bioRxiv preprint doi: https://doi.org/10.1101/517250; this version posted January 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Stem-cell-ubiquitous genes spatiotemporally coordinate division through regulation of stem-cell-specific gene networks

Natalie M Clark[1,2], Eli Buckner[3,†], Adam P Fisher[1,†], Emily C Nelson[1,†], Thomas T Nguyen[1,†], Abigail R Simmons[4], Maria A de Luis Balaguer[1], Tiara Butler-Smith[1], Parnell J Sheldon[1,5], Dominique C Bergmann[4,6], Cranos M Williams[3], Rosangela Sozzani[1,2,*]

1. Department of Plant and Microbial Biology, North Carolina State University, Raleigh, NC 27695
2. Biomathematics Graduate Program, North Carolina State University, Raleigh, NC 27695
3. Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC 27695
4. Department of Biology, Stanford University, Stanford, CA 94305
5. Department of Biology, Denison University, Granville, OH 43023
6. Howard Hughes Medical Institute (HHMI), Stanford University, Stanford, CA 94305

* Correspondence to Rosangela Sozzani, ross_sozzani@ncsu.edu
† These authors contributed equally to this work

Abstract

Stem cells are responsible for generating all of the differentiated cells, tissues, and organs in a multicellular organism and, thus, play a crucial role in cell renewal, regeneration, and organization. A number of stem cell type-specific genes have a known role in stem cell maintenance, identity, and/or division. Yet, how genes expressed across different stem cell types, referred here as stem-cell-ubiquitous genes, contribute to stem cell regulation is less understood. Here, we find that, in the Arabidopsis root, a stem-cell-ubiquitous gene, TESMIN-LIKE CXC2 (TCX2), controls stem cell division by regulating stem cell-type specific networks. Development of a mathematical model of TCX2 expression allowed us to show that TCX2 orchestrates the coordinated division of different stem cell types. Our results highlight that genes expressed across different stem cell types ensure cross-communication among cells, allowing them to divide and develop harmonically together.

Results

Local signaling pathways, fate predetermination of cell lineages, and cell plasticity are mechanisms known to maintain stem cell pluripotency[1–9]. However, while roughly 14% of human transcription factors (TFs) are expressed across 32 different cell and tissue types, most of their known functions are specifically localized to certain cell types, and their potential roles across different cell types are unknown[10]. Thus, whether these stem-cell-ubiquitous genes are upstream of known local and cell-type-specific mechanisms and if such global networks control the cross-communications between different cell populations is still an open question.

To understand how and whether stem-cell-ubiquitous genes contribute to cell identity, maintenance, and/or division, we performed gene expression analysis of the stem cells in the Arabidopsis root, as this offers a tractable system given its 3-dimensional radial symmetry and temporal information encoded along its longitudinal axis. To this end, seven root stem cell markers (Figure 1A), as well as a non-stem cell control, were used to identify stem cell-enriched genes, and among those, stem-cell-ubiquitous and stem-cell-specific genes, as it has been shown

bioRxiv preprint doi: https://doi.org/10.1101/517250; this version posted January 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

that there is a correlation between expression levels and functionality in specific cell types[1,10] (Extended Data Fig 1, see Methods). Notably, we found that the expression profiles of our markers together with known stem cell genes, agree with their known expression domains, supporting that our transcriptional profiles are specific to each stem cell population (Extended Data Fig 1). To measure transcriptional differences between the stem cells and the non-stem cells, we next performed a Principal Component Analysis (PCA.) Looking at the top 3 principal components (50.6% of the variation in the data), the PCA shows that the non-stem cell samples (red) are distant (using the Euclidean distance metric) from all of the stem cell populations, suggesting that the stem cells have a different transcriptional signature than the non-stem cells (Figure 1B). Accordingly, when we performed differential expression analysis on these data, we found that 9266 (28% of genes) are significantly enriched ($q < 0.05$ and fold change > 2) in at least one stem cell population compared to the non-stem cells and considered these genes the stem cell-enriched genes (see M&M and Supplemental Table 1). Thus, this approach allowed us to identify core stem cell genes, as functionally important genes are often enriched in the specific cell populations they control[1,10].

While the PCA gives us a general idea of how many genes are cell-specific vs cell-ubiquitous, it reduces the dimensionality of the problem to the three largest components of variance. Consequently, we would expect some genes been differentially enriched across all of the stem cell populations. Accordingly, when we performed differential expression analysis on the 9266 stem cell-enriched genes (see Methods), we find that 2018 genes (21.8% of the stem cell-enriched genes, hereinafter referred to as the stem-cell-ubiquitous genes) are enriched in at least 4 of the 6 unique stem cell types, with 569 of these 2018 (6.1% of the stem cell-enriched genes) enriched in 5 or 6 cell types (Figure 1C). Moreover, as each stem cell population clusters independently from the others in the PCA, we identified 7248 genes (78.2% of the stem cell-enriched genes), hereinafter referred to as the stem-cell-specific genes, enriched in 3 or less stem cell types, with 4331 of those 7248 genes (46.7% of the stem cell-enriched genes) enriched in only 1 stem cell type. This suggests that each specific stem cell type has its own, unique transcriptional signature. Given the separation between stem-cell-ubiquitous genes and stem-cell-specific genes, we next wanted to know if these two groups of genes have seemingly separated functions or, for example, if stem-cell-ubiquitous genes modulate stem-cell-specific gene expression to orchestrate coordinated processes between different cell types.

To test the latter hypothesis, in which stem-cell-specific genes are important for regulating cell type-specific aspects (e.g cell identity), but are regulated by stem-cell-ubiquitous genes so that stem cell maintenance and divisions are tightly coordinated, we used Gene Regulatory Network (GRN) inference and predicted the relationships between all 9266 genes enriched in the stem cells. We used a machine-learning, regression tree approach to infer dynamic networks from steady state, replicate data[11]. Our inferred GRN found regulations among 2982 (32.2%) of the stem cell-enriched genes and predicted that the stem-cell-ubiquitous (red) genes are located in the center of the network, which represents the beginning of the regulatory cascade, and are highly connected to each other (Figure 2A). Meanwhile, the cell-specific (blue) genes mostly regulate each other within the same cell type and are located on the outside of the network, therefore downstream of the cell-ubiquitous genes(Figure 2A). This suggests that the cell-ubiquitous genes are potentially involved in coordinating processes between different stem cells through the regulation of cell-specific genes.

bioRxiv preprint doi: https://doi.org/10.1101/517250; this version posted January 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

We next wanted to identify if the most biologically important genes in the network were cell-specific, cell-ubiquitous, or both, as most results in animals assume that core TFs must be expressed in a cell-specific manner[1]. To predict biological significance, we developed a Network Motif Score (NMS) to quantifies the number of times each gene appears in certain network motifs, such a feedback and feedforward loops[12]. These motifs were chosen as they were significantly enriched in our biological network versus a random network of the same size, and have been shown to often contain genes that have important biological functions[13–15] (Extended Data Fig 2). In our inferred GRN, we found that 737 (24.7%) of the 2982 genes have an NMS > 0, meaning they appear in at least one of the network motifs. To validate the NMS, we found that 22 known stem cell regulators had scores in the top 50% of genes, with 10 of those 22 (45.5%) in the top 25% of genes, supporting that high NMS scores are correlated with stem cell function (Supplemental Table 2). Further, 510 (69.2%) and 217 (31.8%) of these genes are cell-ubiquitous (4 or more enriched stem cells, red) and cell-specific (3 or less enriched stem cells, blue), respectively (Figure 2A). This result is in contrast to that of animal systems, which often assume that core genes have highly cell-specific expression[1]. However, we reasoned that by assuming that core genes must have cell-specific expression, previous studies may have missed genes related to stem cell maintenance that are expressed in multiple cell types. Given that more cell-ubiquitous genes have higher importance scores in our dataset, we focused our downstream analysis on identifying a stem-cell-ubiquitous gene with characteristics of a functionally important regulator.

When we began to examine the stem-cell-ubiquitous regulators, we found that TESMIN-LIKE CXC 2 (TCX2, also known as SOL2), a known homologue of the LIN54 DNA-binding component of the mammalian DREAM complex which regulates the cell cycle and the transition from cell quiescence to proliferation[16–18], had the ninth highest NMS (top 1.2% of genes). This suggests that TCX2 could have an important role across all of the stem cells. To further support the biological significance of TCX2, we examined the subnetwork of its first neighbors (i.e., genes predicted to be either directly upstream or downstream of TCX2). We found that TCX2 is enriched in 5 out of the 6 stem cell types and predicted to regulate at least one gene in all of those cell types, supporting that TCX2 could be a stem-cell-ubiquitous regulator that controls stem-cell-specific core genes (Figure 2B). In addition, when compared to the genes with the top 10 NMS, TCX2 has the highest outdegree (number of edges going out) and low indegree (number of edges coming in), suggesting that TCX2 could orchestrate coordinated stem cell division as suggested by the function of its mammalian homologue[16–18].

If TCX2 is indeed a key regulator for stem cell maintenance and division, we would expect that a change in its expression would cause a developmental phenotype related to these aspects. To test this hypothesis, we obtained two knockdown (*tcx2-1, tcx2-2*) and one knockout (*tcx2-3*) mutants of TCX2 , which all show similar phenotypes (Figure 3A, Extended Data Fig 3). Importantly, we observed in *tcx2-3* an overall disorganization of the stem cells, including aberrant divisions in the Quiescent Center (QC), columella, endodermis, pericycle, and xylem cells (Figure 3A). Additionally, *tcx2-3* mutants showed longer roots due to a higher number in meristematic cell/ higher proliferation (Figure 3A, Extended Data Fig 3). Notably, similar phenotypes related to cell divisions have been observed also in the stomata of *tcx2* mutants[16]. Taken together, these results suggest that TCX2, as stem-cell-ubiquitous genes, regulates stem cell divisions in a cell-type specific manner.

bioRxiv preprint doi: https://doi.org/10.1101/517250; this version posted January 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

We hypothesized that TCX2 controls stem cell division by regulating important, cell type-specific genes. Notably, all of our stem cell markers, in addition to being expressed only to one stem cell type, are known to have functions in stem cell regulation[19–23]. Thus, we crossed the marker lines for the Quiescent Center (QC; WOX5:GFP), Cortex Endodermis Initials (CEI; CYCD6:GFP), Epidermis/Lateral Root Cap Initials (Epi/LRC;FEZ:FEZ-GFP), and Xylem Initials (Xyl;TMO5:3xGFP) (Figure 1A) into the *tcx2-2* and *tcx2-3* mutant alleles (Figure 3B). Compared to WT, in a *tcx2* mutant the expression pattern of these markers is expanded. Specifically, the QC marker expands into the CEI, the CEI marker expands into the endodermis and cortex layers, the Epi/LRC marker expands into the Columella Stem Cells (CSCs), and the Xyl marker expands into the procambial cells (Figure 3B). This suggests that in the absence of TCX2 coordination of stem cell division and identity is unregulated through an unknown mechanism.

When we examined the predicted upstream regulators and downstream targets of TCX2, we found that 75% are predicted to be cell-specific (expressed in ≤3 stem cell types), suggesting that TCX2 could be regulated and it regulates targets in a cell type-specific manner. (Supplemental Table 3). Thus, to identify additional cell-specific regulators as well as targets of TCX2, we obtained mutants of the transcription factors (TFs) predicted to be TCX2's first neighbors (i.e. directly upstream or downstream) that also had high NMS scores (Figure 3C, Supplemental Table 3). Two of the genes, SHORTROOT (SHR), and SOMBRERO (SMB) have phenotypes in the stem cells of their loss-of-function mutants, while the loss-of-function mutant of STERILE APETALA (SAP) is homozygous sterile[20,22,24–26]. Additionally, a quadruple mutant of REVOLUTA (REV) together with three other xylem regulators results in missing xylem layers[27]. Further, we obtained loss-of-function mutants of GATA TRANSCRIPTION FACTOR 9 (GATA9), AT1G75710, ORIGIN OF REPLICATION COMPLEX 1B (ORC1B), ANTHOCYANINLESS 2 (ANL2), and REPRODUCTIVE MERISTEM 28 (REM28), which showed root stem cell phenotypes (Figure 3C, Extended Data Fig 4). We were able to validate that TCX2 was differentially expressed ($p<0.05$) in *gata9*, *at1g75710*, *rev*, *orc1b*, and *anl2* mutants as well as in the SHR overexpression line[20] (75.0% of the 8 predicted upstream regulators). Additionally, we observed that REM28 and SAP (50% of the 4 predicted downstream TCX2 targets) were differentially expressed in the *tcx2-3* mutant (Figure 3D, Supplemental Table 3, Extended Data Fig 4). Overall these results suggest TCX2 orchestrates coordinated stem cell divisions through stem-cell-specific regulatory cascades. Accordingly, given that most of the validated upstream regulators of TCX2 are stem-cell-type specific (Supplemental Table 3), we propose that these cell-specific regulators modulate the dynamics of TCX2 expression in individual cell types. In turn, changes in TCX2 dynamics correlate with changes in expression of its downstream targets (Figures 3C, 3D). Thus, we hypothesized that dynamics of TCX2 differ in specific stem cells, as well as changes in TCX2 expression could be used to predict when each stem cell population divides.

If TCX2 expression is dynamically changing over time in a cell-specific manner, we would predict that the TCX2 GRN also changes temporally. Specifically, we could expect that TCX2 differentially regulates its targets in specific cell types at certain times depending on its expression levels. Thus, to determine if the TCX2 regulatory network changes over time, we first selected 176 genes of interest that were differentially expressed in the transcriptional data we obtained for *tcx2-3* mutant (Supplementary Table 4) as well as enriched in the stem cells, as these are most likely to be the downstream of TCX2 in the stem cells. We inferred GRNs using a

bioRxiv preprint doi: https://doi.org/10.1101/517250; this version posted January 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

time course of the root meristem that is stem cell-enriched[28] (hereinafter referred to as the stem cell time course) to predict one network per time point (every 8 hours from 4 days to 6 days). We found that genes in the first neighbor network of TCX2 have different predicted regulations depending on the time point. Specifically, most of the regulation to and from TCX2 are predicted to occur between 4 days (4D) and 5 days (5D), which is the developmental time at which many stem cell divisions take place[20]. (Extended Data Fig 5). Thus, since our gene expression data suggest that loss of TCX2 function correlates with an increase in stem cell division, we hypothesized that most of the TCX2-regulated stem cell division is occurring between 4D 16H and 5D, time at which TCX2 expression decreases at least by 1.5 fold-change (Extended Data Fig 5).

To test how these time- and cell-specific GRNs affect TCX2 expression and therefore cell division, we built a mechanistic model of the GRNs predicted every 8 hours from 4D to 5D (see M&M and Supplementary Information). We used our stem cell time course to determine the cell-specific networks at each time point and constructed equations for each gene in the network (Figure 4A, Extended Data Fig 6). Unlike our GRN, which only predicts the regulations in each cell at each time point, our Ordinary Differential Equation (ODE) model converts the network prediction into a quantitative model of gene expression. Thus, this model allowed us to quantify how TCX2 dynamics change over time and to correlate significant changes in expression with cell division. Our model included the possibility of some of the proteins moving between cell types, as this is a known local signaling/ cell-to-cell communication mechanism[25,29]. Specifically, we used scanning Fluorescence Correlation Spectroscopy (Scanning FCS) and observed that TCX2 does not move between cells, thus suggesting a cell-autonomous function, while observed movement of WOX5[28] and CRF2/TMO3 between cells is in line with a non-cell-autonomous function (Extended Data Fig 7). As our sensitivity analysis predicted that the oligomeric state of TCX2 in the Xyl, diffusion coefficient of WOX5 from the CEI to the QC, and diffusion coefficient of WOX5 from the QC to the Xyl were three of the most important parameters in the model, we experimentally determined these parameters (Extended Data Fig 6, Supplemental Table 5). Given that our network and time course data predict that TCX2-mediated cell division is tightly coordinated and controlled between 4D 16H and 5D, we wanted to ensure that we accurately measured TCX2 dynamics in this time period to produce the best predictive model of stem cell division. To this end, we quantified the expression of the TCX2:TCX2-YFP marker in different stem cells every 2 hours from 4D 18H to 4D 22H (hereinafter referred to as the YFP tracking data) (Figure 4B, see M&M). We then used the average expression of TCX2 in each cell at each time point to estimate parameters in our model (Supplemental Table 6). The result of this model is thus a spatiotemporal map of the expression dynamics of TCX2 and its predicted first neighbors. Given that TCX2 expression has previously been shown to disappear 1-2 hours before stomatal division[16], we reasoned that we could use our model of TCX2 expression to predict when stem cell division occurs in the root.

Our model predicts that there is a significant (fold-change > 1.5) increase in TCX2 expression specifically in the QC and Xyl between 4D 8H and 4D 16H. After this time, our model predicts that the expression of TCX2 in the QC does not significantly decrease and is significantly higher than in all of the actively dividing stem cells (Figure 4C, Supplemental Table 7). Given that the QC does not divide at 5D[28], this suggests that high levels of TCX2 correlate with a lack of QC division. This prediction is supported by our YFP tracking data which shows that half of the QC cells have relatively constant TCX2 levels between 4D 16H and 5D (Extended Data Fig 8).

bioRxiv preprint doi: https://doi.org/10.1101/517250; this version posted January 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Meanwhile, TCX2 expression is predicted to significantly decrease between 4D 16H and 5D in both the Xyl and CSCs, suggesting that these cells divide during this time. This prediction is also supported by our YFP tracking data showing that the majority of Xyl and CSCs cells have no TCX2 expression after 4D 20H (Extended Data Fig 8). In contrast, the CEI and Epi/LRC show only a modest decrease in TCX2 expression between 4D 16H and 5D. This could be due to only some of these cells dividing at that time, as our YFP tracking data shows a large amount of variation in TCX2 expression in these cell populations (Extended Data Fig 8). Taken together, our model and experimental data both suggest that TCX2 not only initiates the division of the actively dividing stem cells, but it also inhibits the division of the QC during the same timeframe, through an unknown mechanism Further, our results allow us to narrow the timing of TCX2-induced stem cell division to a 4-hour window, between 4D 20H and 5D.

Overall, our results show TCX2 as an important cell-ubiquitous gene that regulates stem cell division by coordinating cell-specific regulatory networks. We showed that *tcx2* mutants have additional cell divisions in all stem cell populations and misexpression of known cell-specific marker genes. Further, we validated that TCX2 regulates cell-specific genes, supporting that cell-ubiquitous and cell-specific genes work together to coordinate cell division. Our mechanistic model of the TCX2 GRN illustrated that we can use TCX2 expression to predict the timing of stem cell division. Our results provide evidence that cell-ubiquitous genes and global signaling mechanisms are important for maintaining stem cell identity and plasticity.

Materials and Methods

Lines used in this study

A list of T-DNA insertion lines used in this study is provided in Supplemental Table 8. All T-DNA insertion lines were obtained from the Arabidopsis Biological Resource Center (ABRC: https://abrc.osu.edu/). The marker lines used in this study are described as follows: WOX5:GFP[19], CYCD6:GFP[20], J2341:GFP[30], FEZ:FEZ-GFP[22], TMO5:3xGFP[23], CVP2:NLS-VENUS[31], AGL42:GFP[32]. The TCX2:TCX2-YFP translation fusion is described in[16], the WOX5:WOX5-GFP translational fusion is described in[28], and the TMO3:TMO3-GFP translational fusion is described in[33].

Stem cell transcriptional profile

Three to four biological replicates were collected for each marker line. For each biological replicate, 250-500mg of seed were wet sterilized using 50% bleach, 10% Tween and water and stratified at 4°C for 2 days. Seeds were plated on 1x MS, 1% sucrose plates with Nitex mesh and grown under long day conditions (16 hr light/8 hr dark) at 22°C for 5 days. Protoplasting, cell sorting, RNA extraction, and library preparation were performed as described in[36]. For the non-stem cell control, the GFP-negative cells from the AGL42:GFP line were collected. Libraries were sequenced on an Illumina HiSeq 2500 with 100bp single end reads. Reads were mapped and FPKM (fragments per kilobase per million mapped reads) values were obtained using Bowtie, Tuxedo, and Rsubread as described in[34]. Data are available on Gene Expression Omnibus (GEO: https://www.ncbi.nlm.nih.gov/geo/), accession #GSE98204.

Differential expression analysis

Differential expression analysis was performed using PoissonSeq[34,35]. First, stem cell-enriched genes were identified as being enriched (q-value < 0.05 and fold change > 2) in any one stem cell

bioRxiv preprint doi: https://doi.org/10.1101/517250; this version posted January 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

population compared to the non-stem cell control. Then, genes were classified as enriched in each stem cell type if they met the enrichment criteria in that stem cell type versus all other stem cell types. If genes were equally expressed in more than one stem cell type, they were considered enriched in multiple stem cell types. All differentially expressed genes are reported in Supplemental Table . The Venn diagram in Figure 1C displaying the proportions of genes enriched in each stem cell was constructed using InteractiVenn[36] (http://www.interactivenn.net/).

Gene regulatory network inference

The Regression Tree Pipeline for Spatial, Temporal, and Replicate data (RTP-STAR[11]) was used for all network inference. All networks were inferred using the default parameters described in[11]. For the SCN GRN, networks were inferred for each stem cell separately (resulting in 6 networks, one for each stem cell) and then combined to form the final network. For the stem-cell-specific networks, only the genes enriched in that specific stem cell were used in the network inference. If genes were enriched in multiple stem cells, they were included in all of those individual stem cell networks (e.g. TCX2, which is enriched in all of the stem cells except Protophlo, was included in 5 of the 6 stem cell networks). In addition, only the replicates from that specific stem cell and the SCN marker were used for the inference (e.g. for the QC-enriched cells, only the WOX5:GFP and AGL42:GFP replicates were used). Due to the pseudo-random nature of *k*-means clustering (i.e., the first clustering step is always random), 100 different clustering configurations were used for network inference. Edges that appeared in at least 1/3 of the 100 different networks were retained in the final network as this cutoff resulted in a scale-free network. For the time point-specific GRNs, the same parameters were used as for the SCN GRN. The replicates from the stem cell time course[28] were used to construct each network at each time point. All network visualization was performed using Cytoscape (http://cytoscape.org/).

Biological validation

Confocal imaging was performed on a Zeiss LSM 710. Cell walls were counterstained using propidium iodide (PI). The corrected total cell fluorescence (CTCF) was calculated as in[28]. When counting cells with GFP expression, a local auto threshold using the Phansalkar method (REF) was applied in ImageJ to the GFP channel before counting. For qPCR, total RNA was isolated from approximately 2mm of 5 day old Col-0, *gata9-1, gata9-2, at1g75710-1, at1g75710-2, rev-5, orc1b-1, orc1b-2, anl2-2* and *anl2-3,* root tips using the RNeasy Micro Kit (Qiagen). qPCR was performed and analyzed as described in[28]. Differential expression was defined as a $p<0.05$ using a z-test with a known mean of 1 and standard deviation of 0.17 (based on the Col-0 sample). Primers used for qPCR are provided in Supplementary Table 9. SHR regulation of TCX2 was validated using data from[20].

For the *tcx2-3* transcriptional profile, total RNA was isolated from approximately 2mm of 5 day old Col-0 and *tcx2-3* root tips using the RNeasy Micro Kit. cDNA synthesis and amplification were performed using the NEBNext Ultra II RNA Library Prep Kit for Illumina. Libraries were sequenced on an Illumina HiSeq 2500 with 100 bp single-end reads. Reads were mapped and differential expression was calculated as previously described, except the q-value threshold was set to 0.5 based on the q-value of TCX2, which was assumed to be differentially expressed in its own mutant background. Data are available on GEO, accession #GSE123984.

TCX2:TCX2-YFP tracking

bioRxiv preprint doi: https://doi.org/10.1101/517250; this version posted January 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Confocal images of the TCX2:TCX2-YFP line were obtained by imaging roots submerged in agar every 2 hours. A MATLAB-based image analysis software was used to detect, segment, and track individual cells expressing TCX2:TCX2-YFP in 3D time-course fluorescence microscopy images[37]. The average voxel intensity, which is a proxy for YFP expression, was measured as the average voxel value within the set of voxels describing a segmented cell.

Scanning Fluorescence Correlation Spectroscopy (Scanning FCS)

Image acquisition for Scanning FCS was performed on a Zeiss LSM880 confocal microscope. For Number and Brightness (N&B) on the TCX2:TCX2-YFP and 35S:YFP lines, the parameters were set as follows: image size of 256x256 pixels, pixel dwell time of 8.19 µs, and pixel size of 100 nm. The 35S:YFP line was used to calculate the monomer brightness and cursor size as described in[25,38]. For Pair Correlation Function (pCF) on the 35S:GFP, TCX2:TCX2-YFP and TMO3:TMO3-GFP lines, the parameters were set as follows: image size of 32x1 pixels, pixel dwell time of 8.19 µs, and pixel size between 100-500nm. The movement index (MI) of the 35S:GFP line was used as a positive control. All analysis was performed in the SimFCS software as described in[25,38].

Ordinary Differential Equation (ODE) modeling

ODE equations were constructed based on the GRNs shown in Figure 4A. One set of equations was built for each gene in each cell type. The equations changed at 4D 8H and 4D 16H to account for the changes in the predicted network (as shown in Figure 4A). If a sign was not predicted in the network, it was assumed that the regulation was positive (activation) in the model. A schematic showing the location of genes, and what proteins can move between cell types, is presented in Extended Data Fig 6. All equations are provided in Supplemental Equations.

A sensitivity analysis was performed using the total Sobol index[25,38,39]. Sensitive parameters were defined as having a significantly higher ($p<0.05$) total Sobol index than the control parameter using a Wilcoxon Test with Steel-Dwass for multiple comparisons. (Supplemental Table 5) The sensitive diffusion coefficients and oligomeric states were experimentally measured using scanning FCS. The remainder of the parameters were estimated either directly from the stem cell time course, or by using simulated annealing on the stem cell time course as described in[28]. All parameter values, and how they were estimated, are reported in Supplemental Table 6.

Contributions

NMC and RS conceptualized the study and designed the experiments. NMC and APF performed transcriptional profiling. NMC and MAdLB performed differential expression analysis. NMC, ECN, TTN, TBS, and PJS performed biological validation. NMC, ECN, TTN, and PJS collected confocal images. NMC constructed and analyzed the mathematical model. ARS and DCB contributed the TCX2:TCX2-YFP translational fusion. EB and CMW analyzed the YFP tracking data. NMC and RS wrote the paper, and all co-authors edited the paper.

Acknowledgements

We thank Christian Hardtke for providing the CVP2:NLS-VENUS line and Rüdiger Simon for providing the WOX5:WOX5-GFP line. We thank Irena Brglez for her assistance with media preparation. We thank Sarah Schuett and the Flow Cytometry and Cell Sorting Laboratory at

bioRxiv preprint doi: https://doi.org/10.1101/517250; this version posted January 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

North Carolina State University (NCSU) for their assistance with cell sorting. Images in this manuscript were generated using the instruments and services at the Cellular and Molecular Imaging Facility (CMIF) at NCSU. Next-generation sequencing was performed by the Genomic Sciences Laboratory (GSL) at NCSU.

Funding

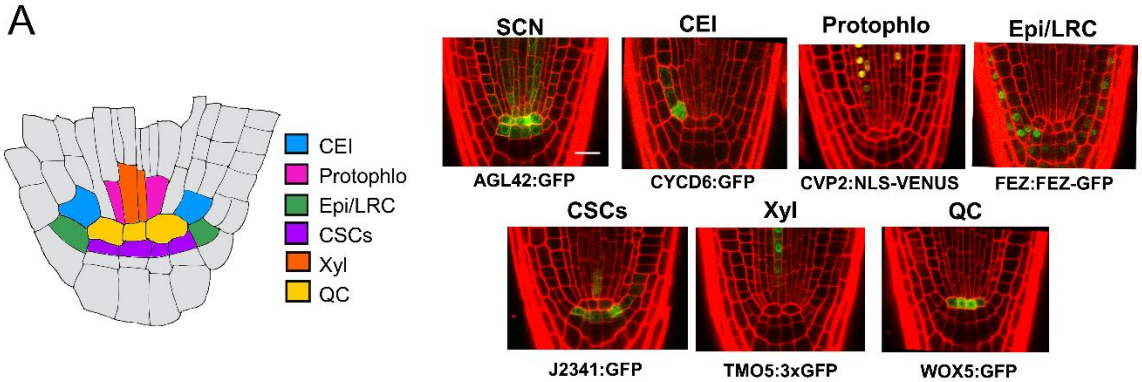
This work was supported by an NSF GRF (DGE-1252376) awarded to NMC and APF. EB is supported by a GAAN Fellowship in Molecular Biotechnology (grant #P200A160061). ARS was supported by the Donald Kennedy Fellowship and NIH graduate training grant NIH5T32GM007276 to Stanford University. PJS was supported by an Integrated Molecular Plant Systems Research Experience for Undergraduates (IMPS REU) grant awarded to NCSU DCB is an investigator of the Howard Hughes Medical Institute. Research in the RS lab was funded by an NSF CAREER grant (MCB-1453130).

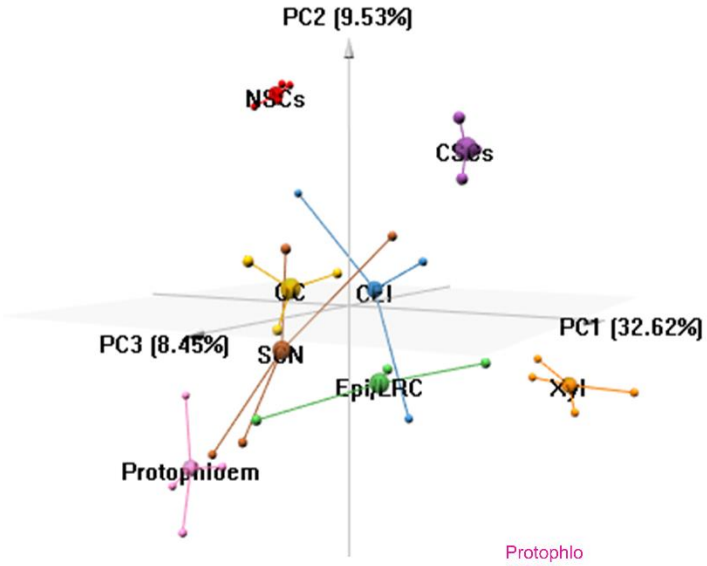
Supplementary Information

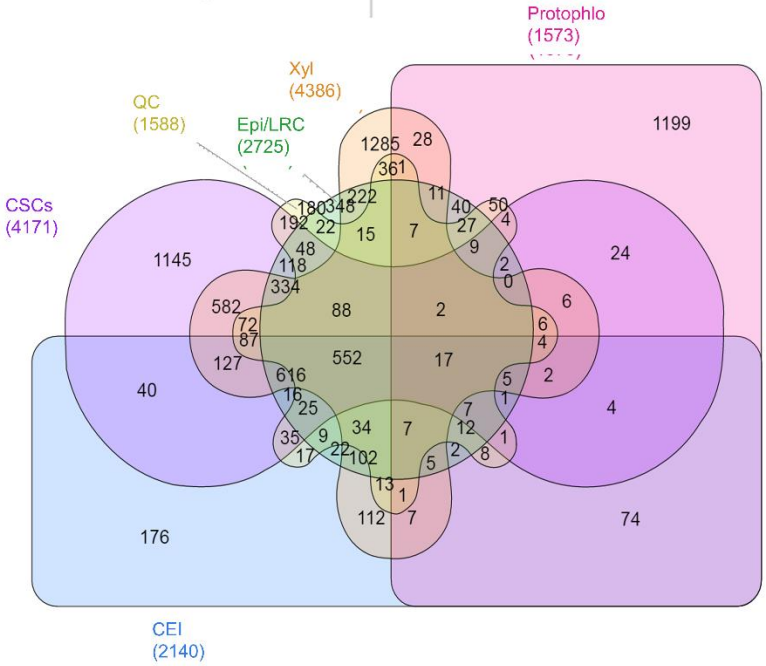
Supplementary figures, tables, and equations are included in the Supplementary Information PDF.

bioRxiv preprint doi: https://doi.org/10.1101/517250; this version posted January 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Main Figures





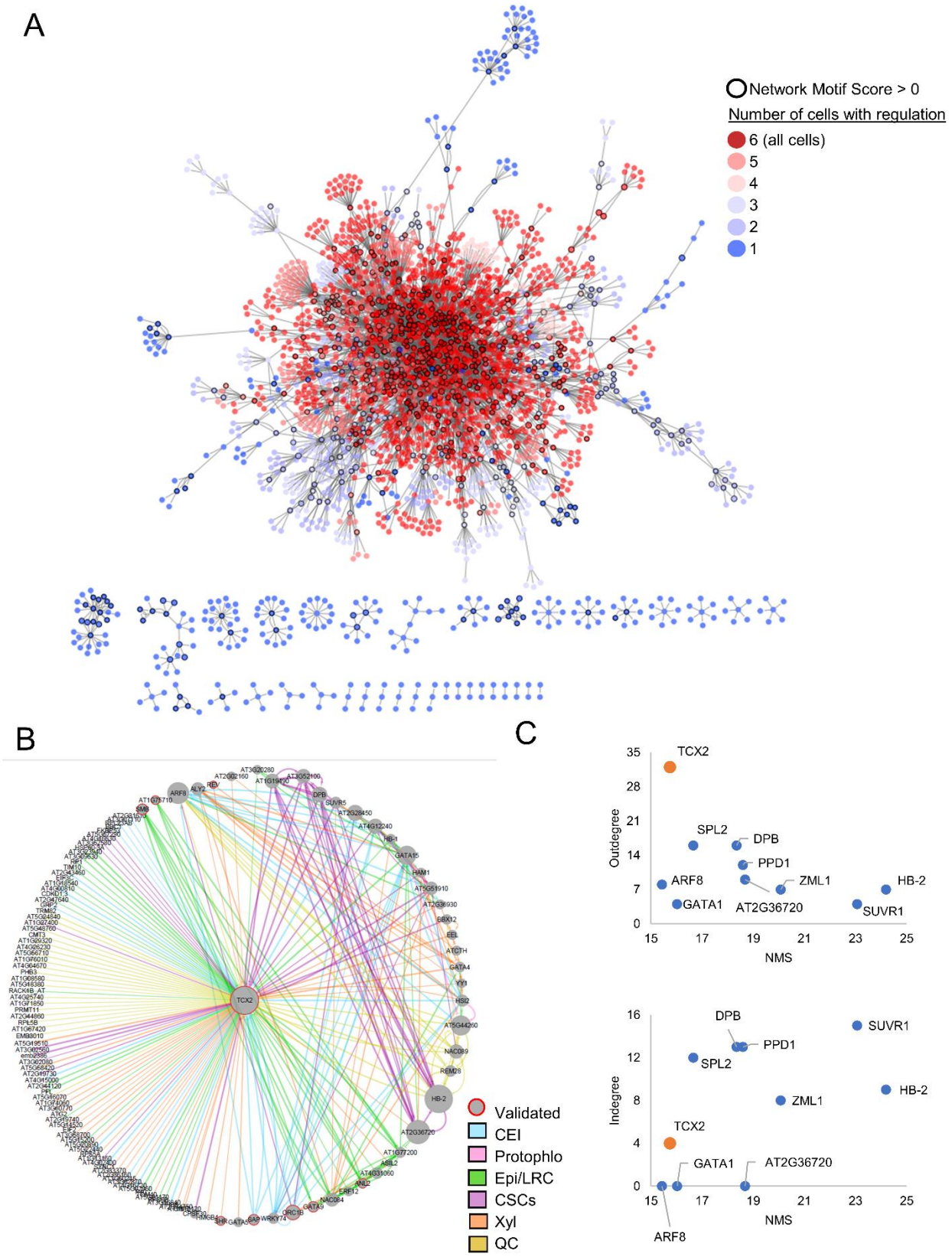


363

bioRxiv preprint doi: https://doi.org/10.1101/517250; this version posted January 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 1. Distribution of cell-specific and cell-ubiquitous genes within the Arabidopsis root stem cell niche. (A) (left) Schematic of the Arabidopsis root stem cell niche. CEI – cortex endodermis initials (blue); Protophlo- protophloem (pink); Epi/LRC – epidermis/lateral root cap initials (green); CSCs – columella stem cells (purple); Xyl – xylem initials (orange); QC – quiescent center (yellow). (left) GFP marker lines used to transcriptionally profile stem cells. SCN – stem cell niche; Scale bar = 20µm. (B) 3D principal component analysis (PCA) of the stem cell transcriptional profiles. The x, y, and z axis represent the three largest sources of variation (i.e. three largest principal components) of the dataset. Small spheres are biological replicates, large spheres are centroids. Red – Non stem cells (NSCs); Brown – SCN; Blue – CEI; Pink – Protophlo; Green – Epi/LRC; Purple – CSCs; Orange – Xyl; Yellow – QC; (C) Distribution of the 9266 stem cell-enriched genes across the stem cell niche. Enrichment criteria are q-value < 0.05 (from PoissonSeq) and fold change in expression > 2.

bioRxiv preprint doi: https://doi.org/10.1101/517250; this version posted January 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



377

bioRxiv preprint doi: https://doi.org/10.1101/517250; this version posted January 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 2. Gene regulatory network (GRN) of the stem cell-enriched genes connects cell-specific and cell-ubiquitous hub genes. (A) Inferred GRN of 2982 out of the 9266 stem cell-enriched genes. Genes are colored based on the number of genes in which they are enriched, with red genes (>3 enriched cells) considered cell-ubiquitous and blue genes (≤ 3 enriched cells) considered cell-specific. Black outlines represent hub genes which have a normalized motif score (NMS) > 0. Arrows represent predicted activation, bars inferred repression, and circles no inferred sign. (B) First-neighbor GRN of TCX2. Gene size represents the NMS score. Red borders represent the genes which were biologically validated. Edge colors represent the cell in which the regulation is inferred. Blue – CEI; Pink – Protophlo; Green – Epi/LRC; Purple – CSCs; Orange – Xyl; Yellow – QC. (C) Outdegree (top plot) and indegree (bottom plot) vs NMS score of the genes with the top 10 NMS scores in (A). TCX2 is highlighted in orange.

bioRxiv preprint doi: https://doi.org/10.1101/517250; this version posted January 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

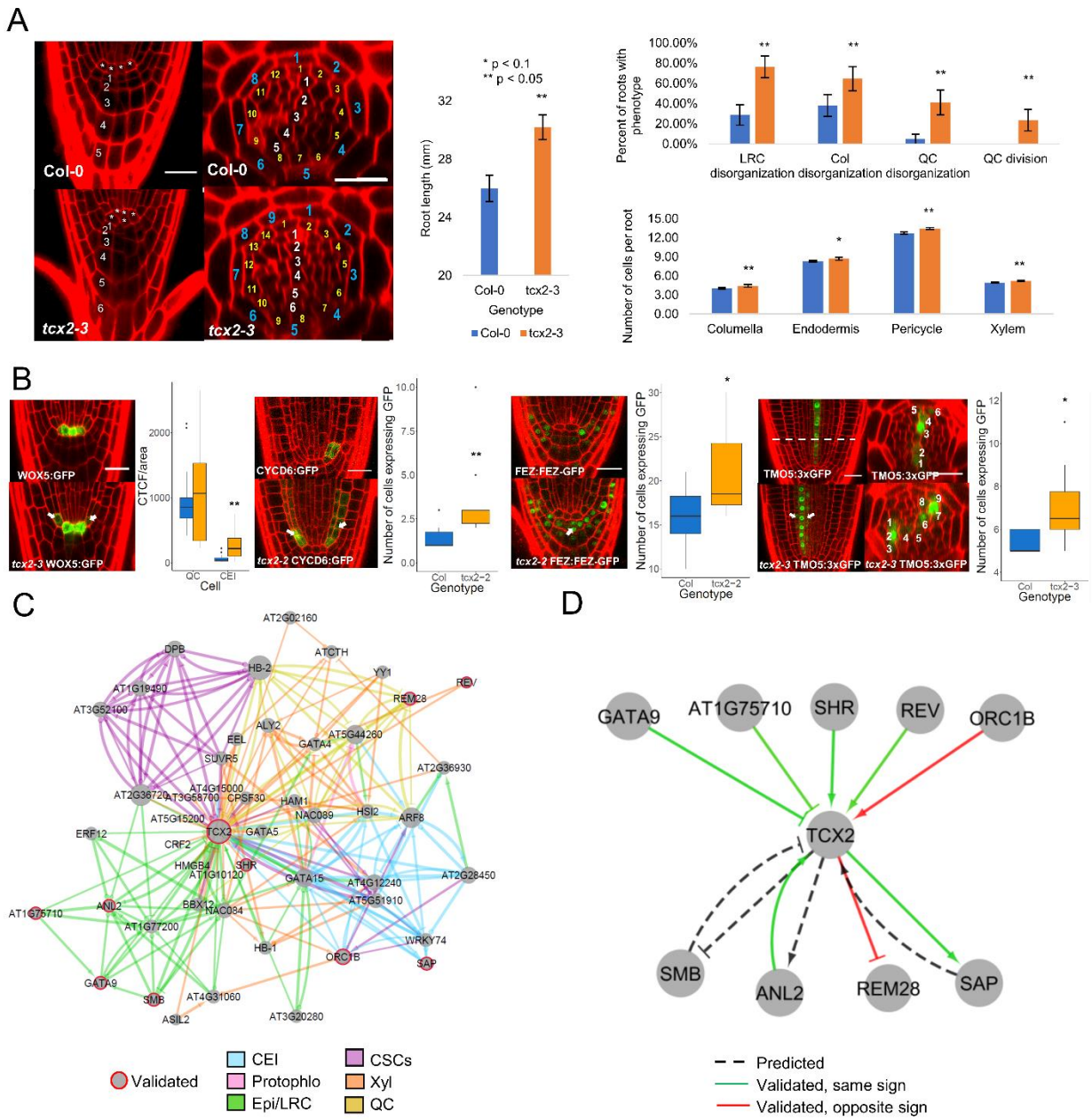


Figure 3. TCX2 controls stem cell division through cell-specific regulators and targets. (A) (Left panel) Medial longitudinal (left) and radial (right) sections of 5 day old WT (top) and *tcx2* mutant (bottom) plants. In medial longitudinal sections, * labels QC cells and numbers denote columella cell files. In radial sections, white numbers denote xylem cells, yellow pericycle, and blue endodermis. (Middle panel) Length of 7 day old WT (blue) and *tcx2* mutant (orange) roots. (Right panel) Quantification of stem cell phenotypes (top plot) and number of cell files (bottom plot) in 5 day old WT (blue) and *tcx2* mutant (orange) roots. * denotes $p < 0.1$, ** denotes $p < 0.05$, Wilcoxon test. Error bars denote SEM. (B) (left panels) Medial longitudinal sections of 5 day old WOX5:GFP (left), CYCD6:GFP (second from left), FEZ:FEZ-GFP (third from left), and TMO5:3xGFP (right) in WT (top) and *tcx2* mutant (bottom) plants. For TMO5:3xGFP, a radial

bioRxiv preprint doi: https://doi.org/10.1101/517250; this version posted January 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

section (middle) is also shown taken at the location of the white, dashed line. (right panels) Quantification of GFP in WT (blue) and *tcx2* mutant (orange) plants. Black dots represent outliers. (C) First neighbor TF network of TCX2. Gene size represents the NMS score. Red borders represent the genes which were biologically validated. Edge colors represent the cell in which the regulation is inferred. Blue – CEI; Pink – Protophlo; Green – Epi/LRC; Purple – CSCs; Orange – Xyl; Yellow – QC. Arrows represent predicted activation, bars inferred repression, and circles no inferred sign. (D) Validated first-neighbor TFs of TCX2. Dashed lines were predicted in the GRN but not validated. Green (red) lines represent validated regulations that were found (were not found) to agree with the predicted sign.

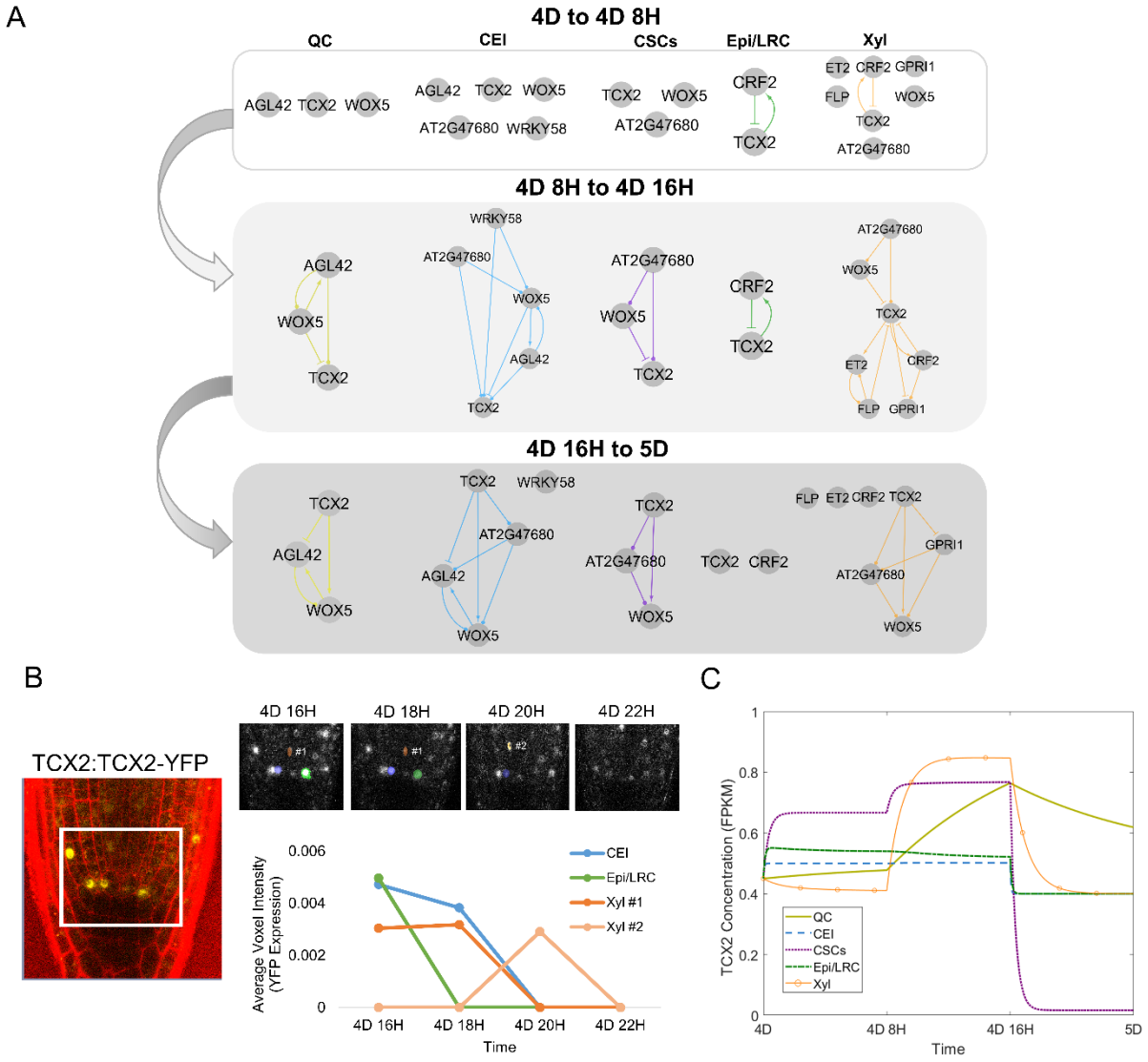


Figure 4. Mathematical modeling of TCX2 network predicts timing of cell division. (A) TCX2 first neighbor TF networks predicted using RTP-STAR on the stem cell time course for 4 day (4D) to 4 days 8 hours (4D 8H) (top), 4D 8H to 4D 16H (middle), and 4D 16H to 5D (bottom). Networks are separated based on the cell type the genes are expressed in: QC (yellow), CEI (blue), CSCs (purple), Epi/LRC (green), Xyl (orange). Arrows represent predicted activation, bars inferred repression, and circles no inferred sign. (B) (left) Representative image

bioRxiv preprint doi: https://doi.org/10.1101/517250; this version posted January 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

of TCX2:TCX2-YFP at 4D 16H. White box represents the stem cell niche were cells were tracked over time. (right, top) YFP-positive cells tracked every 2 hours from 4D 16H (left) to 4D 20H (right). Stem cells that were tracked are marked in blue (CEI), green (Epi/LRC), and orange (Xyl). Two Xyl cells were tracked, #1 and #2. All of these 4 stem cells had no measurable YFP expression at 4D 22H. (right,bottom) Quantification of YFP expression in tracked cells. (C) ODE model prediction of cell-specific TCX2 expression from 4D to 5D. FPKM: fragments per kilobase per million mapped reads.

bioRxiv preprint doi: https://doi.org/10.1101/517250; this version posted January 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

1. Alessio, A. C. D. *et al.* A systematic approach to identify candidate transcription factors that control cell identity. *Stem Cell Reports* **5,** 763–775 (2015).

2. Morris, S. A. & Daley, G. Q. A blueprint for engineering cell fate : current technologies to reprogram cell identity. *Cell Res.* **23,** 33–48 (2013).

3. Bulstrode, H. *et al.* Elevated FOXG1 and SOX2 in glioblastoma enforces neural stem cell identity through transcriptional control of cell cycle and epigenetic regulators. *Genes Dev.* **31,** 1–17 (2017).

4. Vierbuchen, T. & Wernig, M. Molecular Roadblocks for Cellular Reprogramming. *Mol. Cell* **47,** 827–838 (2012).

5. Graf, T. & Enver, T. Forcing cells to change lineages. *Nature* **462,** 587–594 (2009).

6. Sancho-Martinez, I., Baek, S. H., Carlos, J. & Belmonte, I. Lineage conversion methodologies meet the reprogramming toolbox. *Nat. Cell Biol.* **14,** 892–899 (2012).

7. Clevers, H., Loh, K. M. & Nusse, R. An integral program for tissue renewal and regeneration: Wnt signaling and stem cell control. *Science (80-.).* **346,** 1248012 (2014).

8. Es, J. H. Van *et al.* Dll1 + secretory progenitor cells revert to stem cells upon crypt damage. *Nat. Cell Biol.* **14,** 1099–1104 (2012).

9. Takahashi, K. *et al.* Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell* **131,** 861–872 (2007).

10. Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A. & Luscombe, N. M. A census of human transcription factors: function, expression and evolution. *Nat. Rev. Genet.* **10,** 252–263 (2009).

11. de Luis Balaguer, M. A., Spurney, R. J., Clark, N. M., Fisher, A. P. & Sozzani, R. TuxNet: A simple interface to process RNA sequencing data and infer gene regulatory networks. *bioRxiv* 337147 (2018).

12. Fisher, A. P., Clark, N. M. & Sozzani, R. Positive feedback and feedforward loops between PERIANTHIA, WUSCHEL-RELATED HOMEOBOX5 and GRF-INTERACTING FACTOR 1 modulate gene expression and function in the Arabidopsis root. *bioRxiv* 439851 (2018).

13. Alon, U. Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* **8,** 450–461 (2007).

14. Milo, R., Shen-Orr, S., Itzkovitz, S. & Kashtan, N. Network Motifs: Simple Building Blocks of Complex Networks. *Science (80-.).* **298,** (2002).

15. Ingram, P. J., Stumpf, M. P. H. & Stark, J. Network motifs: Structure does not determine function. *BMC Genomics* **7,** 1–12 (2006).

16. Simmons, A. R., Davies, K. A., Wang, W., Liu, Z. & Bergmann, D. C. SOL1 and SOL2 Regulate Fate Transition and Cell Divisions in the Arabidopsis Stomatal Lineage. *bioRxiv*

bioRxiv preprint doi: https://doi.org/10.1101/517250; this version posted January 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

394940 (2018).

17. Sadasivam, S. & Decaprio, J. A. The DREAM complex: master coordinator of cell cycle-dependent gene expression. *Nat. Rev. Cancer* **13,** 585–595 (2013).

18. Schmit, F., Cremer, S. & Gaubatz, S. LIN54 is an essential core subunit of the DREAM / LINC complex that binds to the cdc2 promoter in a sequence-specific manner. *FEBS J.* **276,** 5703–5716 (2009).

19. Sarkar, A. K. *et al.* Conserved factors regulate signalling in Arabidopsis thaliana shoot and root stem cell organizers. *Nature* **446,** 811–814 (2007).

20. Sozzani, R. *et al.* Spatiotemporal regulation of cell-cycle genes by SHORTROOT links patterning and growth. *Nature* **466,** 128–132 (2010).

21. Cruz-Ramırez, A. *et al.* A Bistable Circuit Involving SCARECROW-RETINOBLASTOMA Integrates Cues to Inform Asymmetric Stem Cell Division. *Cell* **150,** 1002–1015 (2012).

22. Willemsen, V. *et al.* The NAC Domain Transcription Factors FEZ and SOMBRERO Control the Orientation of Cell Division Plane in Arabidopsis Root Stem Cells. *Dev. Cell* **15,** 913–922 (2008).

23. De Rybel, B. *et al.* A bHLH Complex Controls Embryonic Vascular Tissue Establishment and Indeterminate Growth in Arabidopsis. *Dev. Cell* **24,** 426–437 (2013).

24. Helariutta, Y. *et al.* The SHORT-ROOT Gene Controls Radial Patterning of the Arabidopsis Root through Radial Signaling. *Cell* **101,** 555–567 (2000).

25. Clark, N. M. *et al.* Tracking transcription factor mobility and interaction in Arabidopsis roots with fluorescence correlation spectroscopy. *Elife* **5,** e14770 (2016).

26. Byzova, M. V *et al.* Arabidopsis STERILE APETALA , a multifunctional gene regulating inflorescence , flower , and ovule development. *Genes Dev.* **2,** 1002–1014 (1999).

27. Carlsbecker, A. *et al.* Cell signalling by microRNA165/6 directs gene dose-dependent root cell fate. *Nature* **465,** 316–321 (2010).

28. Clark, N. M. *et al.* Cell type-specific differences in protein complex stoichiometry and transcriptional regulation affect the timing of stem cell division. *bioRxiv* 439331 (2018).

29. Gallagher, K. L., Paquette, A. J., Nakajima, K. & Benfey, P. N. Mechanisms Regulating SHORT-ROOT Intercellular Movement. *Curr. Biol.* **14,** 1847–1851 (2004).

30. Sabatini, S., Heidstra, R., Wildwater, M. & Scheres, B. SCARECROW is involved in positioning the stem cell niche in the Arabidopsis root meristem. *Genes Dev.* **17,** 354–358 (2003).

31. Rodriguez-Villalon, A., Gujas, B., Wijk, R. Van, Munnik, T. & Hardtke, C. S. Primary root protophloem differentiation requires balanced phosphatidylinositol-4 , 5-biphosphate levels and systemically affects root branching. *Development* **142,** 1–10 (2015).

32. Nawy, T. *et al.* Transcriptional Profile of the Arabidopsis Root Quiescent Center. *Plant*

bioRxiv preprint doi: https://doi.org/10.1101/517250; this version posted January 10, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Cell **17,** 1908–1925 (2005).

33. Schlereth, A. *et al.* MONOPTEROS controls embryonic root initiation by regulating a mobile transcription factor. *Nature* **464,** 913–916 (2010).

34. Clark, N. M., Fisher, A. P. & Sozzani, R. Identifying Differentially Expressed Genes Using Fluorescence-Activated Cell Sorting (FACS) and RNA Sequencing from Low Input Samples. in *Computational Cell Biology: Methods and Protocols* **1819,** 139–151 (2018).

35. Li, J., Witten, D. M., Johnstone, I. M. & Tibshirani, R. Normalization , testing , and false discovery rate estimation for RNA-sequencing data. *Biostatistics* **13,** 523–538 (2012).

36. Heberle, H., Meirelles, G. V, da Silva, F. R., Telles, G. P. & Mignhim, R. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* **16,** 169 (2015).

37. Buckner, E. *et al.* Tracking Gene Expression via Light Sheet Microscopy and Computer Vision in Living Organisms. *40th Int. Conf. IEEE EMBS* 818–821 (2018).

38. Clark, N. M. & Sozzani, R. Measuring Protein Movement, Oligomerization State, and Protein–Protein Interaction in Arabidopsis Roots Using Scanning Fluorescence Correlation Spectroscopy (Scanning FCS). in *Plant Genomics: Methods and Protocols* **1610,** 251–266 (2017).

39. Sobol, I. M. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math. Comput. Simul.* **55,** 271–280 (2001).

40. Saltelli, A. *et al.* Variance based sensitivity analysis of model output . Design and estimator for the total sensitivity index. *Comput. Phys. Commun.* **181,** 259–270 (2010).