology Research Programme, Faculty of Biological and Environmental Sciences, Viikki Plant Science Centre, University of Helsinki; Helsinki, Finland.

[3]Department of Biological Sciences, Graduate School of Science, Osaka University; Osaka, Japan.

[4]Faculty of Science and Engineering, Konan University; Kobe, Japan.

[5]GRA&GREEN Inc., Incubation Facility, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Japan.

[6]Department of Gene Function and Phenomics, National Institute of Genetics; Mishima, Japan

[7]Plant and Microbial Biology Department, North Carolina State University; Raleigh, USA.

[8]Wellcome Trust and MRC Cambridge Stem Cell Institute and Department of Haematology, University of Cambridge; Cambridge, UK.

[9]Computer Science Department, Courant Institute for Mathematical Sciences, New York University; NY, USA.

[10]Faculty of Biochemistry, Biophysics and Biotechnology, Jagiellonian University; Kraków, Poland

[11]Plant Systems Biology, Technical University of Munich; Freising, Germany

[12]Institute of Cell and Interaction Biology, CEPLAS, Heinrich-Heine-University Düsseldorf; Düsseldorf, Germany

[13]Centre for Organismal Studies, Heidelberg University; Heidelberg, Germany

[14] Bioinformatics Training Facility, Department of Genetics, University of Cambridge, Cambridge, UK

[15]Data Mining and Modelling for Biomedicine, VIB Center for Inflammation Research; Ghent, Belgium

[16]Department of Applied Mathematics, Computer Science and Statistics, Ghent University; Ghent, Belgium

[17]Ghent University, Department of Plant Biotechnology and Bioinformatics; Ghent, Belgium

[18]VIB Center for Plant Systems Biology; Ghent, Belgium

[19]Department of Plant Biology, Swedish University of Agricultural Sciences; Uppsala, Sweden.

[20]Department of Genetics, the Graduate University for Advanced Studies, SOKENDAI; Mishima, Japan

[21] Department of Chemical Engineering and Biotechnology, University of Cambridge; Cambridge, UK

[22]The Alan Turing Institute, British Library; London, UK

[23]Department of Plant Sciences, Wageningen University and Research; Wageningen, The Netherlands.

[24]Rijk Zwaan R&D, 4793 Fijnaart, The Netherlands

[25]Center for Genomics and Systems Biology, New York University; NY, USA.

†These authors contributed equally to this work

‡These authors contributed equally to this work

*Corresponding authors

Abstract: In the plant meristem, tissue-wide maturation gradients are coordinated with specialized cell networks to establish various developmental phases required for indeterminate growth. Here, we used single-cell transcriptomics to reconstruct the protophloem developmental trajectory from birth of cell progenitors to terminal differentiation in the Arabidopsis root. PHLOEM EARLY DNA-BINDING-WITH-ONE-FINGER (PEAR) transcription factors mediate lineage bifurcation by activating GTPase signaling and prime a transcriptional differentiation program. This program is initially repressed by a meristem-wide gradient of PLETHORA transcription factors. Only the dissipation of PLETHORA gradient permits activation of the differentiation program that involves mutual inhibition of early vs. late meristem regulators. Thus, for phloem development, broad maturation gradients interface with cell-type specific transcriptional regulators to stage cellular differentiation.

One-Sentence Summary: Single-cell analysis shows how global signals in the root meristem interact with the cell type specific factors to determine distinct phases of phloem development.

Main text: Roots consist of several concentric layers of functionally distinct cell files, which initially bifurcate and establish distinct identities around the quiescent center and its surrounding stem cells. Cells within each file mature through the distinct zones of cell proliferation and differentiation (*1*). For example, in Arabidopsis, the development of the protophloem sieve elements involves a transient period of cell proliferation, during which, in addition to amplification of cells within the file, two lineage-bifurcating events take place (Fig. 1A) (*2*). Soon after the cell proliferation ceases, cells of the protophloem sieve element lineage initiate a differentiation process which culminates in enucleation, an irreversible process that gives rise to the mature conductive cells (*3*). Because of specific modulation of the graded distribution of the key phytohormonal cue auxin, the differentiation of protophloem sieve elements occurs faster than that of the other cell files (*4*). Therefore, protophloem sieve element development offers a tractable scheme to understand how the two processes of cell specialization and maturation interact.

Phloem developmental trajectory at single-cell resolution.

In order to understand the process of protophloem sieve element development at a high resolution, we took a combination of approaches based on time-lapse confocal imaging (*5*) and single cell transcriptomics (*6*). Using phloem-specific marker (*pPEAR1::H2B-YFP pCALS7::H2B-YFP*) we precisely mapped cellular behavior of the on average of 19 cells that constitute the protophloem sieve element developmental trajectory until enucleation, which takes place every 2 hours in the final cell position. The passage of the cell from its "birth" at

the stem cell until its enucleation took a minimum of 79 hours (Fig. S1, movies S1, S2). To dissect the genetic control underlying this temporal progression, we opted for deep profiling of the 19 cells that represent the developmental trajectory of protophloem sieve element, using cell sorting and well-based single cell sequencing over higher throughput but shallower droplet-based profiling (*6–12*). We used fluorescent reporter lines whose expression represent various spatio-temporal domains within the developmental trajectory of protophloem sieve element (Fig. S2A, B). The single-cell profiles allowed us to cluster cells together with known protophloem sieve element markers to identify 758 cells that densely sampled the 19 cell positions and captured the span of protophloem sieve element maturation (Fig. 1B, Fig. S2C-G).

We sought to use the high-resolution profile of the protophloem sieve element lineage to ask how cell passage through stable signaling gradients in the meristem controls the stages of cellular specialization. In particular, while a number of regulators of either phloem cell identity or meristem zonation have been described (*13*, *14*), little is known about how these two regulatory processes interact to control organogenesis. Using Monocle 2 (*15*, *16*), we projected the 758 protophloem sieve element lineage cells into a pseudo-temporal order and investigated transcriptional transitions along the developmental trajectory (Fig. 1B-D). Rather than gradual changes, we observed four transcriptomic domains separated by three narrow transition zones (Fig. 1D, E; Table S1). Based on the alignment with the temporal expression patterns of selected genes, we were able to determine that these domains correspond approximately to cells at positions 1-7 [a], 8-11 [b], 12-15 [c] and 16-19 [d], respectively (Fig. S3). To further understand which aspects of protophloem sieve element maturation these various positions represent, we extended time-lapse confocal imaging with more temporally specific marker lines *pNAC86::H2B-YFP* and *pNEN4::H2B-YFP*, active at later developmental stages (3). We found that the differentiation time, measured from the last cell

division to enucleation takes around 20 hours with some variation up to the final stage defined by expression of *NAC45/86-DEPENDENT EXONUCLEASE-DOMAIN PROTEIN 4* (*NEN4*) (active in positions 18-19), (Fig. 1E, Fig. S1D, H, I, Movies S1-S12). In summary, based on the high congruence of the single-cell transcriptome and live imaging data, we were able to assign seven distinct developmental phases along the protophloem sieve element trajectory: (I) "stem cell", position 1; (II) "transit amplifying", position 2-9; (III) "transitioning", position 8-11; (IV) "early differentiating", position 10-15; (V) "late differentiating", position 16-17; (VI) "very late differentiating", position 18-19; (VII) "enucleating", position 19 (Fig. 1F, G, Fig. S1, Table S2).

PEARs promote lineage bifurcation via GTPase signaling.

Proximal to the stem cell (I) developmental phase, the first distinctive feature of the protophloem sieve element lineage is the bifurcation of the procambial and metaphloem cell files from the progenitor protophloem sieve element lineage through a pair of subsequent periclinal (asymmetric) cell divisions in the domain of transit amplifying cells (II). Using the single-cell lineage and imaging analysis, we sought to precisely map these divisions (Fig. 2A). We observed that the first periclinal division followed exclusively a rare event of phloem stem cell division (Movie S13, Fig. S4A). The second, more frequent, periclinal division was observed predominantly at position 3 (Fig. 1F). We have recently shown that the PEAR transcription factors (transcribed in domains I-IV) mediate early asymmetric divisions in the phloem lineage and laterally adjacent procambial cells in a cell autonomous and cell non-autonomous manner, respectively (*17*). In order to identify potential downstream effector genes for this PEAR function, we focused on the genes enriched in the expression domain of *pPEAR1Δ::erVenus* marker line (Methods) capturing the bifurcation events and the resulting protophloem, metaphloem and procambium cell lineages (Fig. 2B, Fig. S4B).

Among the sieve element enriched genes that were highly expressed in single cell profiles preceding and during the bifurcation (domain II), we identified and validated the protophloem sieve element abundant expression of Rho-related GTPase, Rho of plants 9 (ROP9) (*18*) as well as several genes encoding PRONE-type ROP guanine nucleotide exchange factors (ROPGEF) (Fig. 2B, C, D, Fig. S4B, C, F) (*19*). ROP GTPase signaling controls polarity of the multiple cell types during cell differentiation (*20*-*22*) and specific cell division events (*23*-*25*). Subsequently, we determined that ROPGEF3 and ROPGEF5 expression in the protophloem sieve element lineage is dependent upon PEAR factors, based on the spatio-temporal correlation as well as the analysis of transcriptional reporters in the *pear* sextuple mutant background (Fig. 2E). In addition, functional analysis of the PEAR binding sites previously indicated by the DAPseq technique (*26*) in the promoter region of ROPGEF genes affected their expression level (Fig. S4D) (17), suggesting a direct interaction.

In the dividing cells, ROPGEFs accumulate broadly at the cell membrane but were depleted from the expected position of cortical division zone, which demarcates the future division plane (Fig. 2F) (*25*). Indeed, observed gaps in ROPGEF localization coincided with the position of microtubule array called the preprophase band, the earliest marker of cell division plane in plants (Fig. 2G, Fig. S4E) (*25*). ROPGEFs catalyze disassociation of GDP from inactive ROP-GDP complex that enables quick binding of free cytosolic GTP and thus activates ROP signaling. In the active state, ROP-GTP interacts with a number of different effector proteins to mediate downstream signaling (*27*). In order to detect cellular position of the active ROP signaling in relation to the periclinal and anticlinal cell division planes in phloem, we utilized molecular biosensor of ROP signaling that consist of fluorescently tagged, ROP-GTP binding domain from MICROTUBULE DEPLETION DOMAIN1 (MIDD1ΔN) effector protein (*28*). Similarly to the localization of ROPGEFs, subcellular

localization of active ROP signaling was detected on the cell membrane and was absent in the cortical division zone of protophloem sieve element cells during mitosis (Fig. 2H).

In order to test whether ROP signaling plays a decisive role in the selection of cell division plane, we generated an inducible line expressing the constitutively active form of ROP9 ($ROP9^{CA}$) (Methods) and lines ectopically expressing phloem enriched ROPGEFs. Accumulation of $ROP9^{CA}$-3xYFP on the radial walls of the protophloem sieve element lineage correlated with cell expansion to the radial direction and reorientation of the cell division plane (Fig. 2I, Fig. S4F). Ectopic expression of ROPGEFs resulted in ectopic periclinal cell divisions in the outer root layers and pericycle, which rarely undergo such division (Fig. 2J, K, Fig. S4G, H). Members of PRONE-type ROPGEF gene family in Arabidopsis have been previously proposed to act redundantly in number of processes in which they activate ROP signaling (*29*). On the other hand, loss of *SPIKE1* (*SPK1*), encoding a single copy ROP interacting DOCK family GEF causes phenotypes mimicking the combinatorial *rop* mutants (*30*-*32*). Therefore, we focused on the loss-of-function alleles of *SPK1*, one of which we identified in the genetic screen for factors promoting formative (periclinal) cell divisions (Supplementary Materials). In the *spk1* loss-of-function mutant, we detected a significant reduction in periclinal divisions in several tissues, including protophloem sieve element cell lineage (Fig. 2L, M, Fig. S4I, J, K). We conclude that, in the transit amplifying cells (domain II, position 2-9), PEAR function promotes the bifurcation involving the emergence of the protophloem sieve element cell lineage by switching the orientation of the cell divisions at least partially through the activation of ROPGEF-ROP signaling module.

PLETHORAs stage *APL* expression and phloem differentiation.

Another distinct feature of the early protophloem sieve element developmental trajectory is the transition from cell division to cell differentiation (II-III-IV). This transition mapped closely to the first major change in the protophloem sieve element transcriptome. In the first transcriptomic domain (I-II), we detected transcripts of the PLETHORA gene family (Fig. 1E), whose relatively persistent proteins are known to spread shootward through cell-to-cell movement. This movement, together with a mitotic dilution effect, contributes to the formation of the shootward protein gradient. (*14*). Prior work has shown that PLETHORA transcription factors broadly regulate meristem development, promoting cell division at moderate concentrations, and then permitting elongation and differentiation as levels drop (*14*, *33*, *34*). However, it is not clear how individual cell files interpret the meristem-wide PLETHORA gradient for their own specialized differentiation.

We hypothesized that the PLETHORA gradient might mediate the first transcriptional shift (i.e. domain II to III) towards protophloem sieve element differentiation by permitting a new set of transcripts to be expressed (Fig. 3A). We tested this hypothesis by driving PLETHORA2 (PLT2) under several promoters that extended its expression in the protophloem sieve element in later maturation stages than its native domain (Fig. 3B, Fig. S5A). When using the *pNAC86::XVE* inducible promoter, active in domains V-VII (*3*, *35*), ectopic PLT2 delayed protophloem sieve element enucleation (Fig. 3B, Fig. S5A). Transcriptional profiling of phloem cells expressing the construct showed an upregulation of genes (Table S3) that mapped to early stages of the protophloem sieve element single-cell trajectory (from domains I-II) - the known PLT2 protein gradient (Fig. 3C). These results suggest that extending the PLT2 gradient is sufficient to prolong the early stages of meristem maturation within the protophloem sieve element lineage, providing a connection between the maturation of a specific cell file and a meristem-wide protein gradient. In addition, in the pseudo-time ordered single cells, we could detect complementary oscillatory patterns of the

putative S-phase and G2-M-phase genes that were upregulated PLETHORA targets, apparently corresponding to regular progressions through the cell-cycle (Fig. 3C, Fig. S5B). Furthermore, *ALTERED PHLOEM DEVELOPMENT* (*APL*), *NAC45/86* and *NEN4*, known key regulators of the protophloem sieve element enucleation pathway (*3*), were among the PLT2-downregulated genes (Fig S5C, Table S3). This is consistent with the presence of *APL* in the large set of genes downregulated by PLETHORA overexpression (*33*). We validated the downregulation of *APL* and *NEN4* by ectopic *PLT2* expression with in situ hybridization (Fig. 3D, Fig. S5D). We also monitored a shootward shift of *APL* expression domain in the roots after conditional ectopic induction of *PLT2* expression. The induction of PLT2 in the phloem cells beyond its native domain confirmed that activation of APL-dependent genetic program requires dissipation of the PLETHORA gradient (Fig. 3E). In order to test the role of PLETHORAs in controlling the transition between transit amplification and differentiation in phloem, we used an inducible, tissue specific CRISPR/Cas9 approach to mutate *PLT2* specifically in protophloem sieve element cell file (*36*). We observed an acceleration of the protophloem sieve element differentiation as well as the expression of *pAPL::erTurq* reporter towards the QC without affecting the broader meristem size or root growth, showing that loss of PLETHORA function in its native domain allows precocious expression of mid- to late-stage protophloem sieve element differentiation regulators (Fig. 3F, Fig. S5F-H).

We sought to further test whether PLT2 directly regulates the protophloem sieve element-specific differentiation program, as we found AP2 (a member of the PLETHORA family) family binding sites in the *APL* promoter region, as defined by the DAPseq technique (*26*). Indeed, we confirmed the direct binding of PLT2 to several regions of the *APL* promoter by ChIP-qPCR (Fig. 3G). Furthermore, along with AP2 sites, the *APL* promoter is also enriched for binding sites of HANABA TANARU (HAN), a GATA transcription factor. In turn, *HAN* is a PLETHORA target (*33*) and accordingly, upon ectopic PLT2 expression we detect *HAN*

transcripts expressed in late protophloem sieve element development (Fig. S5C, I). Ectopic HAN expression under *pNAC86:XVE* led to a delay in enucleation (Fig. S5J), similar to PLT2 overexpression in the same domain. We conclude that the PLETHORA gradient directly (and possibly in a feedforward manner with HAN) orchestrates protophloem sieve element differentiation by cell autonomously repressing transcription of the phloem regulator *APL*. Overall, the results show how the PLETHORA gradient first promotes cell proliferation in the protophloem sieve element lineage and then helps to time the later stages of cellular maturation.

PEARs promote *APL* to orchestrate phloem differentiation.

Given the results above, we reasoned that an early phloem-specific transcription factor must activate *APL* expression. In order to identify genes that could fill that role, we first generated a list of sieve element genes enriched in our bulk-sorted cells from that tissue compared to published data profiling other tissue types of the root meristem (*37*, Fig. S6A, Table S4). We further narrowed the list by intersecting it with sieve element enriched genes identified in the cluster analysis of single-cell RNAseq profiles of the *pPEAR1Δ::erVenus* reporter line (Table S5; Fig. 4A, B, Fig. S6B-H). From this analysis, we identified 542 sieve element enriched genes (Table S6) and corroborated their specificity in the published whole-root scRNAseq atlas (Table S7) (*12*). We modeled gene regulation using a machine learning approach on the pseudotime-ordered 758 single-cell profiles and 4924 highly variable genes. Among 208 TFs in this dataset, the majority of known protophloem sieve element transcription factors (such as *APL*, *NAC045* and *NAC086*) were among the top 20 regulators (Table S8). We validated the model by comparing predicted targets with genes induced by *in vivo* ectopic expression of the same TFs, confirming a significant overlap of targets in 3 out of 5 cases (Table S8). Among the top 20 regulators we also identified four related genes that encode early sieve

element abundant PEAR transcription factors (*PEAR1, PEAR2, DNA BINDING WITH ONE FINGER6, TARGET OF MONOPTEROS6*) (Fig. 4C). We recently showed that simultaneous loss of six PEAR genes results in defects in protophloem sieve element differentiation (*17*). We subsequently profiled the transcriptomes of wildtype and *pear* sextuple mutant (Fig. 4D) root meristems and identified 203 downregulated genes overlapping with our protophloem sieve element specific gene list (Table S9). The expression of *APL* as well as its downstream targets – *NAC045, NAC086* and *NEN4* was lost in protophloem tissue of *pear* sextuple mutant (Fig. 4E, F Fig. S7A). Subsequently, expression of *APL* and *NAC086* reporter lines was restored in the pear sextuple mutant upon induction of *PEAR1*, corroborating that transcriptional activation of *APL* in the protophloem sieve element is dependent on activity of PEAR factors (Fig. 3F).

To test whether PEAR1 can directly regulate expression of *APL* in its endogenous expression domain (cells 1-14), we performed chromatin immunoprecipitation (ChIP) followed by quantitative PCR (qPCR) using *pPEAR1::PEAR1-GFP* protein fusion and identified multiple PEAR1 binding sites within *APL* promoter (*pAPL*) (Fig. 4G). Truncation analysis of *pAPL* indicated presence of an enhancer element, responsible for expression of *APL* in the cells transitioning from cell division to cell differentiation, within 2039 bp to 2962 bp region upstream of *APL* open reading frame (ORF) (Fig. 4H). Our ChIP analysis detected a single strong PEAR1-GFP peak in the promoter sequence beyond 2039 bp distance from the ORF and another strong peak at the upstream end of the 2 kb region, both of which were also detected in the publicly available DAP-Seq data (Fig. 4G, Fig. S7C) (*26*). Furthermore, within the detected regions (-2672 to -2512 and -1946 to -1844) we identified multiple clusters of DOF binding motifs (AAAG) (*26*) that constitute an enhancer element required for the transcriptional activation of *APL* in the phloem transition zone (domain III) (Fig. 4H, I, Fig. S7C). Although the expression of *APL* in the protophloem sieve element is dependent on

PEARs (Fig. 4F), *APL* expression domain extends beyond *PEAR* domain (cells 15-19; Fig. 1E, Fig. S3A). It is possible that either the PEAR proteins and/or *APL* mRNA persist this period of some 10 hours before enucleation. Alternatively, there may be intermediate factors acting downstream of PEARs to promote *APL* expression during late stages of phloem development. Collectively, the data supports a role for PEARs controlling the onset of APL expression to regulate a transition in phloem differentiation. The transition is controlled by the PLETHORAs, whose role in promoting division ultimately dissipates its own gradient. When PLETHORA levels decline sufficiently, PEARs can then effectively upregulate *APL*. The opposing regulation of *APL* by positively regulating PEARS and inhibitory PLETHORAs illustrates how antagonistic mechanisms – one forming a morphogen-like gradient across the meristem – orchestrate developmental timing within a cell file.

Sequential mutual inhibition directs developmental transitioning.

The final major transcriptional transition in the phloem lineage occurs between the domains IV-V. To explore this transition, we ectopically expressed NEN4 and PLT2 at various developmental stages. When expressed in early ectopic domains, NEN4 expression causes cell death, while PLT2 expression forces cells back into the cell cycle. However, later expression of these two transcription factors, have little or no visible effect on cells, showing the developmental program of domain V appears resilient to these perturbations (Fig. 3B, Fig. S5A, Fig. S9). This indicates that the high number of protophloem sieve element specific genes during the final 8 hours of differentiation remodel the cellular behavior in an irreversible manner. We next sought to explore how widely the PEARs control transcriptional programs related to this final stage of sieve element development. We combined a gene regulatory analysis in the *pear* mutant with systematic overexpression and modelling approaches (Fig. S7A, B, Fig. S8). Our analysis revealed that - in addition to known phloem

regulators *APL, NAC045, NAC086* and *NAC028* - 10 out of 13 newly validated phloem enriched transcription factors are dependent on PEARs (Fig. S7A, B, Fig. 4F). Overexpression of two of these, *ZAT14* (AT5G03510), which was also the 3rd most important TF in the machine learning model, and its close homolog *ZAT14L* (AT5G04390) led to arrest of cell cycle and premature cell elongation (Fig. 4J, K). Transcriptional profiling provided further evidence for a putative dual role in timing cell division and cell expansion (that occurs largely after enucleation in this cell lineage) (Tables S10-S14). In addition, the gene regulatory network model predicted a pattern of sequential mutual inhibition in the target sets of high-scoring transcriptional regulators (Table S15); for example, genes repressed by ZAT14 significantly overlap with genes activated by the earlier expressed PEARs and *vice versa* (Fig. 4L). Overexpression analysis confirmed a significant over-representation in the overlap between genes up-regulated by PEARs and down-regulated by ZAT14 (Table S16) (*17*).

By combining single-cell transcriptomics with live imaging, here we have mapped the cellular events from the birth of the phloem cell to its terminal differentiation into phloem sieve element cells spanning a timeframe of 79 hours. In the early part of the developmental trajectory, where cells are proliferating, the PEAR factors promote the asymmetric periclinal divisions that result in lineage bifurcation. We pinpoint the ROPGEF-ROP regulatory module as an effector of early PEAR function in promoting the periclinal cell divisions central to vascular development. In addition, the PEARs activate the final 20-hour terminal differentiation program, which highlights them as central integrators that connect early and late phloem development. Our high-resolution phloem developmental trajectory reveals three abrupt transitions in the gene expression program. The late, PEAR-regulated protophloem sieve element program is directly and antagonistically controlled by the broad PLETHORA gradient, which connects this morphogen-like gradient to cellular maturation. We propose

that mutual inhibition of target genes by sequentially expressed transcription factors represents a “seesaw” mechanism (Fig. S10) that allows rapid transitions and prevent gene expression programs with conflicting effects on cellular physiology (e.g., division vs. enucleation). Similar models have been implicated in so-called attractor states in cell fate decisions in animals (*38*). In the future it will be interesting to determine how conserved these principles of sieve element differentiation are in an evolutionary context, as well as how extensively they apply to other differentiation trajectories in plants.

Methods summary

Single-cell transcriptomic data described in the manuscript were generated from the protophloem/metaphloem sieve element and procambial cells sorted with a use of tissue specific fluorescent reporter lines. Root tips of 5 days old Arabidopsis plants were used as a tissue material for protoplasting. RNA sequencing of the sorted cells was performed following well-based Smart-seq protocol. Obtained transcriptomes, corresponding to the cells from protophloem cell lineage, were ordered in pseudotime using Monocle2 package which generated a single linear protophloem developmental trajectory. Expression profiles and pseudotime coordinates of the known phloem-expressed genes were further confirmed with in situ and reporter lines analysis.

Gene regulatory network was modelled using a random forest machine learning approach. Selected interactions, representing mutual inhibition (the “seesaw” model), were confirmed by the transcriptome analysis of lines overexpressing a candidate gene or profiling of the loss-of-function lines.

To understand cell behaviour at different developmental phases, confocal long-term live imaging was performed with the protophloem sieve element specific and nuclear localised

reporter line. Up to 5-days long movies were recorded and cell behaviour, including number and position of cell divisions, enucleation as well as the time of these events were recorded. All the details of methods including those summarized above are provided in the supplementary materials.

References

1. L. Dolan, K. Janmaat, V. Willemsen, P. Linstead, S. Poethig, K. Roberts, B. Scheres, Cellular organisation of the Arabidopsis thaliana root. *Development*. **119**, 71–84 (1993).
2. A. P. Mähönen, M. Bonke, L. Kauppinen, M. Riikonen, P. N. Benfey, Y. Helariutta, A novel two-component hybrid molecule regulates vascular morphogenesis of the Arabidopsis root. *Genes Dev.* **14**, 2938–2943 (2000).
3. K. M. Furuta, S. R. Yadav, S. Lehesranta, I. Belevich, S. Miyashima, J. Heo, A. Vatén, O. Lindgren, B. D. Rybel, G. V. Isterdael, P. Somervuo, R. Lichtenberger, R. Rocha, S. Thitamadee, S. Tähtiharju, P. Auvinen, T. Beeckman, E. Jokitalo, Y. Helariutta, Arabidopsis NAC45/86 direct sieve element morphogenesis culminating in enucleation. *Science*. **345**, 933–937 (2014).
4. P. Marhava, A. E. L. Bassukas, M. Zourelidou, M. Kolb, B. Moret, A. Fastner, W. X. Schulze, P. Cattaneo, U. Z. Hammes, C. Schwechheimer, C. S. Hardtke, A molecular rheostat adjusts auxin flux to promote root protophloem differentiation. *Nature*. **558**, 297–300 (2018).
5. R. Rahni, K. D. Birnbaum, Week-long imaging of cell divisions in the Arabidopsis root meristem. *Plant Methods*. **15**, 30 (2019).

6. I. Efroni, A. Mello, T. Nawy, P.-L. Ip, R. Rahni, N. DelRose, A. Powers, R. Satija, K. D. Birnbaum, Root Regeneration Triggers an Embryo-like Sequence Guided by Hormonal Interactions. *Cell*. **165**, 1721–1733 (2016).
7. T. Denyer, X. Ma, S. Klesen, E. Scacchi, K. Nieselt, M. C. P. Timmermans, Spatiotemporal Developmental Trajectories in the Arabidopsis Root Revealed Using High-Throughput Single-Cell RNA Sequencing. *Dev. Cell*. **48**, 840-852.e5 (2019).
8. C. N. Shulse, B. J. Cole, D. Ciobanu, J. Lin, Y. Yoshinaga, M. Gouran, G. M. Turco, Y. Zhu, R. C. O'Malley, S. M. Brady, D. E. Dickel, High-Throughput Single-Cell Transcriptome Profiling of Plant Cell Types. *Cell Reports*. **27**, 2241-2247.e4 (2019).
9. K. Jean-Baptiste, J. L. McFaline-Figueroa, C. M. Alexandre, M. W. Dorrity, L. Saunders, K. L. Bubb, C. Trapnell, S. Fields, C. Queitsch, J. T. Cuperus, Dynamics of Gene Expression in Single Root Cells of Arabidopsis thaliana. *Plant Cell*. **31**, 993–1011 (2019).
10. T.-Q. Zhang, Z.-G. Xu, G.-D. Shang, J.-W. Wang, A Single-Cell RNA Sequencing Profiles the Developmental Landscape of Arabidopsis Root. *Mol Plant*. **12**, 648–660 (2019).
11. K. H. Ryu, L. Huang, H. M. Kang, J. Schiefelbein, Single-Cell RNA Sequencing Resolves Molecular Relationships Among Individual Plant Cells. *Plant Physiol.* **179**, 1444–1456 (2019).
12. J. R. Wendrich, B.-J. Yang, N. Vandamme, K. Verstaen, W. Smet, C. Van de Velde, M. Minne, B. Wybouw, E. Mor, H. E. Arents, J. Nolf, J. Van Duyse, G. Van Isterdael, S. Maere, Y. Saeys, B. De Rybel, Vascular transcription factors guide plant epidermal responses to limiting phosphate conditions. *Science*. **370**, 6518 (2020)
13. M. Bonke, S. Thitamadee, A. P. Mähönen, M.-T. Hauser, Y. Helariutta, APL regulates vascular tissue identity in Arabidopsis. *Nature*. **426**, 181–186 (2003).

14. A. P. Mähönen, K. Ten Tusscher, R. Siligato, O. Smetana, S. Díaz-Triviño, J. Salojärvi, G. Wachsman, K. Prasad, R. Heidstra, B. Scheres, PLETHORA gradient formation mechanism separates auxin responses. *Nature*. **515**, 125–129 (2014).
15. C. Trapnell, D. Cacchiarelli, J. Grimsby, P. Pokharel, S. Li, M. Morse, N. J. Lennon, K. J. Livak, T. S. Mikkelsen, J. L. Rinn, The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature Biotechnology*. **32**, 381–386 (2014).
16. X. Qiu, Q. Mao, Y. Tang, L. Wang, R. Chawla, H. A. Pliner, C. Trapnell, Reversed graph embedding resolves complex single-cell trajectories. *Nature Methods*. **14**, 979–982 (2017).
17. S. Miyashima, P. Roszak, I. Sevilem, K. Toyokura, B. Blob, J. Heo, N. Mellor, H. Help-Rinta-Rahko, S. Otero, W. Smet, M. Boekschoten, G. Hooiveld, K. Hashimoto, O. Smetana, R. Siligato, E.-S. Wallner, A. P. Mähönen, Y. Kondo, C. W. Melnyk, T. Greb, K. Nakajima, R. Sozzani, A. Bishopp, B. D. Rybel, Y. Helariutta, Mobile PEAR transcription factors integrate positional cues to prime cambial growth. *Nature*. **565**, 490–494 (2019).
18. P. Winge, T. Brembu, R. Kristensen, A. M. Bones, Genetic structure and evolution of RAC-GTPases in Arabidopsis thaliana. *Genetics*. **156**, 1959-1971 (2000).
19. A. Berken, C. Thomas, A. Wittinghofer, A new family of RhoGEFs activates the Rop molecular switch in plants. *Nature*. **436**, 1176-1180 (2005).
20. Y. Fu, Y. Gu, Z. Zheng, G. Wasteneys, Z. Yang, Arabidopsis interdigitating cell growth requires two antagonistic pathways with opposing action on cell morphogenesis. *Cell*. **120**, 687-700 (2005).

21. A. J. Molendijk, F. Bischoff, C. S. Rajendrakumar, J. Friml, M. Braun, S. Gilroy, K. Palme, Arabidopsis thaliana Rop GTPases are localized to tips of root hairs and control polar growth. *EMBO J*. **20**, 2779-2788 (2001).

22. Y. Oda, H. Fukuda, Initiation of cell wall pattern by a Rho- and microtubule-driven symmetry breaking. *Science*. **337**, 1333-1336 (2012).

23. J. A. Humphries, Z. Vejlupkova, A. Luo, R. B. Meeley, A. W. Sylvester, J. E. Fowler, L. G. Smith, ROP GTPases act with the receptor-like protein PAN1 to polarize asymmetric cell division in maize. *Plant Cell*. **6**, 2273-2284 (2011).

24. P. Yi, G. Goshima, Rho of Plants GTPases and cytoskeletal elements control nuclear positioning and asymmetric cell division during Physcomitrella patens branching. *Current Biology*. **30**, 2860-2868 (2020).

25. D. Stockle, A. Herrmann, E. Lipka, T. Lauster, R. Gavidia, S. Zimmermann, S. Müller, Putative RopGAPs impact division plane selection and interact with kinesin-12 POK1. *Nature Plants.* **2**, 16120 (2016).

26. R. C. O'Malley, S. C. Huang, L. Song, M. G. Lewsey, A. Bartlett, J. R. Nery, M. Galli, A. Gallavotti, J. R. Ecker, Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. *Cell*. **165**, 1280–1292 (2016).

27. G. Feiguelman, Y. Fu, S. Yalovsky, ROP GTPases structure-function and signaling pathways. *Plant Physiology*. **176**, 57-79 (2018).

28. Y. Sugiyama, M. Wakazaki, K. Toyooka, H. Fukuda, Y. Oda, A Novel Plasma Membrane-Anchored Protein Regulates Xylem Cell-Wall Deposition through Microtubule- Dependent Lateral Inhibition of Rho GTPase Domains. *Current Biology*. **27**, 2522-2528 (2017).

29. P. Denninger, A. Reichelt, V. A. F. Schmidt, D. G. Mehlhorn, L. Y. Asseck, C. E. Stanley, N. F. Keinath, J. F. Evers, C. Grefen, G. Distinct RopGEFs successively

drive polarization and outgrowth of root hairs. *Current Biology*. **29**, 1854-1865 (2019).

30. J. L. Qiu, R. Jilk, M. D. Marks, D. B. Szymanski, The Arabidopsis SPIKE1 gene is required for normal cell shape control and tissue development. *The Plant Cell*. **14**, 101-118 (2002).
31. D. Basu, J. Le, T. Zakharova, E. L. Mallery, D. B. Szymanski, A SPIKE1 signaling complex controls actin-dependent cell morphogenesis through the heteromeric WAVE and ARP2/3 complexes. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 4044-4049 (2007).
32. H. Ren, X. Dang, Y. Yang, D. Huang, M. Liu, X. Gao, D. Lin, SPIKE1 activates ROP GTPase to modulate petal growth and shape. *Plant Physiology*. **172**, 358-371 (2016).
33. C. Galinha, H. Hofhuis, M. Luijten, V. Willemsen, I. Blilou, R. Heidstra, B. Scheres, PLETHORA proteins as dose-dependent master regulators of Arabidopsis root development. *Nature*. **449**, 1053–1057 (2007).
34. L. Santuari, G. F. Sanchez-Perez, M. Luijten, B. Rutjens, I. Terpstra, L. Berke, M. Gorte, K. Prasad, D. Bao, J. L. P. M. Timmermans-Hereijgers, K. Maeo, K. Nakamura, A. Shimotohno, A. Pencik, O. Novak, K. Ljung, S. van Heesch, E. de Bruijn, E. Cuppen, V. Willemsen, A. P. Mähönen, W. Lukowitz, B. Snel, D. de Ridder, B. Scheres, R. Heidstra, The PLETHORA Gene Regulatory Network Guides Growth and Cell Differentiation in Arabidopsis Roots. *The Plant Cell*. **28**, 2937–2951 (2016).
35. R. Siligato, X. Wang, S. R. Yadav, S. Lehesranta, G. Ma, R. Ursache, I. Sevilem, J. Zhang, M. Gorte, K. Prasad, M. Wrzaczek, R. Heidstra, A. Murphy, B. Scheres, A. P. Mähönen, MultiSite Gateway-Compatible Cell Type-Specific Gene-Inducible System for Plants. *Plant Physiology*. **170**, 627–641 (2016).

36. X. Wang, L. Ye, M. Lyu, R. Ursache, A. Löytynoja, A. P. Mähönen, An inducible genome editing system for plants. *Nat. Plants* **6**, 766–772 (2020).

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

38. J. Shu, H. Deng, Lineage Specifiers: New Players in the Induction of Pluripotency. *Genomics, Proteomics & Bioinformatics*. **11**, 259–263 (2013).

39. A. Lampropoulos, Z. Sutikovic, C. Wenzl, I. Maegele, J. U. Lohmann, J. Forner, GreenGate - a novel, versatile, and efficient cloning system for plant transgenesis. *PLoS One.* ***8***, e83043 (2013).

40. W. Decaestecker, R. A. Buono, M. L. Pfeiffer, N. Vangheluwe, J. Jourquin, M. Karimi, G. Van Isterdael, T. Beeckman, M. K. Nowack, T. B. Jacobs, CRISPR-TSKO: A Technique for Efficient Mutagenesis in Specific Cell Types, Tissues, or Organs in Arabidopsis. *Plant Cell.* ***31***, 2868–2887 (2019).

41. N. M. Clark, A. P. Fisher, R. Sozzani, "Identifying Differentially Expressed Genes Using Fluorescence-Activated Cell Sorting (FACS) and RNA Sequencing from Low Input Samples" in: *Computational Cell Biology: Methods and Protocols*, L. von Stechow, A. Santos Delgado, Eds. (Springer New York, 2018) pp. 139–151.

42. F. Zambelli, F. Mastropasqua, E. Picardi, A. M. D'Erchia, G. Pesole, G. Pavesi, RNentropy: an entropy-based tool for the detection of significant variation of gene expression across multiple RNA-Seq experiments. *Nucleic Acids Res*. **46**, e46–e46 (2018).

43. S. Picelli, O. R. Faridani, Å. K. Björklund, G. Winberg, S. Sagasser, R. Sandberg, Full-length RNA-seq from single cells using Smart-seq2. *Nature Protocols*. **9**, 171–181 (2014).

44. J. S. Dahlin, F. K. Hamey, B. Pijuan-Sala, M. Shepherd, W. W. Y. Lau, S. Nestorowa, C. Weinreb, S. Wolock, R. Hannah, E. Diamanti, D. G. Kent, B. Göttgens, N. K. Wilson, A single-cell hematopoietic landscape resolves 8 lineage trajectories and defects in Kit mutant mice. *Blood*. **131**, e1–e11 (2018).

45. F. A. Wolf, P. Angerer, F. J. Theis, SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology*. **19**, 15 (2018).

46. D. Kim, B. Langmead, S. L. Salzberg, HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*. **12**, 357–360 (2015).

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

48. C. Trapnell, B. A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M. J. van Baren, S. L. Salzberg, B. J. Wold, L. Pachter, Transcript assembly and abundance estimation from RNA-Seq reveals thousands of new transcripts and switching among isoforms. *Nat Biotechnol*. **28**, 511–515 (2010).

49. A. Fernandez, A. Drozdzecki, K. Hoogewijs, A. Nguyen, T. Beeckman, A. Madder, P. Hilson, Transcriptional and Functional Classification of the GOLVEN/ROOT GROWTH FACTOR/CLE-Like Signaling Peptides Reveals Their Role in Lateral Root and Hair Formation. *Plant Physiology*. **161**, 954–970 (2013).

50. V. D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks. *J. Stat. Mech.* P10008 (2008).

51. J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, Fiji: an open-source platform for biological-image analysis. *Nature Methods*. **9**, 676–682 (2012).

52. E. K. Yoon, S. Dhar, M.-H. Lee, J. H. Song, S. A. Lee, G. Kim, S. Jang, J. W. Choi, J.-E. Choe, J. H. Kim, M. M. Lee, J. Lim, Conservation and Diversification of the SHR-SCR-SCL23 Regulatory Network in the Development of the Functional Endodermis in Arabidopsis Shoots. *Mol Plant*. **9**, 1197–1209 (2016).

53. C.-H. Yang, W.-T. Kuo, Y.-T. Chuang, C.-Y. Chen, C.-C. Lin, Cyclin B1 Destruction Box-Mediated Protein Instability: The Enhanced Sensitivity of Fluorescent-Protein-Based Reporter Gene System. *BioMed Research International*. **2013** (2013)

Acknowledgements: We thank Yuki Kondo, Philip Benfey and Ottoline Leyser for providing seeds of *pNAC57::2GFP*, *pSHR::erGFP* and *pPIN4::PIN4-GFP* reporters, respectively and the staff of the Flow Cytometry Core Facility at CIMR for their technical support with cell sorting. We thank Satoshi Fujita (National Institute of Genetics, Japan) for providing pDONRP2rP3_MIDD1ΔNcds_stop vector. We are thankful to Raymond Wightman and Gareth Evans for technical support with microscopy experiments.

This research was funded in whole, or in part, by the Wellcome Trust [203151/Z/16/Z] and the UKRI Medical Research Council [MC_PC_17230]. For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Funding: This work was supported by

Finnish CoE in Molecular Biology of Primary Producers (Academy of Finland CoE program 2014-2019) decision #271832 (YH)

Gatsby Foundation GAT3395/PR3 (YH)

University of Helsinki award 799992091 (YH)

ERC Advanced Investigator Grant SYMDEV No. 323052 (YH)

NSF-BBSRC MCSB grant 1517058 (RS, YH)

NSF CAREER MCB grant 1453130 (RS)

NIH grant GM078279 (KDB)

NFS IOS grant 1934388 (KDB, DS)

NIH grant R35GM136362 (KDB)

MRC Clinical Research Infrastructure award MR/M008975/1 (BG)

Core funding from the Wellcome and MRC to the Cambridge Stem Cell Institute (BG)

Academy of Finland grants 266431 and 307335 (APM, XW)

NWO Horizon grant 050-71-054 (RH)

ERC Starting Grant TORPEDO; 714055 (BDR)

Research Foundation - Flanders (FWO; Odysseus II G0D0515N) (BDR)

Gatsby Foundation CDF grant (SEA)

BOF postdoctoral fellowship from Ghent University (JRW)

Wallenberg Academy Fellowship KAW 2016.0274 (CWM)

Wellcome Strategic Award 105031/D/14/Z (FH)

JSPS Research Fellowship for Young Scientists (KT)

JSPS KAKENHI grant JP16J00131 (KT)

JSPS Overseas Research Fellowship (YS)

MEXT KAKENHI grant 16H06280, 19H05677 (YO)

The Finnish Academy of Science (JH)

Authors contributions:

Conceptualization: PR, JH, BB, KT, YS, MAdLB, JC, PD, GG, RH, BS, CWM, BG, APM, RS, KDB, YH

Methodology: JC, XW, YO, AMB, MG

Investigation: PR, JH, BB, KT, YS, WWYL, RU, EM

Data analysis: PR, JH, BB, KT, YS, MAdLB, FH, JC, HT, KV, JW, YS, BDR, SEA, DS, KDB

Writing: PR, JH, BB, KDB, YH

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: All data are available in the supplementary material and raw RNA sequencing data for pooled (bulk sorted) protophloem sieve element cells, single protophloem sieve element cells and overexpression and mutant profiling is deposited at GEO-NCBI (GEO submissions GSE142259, GSE140778, GSE140977, respectively).

List of supplementary materials

References (39-53) are cited only in the supplementary material

Materials and Methods

Figs. S1 to S10

Tables S1 to S20

Movies S1 to S13

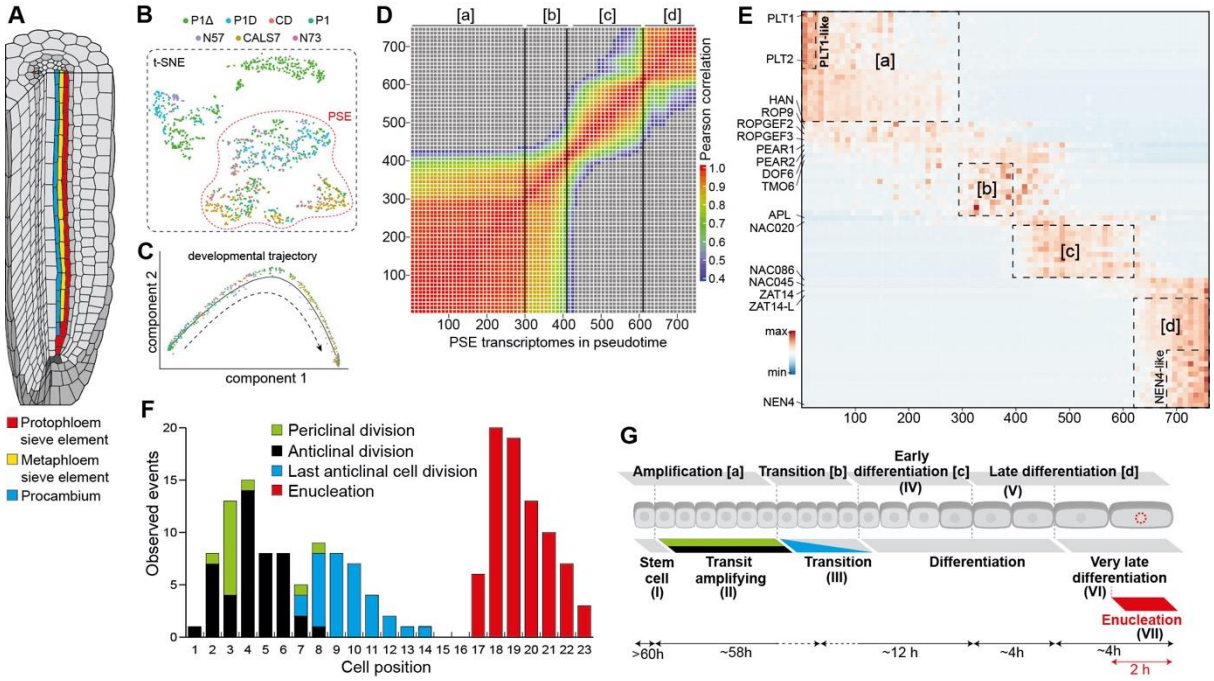

Figure 1. Phloem development at single-cell resolution.

(A) Schematic of the Arabidopsis root tip depicting position of protophloem sieve element, metaphloem sieve element and procambium cell lineages originating from a single phloem stem cell. **(B)** t-SNE plot of 1242 transcriptomes of cells sorted with P1Δ, P1D, CD, P1, N57, CALS7 and N73 reporter lines specific to different domains of the developing phloem. Indicated protophloem sieve element cells were used for the pseudotime trajectory analysis (Fig. S2, Supplementary Material). **(C**) protophloem sieve element transcriptomes ordered along developmental trajectory using Monocle 2. **(D)** Heatmap of Pearson correlation along the pseudotime trajectory. Vertical lines indicate 3 strongest correlation drops and separate four groups of transcriptomes with higher similarity [a], [b], [c] and [d]. **(E)** Gene expression heatmap of protophloem sieve element regulators and 10 most specific genes from the 4 groups defined in D) and the nested *PLT1* ("PLT1-like") or *NEN4* ("NEN4-like") expression domains in pseudotime-ordered protophloem sieve element transcriptomes. **(F)** Histogram of

cell behavior based on long-term live imaging. **(G)** Seven domains and the time cells spend in each position of the developing protophloem sieve element as determined by the transcriptomics (above) and live imaging (below): (I) “stem cell”, position 1 [a], t>60h; (II) “transit amplifying”, position 2-9 [a], t=58h, SD+8.1h, (III) “transitioning”, position 8-11 [b]; (IV) “early differentiating”, position 10-15 [c], t=12h; (V) “late differentiating”, position 16-17 [d], t=4h; (VI) “very late differentiating - NEN4-like”, position 18-19 [d], t=4h; VII “enucleating”, position 19 [d], t= 2h (Movie S1, S2).

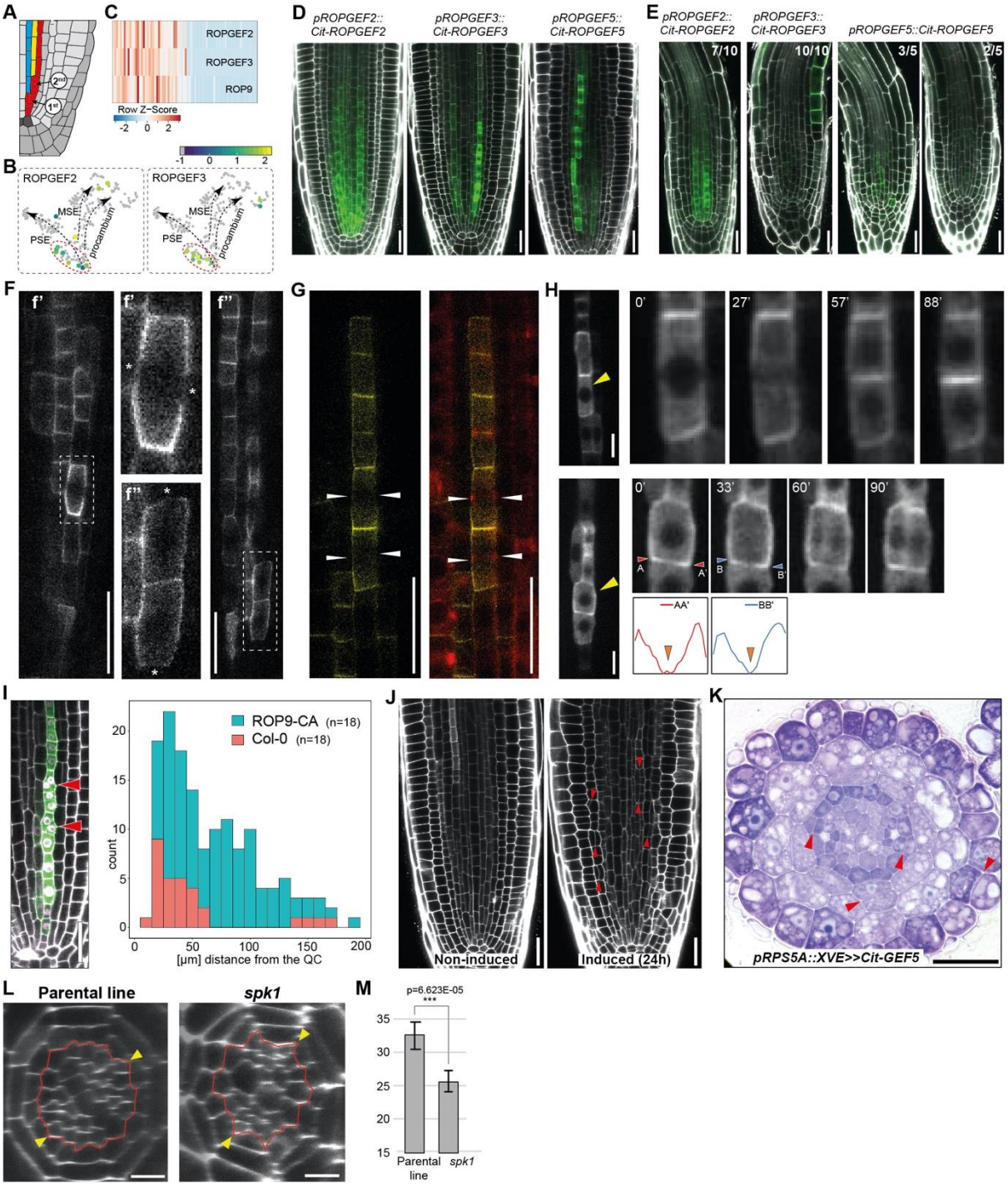

Figure 2. PEARs control asymmetric divisions by promoting ROP signaling in the phloem pole.

(A) Schematic indicating position of the two periclinal divisions in the phloem cell lineage. **(B)** Expression of ROPGEF2 and ROPGEF3 at the time of phloem lineage bifurcation. **(C)** Peak expression of ROPGEF2, 3 and ROP9 in the early phloem cells as detected in the pseudotime-ordered single cell protophloem sieve element transcriptome data. **(D)**

Expression pattern of phloem enriched ROPGEFs. ROPGEF3 and 5 share similar expression domain – enriched in protophloem sieve element and adjacent vascular cell files; ROPGEF2 is expressed in protophloem sieve element but also in other outer procambial cells and pericycle (Fig. S4D). Scale bars: 25 µm. **(E)** Expression of ROPGEF2, 3 and 5 in the *pear* sextuple mutant background. Scale bars: 25 µm. **(F)** Protein localization of *pROPGEF5::Cit-ROPGEF5* during anticlinal (f') and periclinal (f'') cell division. Gaps in ROPGEF5 signal are indicated with an asterisk. Scale bars: 25 µm. **(G)** Depletion of Cit-ROPGEF5 membrane signal at the cortical division zone (CDZ) during cell division. CDZ is marked by accumulating cortical microtubules (mCherry-TUA5) forming pre-prophase band (white arrowheads). Scale bars: 25 µm. **(H)** Time course analysis of the dynamic pattern of active ROP signaling in the dividing phloem cells. Depletion of *pPEAR1::mScarlet-I-MIDD1ΔN* signal at the CDZ in the anticlinally (upper row) and periclinally (lower row) dividing cells (yellow arrowheads). Quantification of fluorescent signal intensity in the periclinally dividing cells. Scale bars: 10 µm. **(I)** Quantification of asymmetric cell divisions (red arrowheads) in the protophloem sieve element cell lineage after expression of constitutively active ROP9 (Q64L) (*pPEAR1::XVE>>ROP9*^{CA}). Scale bars: 25 µm. **(J)** Ectopic asymmetric cell divisions (red arrowheads) 24h after induction of ectopic Cit-GEF5 expression (*pRPS5A::XVE>>Cit-GEF5*). Scale bars: 25 µm. **(K)** Toluidine blue stanning of resin sections of Cit-GEF5 overexpressing line (*pRPS5A::XVE>>Cit-GEF5*) 24h after induction. Red arrowheads indicate ectopic periclinal cell divisions in epidermis, endodermis and pericycle. Scale bars: 25 µm. **(L)** Identification of *spk1* allele in the mutant screen of *pRPS5A::PEAR1-GR* parental line. Presented are images from non-induced plants. Scale bars: 10 µm. **(M)** Quantification of vascular cell files in the *spk1* mutant and its parental line *pRPS5A::PEAR1-GR*. Both lines were not induced.

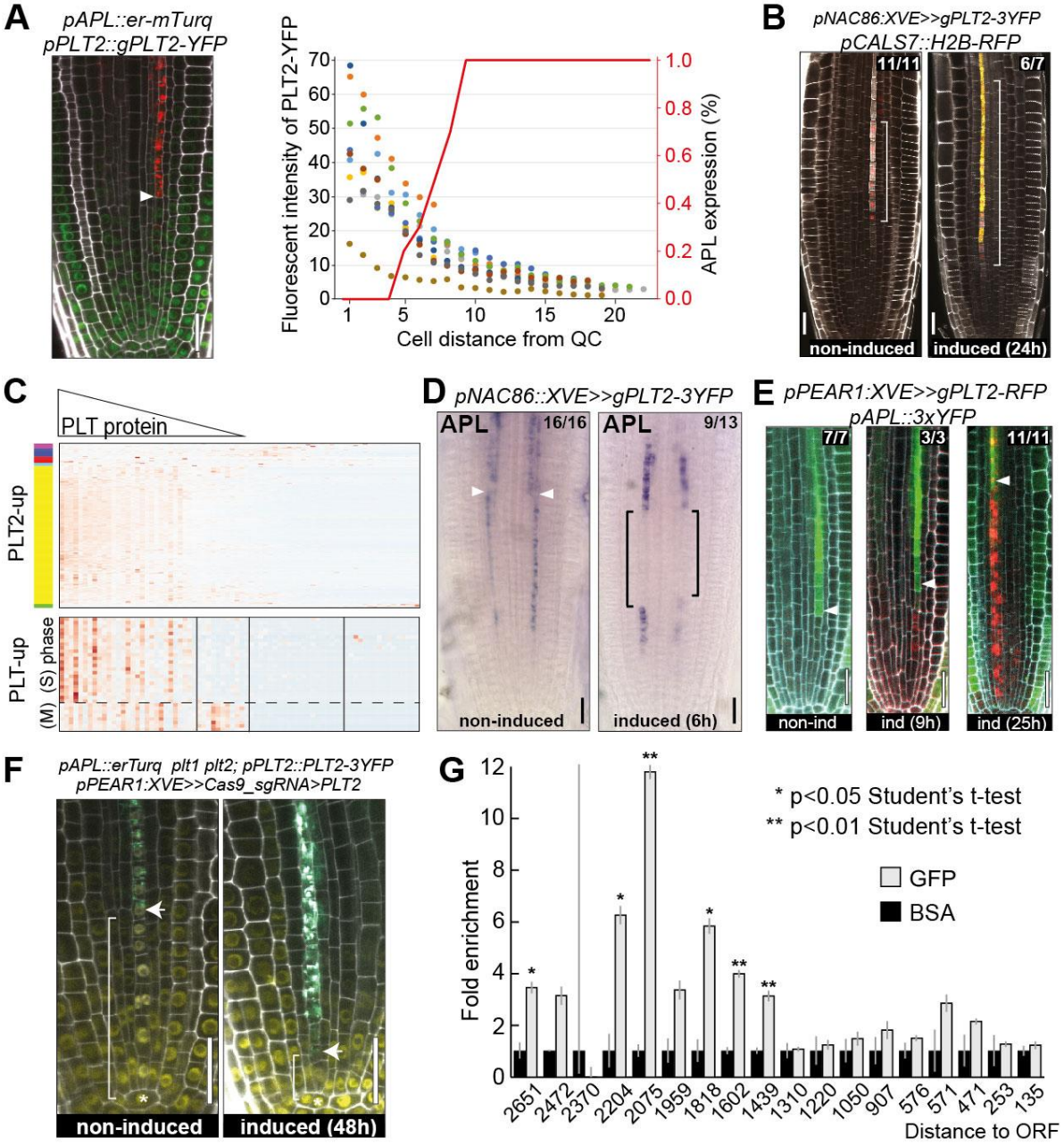

Figure 3. PLT2 inhibits phloem differentiation by directly repressing APL expression.

(A) Quantification of fluorescent intensity of PLT2-YFP in protophloem sieve element cells of 9 roots indicated with dots of different colours. Percentage of roots expressing *APL* in a given protophloem sieve element cell is indicated as a red line (n=9). Onset of *APL* expression coincides with diminishing level of PLT2 protein. Arrowhead indicates onset of *APL* expression in protophloem sieve element. **(B)** Ectopic expression of *PLT2* under *pNAC86::XVE* promoter delays protophloem sieve element enucleation. Square brackets

indicate extended expression domain of *pCALS7::H2B-RFP*, a reporter used for monitoring enucleation. **(C)** Native expression profile of PLT2 targets in protophloem sieve element cells ordered in pseudotime. Genes upregulated after 6 hours of induction of the line shown in B) are plotted. Upper panel shows gradually diminishing expression of target genes which reflects the PLT2 protein gradient. Lower panel shows PLT2 upregulated cell cycle genes with oscillatory expression pattern. **(D)** In situ hybridization of *APL* before and 6h after ectopic expression of PLT2-3xYFP. Arrowheads indicate position of protophloem sieve element enucleation beyond which point *APL* is expressed in phloem pole pericycle, companion cells and metaphloem sieve element (Fig. S5E). Brackets indicate *pNAC086* activity domain. **(E)** Time course of transcriptional repression of *APL* in cells ectopically expressing PLT2-RFP under inducible *pPEAR1::XVE* promoter. **(F)** Early activation of *APL* expression 48h after phloem specific knock-out of *PLT2*. **(G)** ChIP-qPCR of PLT2-3xYFP on *APL* promoter revealed PLETHORA binding region -2204 to -1439 bp upstream of *APL* ORF. All scale bars, 25 µm.

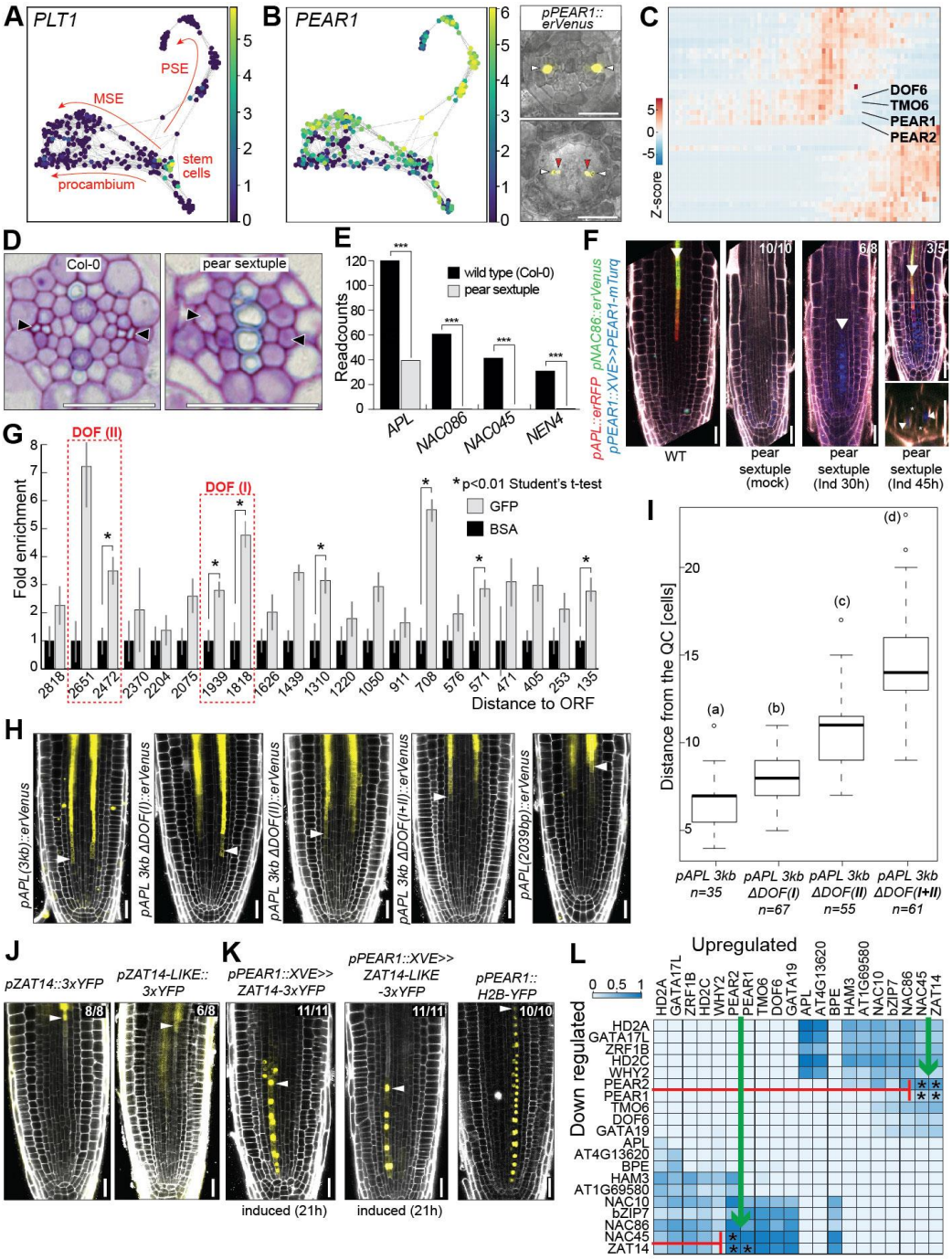

Figure 4. PEARs orchestrate phloem differentiation.

(A) Force-directed clustering of 272 single-cell transcriptomes obtained using the *pPEAR1Δ::erVenus* reporter. Plotted is expression of stem cell abundant *PLT1*. Arrows: cellular trajectories inferred from known gene expression patterns (Fig. S6). **(B)** Strong enrichment of *PEAR1* expression in protophloem sieve element and metaphloem sieve element trajectories confirmed by *pPEAR1::erVenus* reporter line. White arrowheads:

protophloem sieve element, red arrowheads: metaphloem sieve element. **(C)** Expression heatmap: PEAR genes among the earliest phloem specific transcription factors. **(D)** Lack of protophloem sieve element differentiation in the mature part of the *pear* sextuple mutant root. Arrowheads: protophloem sieve element position. **(E)** Lack of *APL* pathway activation in the roots of *pear* sextuple mutant based on RNASeq analysis. **(F)** Inducible expression of *PEAR1-mTurq* is sufficient to activate transcription of *APL* and *NAC86* reporters in *pear* sextuple mutant background. **(G)** ChIP-qPCR of PEAR1-YFP shows direct interaction of PEAR1 with *APL* promoter at multiple positions. Two prominent PEAR1 binding sites are indicated with red dashed rectangles. **(H)** Expression patterns of modified *pAPL* reporter lines. Length of "3kb" promoter equals 2962 bp. DOF(I) and DOF(II) correspond to two enhancer elements indicated in panel G. Details of modification are provided in Fig. S7C. **(I)** Quantification of the onset of *pAPL* expression after modification of DOF binding motives. Statistically significant differences between groups were tested using Tukey's HSD test $P < 0.05$. Different letters indicate significant difference at $P < 0.05$. **(J)** Expression of ZAT14 and ZAT14L during late differentiation of protophloem sieve element. Arrowheads: last cell before enucleation. **(K)** Ectopic expression of *ZAT14* and *ZAT14L* under *pPEAR1::XVE* results in cell elongation and inhibition of cell division. Arrowheads: last cell before enucleation. *pPEAR1::H2B-YFP* line shows regular number of protophloem sieve element cells. **(L)** Heatmap shows significantly overlapping and oppositely regulated target sets of the 20 most important TFs from the GRN model. Color intensity shows a fraction of overlapping target sets. The colormap represents significantly overlapping sets (Fisher Exact Test, if $p<0.05$, val=1) multiplied by the fraction of overlap. Asterisk indicates experimental validation of up and downregulated sets from TF OE *in vivo* (Tables S15, S16). All scale bars, 25 µm.

Supplementary Materials

Materials and Methods

Plant materials and growth condition

Previously published and newly generated plant materials are listed in Table S18.

Arabidopsis seeds were surface sterilized in either a bleach solution (5% sodium hypochlorite with 0.02% tween 20) for 5 minutes with gentle rotation followed by 8 washing steps with sterile distilled water, or by chlorine gas treatment (3ml HCl added to 100ml of bleach enclosed in an airtight container) for a minimum of 4h. After sterilization seeds were imbibed in sterile water for 2 – 5 days at 4 °C in the dark. Seeds were germinated on 1/2 strength Murashige and Skoog (½ MS) agar medium containing 1% sucrose and 0.8% difco agar. Seedlings and adult plants were grown under long-day condition – 16 hours light (188 μmol m^{-2} s^{-1}) and 8 hours dark at 23 °C. For confocal imaging, anatomical analysis, and RNA-seq experiments, 5-day-old vertically grown seedlings were used, unless otherwise specified.

Methods

Generation of fluorescent reporter lines.

All the reporter lines (except for ROPGEF protein fusion reporters described below) were generated according to the previously described method which is based on multisite gateway system (Invitrogen)(*3*). To generate transcriptional reporters, promoters were cloned into pDONR P4-P1R entry vector. In general, up to 3 kb upstream region of each gene was PCR amplified using Col-0 genomic DNA as a template and cloned by BP cloning method (Invitrogen). Inducible constructs for *pHAN::XVE*, *pNEN4::XVE*, *pNAC73::XVE*, *pPEAR1::XVE and RPS5A::XVE* were generated by a classical restriction/ligation approach in the *p1R4-ML::XVE* backbone vector using T4 ligase as described in (*35*). Constructs with small (-1039 to -853) deletion in the promoter sequence of *PEAR1 (pPEAR1Δ*) and modifications in

DOF(I) fragment of *pAPL* (see fig. S7C for details) were generated according to the NEBuilder HiFi DNA Assembly Master Mix protocol (New England Biolabs). Modified DOF(II) sequence of *pAPL* was synthetized by Twist Bioscience, USA, and cloned into entry clone with *pAPL*, following NEBuilder protocol. Entry clones with modified sequences of *pROPGEF2, 3* and 5 were synthetized by Twist Bioscience, USA. Genomic fragments for *GIS*, *AT2G45120*, *NAC010*, *NTL8*, *BPEp2*, *NAC048*, *NAC075*, *WRKY21*, *ZAT14* (AT5G03510) and *ZAT14L* (AT5G04390) were cloned into pDONR221. To generate ROPGEF ectopic overexpressor lines, Citrine-GEF2, 3 an 5 coding sequences were amplified from ROPGEF translational fusion constructs (see description below), subcloned into pDONR221 and used for LR reaction with pRPS5A::XVE promoter. Histone 2B (H2B), DBOX domain from *A. thaliana* Cyclin B1;1 or fluorescent proteins were used in the second entry vector, followed by a fluorescent protein or terminator in the third entry vector, respectively. CRISPR knock-out construct for *PLT2* in pDONR221 is a courtesy of Ari-Pekka Mahonen's group (*36*). Constitutively active form of ROP9 (Q64L) was generated by introduction of point mutation with QuickChange II site-directed mutagenesis kit (Agilent). To generate *pPEAR1:mScarlet-I-MIDD1ΔNcds_FR, pGEMTP4P1r_pPEAR1*, *pDONR221_mScarlet-I* and *pDONRP2rP3_MIDD1ΔNcds_stop* were recombined with *pFR7m34GW* using LR Clonase II (Thermo Scientific). Primer sequences are listed in Table S19. Agrobacterium-mediated floral dip method was used to generate transgenic Arabidopsis plants.

GreenGate cloning of ROPGEF protein fusion reporter lines

ROPGEF constructs under their native promoter fragments were generated using the GreenGate cloning system with modified protocols (*39*). ROPGEF3 in pGGZ003 was previously described in (*29*). ROPGEF promoter fragments were cloned into pGGA modules. Promoter fragments used for ROPGEF1 (-1513 bp), ROPGEF2 (-1513 bp), ROPGEF2 (-

1475 bp), ROPGEF3 (-1742 bp), ROPGEF5 (-2500 bp), ROPGEF6 (-2379 bp), ROPGEF7 (-2501 bp). ROPGEF-ORFs were cloned into pGGC modules including START and STOP codon and a "ca" to keep the frame. To remove internal BsaI restriction sites, multiple PCRs fragments with mutated sites were fused seamless using flanking BsaI-Sites. To generate expression constructs, entry vector modules of ROPGEF promoters (A), N-mCitrine (B), ROPGEF-ORFs (C), D-Decoy (D), HSP18.2-Terminator (E), and pGGF009-BastaR (F) were fused into the destination vectors pGGZ003, or pFASTR-AG / pGGYR000 (*40*).

Fluorescence-activated cell sorting.

The destruction box (DBOX) sequence of *AtCYCB1;1* was fused to yellow fluorescent protein (YFP) to readily shut down YFP expression in dividing cells beyond promoter activity (*53*). Expression of this fusion protein using *PEAR1* and *CVP2* promoters resulted in predominant expression in the 2-6th and 3-9th cells, respectively (fig. S2A, B). However, these reporter constructs showed a cell-cycle biased expression pattern, in that the signal disappears right after cell division and comes back first in the cytoplasm and then to nucleus right before division (Movie S2). *pPEAR1* driven *3YFP* expression spans 10 cells from +2 position. For three other reporter lines, we combined two phloem reporters to limit the sorting population. For instance, *pNAC057::2GFP* was combined with *pCALS7::H2B-RFP*, which restricts the only GFP positive cell population to cells 4-14. (fig. S2A, B). The expression of *pCALS7::H2B-YFP* is observed in 7 cells prior to the enucleation, whereas *NAC073* promoter is active in a narrow domain consisting of 4 cells before enucleation. However, YFP fluorescence could be found also in phloem pole pericycle cells neighboring enucleated protophloem sieve element in both lines (fig. S2B). We've combined these with *pS32::erRFP*, a reporter broader in the stem cell niche and exclusive in differentiating protophloem sieve element, creating a refined overlapping expression in protophloem sieve elements.

Protoplast isolation and cell sorting

Arabidopsis root protoplasts were isolated according to the published protocol (*41*). In short, sterilized seeds were plated onto the nylon mesh (57-103, Nitex) placed on the top of MS agar media. Two rows of seeds comprising approximately 100-150 seeds per row were then grown vertically for 5 days in long-day growth condition. To isolate root protoplasts, ⅓ of the root was cut and chopped on the surface of the mesh and transferred to the buffer containing cell wall degrading enzymes - 1.5 g Cellulase (C1794, Sigma) and 0.1 g Pectolyase (P3026, Sigma) in 100 ml protoplasting solution. After 1 hour of incubation with occasional stirring, the protoplasts were spun down (200 rcf for 6 min) in 15 ml Falcon tube and after removal of supernatant, cells were resuspended in 700 µl protoplasting solution without enzymes. After filtering with 70 µm and 40 µm cell strainers sequentially, FACS was performed using 70 micron nozzle at 10 psi. Fluorescence positive cells were collected in a round-bottom polystylene tube that contains 150 - 200 µl RLT buffer with beta-mercaptoethanol (10 µl / 1 ml RLT buffer, RNeasy micro kit protocol, Qiagen). Cell sorting was performed for about 15 min and collected cells were immediately frozen on dry ice and kept in -80 °C for up to one week. For single cell transcriptomics, protoplasts were produced and isolated as described here, but sorted individually into 96-well plates.

The only exception are the cells isolated with the *pPEAR1Δ::erVenus* reporter line. To enrich in meristematic cells, whole roots were submerged in 1.5 ml tubes of 0.5 ml cell wall degrading buffer. The roots were picked up and shaken with forceps every 2 minutes for 12 minutes. This proved to be sufficient for the enzymes to degrade the cell wall in the transition zone of the root. The shaking helped to separate the meristems from the remains of root. The remains of the root were then removed with forceps and the contents of the tubes were united and incubated for another 50 min following the same steps as for the other samples.

RNA sequencing of FACS isolated cells

For bulk sorted samples, including sorted cells ectopically expressing PLT2.

Frozen cells were quickly thawed by adding the same volume of EtOH (150 - 200 µl), and RNA extraction was done using RNeasy Micro Kit (Qiagen) as previously described by (*41*). RNA integrity was measured with Agilent HS RNA TapeStation system following the manufacturer's instructions. Samples of RNA integrity value (RIN) 6.3 and above were taken to the following cDNA synthesis step. Clontech Smarter Ultra Low Input Library Kit V4 was used for cDNA synthesis and amplification (18 PCR cycles). Samples were then submitted to Novogene for library construction and RNA sequencing on an Illumina HiSeq SE50 run.

Analysis of bulk RNA-seq data for the identification of phloem-abundant genes.

Raw counts have been aligned and mapped according to a standardized Bowtie 2 suit using the TAIR10 genome. Feature Counts was used to obtain read counts and differential gene expression analysis was performed using edgeR. In order to find phloem-abundant genes (925, fig. S6A, Table S4), we've adopted Shannon-entropy based selection of tissue specific genes as described by (*42*). As a comparison dataset, we used published root map RNA-seq data of other tissue types (*37*).

Smart-seq Single-cell RNA sequencing.

Single protoplast cells were sorted into lysis buffer and sample preparation for RNA-sequencing was performed as previously described (*43*, *44*). Briefly, cDNA was reverse transcribed with SuperScript II Reverse Transcriptase (Invitrogen) and amplified with 23 PCR cycles using KAPA HiFi HotStart polymerase (Roche). Amplified PCR products were purified with Ampure XP Beads (Agencourt) at a volume ratio of 1:0.6 DNA:beads, and quantified

using Quant-iT™ PicoGreen™ dsDNA Assay Kit. Libraries were prepared using Illumina Nextera XT DNA preparation kit and either Nextera XT index kit – 96 indexes or Nextera XT index kit v2. Pooled libraries were sequenced on the Illumina HiSeq4000 (single-end reads, 50-bp).

Identification of protophloem sieve element specific genes.

Differentially expressed genes between specific pairs of clusters were calculated using a t-test with overestimated variance and Benjamini-Hochberg correction for adjusted p-values as implemented in the scanpy (*45*) rank_genes_groups function. All genes significantly enriched (q-value <0.05) in protophloem sieve element cluster 8 or 9 or 10 or 12 when compared to clusters 3 and 6 and 11 formed a list of 1192 genes enriched in SEs in this data set (Table S5). By intersecting these with the 925 phloem enriched genes, 542 sieve element specific genes were identified (fig S5H, Table S6).

Single-cell data analysis.

The transcriptome raw data of individual cells were mapped using HISAT2 (*46*), SAM files converted to BAM files using SAMTools (*47*). Subsequently Cuffquant and Cuffnorm were used to obtain read counts and FPKM values (*48*). These were then analysed using Monocle2 (*15*, *16*). We filtered out data of cells with <10.000 reads mapped to a unique locus, <20% reads mapped to TAIR10 genome and <2000 genes showing an expression above 0.1 (1242 of 1375 transcriptome). Only genes with expression in at least 50 transcriptomes passing these thresholds were considered in the further downstream analysis. Using Monocle2's tSNE and clustering function, among 14 clusters, we removed 6 clusters (2, 3, 5, 7, 11, 12) that do not represent protophloem sieve element cells (Fig. S2). Columella cells were sorted with *pPEAR1Δ::erVenus* reporter that shows strong ectopic fluorescence in those cells. Those cells

were identified according to expression of *GLV5* and *GLV7* (*49*) (fig S2, Clusters 2, 5, 11). These cells were then also removed from further analysis. The remaining transcriptomes (1026) were plotted after dimensional reduction using tSNE again. It revealed a cluster of mainly *pNAC057::2xGFP* and *pPEAR1::dBOX-3xYFP* originating transcriptomes (Cluster 7) as well as two clusters of *pPEAR1Δ::erVenus* cells that were separated from the other cells's transcriptomes (fig. S2, Clusters 3, 11). Using an unsupervised approach by clustering these cells into 10 clusters and looking for differentially expressed genes (DEG) between these clusters, we created a developmental trajectory based on the top 4000 DEG. These resulted in a protophloem sieve element trajectory with a single branching point where transcriptomes of the three outstanding clusters of the original t-SNE were located (Clusters 3, 7, 12, Fig S2). We suspected that the *pNAC057::2xGFP* and *pPEAR1::dBOX-3xYFP* transcriptomes would likely originate from contaminating cells, as this cluster also showed expression of AT2G22850/bZIP6, whose reporter line (S17) is specifically expressed in the phloem pole pericycle. Based on the lineage branching analysis, original clusters 3 and 12 consisting mostly of cells originating from *pPEAR1Δ::erVenus* reporter were assigned metaphloem sieve element and procambium identity.

To further investigate the lineage branching, filtered *pPEAR1Δ::erVenus* transcriptomes were analysed in scanpy v1.4 (*45*). Genes detected in fewer than 10 cells were filtered out, and data were normalised using the scanpy normalize_per_cell function with default parameters. Normalised counts were log(x+1) transformed and highly variable genes identified with the scanpy highly_variable_genes function with parameters max_mean=8 and min_mean=0.5. Principal component analysis was calculated on scaled gene expression, and a force directed graph visualisation calculated using the scanpy draw_graph function with a k=10 nearest neighbours calculated on the top 50 principal components. Cells were clustered using Louvain clustering (*50*) on the k=10 nearest neighbor graph with resolution=2. Looking at the

expression of genes known to be differentially expressed between the procambial and sieve element lineages (Fig S5) allowed the identification of the protophloem sieve element, metaphloem sieve element and procambial cell clusters.

Pseudotime inference for protophloem sieve element trajectory

After removing the outstanding clusters containing the columella, lateral root cap, phloem pole pericycle, metaphloem sieve element and procambium cells in the Monocle2 analysis, the remaining 758 cells were placed on a linear developmental trajectory using unsupervised tSNE clustering into 10 clusters and DDRTree dimensional reduction on the top 4000 DEGs between clusters. This step assigned a pseudotime value to each cell and ordering the transcriptomes according to these values was confirmed by the observed expression patterns of known protophloem sieve element expressed genes (Fig. 1E, Table S17).

Pearson correlation calculation.

The previously obtained pseudo-time order of 758 protophloem cells was used to calculate the Pearson correlation along pseudo-time. We applied a moving average window to reduce the effect of noise to this analysis. We tested window sizes of 5, 10, 15, 20, 30, and 50 cells moved by 5, 10, 15, 20 or 30 cells. Averaging 30 cells and moving 10 cell positions per step gave enough simplification while keeping a good level of detail. The Pearson correlation between the averaged transcriptomes was then calculated for all expressed genes as well as the 4000 genes from the pseudo-time ordering. By plotting the correlation matrix in a heatmap, we could identify the points in pseudo-time where the transcriptome changes rapidly. These changes were more pronounced looking only at the 4000 ordering genes, as these would also be among the top DEG along pseudo-time. These rapid changes could also be explained by an underrepresentation of these cells in our data set. However, the distribution of cells in pseudo-

time shows that these points of big transcriptome changes do not overlap with the end or beginning of sorted populations in pseudo-time (fig S2G). Thus an equal sampling before and after these changes can be assumed, making the underrepresentation unlikely.

After defining the 4 major domains on 30-averages, 10-cell averages in that range of 40 cells (2 adjacent 30-cell windows with step size 10) were used to define the borders more precisely, with a final step using 5-cell averages that were split in the middle.

Domain specific gene identification

We manually determined the border between stem cell and the second cell in pseudotime by the first noticeable drop of *PLT1* expression at pseudotime position 27. The *NEN4* expression domain was determined by the increase of *NEN4* transcripts beyond pseudotime position 691. As these are nested within domains 1 and 4, respectively, only the definitions for domain 1-4 were used to identify domain specific genes. These domains were compared using the same approach as for the cluster comparisons in the force-directed analysis. The highest q-value (qval_max) for each gene in the three possible comparisons were used for thresholding. For PLT1-like and NEN4-like domains, comparisons to all 5 other domains were used, respectively. Only genes with qval_max<0.05 were considered significantly enriched (Table S1).

RNA extraction and sequencing of whole meristem samples

For the transcriptome profiling of PEAR sextuple mutant (*pear1 pear2 dof6 tmo6 obp2 hca2*) and inducible overexpression lines of ZAT14 and ZAT14L, primary root meristems were dissected under stereomicroscope using conventional needles at 5 days after germination. To facilitate cutting, plants were transferred from the ½ MS 1% agar plates to ½ MS 5% agar plates in bulk of 10. Dissected meristems (in bulk of 10) were then transferred to the droplet of

RNlater solution kept in the cap of the test tube at room temperature. Tubes were stored in the fridge until RNA extraction with RNeasy Plant Mini Kit (Qiagen). RNA extraction from 100 root tips of wild type Arabidopsis would typically yield 2.5µg of good quality (RIN>8.5) total RNA of which 2µg were submitted for RNA-Seq analysis on the HiSeq-PE150 platform at Novogene. Overexpression of ZAT14 and ZAT14L was induced by transferring *pWOL:XVE>>ZAT14-mTurq* and *pWOL:XVE>>ZAT14L-mTurq* plants onto the media plates supplemented with 5µM 17-b-estradiol for 10 or 14h. Plants transferred to media plates supplemented with DMSO we used as controls.

Quantification of fluorescent reporter lines.

For quantification, confocal microscopy image files were directly imported into the image analysis software FIJI (*51*). Cell walls were outlined using the polygon selection tool in the red channel of PI stained root images. In addition to the protophloem sieve element cells, four large areas within the root but not covering any protophloem sieve element cells were drawn and the detected fluorescence considered as background levels. The average fluorescence in the YFP channel was then measured and normalized to the background fluorescence. The presence of nuclei in images of *pCALS7::H2B-RFP, pCALS7::H2B-YFP* and *pNAC073::H2B-YFP* was judged in the corresponding channel. In case of *pCALS7::H2B-YFP* and *pNAC073::H2B-YFP* reporter lines, the last nucleated cell was defined as position 19 in each root corresponding to the average enucleation position obtained from live imaging. This way, the cells in similar developmental stages were placed closer together and thus the variance was reduced.

Histological analyses and confocal microscopy.

Cross-sections of hard resin-embedded Arabidopsis roots were obtained as described (*2*), stained with toluidine blue and imaged on Zeiss Axio Imager microscope equipped with 64

MP colour SPOT Flex high resolution camera. DAPI staining was performed as described (*3*). Confocal imaging was carried out on a Leica TCS SP8 equipped with 405 nm (DAPI stained samples), 442 nm (mTurq), 488 nm (GFP), 514 nm (YFP) and 561 nm (RFP and propidium iodide stained samples) lasers. Live imaging was performed as described (*5*) on a Leica SP8 inverted scanning confocal microscope that uses solid-state 514 nm and 552 nm lasers for imaging YFP and mCherry respectively. Time-lapse imaging of *pPEAR1:mScarlet-I-MIDD1ΔNcds* line was performed with inverted fluorescent microscope (Ecripse Ti, Nikon) fitted with a confocal unit (CSU-X1, Yokogawa). The images were acquired approximately every 30 minutes using MetaMorph (Molecular Devices).

Whole mount in situ hybridization.

Target mRNA sequences were PCR amplified using whole seedling cDNA library of the Col-0 as a template (primer sequences are listed in Table S19). PCR fragments were cloned into pCR-Blunt II-TOPO vector (Invitrogen) that encodes both T7 and SP6 promoter priming sites. XbaI and HindIII restriction enzymes were used to linearize the APL and NEN4 vectors, respectively, leaving only one of the promoters intact. In-vitro transcription reactions were performed using DIG RNA Labelling Kit (Roche). Hybridization and detection were carried out as described (*17*).

Chromatin immunoprecipitation and qPCR.

1/3 of 7-day-old roots was cut and 1.5g of root materials was collected in chilled cross-linking solution (1% formaldehyde and 5 mM EDTA in 1× PBS). Cross linking was done by applying vacuum for 30s, five times at room temperature. ChIP experiment was performed according to the published protocol (*36*) except for the DNA shearing step. Shearing was done in the

BioRuptor, 30s on/off at 10 cycles. Primers used for the real-time PCR are listed in Table S19. Real-time PCR conditions were as follows: Initial denaturation 50°C 2 min followed by 95°C 10 min. Amplification was done with 45 cycles of 95°C 15 s, 60°C 1 min, and 72°C 15s. Reaction mix with SYBR was prepared as described in the manufacturer's instruction (ROCHE). 3 independent biological replicates were used to calculate the fold-change with the BSA control set to 1.

PLT2 induction experiment.

To investigate protophloem sieve element abundant PLT2 targets, *pCALS7::H2B-RFP* was crossed with *pNAC86::XVE>>PLT2-3xYFP*. For the induction of PLT2, 5-day-old seedlings grown on the standard ½ media were transferred onto 10 µM 17-b-estradiol containing ½ MS media for 6 hours. In short, roots were cut and collected in the protoplasting solution containing the same volume of estradiol and DMSO respectively, and incubated for 12 min with vigorous swirling with forceps. After additional 50 min of incubation in the same solution, protoplasts were precipitated and re-suspended for the following cell sorting.

Inducible and tissue specific manipulation of PLT2 expression with CRISPR-Cas9 technology.

We used *plt12;pPLT2::PLT2-3xYFP*, partially complementing *plt1;2* double mutant phenotype, as a genetic background to CRISPR PLT2. Estradiol inducible *pPEAR1::XVE* promoter was used to drive CAS9p-t35S and four guide RNAs were designed for targeting *PLT2* as described (*52*).

Fluorescence recovery after photobleaching to study precise expression domains.

To precisely determine the expression domain of *pGIS::3xYFP*, Fluorescence recovery after photobleaching (FRAP) was used to abolish the fluorescence in all phloem cells and images of new translated proteins were taken after 90-120min of recovery. FRAP was done on a Leica TCS SP8 using the FRAP software module with activated FRAP booster. The laser lines 458 nm, 476 nm, 488 nm, 496 nm and 514 nm were simultaneously used at 70% laser power to bleach the phloem with a defined ROI at 256 x 256 pixel with 600 Hz scanning speed and with a x63 lens. 20-40 repetitions of the bleaching were applied for nearly complete bleaching. After recovery, regular confocal microscopy images were taken and the fluorescence quantified using the Fiji software package of ImageJ (*51*). The background fluorescence in the ground tissues was measured and subtracted from the measured fluorescence in the protophloem sieve element cells. For each observed root, the mean fluorescence in the brightest cell was set to 1 and the others cell as a ratio of that.

Cell behavior analysis.

Long-term live imaging was performed with a nuclear localised reporter line (*pPEAR1::H2B-YFP pCALS7::H2B-YFP)* expressed in every cell of the protophloem sieve element lineage. This reporter line was obtained by crossing *pPEAR1::H2B-YFP*, expressed strongly in positions 1-15th, with *pWOX5-mCherry/PET111-GFP*. F1 generation was then crossed to *pCALS7::H2B-YFP*, whose expression is predominant in the last 6 cells before enucleation (*3*). The imaging chamber was prepared according to the published method (*5*) with 1.5% agarose in the media.

Random forest model for gene regulatory network inference

To model gene regulatory connections, we first selected the 15% most variable genes among the 758 cells using the genevarfilter function in Matlab ('Percentile', 85), leaving 4,924 genes

for model inference. We modeled gene regulation using a Random Forest machine learning approach on the 758 single-cell expression profiles and the 4,924 highly variable genes, which included 208 transcription factors (TFs). In general, the Random Forest model allows for non-linear dependencies of target genes on causal transcription factors. Each single-cell expression profile is treated as a steady-state condition, allowing the model to learn a function that maps expression values of TFs to the expression value of each target gene. In the Random Forest approach, the TF-to-target association is described with a "score" that reflects the contribution of the TF to the expression of its target according to the model.

To address drop-out effects and other noise in single-cell data in the pre-processing stage, we merged the expression of consecutive cells to generate pseudo-cells using the following procedure to optimize the "bin size" (number of consecutive merged cells): we subdivided the 758 single-cell expression profiles into varying bin sizes, taking the median of the expression value of each gene in each bin or pseudo-cell as the value of that pseudo-cell. The Random Forest approach uses bootstrap aggregation, where each new tree is trained on a bootstrap sample of the training data. The remaining out-of-bag error is estimated as the average error for each training data point p_i evaluated on predictions from trees that do not include p_i in their corresponding bootstrap sample. For the dataset, the optimal bin size that minimized the out-of-bag error was 12 cells, providing our steady-state inference model a total of 64 pseudo-cells. Finally, the Random Forest model ranks TFs based on their influence (score) on target gene expression, generating a predicted gene regulatory network (GRN) based on TF causality. To refine these TF–target predictions, we retained the top-10 highest scoring transcription factors for each gene target, resulting in 49,240 (TF-to-target) edges. The code is available at https://github.com/jacirrone/OutPredict. TFs were then ranked by their number of targets to derive the ranked list of the most important TFs (Table S8).

Correlation analysis of the model in pseudotime

TF targets were classified into positive vs. repressive downstream sets in both the Random Forest model: Pearson correlation between each TFs and individual targets was used to determine regulatory effects (negative correlation, r<0, was classified as repressive regulation, and r>0 was classified as positive regulation, Table S8). We evaluated significant overlap between all pairwise positive and negative regulatory sets for each transcription factor (seesaw model) using the Fisher Exact test in Matlab. For the heatmap in Figure 4L, the binary output of the Fisher Exact test (p<0.05=1, p>0.05=0) was multiplied by the fraction of overlap between the two TF target sets (Table S15). Similarly, the overlap of up and down regulated targets in RNA sequencing upon ectopic overexpression of PLT2, PEAR1/PEAR2 (*17*) and NAC45 (*3*), ZAT14 or ZAT14-like was evaluated by systematically testing the pairwise intersection between the up- and down-regulated sets for each TF in the in vivo overexpression results. The criteria for identifying the up- and down-regulated sets (Set 1, Set 2) was an adjusted p value of 0.05 accounting for multiple testing and a two-fold change in expression over control either up- or down-regulated. The intersection of the two sets was tested for significant overlap using the Fisher Exact test, with a pval cutoff at 0.05. (Table S16). In the Fisher Exact tests, the overlap of Set 1 with all eligible transcripts was taken as outcome 1 (background intersection), while the overlap of Set 1 with Set 2 was taken as outcome 2 (test intersection).

Live imaging of the late differentiating cells.

pNAC86::H2B-YFP and *pNEN4::H2B-YFP* lines were crossed with *pWOX5-mCherry*. F1 seeds were grown vertically for 3 days on ½ MS plate and subsequently transferred to the imaging chambers described above to record movies S3-S12. Movies were analyzed to

determine the time from the onset of H2B-YFP expression until the time given cell enucleates (fig. S1I).

Genetic screen for factors promoting periclinal cell divisions.

The seeds of PEAR1 overexpressing line, *pRPS5A::PEAR1-GR* were mutagenized with ethyl methanesulfonate (EMS) and germinated on MS media supplemented with dexamethasone (+DEX). The seedlings with narrow root were selected and reduced number of vascular cell files was confirmed with confocal microscopy. Because the *spk1* mutant was seedling lethal, heterozygous line was recovered which produces quarter of seedlings showing the narrow root phenotype. The seedlings with the narrow root phenotype also show the pleiotropic developmental defects including loss of trichome branching. Because of this phenotype we sequenced the *SPK1* gene and found a mutation in the splicing acceptor site of 29th intron. Another allele of SPK1 - *spk1-3*, a TILLING (*31*) line showed similar reduction of vascular cell number and abnormal cell alignment.

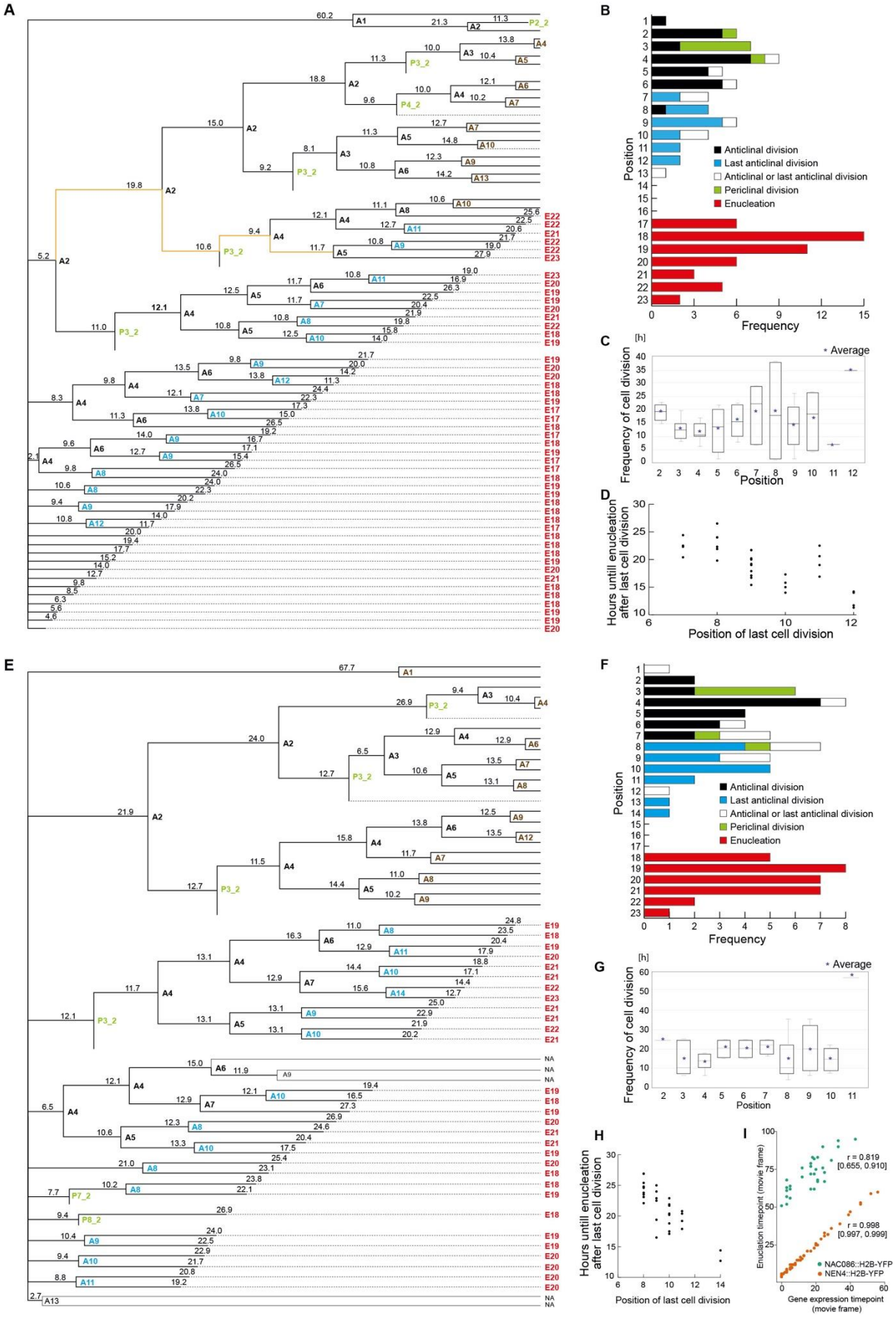
A
B
C
D
E
F
G
H
I
Anticlinal division
Last anticlinal division
Anticlinal or last anticlinal division
Periclinal division
Enucleation
Position
Frequency
Frequency of cell division
Average
Position
Hours until enucleation after last cell division
Position of last cell division
r = 0.819
[0.655, 0.910]
r = 0.998
[0.997, 0.999]
NAC086::H2B-YFP
NEN4::H2B-YFP
Enucleation timepoint (movie frame)
Gene expression timepoint (movie frame)

Fig. S1. Live imaging of protophloem sieve element development.

(A) and **(E)** Cell division trees extracted from Supplementary movies 1 and 2, respectively. Time between cell divisions is indicated above the line connecting two nodes that demark type (A=anticlinal, P=periclinal) and position of the cell at the time of division. Cell division annotation in black describes anticlinal divisions in Transit Amplifying zone; in blue, last anticlinal divisions (Transition zone); in brown, last divisions observed in the movie where fate of the daughter cells can't be determined; in green, periclinal divisions (no first periclinal division has been observed). Annotation in red informs about the position of enucleation. Division path in orange depicts shortest passage through transit amplifying cells observed. **(B)** and **(F)** Histograms of cellular events from movies 1 and 2, respectively. **(C)** and **(G)** Time between cell divisions in different positions of protophloem sieve element cell lineage. Dots indicate the average. **(D)** and **(H)** The time from last cell division till enucleation plotted for different cell positions. **(I)** Pearson's correlation (r) between gene expression activation time point and the time point this cell enucleated. 95% confidence interval is provided in square brackets. The confidence interval is wider for *NAC086*, reflecting a higher variation in the interval between the time the gene expression is activated and the time the cell enucleates.

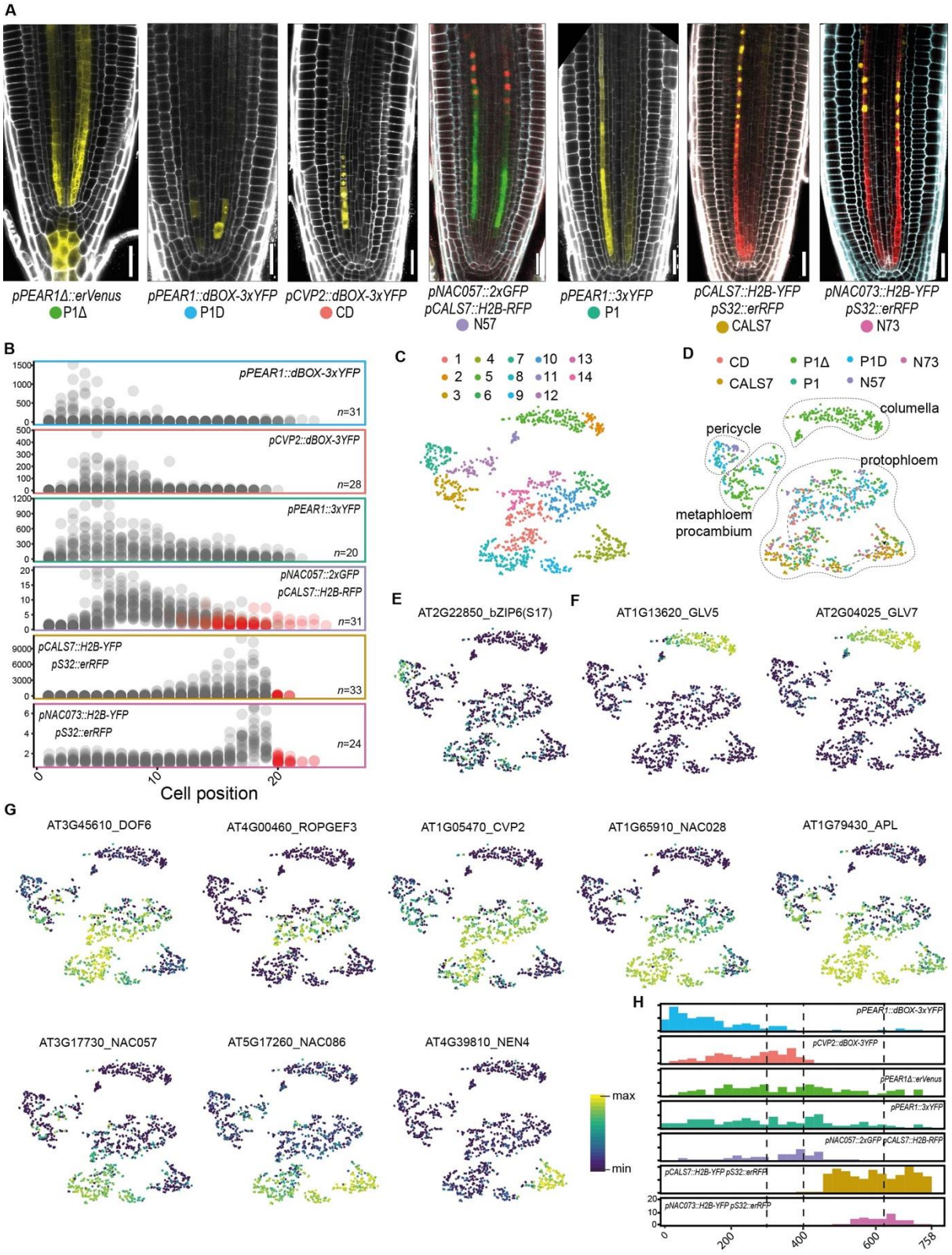

Fig. S2. Sorting of phloem specific cells.

(A) Reporter lines used for protoplasting and FACS based isolation of cells. Modified *PEAR1* promoter (P1Δ) shows ectopic expression in columella reflected in additional columella cluster

in (D, F). Scale bars, 25 µm. **(B)** Fluorescent signal intensity distribution in the phloem reporter lines from (A). Averaged and background-normalised fluorescence per protophloem sieve element cells is given from the stem cell to the enucleating cell. Red dots indicate cells with RFP fluorescent nucleus (*pNAC057::2xGFP pCALS7::H2B-RFP*) or that are enucleated (*pCALS7::H2B-YFP* and *pNAC073::H2B-YFP*). **(C)** t-SNE plot reveals 14 different clusters of the analysed cells. **(D)** t-SNE plot with transcriptomes color-coded based on the reporter line they originate from. Cluster annotation based on the expression of known, tissue-specific genes. **(E)** Expression of pericycle specific and **(F)** columella specific genes indicate cluster identities. **(G)** Identification of protophloem sieve element cluster(s) based on the expression of sieve element specific genes. Subsequent t-SNE plots indicate a developmental gradient within protophloem sieve element cluster(s). **(H)** Histogram of cell distribution along pseudotime of 758 ordered protophloem sieve element transcriptomes based on the reporter line used for isolation. Correlation drop points are indicated with dashed lines.

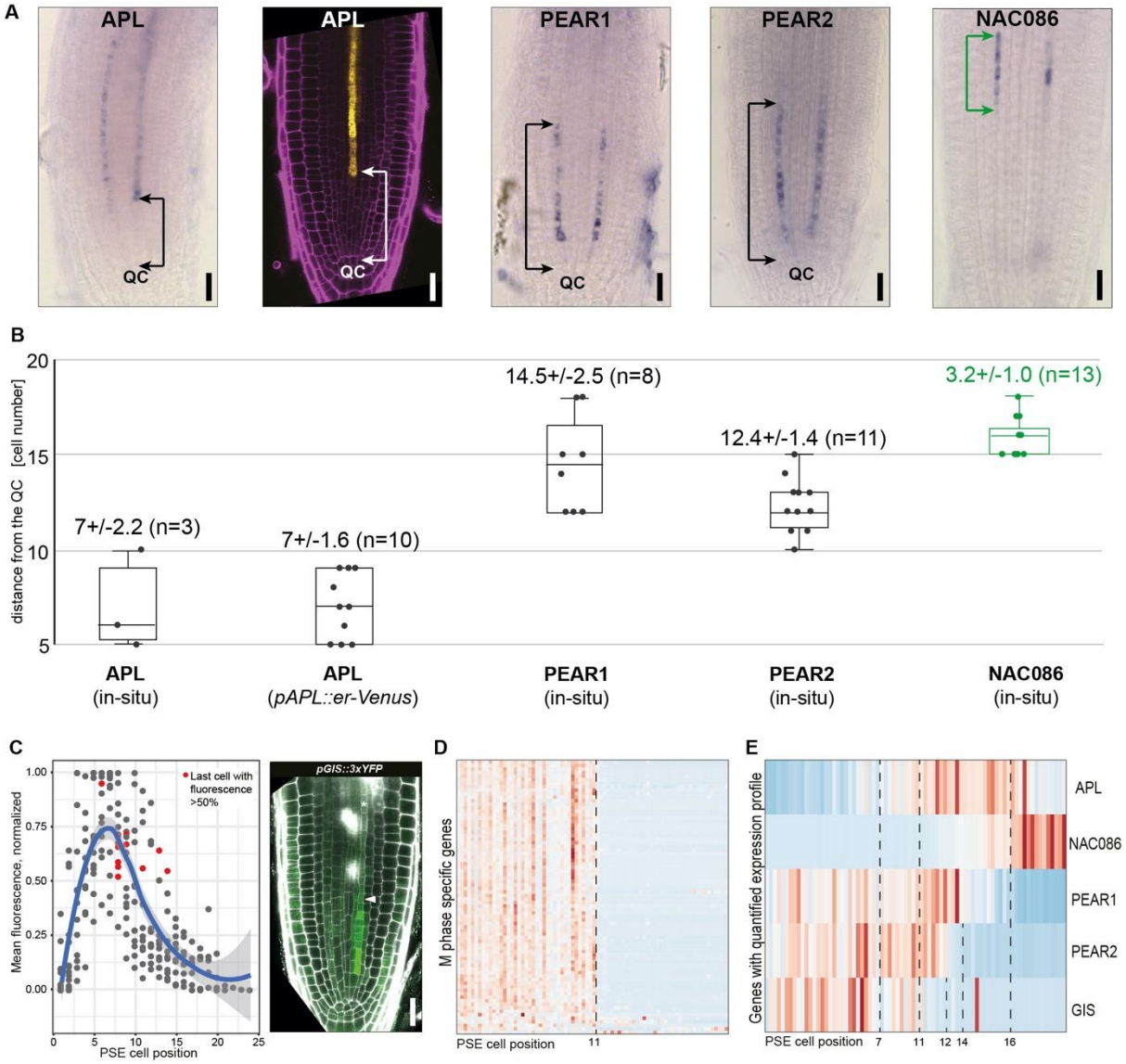

Fig. S3. Mapping of the single cell transcriptome onto 19 cells of the protophloem sieve element cell lineage.

(A) Representative images of in situ RNA hybridization and reporter line indicating the onset or termination of expression of protophloem sieve element specific genes. The quantified domains are indicated with arrows. Scale bars, 25 µm. **(B)** Quantification of domains indicated in panel (A) representing onset of *APL* and *NAC086* and termination of *PEAR1* and *PEAR2* expression. Expression of *NAC086* (in green) terminates upon cell enucleation at position 19. **(C)** Quantification of relative fluorescence per protophloem sieve element cell position of *pGIS::3xYFP* lines in FRAP experiments (n=12 phloem poles). Representative image of

pGIS::3xYFP after FRAP recovery. Arrowhead points at the last cell with fluorescent level above 50%. Scale bar, 25 µm. **(D)** Gene expression heatmap of genes specifically expressed during M phase of the cell cycle (Menges et al., 2005). The end of expression can be linked to the position 11 in the live imaging data of cell division in figure S1. **(E)** Heatmap showing expression of genes quantified in (A to C) in pseudotime ordered protophloem sieve element transcriptomes. Approximate start or end points of gene expression are indicated with dashed lines, linking quantified positions in (B, C, D) to pseudotime positions.

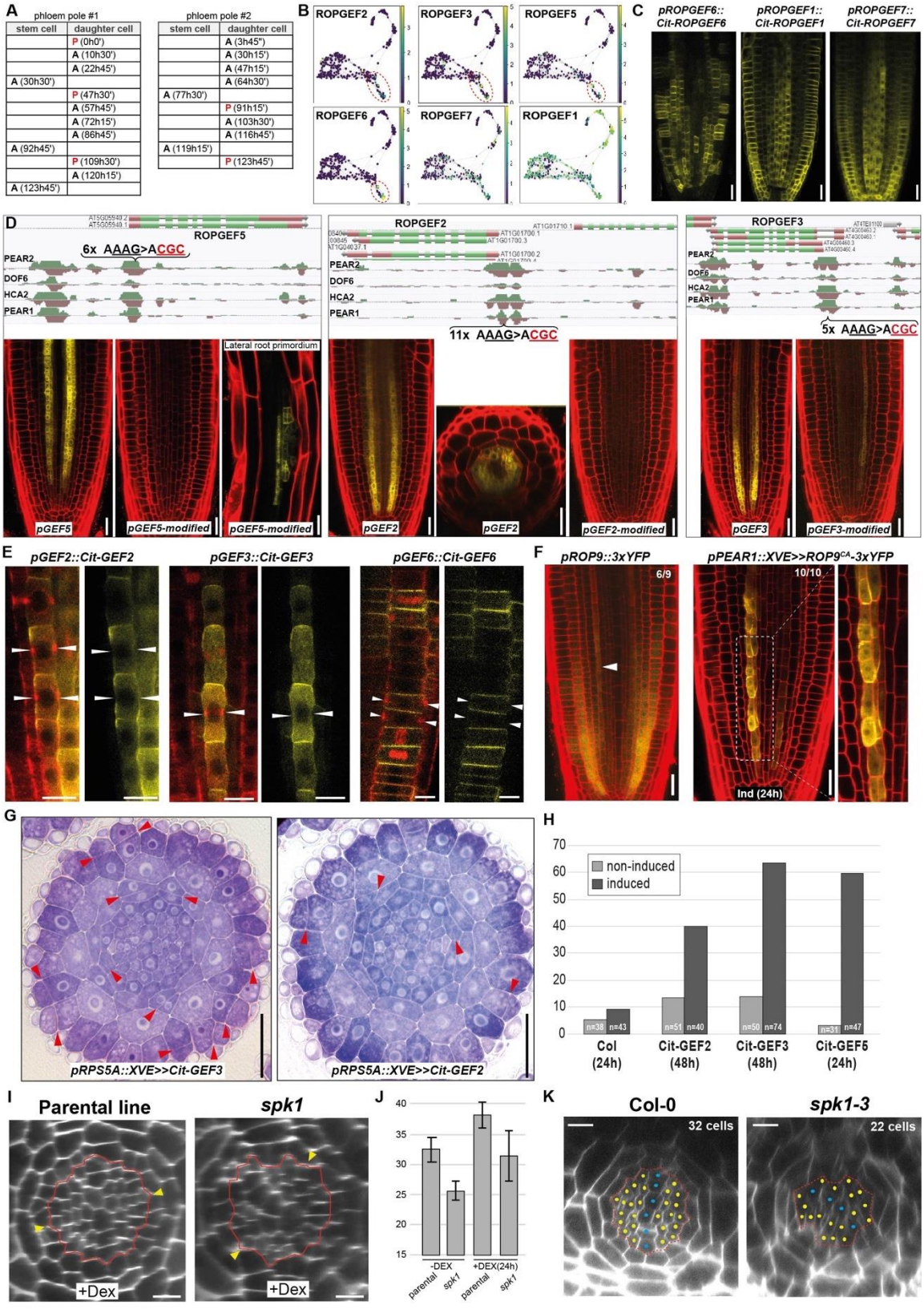

Fig. S4. PEARs control asymmetric divisions by promoting ROPGEF-ROP signaling in the phloem pole.

(A) Cell division pattern of phloem stem cell (position 1) and its first daughter cell (position 2). Data extracted from Movie S13**. (B)** Expression pattern of 6 ROPGEFs in the phloem single cell data. Dashed-line circles indicate narrow expression domain of *ROPGEF2, 3* and *5* and M-phase specific expression of *ROPGEF6*. **(C)** Broad expression of *ROPGEF1, 6* and *7* in the root meristem. Scale bars: 25 µm. **(D)** PEAR dependent vascular expression of phloem enriched *ROPGEFs*. In the modified promoters DOF biding sites (AAAG) were changed to (ACGC). PEAR DAP-Seq data was extracted from Plant Cistrome Database (*26*). All reporters use promoter driven *erVenus* fluorescent tag. Scale bars: 25 µm. **(E)** Depletion of *Cit-ROPGEF2, 3 and 6* membrane signal at the cortical division zone (CDZ) during cell division. CDZ is marked by accumulating cortical microtubules (mCherry-TUA5) forming preprophase band (white arrowheads). Scale bar: 10 µm. **(F)** ROP9 expression in protophloem sieve element (*pROP9::3xYFP*); accumulation of constitutively active (Q64L) form of ROP9 ($ROP9^{CA}$) on the radial walls of the protophloem sieve element cells. Scale bars: 25 µm. **(G)** Toluidine blue stanning of resin sections of *pRPS5A::XVE>>Cit-GEF2* and *pRPS5A::XVE>>Cit-GEF3* lines, 48h after induction. Red arrowheads indicate ectopic periclinal cell divisions in epidermis, endodermis and pericycle. Scale bars: 25 µm. **(H)** Percentage of plants that show ectopic periclinal cell divisions in pericycle layer before and 24 h or 48 h after induction. **(I)** Ectopic periclinal cell divisions induced by overexpression of PEAR1-GR in the wild type and *spk1* background. Phloem poles are indicated with yellow arrowheads. Red dashed line indicates vascular cells. Scale bar: 10 µm. **(J)** Quantification of vascular cell files after overexpression of PEAR1 in wt and *spk1* mutant background**. (K)** Reduced number of vascular cell files (indicated on the image) in *spk1-3* (Tilling) mutant allele (*31*). Red dashed line indicates vascular cells. Phloem and procambial cells (yellow dots), xylem cells (blue dots). Scale bar: 10 µm.

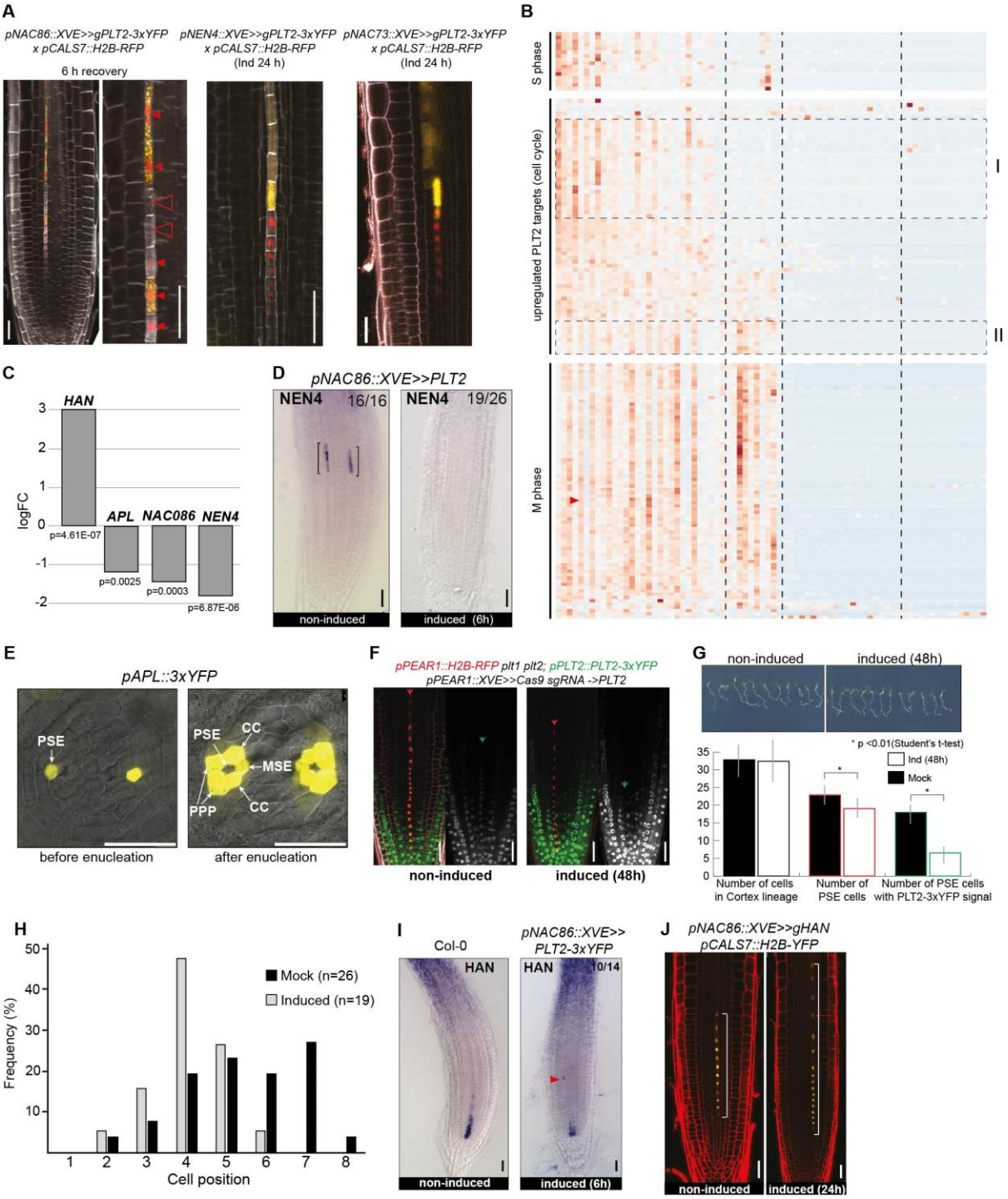

Fig. S5. PLT2 inhibits expression of late protophloem sieve element genes and the key regulator *APL*. (A) The effect of ectopic expression of *PLT2* under p*NAC086* is reversible as indicated by the enucleation of PLT2-free protophloem sieve element cells after recovery. During very late differentiation of phloem (domain VI) expression of *PLT2* under *pNAC073* and *pNEN4* does not delay enucleation, neither promotes cell division. **(B)** Gene expression

heatmap highlighting oscillatory patterns of a subset of PLT2 upregulated cell cycle related targets (clusters I and II), coinciding with expression of either S or M phase specific genes. **(C)** Gene expression changes (RNA-Seq) in FACS-sorted phloem cells of *pNAC086::XVE>>PLT2-3xYFP x pCALS7::H2B-RFP* line, 6h after induction. See Table S3 for full data set. **(D)** In situ hybridization of *NEN4* before and 6h after ectopic expression of PLT2-3xYFP. Brackets indicate in situ signal. **(E)** protophloem sieve element specific expression of *APL* before enucleation and its expression in companion cells (CC), metaphloem sieve element (MSE) and phloem pole pericycle (PPP) after protophloem sieve element enucleation. **(F)** Premature enucleation of phloem cells, monitored by *pPEAR1::H2B-RFP*, 48h after induction of protophloem sieve element specific CRISPR construct targeting *PLT2* (including complementing construct) in *plt1 plt2 pPLT2::PLT2-3xYFP* background. Last nucleated protophloem sieve element cell and last protophloem sieve element cell expressing PLT2-3YFP are indicated with red and green arrows, respectively. **(G)** No change to meristem size was observed 48h after induction of CRISPR construct. Number of protophloem sieve element cells was monitored based on *pPEAR1::H2B-RFP*. **(H)** Quantification of p*APL::erTurq* expression after induction of tissue specific CRISPR construct targeting *PLT2*. **(I)** Native and ectopically activated expression of *HAN*. Arrowhead indicates ectopic expression of *HAN* promoted by ectopic expression of PLT2. **(J)** Ectopic expression of HAN delays enucleation in the late domain of differentiating phloem. Square brackets indicate extended expression domain of *pCALS7::H2B-RFP*, a reporter used for monitoring enucleation. All scale bars: 25 µm.

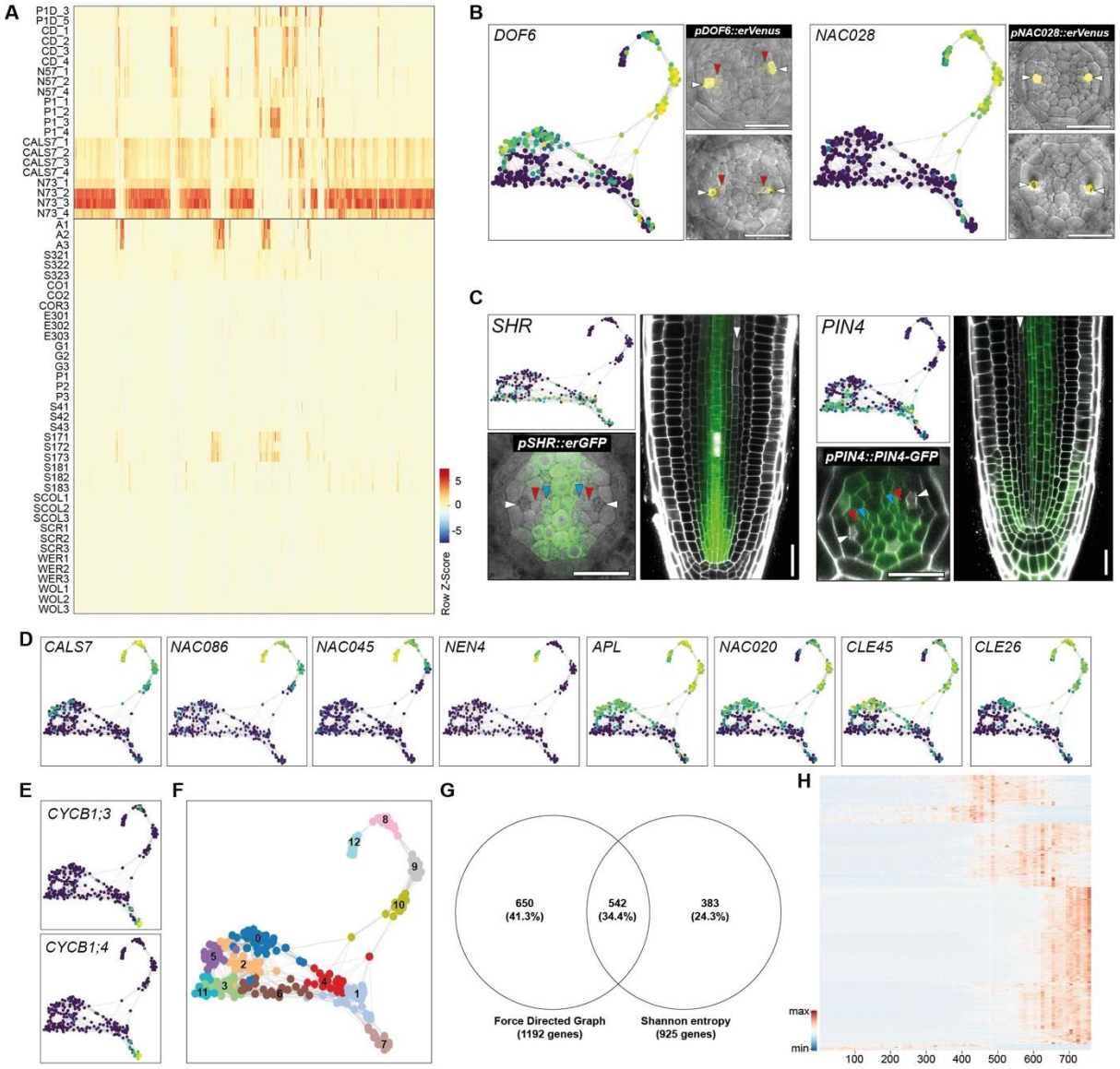

Fig. S6. Identification of protophloem sieve element enriched genes using bulk and individually sorted cells.

(A) Expression heatmap of 925 phloem enriched genes in the combined transcriptome data set of bulk sorted protophloem sieve element domains and other cell types from (*37*). **(B)** Force-directed graph of 272 single-cell transcriptomes obtained using the *pPEAR1Δ::erVenus* reporter. Expression of known sieve element genes indicates protophloem and metaphloem sieve element trajectories. Vibratome cross sections of fluorescent reporter lines confirm SE specificity of both genes. **(C)** Identification of procambial ("non-SE") cell trajectory based on the expression pattern of *SHORT ROOT* (*SHR*) and *PIN-FORMED 4* (*PIN4*). Both genes are

expressed broadly in the very early cells but quickly become excluded from the protophloem sieve element and metaphloem sieve element cell files. **(D)** Expression pattern of known sieve element specific genes in the force-directed graph analysis confirms cluster identity. **(E)** Expression of M-phase specific cyclins indicates that the separation of cluster 7 (F) is based on the stage of cell cycle. **(F)** Distribution of 12 Louvain clusters of the single cell transcriptomes from cell sorting of *pPEAR1Δ::erVenus* reporter. **(G)** Venn diagram of phloem enriched (Shannon Entropy analysis of bulk sorted cells) and sieve element specific genes (based on the comparisons between Louvain clusters) shows the large overlap of 542 common genes. **(H)** Expression heatmap of 542 sieve element enriched genes from (G) in pseudotime. SE enriched genes show predominant expression during late stages of protophloem sieve element development.

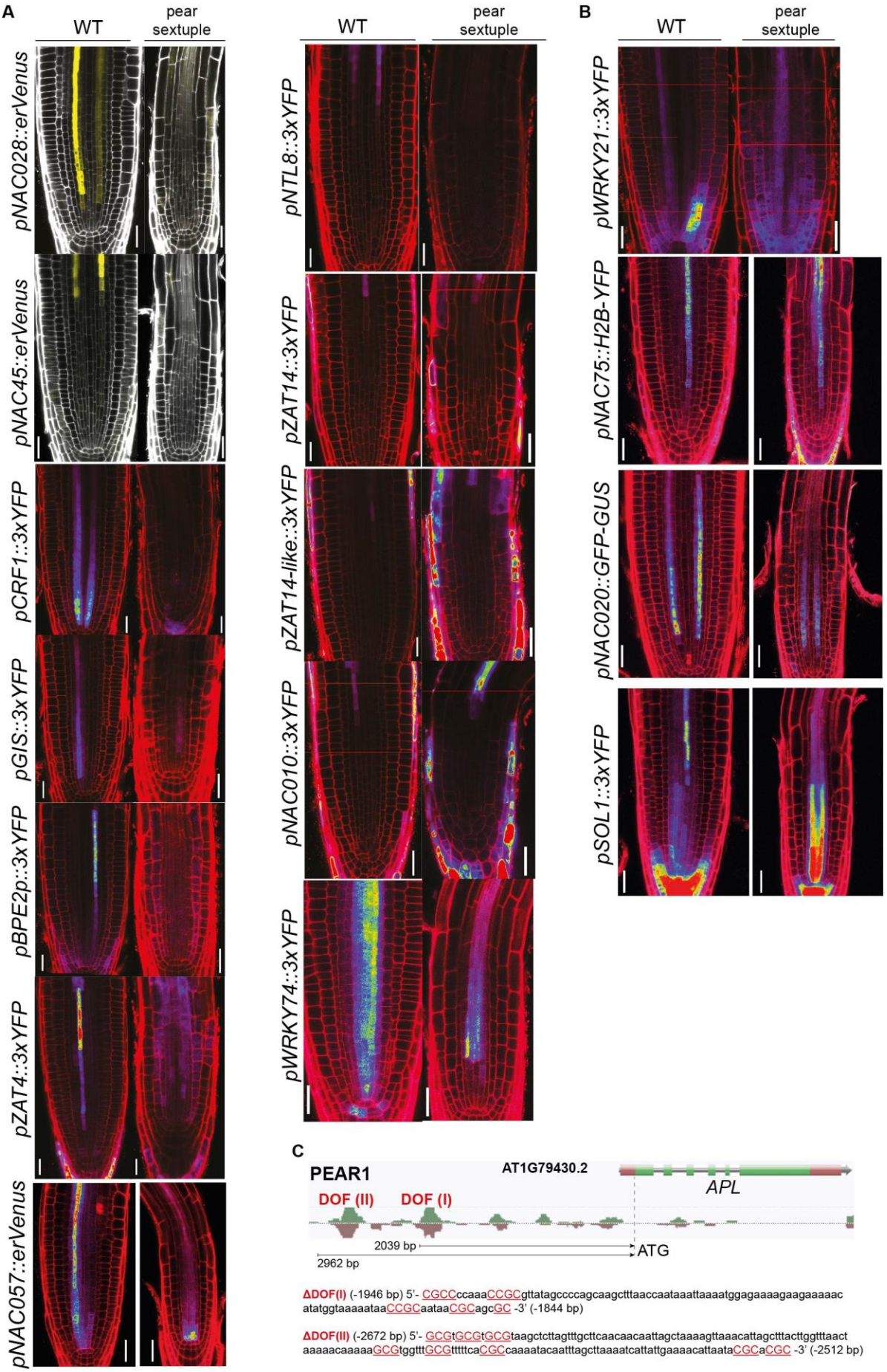
A
WT
pear sextuple
pNAC028::erVenus
pNAC045::erVenus
pCRF1::3xYFP
pGIS::3xYFP
pBPE2p::3xYFP
pZAT4::3xYFP
pNAC057::erVenus
pNTL8::3xYFP
pZAT14::3xYFP
pZAT14-like::3xYFP
pNAC010::3xYFP
pWRKY74::3xYFP
B
pWRKY21::3xYFP
pNAC75::H2B-YFP
pNAC020::GFP-GUS
pSOL1::3xYFP
C
PEAR1
AT1G79430.2
APL
DOF (II)
DOF (I)
2039 bp
2962 bp
ATG
ΔDOF(I) (-1946 bp) 5'- CGCCccaaaCCGCgttatagccccagcaagctttaaccaataaattaaaatggagaaaagaagaaaaac
atatggtaaaaataaCCGCaataaCGCagcGC -3' (-1844 bp)
ΔDOF(II) (-2672 bp) 5'- GCGtGCGtGCGtaagctcttagtttgcttcaacaacaattagctaaaagttaaacattagctttacttggtttaact
aaaaacaaaaaGCGtggtttGCGtttttcaCGCcaaaatacaatttagcttaaaatcattattgaaaacattaataCGCaCGC -3' (-2512 bp)

Fig. S7. Expression pattern of PEAR-dependent and -independent, protophloem sieve element enriched TFs.

(A) Known and newly identified protophloem sieve element enriched TFs whose expression during phloem development depends on PEARs. **(B)** 4 of the tested promoters remained active in pear sextuple mutant, indicating that expression of these transcription factors does not depend on PEARs. Images were taken using the same settings, ensuring comparable intensities between genetic backgrounds. All scale bars: 25 μm. **(C)** An image extracted from Plant Cistrome Database (http://neomorph.salk.edu/dev/pages/shhuang/dap_web/pages/index.php) showing PEAR1 binding to *APL* promoter. DOF binding sites of the two prominent peaks have been modified as listed below the image. Coordinates indicate position upstream of APL ORF.

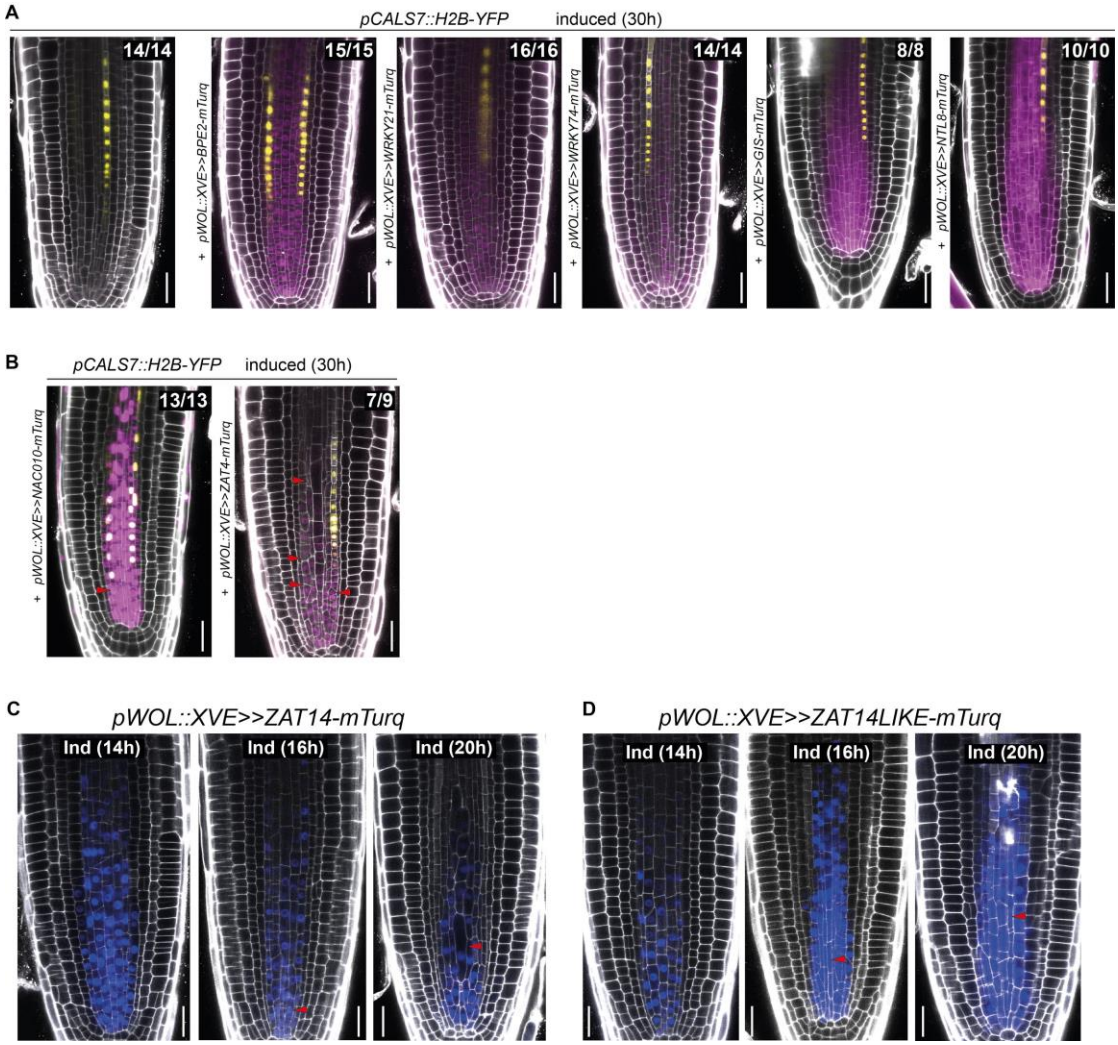

Fig. S8. Ectopic overexpression studies of newly identified protophloem sieve element enriched transcription factors.

(A) Overexpression of *BPE2*, *WRKY21*, *WRKY74*, *GIS* and *NTL8* under *pWOL::XVE* promoter (active broadly in the vascular tissue) did not result in any obvious phenotypes. Transgenic lines carry *pCALS7::H2B-YFP* to monitor phloem enucleation. **(B)** Overexpression of *NAC010* and *ZAT4* resulted in shortening of the root meristem and defects in the cell shape, respectively, indicated by red arrowheads. **(C)** and **(D)** Extensive cell elongation and inhibition of cell division phenotypes in the lines overexpressing *ZAT14* and *ZAT14-like*. Red arrowheads indicate unusual cell expansion. First signs of the phenotype appear at 16h after induction. Lines were profiled (RNA-Seq of root meristems) at 10h and 14h after induction. Numbers indicate observations with similar expressions and phenotype. All scale bars: 25 µm.

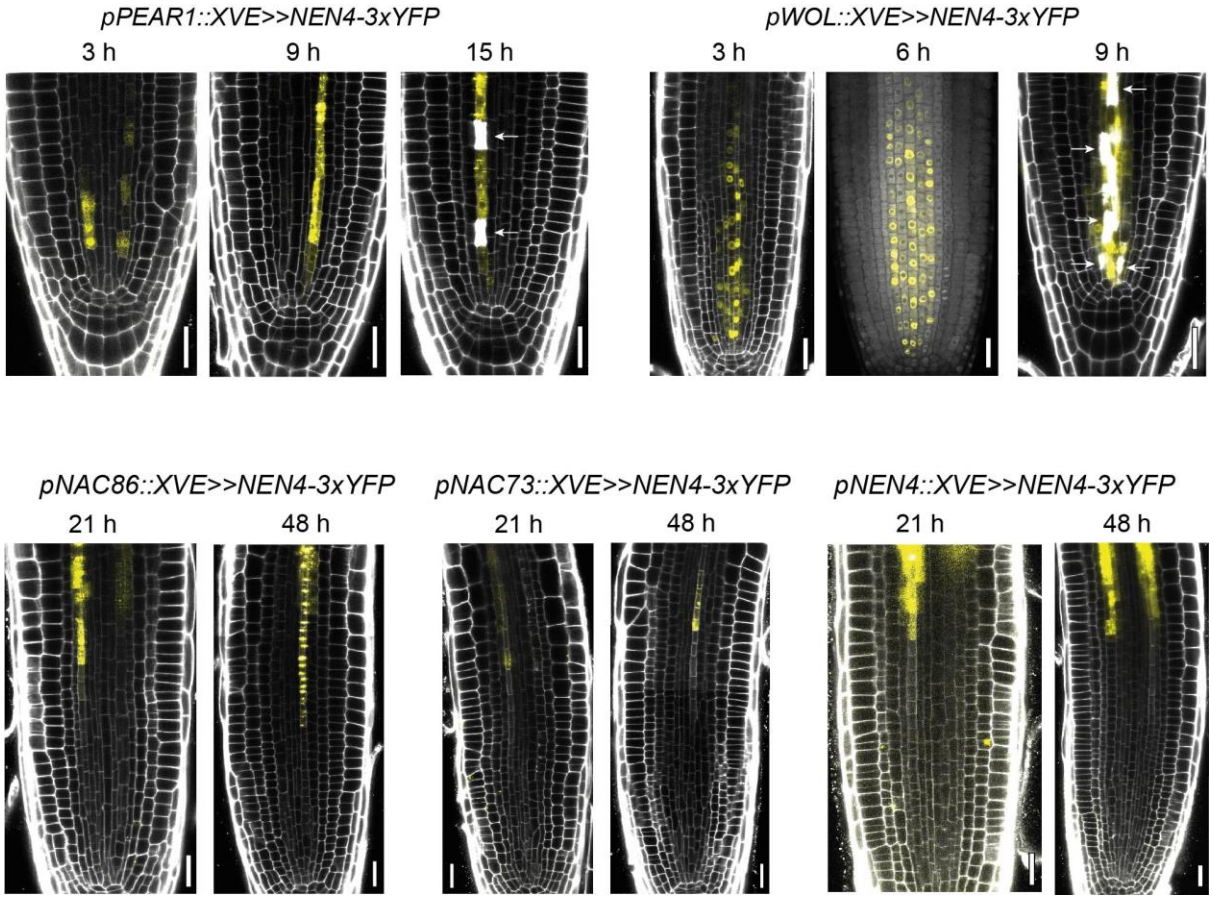

Fig. S9. Ectopic expression of NEN4 reveals stage of fixed protophloem sieve element cell identity.

Premature expression of *NEN4* under protophloem sieve element specific *pPEAR1* or the broadly active *pWOL* promoters results in cell death after 15 or 9 hours, respectively. On the other hand, expression of *NEN4* under late protophloem sieve element promoters (*pNAC86, pNAC73, pNEN4*) reveals resistance of cells close to maturation to NEN4-induced cell death. White arrows point to dead cells. All scale bars: 25 µm.

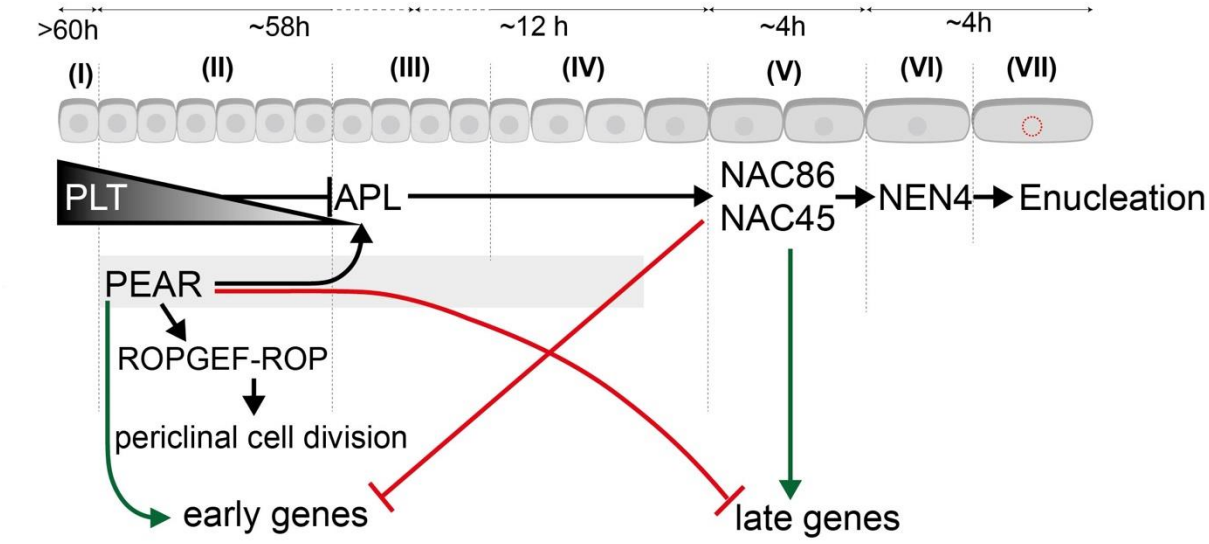

Fig. S10. Gene regulation along the protophloem sieve element developmental trajectory.

Genetic interactions of transcription factors orchestrating protophloem sieve element development. The developmental phases (I-VII) and timing of phase transitions were determined in this study. The seesaw concept is indicated by the validated up- and down-regulation of early and late expressed genes by PHLOEM EARLY EXPRESSED DOFs (PEARs) and NAC DOMAIN CONTAINING PROTEIN45/86 (NAC45/86), respectively and *vice versa*. Expression of ALTERED PHLOEM DEVELOPMENT (APL) gene in the protophloem sieve element is promoted by PEARs but counteracted, around the stem cell niche, by PLETHORA (PLT) factors. NAC86/45 and NAC45/86-DEPENDENT EXONUCLEASE-DOMAIN PROTEIN 4 (NEN4) promote protophloem sieve element enucleation downstream of APL. In the early phloem, PEAR factors control lineage bifurcation by activating transcription of RHO OF PLANTS GUANINE NUCLEOTIDE EXCHANGE FACTORs (ROPGEFs) that in turn locally activate ROP signaling.

Supplementary tables

All Supplementary Tables can be found in the separate work sheets of a single Supplementary_Tables.xlsx file.

Table S1: List of temporal specific genes. These were identified by comparing the transcriptomes in this domain to those of all others in pairwise comparisons. The highest qval of all these comparisons was then used for further selection (only genes with qval <0.05 are listed). Genes specific to the nested PLT1-like and NEN4-like domains can also be part of domain [a] or domain [d], respectively.

Table S2: List of developmental phases of protophloem sieve element

Table S3: List of PLT2 targets after ectopic expression in late protophloem sieve element cells. The expression in each replicate (RPKM), logarithm of the fold change (logFC, between EST/induced and DMSO/non-induced), p values (PValue), false discovery rates (FDR) are given. Upregulated genes that have GO term annotations as “cell cycle” are indicated as well as here detected specificity to M or S phase of those genes.

Table S4: List of 925 phloem enriched genes based on Shannon entropy analysis of bulk sorted cells. The gene IDs, names and description are given. RPKM values for all used replicates, including the published data run through the same bioinformatics pipeline are given (Columns D-BK). Statistics of the Shannon entropy calculations can be found in Columns BL-BS.

Table S5: List of 1192 SE enriched genes based on cluster comparisons of 272 *pPEAR1Δ::erVenus* single cells from the protophloem sieve element, metaphloem sieve element and procambial lineage. For further description of the applied comparisons, see M&M.

Table S6: List of the protophloem sieve element enriched genes that form the overlap between Table S4 and Table S5.

Table S7: Average expression per cluster of 467 out of 542 phloem enriched genes (Table S6) that were detected in published root expression atlas (Wendrich et al., submitted). These genes are not broadly expressed throughout the tissue types. While many genes show either protophloem sieve element or phloem specific expression, some show specific expression in other cell types. Several genes confirmed to be expressed during late protophloem sieve element cells such as *NAC010*, *NAC73*, *ZAT14*, *RMT2* and *NEN4* appear not to be expressed in the sieve element cluster. This could potentially indicate the underrepresentation of late protophloem sieve element cells within the sieve element cluster (<30 cells).

Table S8: Core predictions of the random forest GRN. The top10 most important transcription factors for each gene were selected, resulting in 49240 edges in the model (ranked by importance). The gene ID for transcription factor (TF) and the predicted target (Target), the importance of this edge (Importance), the correlation (Correlation) and the regulation based on the correlation is given (Correlation sign). These core edges of the model were in turn used to determine the most important transcription factors by the number of their target genes. The gene ID for each transcription factor (gene), number of target genes (target #), combined importance of these edges (target importance) and a short annotation for each transcription

factor (annotation) is given. The model was tested for enrichment of targets determined in available overexpression or mutant data sets. The results of this enrichment test also given.

Table S9: List of differentially expressed genes in the pear sextuple root meristem. Read counts, log2 fold change, pval and padj are given. Only significantly differentially expressed genes (padj<0.05) are shown.

Table S10: List of differentially expressed genes after 10h induction of *pWOL::XVE>>ZAT14*. Gene IDs, mean read counts of 3 replicates for 10h induction and DMSO control (Z10 and Z_DMSO), log2foldchange, pvalue, padjust, gene loci and gene names are given for all significant differentially expressed genes (padj<0.05).

Table S11: List of differentially expressed genes after 14h induction of *pWOL::XVE>>ZAT14*. Gene IDs, mean read counts of 3 replicates for 14h induction and DMSO control (Z14 and Z_DMSO), log2foldchange, pvalue, padjust, gene loci and gene names are given for all significant differentially expressed genes (padj<0.05).

Table S12: List of differentially expressed genes after 10h induction of *pWOL::XVE>>ZAT14L*. Gene IDs, mean read counts of 3 replicates for 10h induction and DMSO control (L10 and L_DMSO), log2foldchange, pvalue, padjust, gene loci and gene names are given for all significant differentially expressed genes (padj<0.05)

Table S13: List of differentially expressed genes after 14h induction of *pWOL::XVE>>ZAT14L*. Gene IDs, mean read counts of 3 replicates for 10h induction and

DMSO control (L14 and L_DMSO), log2foldchange, pvalue, padjust, gene loci and gene names are given for all significant differentially expressed genes (padj<0.05)

Table S14: PANTHER Gene ontology analysis of ZAT14 and/or ZAT14L targets (14h induction). Details such as test type and correction are given alongside the expected and detected number of genes for each GO category, the fold enrichment and p value.

Table S15: Significant overlap of up and down regulated, predicted targets for the Top20 transcription factors in the model, respectively.

Table S16: Summary of statistical tests for overlapping gene sets of downstream targets in over-expression data sets. Overrepresentation ratio and pval (Fisher test) are given as well as genes that are both up and down regulated by early or late transcription factors. PLT2 and ZAT14L were included as those are also important regulators in this study with very specific early or late expression and overexpression target data was available.

Table S17: List of single cell transcriptomes used in this study. Cell identifiers (ID), reporter line used to isolate the cell (sample), number of total reads, unmapped reads, reads mapped to a single locus and reads mapped to multiple loci, percent of reads mapped to single locus, reads mapped to multiple loci and generally mapped to A. thaliana genome, average library size (lib_size_avr), Cluster number when all cells passing quality thresholds were clustered (Cluster_all_cells), cluster number when only 758 protophloem sieve element were clustered (Cluster_PSE) and assigned pseudotime for protophloem sieve element cells are given (Pseudotime_PSE).

Table S18: List of fluorescence reporters and mutant lines of *A. thaliana* analyzed in this publication. Genetic background, introduced transgenic construct, abbreviation and origin are given.

Table S19: List of oligonucleotide sequences used in this study. Primer names, including gene names and positions and amplified fragment sizes are given together with the sequence.

Table S20: Live imaging quantification data. The time between onset of expression and enucleation in *pNAC86::H2B-YFP* and *pNEN4::H2B-YFP* line and between enucleation was quantified. For the expression length, the root and corresponding movie, the observed phloem pole (pole), movie frame with onset and enucleation and the difference (frame gene ON, frame enucleation, frames gene expression, respectively), the time of expression before enucleation in hours (time cell enucleation [h]) and the corresponding movies are given. The enucleation quantification is based only on movie S1 and movie S2 and their analysis in Fig S1A, E. Time between consecutive enucleation events and the corresponding movie are given.

Following references are cited only in the Supplementary Materials

39. A. Lampropoulos, Z. Sutikovic, C. Wenzl, I. Maegele, J. U. Lohmann, J. Forner, GreenGate - a novel, versatile, and efficient cloning system for plant transgenesis. *PLoS One.* ***8***, e83043 (2013).
40. W. Decaestecker, R. A. Buono, M. L. Pfeiffer, N. Vangheluwe, J. Jourquin, M. Karimi, G. Van Isterdael, T. Beeckman, M. K. Nowack, T. B. Jacobs, CRISPR-

TSKO: A Technique for Efficient Mutagenesis in Specific Cell Types, Tissues, or Organs in Arabidopsis. *Plant Cell*. ***31***, 2868–2887 (2019).

41. N. M. Clark, A. P. Fisher, R. Sozzani, in *Computational Cell Biology: Methods and Protocols*, L. von Stechow, A. Santos Delgado, Eds. (Springer New York, New York, NY, 2018; https://doi.org/10.1007/978-1-4939-8618-7_6), *Methods in Molecular Biology*, pp. 139–151.
42. F. Zambelli, F. Mastropasqua, E. Picardi, A. M. D'Erchia, G. Pesole, G. Pavesi, RNentropy: an entropy-based tool for the detection of significant variation of gene expression across multiple RNA-Seq experiments. *Nucleic Acids Res*. **46**, e46–e46 (2018).
43. S. Picelli, O. R. Faridani, Å. K. Björklund, G. Winberg, S. Sagasser, R. Sandberg, Full-length RNA-seq from single cells using Smart-seq2. *Nature Protocols*. **9**, 171–181 (2014).
44. J. S. Dahlin, F. K. Hamey, B. Pijuan-Sala, M. Shepherd, W. W. Y. Lau, S. Nestorowa, C. Weinreb, S. Wolock, R. Hannah, E. Diamanti, D. G. Kent, B. Göttgens, N. K. Wilson, A single-cell hematopoietic landscape resolves 8 lineage trajectories and defects in Kit mutant mice. *Blood*. **131**, e1–e11 (2018).
45. F. A. Wolf, P. Angerer, F. J. Theis, SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology*. **19**, 15 (2018).
46. D. Kim, B. Langmead, S. L. Salzberg, HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*. **12**, 357–360 (2015).
47. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. **25**, 2078–2079 (2009).

48. C. Trapnell, B. A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M. J. van Baren, S. L. Salzberg, B. J. Wold, L. Pachter, Transcript assembly and abundance estimation from RNA-Seq reveals thousands of new transcripts and switching among isoforms. *Nat Biotechnol*. **28**, 511–515 (2010).
49. A. Fernandez, A. Drozdzecki, K. Hoogewijs, A. Nguyen, T. Beeckman, A. Madder, P. Hilson, Transcriptional and Functional Classification of the GOLVEN/ROOT GROWTH FACTOR/CLE-Like Signaling Peptides Reveals Their Role in Lateral Root and Hair Formation. *Plant Physiology*. **161**, 954–970 (2013).
50. V. D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks. *J. Stat. Mech.* **2008**, P10008 (2008).
51. J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, Fiji: an open-source platform for biological-image analysis. *Nature Methods*. **9**, 676–682 (2012).
52. E. K. Yoon, S. Dhar, M.-H. Lee, J. H. Song, S. A. Lee, G. Kim, S. Jang, J. W. Choi, J.-E. Choe, J. H. Kim, M. M. Lee, J. Lim, Conservation and Diversification of the SHR-SCR-SCL23 Regulatory Network in the Development of the Functional Endodermis in Arabidopsis Shoots. *Mol Plant*. **9**, 1197–1209 (2016).
53. C.-H. Yang, W.-T. Kuo, Y.-T. Chuang, C.-Y. Chen, C.-C. Lin, Cyclin B1 Destruction Box-Mediated Protein Instability: The Enhanced Sensitivity of Fluorescent-Protein-Based Reporter Gene System. *BioMed Research International* (2013), , doi:10.1155/2013/732307.

Supplementary Movies:

Movie S1: Live-imaging of *pPEAR1::H2B-YFP pCALS7::H2B-YFP pWOX5-mCherry*; quantification in Fig S1A.

Movie S2: Live-imaging of *pPEAR1::H2B-YFP pCALS7::H2B-YFP pWOX5-mCherry*; quantification in Fig S1E.

Movie S3: Promoter activity of *pNAC86::H2B-YFP*, quantification in Fig S1I and Table S20.

Movie S4: Promoter activity of *pNAC86::H2B-YFP*, quantification in Fig S1I and Table S20.

Movie S5: Promoter activity of *pNAC86::H2B-YFP*, quantification in Fig S1I and Table S20.

Movie S6: Promoter activity of *pNEN4::H2B-YFP*, quantification in Fig S1I and Table S20.

Movie S7: Promoter activity of *pNEN4::H2B-YFP*, quantification in Fig S1I and Table S20.

Movie S8: Promoter activity of *pNEN4::H2B-YFP*, quantification in Fig S1I and Table S20.

Movie S9: Promoter activity of *pNEN4::H2B-YFP*, quantification in Fig S1I and Table S20.

Movie S10: Promoter activity of *pNEN4::H2B-YFP*, quantification in Fig S1I and Table S20.

Movie S11: Promoter activity of *pNEN4::H2B-YFP*, quantification in Fig S1I and Table S20.

Movie S12: Promoter activity of *pNEN4::H2B-YFP,* quantification in Fig S1I and Table S20.

Movie S13: Live-imaging of *pPEAR1::H2B-YFP pCALS7::H2B-YFP pWOX5-mCherry*; quantification in Fig S4A.